



# Application of chemometrics to hyperspectral imaging analysis of environmental and agricultural samples

Xin Zhang

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

# Application of chemometrics to hyperspectral imaging analysis of environmental and agricultural samples

Xin Zhang







Doctoral program:

Química Analítica del Medi Ambient i la Pol·lució

# **Application of chemometrics to hyperspectral imaging analysis of environmental and agricultural samples**

**Xin Zhang**

Director: Dr. Roma Tauler Ferré (Institute of Environmental Assessment and Water - Council for Scientific Research IDAEA -CSIC)

Tutor: Dr. Anna de Juan (Department of Analytical Chemistry, University of Barcelona)



## **Acknowledgements**

During the work in this thesis, many people have contributed or been helped me in different ways. I am very grateful to all of you and would like to thank everybody for your time and support. There are some in particular that I would like to take the opportunity to express my gratitude.

First of all, I would like to thank my supervisor, Professor Roma Tauler, who offered me the possibility to do this PhD. He supported me for sharing his knowledge and for guiding me throughout this work all these years. Thanks for reading all the papers and thesis with much care, and a lot of times of revises. Thanks a lot for always taking the time to discuss the work, even when he is very busy. His scientific guidance, his generous way of unconstrained research activities and his great enthusiasm created the optimal working environment. He has not only helped me to develop as a new chemometrist, but also helped me to become a better person with his unique character.

I would like to thank my tutor, Dr. Anna de Juan for accepting to supervise this thesis but also for helping me with her valuable experience and advice. Thanks for her much-valued feedback on my papers and thesis that improved this work. Thanks for always to help me and for sharing her knowledge.

I would like to thank Professor Duponchel, for allowing me participated projects using EPR, SIMS, XPS analysis and for guiding me to visit many interesting groups when I stayed in Lille. The experience enriched my knowledge and broad my horizons in the area we are studying.

Thanks to all staff and all who have come to stay for a good working in group of chemometrics of IDAEA-CSIC, I want to thank all my colleagues and friends for the good times we spent together, for their kindness, friendship, help and interesting discussions inside and beyond the field of Chemometrics. Special thanks to S. Platikanov, J.Jaumot, M. Farrés et al. for their help with various practical problems.

Even though we did not spend much time together I would like to thank my PhD student colleagues at Lille 1 University for always making me feel welcome during my short visits to Lille.

Special thanks to my wonderful friends, those Chinese students studying in Barcelona, give me an incredible amount of energy and happiness. You all inspire me, both professionally and socially.

I also thank China Scholarship Committee, for supporting me to finish my work in Barcelona.

Last but not least, I am deeply grateful to my parents and to my brother for their long lasting help, love and support throughout my entire life.

## Abbreviation

2 D	two dimensional
3 D	three dimensional
AFS	Area of Feasible Solutions
ALS	Alternating Least Squares
AsLS	Asymmetric Least Squares
AVIRIS	Airborne Visible Infrared Imaging Spectrometer (sensor)
CCD	Charge-Coupled Device
CLS	Classical Least Squares
FA	Factor Analysis
FT	Fourier Transform
GPS	Global Positioning System
HPLC	High Performance Liquid Chromatography
ICA	Independent Component Analysis
IR	Infrared
MCR	Multivariate Curve Resolution
MSC	Multiplicative Scatter Correction
MVSA	Minimum Volume Simplex Analysis
NIR	Near Infrared
NMF	Nonnegative Matrix Factorization
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PLS	Partial Least Squares
SIMPLISMA	Simple-to-use Interactive Self-Modeling Mixture Analysis
SNR	Signal to Noise Ratio
SVD	Singular Value Decomposition
TOF-SIMS	Time-of-Flight-Secondary Ion Mass Spectrometry
UV	Ultraviolet
VCA	Vertex Component Analysis



## Content

Abstract .....	1
Chapter 1.....	3
Objectives and structures of the Thesis .....	3
1.1 Objectives.....	5
1.2 Structure of the Thesis .....	7
1.3. List of scientific papers presented in this Thesis.....	8
Chapter 2.....	9
Introduction .....	9
2.1 Hyperspectral imaging applied to environmental and food analysis.....	11
2.2 Chemometric methods.....	33
Chapter 3.....	69
Application of MCR-ALS on remote sensing hyperspectral imaging.....	69
Introduction .....	71
Discussion.....	104
Chapter 4.....	107
Application of hyperspectral imaging combined with chemometrics on food analysis .....	107
Introduction .....	109
Discussion.....	130
Chapter 5.....	133
Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods .....	133
Introduction .....	135
Discussion.....	160
Chapter 6.....	161
Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution .....	161
Introduction .....	163
Discussion.....	178
Chapter 7.....	181

Conclusion .....	181
Resumen de la Tesis .....	187
Resumen.....	189
Objetivos generales de la Tesis .....	191
Introducción .....	195
Resúmenes .....	199
Conclusiones .....	207
References.....	211

# Abstract

Hyperspectral imaging is an active research area in food and environment analysis. Hyperspectral images allow obtaining accurate and reliable knowledge about the chemical composition and distribution of the chemical components on the investigated sample surface. Results of hyperspectral image analysis can be used to acquire fundamental understanding of complex chemical systems for research and development, for commercial testing and adulteration studies, in particular in food and environment analysis and in industrial process analysis and control. Hyperspectral imaging datasets are challenging because of their very large size and complexity. Chemometric methods are proposed to reveal the information contained in the analyzed images as much as possible.

This Thesis deals with the resolution of hyperspectral imaging data by using chemometric methods, in particular by using appropriate data pretreatment methods and by using Multivariate Curve Resolution (MCR) methods. The main contribution of the present Thesis is the study and implementation of the MCR-ALS (Multivariate Curve Resolution Alternating Least Squares) method for the resolution of hyperspectral images, collected by remote sensing (airborne or space borne Earth observation instrument) and by micro-spectroscopy imaging. Specifically, in this Thesis work, we explore the combination of chemometric and hyperspectral imaging methods for the resolution of spectra (signatures) and spatial distribution maps of the chemical constituents of a sample. The ultimate goal of this study is to improve the analysis and interpretation of hyperspectral imaging data by taking advantage of different chemometric powerful tools. Local rank/selectivity properties describing the spatial information of spectroscopic images can be used as a constraint to increase the performance of MCR methods significantly, decreasing rotation ambiguity uncertainties. Different multivariate resolution methods were compared, such as MCR-ALS, Independent Component Analysis (ICA), Principal Component Analysis (PCA), and Minimum Volume Simplex Analysis (MVSA), Multivariate Curve Resolution-Function Minimization (MCR-FMIN), MCR-BANDS and FAC-PACK. All these approaches have been used for the evaluation of the extension of rotation ambiguities remaining in the results after their application. Several hyperspectral images provided by standard and widely used instruments such as NASA's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS), Raman hyperspectral imaging Spectrometer, and Infrared hyperspectral imaging Spectrometer have been used as example of data sets to test the different methods, in particular to test the MCR-ALS method. The effectiveness of MCR-ALS is illustrated by providing exhaustive comparisons with state-of-the-art methods for spectral unmixing using both simulated and real hyperspectral data sets.



# **Chapter 1**

## **Objectives and structures of the Thesis**



## **1.1 Objectives**

The main objective of this Thesis has been the development and implementation of chemometric methods for the analysis of hyperspectral images obtained from remote sensing or from micro-imaging for the analysis of environment and food samples. Different simulated and experimental data sets obtained from public remote sensing data repositories and from experimental measurements have been studied in detail. Especially important has been the extension and application of the MCR-ALS method for the resolution of hyperspectral imaging data with the goal of obtaining the signatures (pure spectra) and distribution maps of the constituents of the analyzed samples.

### **Objectives of the analysis of hyperspectral imaging**

- Apply MCR-ALS to the analysis of hyperspectral image from simulated dataset, experiment datasets including remote sensing hyperspectral image from environment, and micro-hyperspectral image of food commercial sample from chocolate. In all these cases, MCR-ALS was proposed to resolve the spectra of the components in the mixture for their characterization and to estimate their contributions and distribution maps.
- Apply MCR-ALS method to hyperspectral remote sensing image to resolve different objects in the studied image, such as lakes, hard ground, vegetation, buildings etc. and to determine their locations in the images studied.
- Apply MCR-ALS method to analyze hyperspectral image of the chocolate for resolution of their particle shape and size at the micro level, which are the important factor related to the product quality control in the case of food study.
- Apply the MCR-ALS method to the analysis of fluorescence data to track the sources of Dissolved Organic Matter (DOM) in the Ter river for resolving their various contributions, concentrations, geographical distribution and the relationship between human activity along the river and its reservoirs.

### **Objectives of the chemometric analysis**

- Develop and apply MCR-ALS method for data analysis on hyperspectral imaging to obtain pure spectral and constituent distribution in the image.

- Adapt MCR-ALS method with selectivity/local rank constraints to improve the results resolution. FSIW-EFA and correlation coefficient method was proposed for the application of the selectivity constraint in spectroscopic images.
- Apply spectral/signature pretreatment methods to reduce the light scattering influence in NIR, the fluorescence background in Raman spectroscopy when the sample is irradiated, the presence of noise contributions in background and the baseline contribution.
- Apply MCR-ALS method to simultaneously analyze multiple data set arranged in arrays or increased multidirectional structures (multiway) and application of the associated constraint to trilinear models.
- Discuss and compare different ways of calculating the extension of rotation ambiguity of MCR methods, such as MCR-BANDS and FAC-PACK approaches, which allow the evaluation of the quality of the results obtained by MCR-ALS.

## 1.2 Structure of the Thesis

The Thesis is presented in two sections. In the first introductory part, hyperspectral imaging technique, their application in different area and chemometric methods are introduced. In the second part, discussion part of the results, scientific articles and their detail background introduction and result dissection, and finally, references employed are included. This Thesis consists of seven chapters that are described below.

- In the first chapter, the objectives of the Thesis are presented. Furthermore, detailed structure and relationship of scientific work of this report are given.
- In the second chapter, introduction part, background and state of the art techniques of hyperspectral imaging and their application in environment and food area are reviewed. The development of chemometric in hyperspectral imaging and theories of chemometric methods applied in the Thesis are introduced.
- The third chapter, the application of MCR-ALS methods is demonstrated on remote sensing hyperspectral imaging. The simulated data using spectra from USGS library and the public remote sensing data (AVIRIS, Airborne Visible/Infrared Imaging Spectrometer) from NASA are applied for evaluating the chemometric models. The second part shows the effect of using local rank and selectivity constraints based on spatial information of spectroscopic images to increase MCR methods performance and to decrease ambiguity.
- In Chapter 4, it shows the application of Raman and Infrared hyperspectral imaging combined with pretreatment methods and MCR with selectivity constraint to the analysis of the constituents of commercial chocolate samples.
- In Chapter 5, several chemometric resolution methods using bilinear models describing the data are compared. Various ways to calculate the extension of rotation ambiguities of MCR are discussed and compared.
- In Chapter 6, MCR-ALS with the trilinearity constraint is proposed for the analysis of excitation–emission fluorescence data from Dissolved Organic Matter (DOM) in fresh water natural systems, and the results obtained are compared with those obtained with PARAFAC.
- In Chapter 7, the conclusions of the Thesis are presented.

### 1.3. List of scientific papers presented in this Thesis

- Zhang, X.; Tauler, R., Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. **Analytica Chimica Acta** 2013, 762, 25-38.
- Zhang, X.; Marcé, R.; Armengol, J.; Tauler, R., Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution. **Chemosphere** 2014, 111, 120-128.
- Zhang, X.; Juan, A; Tauler, R, Multivariate Curve Resolution applied to hyperspectral imaging analysis of chocolate samples. **Applied Spectroscopy**. 2015, 69(8).
- Zhang, X.; Juan, A; Tauler, R, Local rank-based spatial information for improvement of remote sensing hyperspectral imaging resolution. Submitted to **Talanta**.
- Zhang, X.; R. Tauler. Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods. Submitted to **Journal of Chemometrics**.

## **Chapter 2**

### **Introduction**



This introductory chapter of the Thesis has two main parts. In the first part, the theoretical background and applications of hyperspectral imaging are introduced. In the second part the chemometric methods used in this Thesis are introduced.

## **2.1 Hyperspectral imaging applied to environmental and food analysis**

The theoretical concepts corresponding to hyperspectral imaging and an overview of the applications of hyperspectral imaging on food and environment analysis are introduced in this first part.

Hyperspectral imaging is an integrated technology composed of detector and optical instrumentation on one side, and of computer technology and data processing on the other side <sup>1</sup>. It is also known as chemical or spectroscopic imaging, and it integrates conventional imaging and spectroscopy to attain both spatial and spectral information from an object or sample <sup>2</sup>.

Spectroscopic methods can provide detailed fingerprints of the analyzed food and environment samples using the physical and chemical characteristics of the interaction between electromagnetic radiation and the material of the sample, such as reflectance, transmittance, absorbance, phosphorescence, fluorescence, and radioactive decay<sup>3</sup>, which are either absorbed or emitted. Spectroscopic analysis can provide qualitative and quantitative chemical and physical information <sup>4,5</sup>. By scrutinizing the changes in spectra, one can obtain physical, chemical, and biological information of the analyzed products. However, direct application of spectrometers to samples can only usually detect a small portion of them; therefore, the spectra, strictly speaking, are sometimes not representative of the whole sample, especially when the ingredients are not evenly distributed.

In 1985, the term ‘hyperspectral imaging’ originated from works in remote sensing (the observation of a target by a device without physical contact). It was also called spectroscopic imaging <sup>6</sup>. It was used to make a direct identification of surface materials in the form of images. Hyperspectral imaging studies began with the mapping analysis of airborne minerals in the late 1970s and early 1980s. The invention of the CCDs (charge-coupled device) in 1969 by George Smith and Willard Boyle was the foundation of hyperspectral technology who made its initial development<sup>7</sup>. Progress occurred in the development of the required electronics, hardware, computing, and software throughout the 1980s and until the 1990s. At the beginning of the 1980s, hyperspectral imaging was introduced with the airborne imaging spectrometer (AIS) developed by Alexander Goetz

and his colleagues at NASA's Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, California<sup>8</sup>. In 1983, JPL proposed and developed the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to extend ground-based spectrometers into the air on moving platforms. AVIRIS measured the first spectral images in 1987 and it was the first imaging spectrometer to measure the solar reflected spectrum from 400 nm to 2500 nm with 10 nm intervals<sup>9</sup>. Improvements in sensor, calibration, and data systems were developed to introduce other multispectral and hyperspectral instruments in ground-based and airborne systems during the next years. Although it was first developed for remote sensing applications, hyperspectral imaging has been applied to other areas such as agriculture<sup>10</sup>, biology<sup>11</sup>, environmental<sup>12</sup> and other earth scientific fields<sup>13</sup>.

Hyperspectral images can be considered to be an extension of the concept of digital images but differently to the later, they collect full spectral information on the spatial dimension. Digital images have a finite number of digital values, called pixels, arranged in certain order to describe the color or grayscale information on a surface. The digital image contains a fixed number of rows and columns of pixels. Pixels are the smallest individual elements in an image, holding quantized values that represent the brightness of the object<sup>14</sup>. Therefore, a digital image is an array of  $I$  rows and  $J$  columns, with  $I \times J$  intensity values, also called pixels. A pixel is an intensity value with an associated coordinate in the image. An  $I \times J$  image having  $K$  detected features (variables, wavelengths) would form a three-way data array of size  $I \times J \times K$ . The  $I \times J \times K$  image can be represented as  $K$  slices, where each slice is an image at a single feature. The  $I \times J \times K$  image can also be presented as a two-way array of vectors. In special cases, these vectors can be shown and interpreted as spectra.

In a hyperspectral image, each pixel has not the same single discrete value, but it may have a wide range of values recorded by a sensor or spectroscopy instrument. Like in ordinary spectroscopy, hyperspectral imaging also collects and processes information from across the electromagnetic spectrum. The only difference is that hyperspectral imaging collects the data points in an order as pixels of an image which show the spatial information. Hyperspectral imaging provides much more detailed information on the scene than a color camera. An ordinary color camera only acquires three different spectral channels corresponding to the visual primary colors, red, green and blue. On the contrary, a hyperspectral image is made up of hundreds of contiguous wavebands for each spatial position of a target sample. Each pixel in a hyperspectral image contains the spectrum of a specific spatial position.

Two aspects characterize hyperspectral images:

- They can have many wavelengths (hundreds or thousands).
- Every pixel has a spectrum which can provide chemical information.

Nowadays, different techniques exist for acquiring hyperspectral data which depend on the specific application. Spatial and spectral scanning were mostly used in the previously reported studies<sup>15</sup>. Spatial scanning obtains slit spectra by projecting a strip of the scene onto a slit and dispersing the slit image with a prism or a grating. With these line-scan systems, the spatial dimension is collected by platform movement or scanning. Spectral scanning is typically based on optical band-pass filters. The scene is spectrally scanned by exchanging the filter to change the wavelength of the scanning systems in the spectrometer<sup>16</sup>. Apart from spatial and spectral scanning, non-scanning procedures are in development at present. Non-scanning methods using a single 2-D sensor output produce simultaneously spatial and spectral data. The most important benefit of snapshot hyperspectral imaging is the higher light throughput and the shorter acquisition time. The tremendous amount of memory needed to store the entire dataset makes the computational effort and manufacturing cost high<sup>17, 18</sup>.

Hyperspectral imaging techniques based on Raman, infrared, near infrared and fluorescence spectroscopy are useful methods in different areas, such as polymer research, material science, biomedical diagnostic, pharmaceutical industry, analytical chemistry, process control and environmental analysis<sup>19, 20</sup>. Materials or processes can be also analyzed using multiple imaging techniques across all wavelength and time scales. Hyperspectral imaging can also be used to gain a fundamental understanding of complex chemical systems in industrial processes and use this knowledge to control them. Nowadays, application of hyperspectral imaging is developing very fast, and it allows the in situ acquisition of hyperspectral images (for example, inside the human body, or inside high pressure chemical reactors). The ability to control complex processes in food industry or in environmental monitoring studies will require the same techniques to those used for imaging analysis in other fields<sup>21-23</sup>.

For research, development, commercial testing, or adulteration validation, reference methods for food or environment samples are necessary. Reference methods for food safety and quality control analysis often have limitations, in terms of their adequacy of implementation at the different steps of the food chain<sup>24</sup>. In the traditional analytical chemistry approach, sampling, sample preparation, measuring procedures, and waste

disposals are necessary steps. Normally, these works include laboratory sample pretreatment, dissolution, digestion, separation, enrichment, and other slow and laborious steps. Traditional analytical chemistry methods have the following disadvantages:

- They are time-consuming. Natural samples are usually mixtures and have many complex constituents; therefore they usually need many pretreatment steps before pure components are isolated. However there are some alternative analytical techniques that can produce instantaneous results with good accuracy, but they need the simultaneous application of powerful data analysis techniques<sup>25</sup>;
- They are expensive. For example, chromatographic determinations need many different sensors or expensive columns to analyze different constituents in one sample. Moreover, the control at any crucial link in the process chain may require a large number of independent analyses to be performed<sup>26</sup> and a lot of samples for analysis;
- They should be performed in the laboratory. Because of the laborious and fine operations needed for sample treatment or because of the size of the analytical instruments or because they require instrument stability, the analytical determinations should be done in the laboratory. However in many circumstances, measurements control and management should be on-line or in-field;
- They are not flexible and they are only useful as a single purpose (one method/one parameter analytical tools corresponding for a particular analyte);
- They are not always respectful with the environment (toxic reagents). Usually, sample is destroyed. Also, for example, mobile phases and solvents in chromatography are usually toxic. They contaminate the environment and they are harmful to the health of laboratory workers.

In recent years, the international community is paying more attention on environmental issues and on green chemistry. Green chemistry is the design of chemical products and processes that reduce or eliminate the use and generation of hazardous substances<sup>27</sup>. Analytical laboratories are essential to the implementation of the principles of green analytical chemistry as illustrated in<sup>27</sup>:

- Elimination of reagents and solvents from the analytical laboratory.
- Reduction of the amounts of the reagents used.
- Elimination or reduction of the amount of solvents used in analytical procedures.
- Reduction of emission of vapors and gases.

- Reduction of labor and energy consumption.

### **Hyperspectral imaging instruments**

Considering the limitations of classical analytical reference methods mentioned above, researchers have developed new analytical methods that try to avoid these disadvantages, many of them based on spectroscopic technologies, such as fluorescence spectroscopy<sup>28</sup>, near-infrared spectroscopy (NIR)<sup>29</sup>, mid-infrared spectroscopy (MID)<sup>30</sup> and Raman spectroscopy<sup>31</sup>. Spectroscopic methods enable a much higher level and frequency of sample analysis for quality control in industry and in real-time outdoor monitoring, leading to an improved food safety and environment quality control system<sup>32,33</sup>. The development of robust and flexible spectroscopic instrumentations adapted for on-line or in-field control of the production chain is well suited for the continuous monitoring of processes from raw materials to finished products, and also suited for fluid continuous flow in rivers or air<sup>34</sup>. Such systems provide the possibility for real-time analyses, and they are feasible for large amounts of sample throughput analysis. The other advantages of spectroscopic techniques are their ability to determine multiple parameters or analytes simultaneously, reducing considerably the use of reagents and sample preparation steps.

Because these methods provide detailed spectral data, hyperspectral imaging instruments can be considered an extension of classical spectrometric instruments, allowing also the extraction of quantitative information such as the concentration or relative amounts of material constituents. Even more, hyperspectral imaging spectrometers can also be considered as a particular type of spectrometers, so that image-specific data analytical issues can be addressed as well. Analyzed objects will be shown in a very different way compared to ordinary spectroscopy techniques. Hyperspectral imaging can provide spectral measurements of the whole surface area of the product while conventional spectrometers only can give point measurements. Hyperspectral imaging technology contributes to improve the quality of analysis in environmental monitoring studies and in food processing industry. In addition, it represents a huge increase of the speed of analysis and therefore, a drastically reduction in their costs.

The most used hyperspectral imaging instruments are based on Raman, mid-infrared, near-infrared and fluorescence. They are briefly introduced here. The state of the art and recent applications are also given for each one of these applications.

### **Raman hyperspectral imaging**

Raman hyperspectral imaging was already proposed by Delhaye & Dhamelincourt in 1975<sup>35</sup>. But this technique did not become really feasible as a chemical imaging method

until cooled slow-scan CCD-cameras became available. Now, the Raman microscope is one of the most used instruments of imaging. It normally includes a standard optical microscope, and adds an excitation laser, laser rejection filters, a spectrometer, and an optical sensitive detector such as a charge-coupled device (CCD), or photomultiplier tube, (PMT).

Raman imaging techniques have improved and transformed in recent years because of a growing interest for acquiring multidimensional analytical information. Also, Raman spectroscopy, a complementary vibrational spectroscopic technique for molecular analyses, continues to benefit enormously from advances in laser and array detector developments<sup>36</sup>. Raman imaging has the advantage of its high selectivity, low sensitivity to water, and minimal sample preparation requirements.

The low sensitivity of conventional Raman techniques has led to recent advances in nonlinear Raman microscopic methodologies. Coherent Raman scattering techniques, which include coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) microscopy, have become more prolific in recent years; therefore, the methodologies, instrumentation, and technology involved continues evolving at present. The inherently low scattering cross section of Raman spectroscopy, as well as its diffraction limited lateral resolution, has been overcome by new Raman microscopy techniques. Nonlinear methods such as coherent anti-Stokes Raman spectroscopy and stimulated Raman spectroscopy reduce measurement times and improve resolution, allowing for three-dimensional spectroscopic imaging. Tip-enhanced Raman spectroscopy, offered by surface-enhanced Raman scattering, enables Raman spectroscopic imaging far below the optical diffraction limit<sup>37</sup>.

Yookyung Jung et al. (38) reported recently a longitudinal, real-time alternative for the in vivo, label free imaging of sebaceous glands using Coherent Anti-Stokes Raman Scattering (CARS) microscopy, which has been used for selectively visualize lipids<sup>38</sup>. Anti-Stokes Raman scattering (CARS) microscopy provides a label-free means for visualizing biological samples, but it can suffer from a strong non-resonant background in samples that are prepared using aldehyde-based fixatives<sup>39</sup>.

Label-Free biomedical imaging with high sensitivity was performed by stimulated Raman scattering microscopy (SRS)<sup>40</sup>. A variety of biology and medical applications have been reported, such as differentiating distributions of omega-3 fatty acids and of saturated lipids in living cells, imaging of brain and skin tissues based on intrinsic lipid contrast, and monitoring drug delivery through the epidermis<sup>41, 42</sup>.

The ability to control the size, shape, and material of a surface has been developed in the field of surface-enhanced Raman spectroscopy (SERS). Surface-enhanced Raman scattering (SERS) imaging has widely applied on rapid and sensitive analysis of material surfaces<sup>43</sup>, phenotypical marker detection, cancer cells<sup>44, 45</sup>, organisms<sup>46</sup>, nanocubes of metal<sup>47</sup> and so on. Excitation of the localized surface plasmon resonance of a nanostructured surface or nanoparticle lies at the heart of SERS. The ability to reliably control the surface characteristics has taken SERS from an interesting surface phenomenon to a rapidly developing analytical tool with many possible applications.

Raman can also be applied to the detection of chemical threat agents from a marked stand-off distance<sup>48</sup>. A modern Raman technique that enables recording spectra from layers several millimeters below the sample surface is spatially offset Raman spectroscopy (SORS). Spatially offsets hyperspectral stand-off Raman imaging was used for explosive detection inside containers<sup>49, 50</sup>. SORS also is a powerful new technique for the non-invasive detection and identification of packaged products and drugs<sup>51</sup>.

In this Thesis (Chapter 4) Raman hyperspectral imaging has been used in analysis of chocolate ingredients to investigate the features and distribution of these constituents in chocolate products. The details of this application will be discussed in Chapter 4 and in the associated published paper<sup>52</sup>.

### **Mid-Infrared hyperspectral imaging**

Mid-infrared (MIR, 3-20  $\mu\text{m}$ ) spectroscopy is associated with most organic and inorganic molecules absorbing MIR photons, and it provides inherent molecular selectivity. It was already applied commercially in the early 1990s<sup>53</sup>.

Waveguide-based MIR sensing systems generally comprise four major components: 1) an MIR radiation source, 2) waveguides for propagating the radiation and frequently also serving as the transducer, 3) a wavelength selection device, and 4) an MIR detector. In its most common form, an interferometer is coupled to a multichannel, liquid nitrogen-cooled mercury cadmium telluride (MCT) detector<sup>54</sup>. The detector in commercial imaging spectrometers is typically a linear array (LA) or a focal plane array (FPA).

The combination of an infrared focal-plane array detector and a step-scanning Fourier transform infrared (FT-IR) microscope has proven to be a powerful approach for obtaining spectroscopic images with unprecedented image fidelity. Today, the most popular configuration for IR chemical imaging is the Fourier transform infrared (FT-IR) imaging spectrometer, which employs multiplex detection of wavelengths via interferometry.

Today, FT-IR microscopes are designed to allow the spectra of physically small samples, or regions of small samples, to be measured as quickly and easily as possible. Using the new FT-IR instrument (HYPERION) of Bruker as an example (<https://www.bruker.com/products.html>), the recent development of FTIR spectroscopic imaging has enhanced our capability to examine, on a microscopic scale, the spatial distribution components in physical, chemical and biomedical samples. Recent activity in this emerging area has concentrated on instrument development, theoretical analyses to provide guidelines for imaging practice, novel data processing algorithms, and on the introduction of this technique to new application fields.

L. M. Kehlet et al. has developed a new hyperspectral imaging spectrometer in the mid-IR spectral region which is based on nonlinear frequency up-conversion and subsequent imaging using a standard Si-based CCD camera. A series of up-converted images are acquired with different phase match conditions for the nonlinear frequency conversion process. From this, a sequence of monochromatic images in the 3.2–3.4  $\mu\text{m}$  range are generated<sup>55</sup>.

Kevin Yeh et al. published fast infrared chemical imaging with a quantum cascade laser<sup>56</sup>. The advent of high-intensity, broadly tunable quantum cascade lasers (QCL) has now accelerated IR imaging, but using a fundamentally different type of instrument and approach, namely discrete frequency IR (DFIR) spectral imaging. These advances offer new opportunities for high throughput IR chemical imaging, especially for the measurement of cells and tissues.

In this Thesis, Infrared hyperspectral imaging has been used for the analysis of chocolate samples and its constituents. Different from Raman hyperspectral imaging, infrared hyperspectral imaging can provide the ingredients and distribution of milk chocolate constituents without the undesired effect of strong interference of fluorescent contributions. The details of this application will be discussed in Chapter 4 and published paper<sup>52</sup>.

### **Near-Infrared Hyperspectral imaging**

Near-infrared (NIR, 900–1700 nm) spectroscopy has been proposed after the NIR region was discovered in 1800, revived and developed in the early 1950s and put into practice later in the 1970s<sup>57</sup>. NIR spectroscopy is now a very prominent major analytical technology. In this instrument, the radiation from a broadband source of NIR radiation is passed through a liquid crystal tunable filter (LCTF) so that a narrow region of the NIR spectrum is isolated<sup>58</sup>. The instruments for NIR hyperspectral imaging are more versatile and rugged

than the corresponding instruments used to measure mid-IR and Raman spectra. The initial imaging systems required the samples to be stationary. Newer systems (push broom) collect images from moving samples, enabling online analysis.

The use of NIR hyperspectral imaging has been and it is still being improved extensively for the determination of quality and safety of agricultural and food products. Other fields of interest and research areas where NIR hyperspectral imaging is increasingly applied include pharmaceuticals, medical applications, archaeology and forensic.

Increasing processing capacity of industrial lines raises the demand for strict quality control and optimization of the analyzed samples and for product inspection. Rapid quality assessment for on-line analysis of some food products has been established using NIR based equipment<sup>59</sup>. The potential application of NIR hyperspectral imaging in this approach is still under investigation.

In this Thesis, in Chapter 3, the remote sensing data from AVIRIS satellite from NASA were analyzed. Hyperspectral solar reflected ground images were measured between 400 nm and 2500 at 10 nm intervals. Most of these wavelengths are located in the near infrared region (Near infrared is from about 800 nm to 2500 nm). In this case near infrared hyperspectral imaging was used for remote sensing of ground objects analysis (including minerals, airport, and lakes).

### **Secondary ion mass spectrometry (SIMS)**

Secondary ion mass spectrometry (SIMS) is a imaging technique for surface and thin-film analysis with already a long history and a mature instrumental base, dating from the early 1960s<sup>60</sup>. It is an analytical technique which detects ions ejected from a surface after the surface has been bombarded with high energy ions. Bombarding ions are referred to as primary ions or probing ions and the ions ejected from the surface are referred to as secondary ions. These secondary ions contain analytical information about the sample and they are detected and measured based on their mass/charge ratio<sup>61</sup>.

It is a fast analytical methodology and it has been used extensively to characterize a range of materials including electronics, metallic, polymers and biological samples. SIMS has been used extensively to examine a wide range of samples and it is routinely used in the micro-electronics industry for probing the distribution of dopants in silicon<sup>62</sup>. In materials science area, in 1970s, there were already many reports about application of SIMS on sputtering process of a silicon carbide surface<sup>63</sup>. Its capability has had a revolutionary improved by development of TOF-SIMS. And still at present there are many reports

published in SIMS materials surface analysis. A recent paper about surface characterization of dialyzer polymer membranes using imaging TOF-SIMS and quantitative XPS (X-ray photoelectron spectroscopy) line scans has been recently published<sup>64</sup>.

Application of SIMS to medical and biological research is a relatively new field. SIMS can be used in biometrical and tissue analysis, and many papers have been published about these applications in the two last years. For example, Louise Carlred et al. reported the spatial localization of the major component of senile plaques in Alzheimer's disease (AD), which were mapped in transgenic AD mouse brains using TOF-SIMS<sup>65</sup>. SIMS imaging has been proposed by C. Bich et al. as a tool for micrometric histology of lipids in tissues<sup>66</sup>. N Desbenoit et al. reported using TOF-SIMS imaging for localization and quantification of benzalkonium chloride in eye tissue<sup>67</sup>.

### **Electron paramagnetic resonance (EPR) spectroscopy imaging**

Electron paramagnetic resonance (EPR) spectroscopy is a technique for studying materials with unpaired electrons. EPR was first observed in Kazan State University by Soviet physicist Yevgeny Zavoisky in 1944<sup>68</sup>, and it was developed independently at the same time by Brebis Bleaney at the University of Oxford<sup>69</sup>. In comparison to nuclear magnetic resonance (NMR) spectroscopy, EPR is expected to show a higher sensitivity per unit volume due to the higher gyromagnetic ratio of electron spins due to roughly ten times smaller skin depth of EPR microwaves in comparison to NMR radio waves.

Although EPR is a technique developed many years ago, recently it has been extended to spectroscopic imaging. In chemical industry, electron paramagnetic resonance imaging has been used for real-time monitoring of Li-ion batteries<sup>70</sup>. This is an efficient way to locate 'electron'-related phenomena and opens a new area in the field of battery characterization that should enable future breakthroughs in battery research. In medical area, P. Kuppusamy's research has demonstrated that electron paramagnetic resonance imaging can perform noninvasive anatomical analysis as well as functional imaging analysis and that it provides in vivo physiological information regarding cellular metabolism in tumor and normal tissues. It is not a popular technique in environmental or food sample analysis yet.

Using EPRI to identifying defects on a CaF<sub>2</sub> surfaces due to a laser beam was reported in<sup>71</sup>. In this work, MCR-ALS has been proposed to identify the spectral signatures of all components present in the CaF<sub>2</sub> samples and locate the distribution maps of them in the acquired EPR images.

### **Fast detection imaging instruments based on other concepts and current progress of them**

Fluorescence spectral imaging was not applied in this Thesis, although it could have been a good choice as an alternative technique other than Raman or Infrared techniques in future research. Hyperspectral fluorescence imaging or fluorescence spectroscopy has not received much attention in research until recently<sup>72</sup> apart from kinetic analysis. Continuous development of advanced microscopy systems provides fluorescence spectroscopy imaging data with very high spatial-temporal resolution and multiplexing capabilities.

During the last decade, fluorescence microscopy has evolved as a fundamental tool in basic and applied biomedical research, due to its optical sensitivity and molecular specificity<sup>73</sup>. Guido Zavattini et al. have developed a hyperspectral fluorescence system for 3D *in vivo* optical imaging, which generates surface contour maps of animal tissues. This system is flexible enough to allow the testing of a wide range of illumination and detection geometries.<sup>74</sup> Wim F. J. Vermaas et al. have used hyperspectral fluorescence systems to determine localization and distribution of pigments in cyanobacterial cells *in vivo*<sup>75</sup>. Seong G. et al. presented a hyperspectral fluorescence imaging system with a fuzzy inference scheme for detecting skin tumors on poultry carcasses<sup>76</sup>. Technical developments in near-infrared fluorescence (NIRF) imaging and tomography have enabled recent applications on humans. Future NIRF imaging agents, which consist of bright fluorescent dyes conjugated to disease-targeting moieties, can be used for molecular imaging and image-guided surgery<sup>77</sup>.

Hyperspectral imaging techniques are developing fast; some new instrument can cope with the problem of long data collection time. Snapshot hyperspectral imaging is a class of hyperspectral imaging systems, which utilize a novel optical processor that provides video-rate hyperspectral data cubes<sup>78</sup>. These systems have no moving parts and do not operate by scanning in either the spatial or in spectral dimension. They are capable of recording a full three-dimensional (two spatial, one spectral) hyperspectral datacubes in each video frame, ideal for recording data on transient events, or from unstabilized platforms<sup>79</sup>. They provide new real-time hyperspectral image instruments<sup>80</sup>.

Acousto-optic tunable filter (AOTF) is a novel device for spectrometers. Their electronic tunability qualifies them with the most compelling advantages of higher wavelength scan rate over conventional spectrometers that are mechanically tuned, and the feature of large angular aperture makes the AOTF system particularly suitable in imaging

applications<sup>81</sup>. New e applications of these instruments have been reported in nerve morphometry<sup>82</sup>, tenderness assessment of meat<sup>83</sup>, skin stratum configuration, citrus fruit decay<sup>84</sup> and some other areas.

Imaging techniques are not limited to previously mentioned Raman, Infrared, near-infrared and fluorescence spectroscopies, but also to some other principles, such as those given bellow:

Functional near-Infrared Spectroscopy imaging (fNIRI or fNIRSI) is a technique involving the quantification of chromophore concentration resolved from the measurement of near infrared (NIR) light attenuation, temporal or phasic changes. It was initially introduced by F. Scholkmann, and it is based in the use of NIRS (near-infrared spectroscopy) for the purpose of functional neuroimaging. Using fNIR, brain activity can be measured through hemodynamic responses associated with neuron behavior<sup>85</sup>.

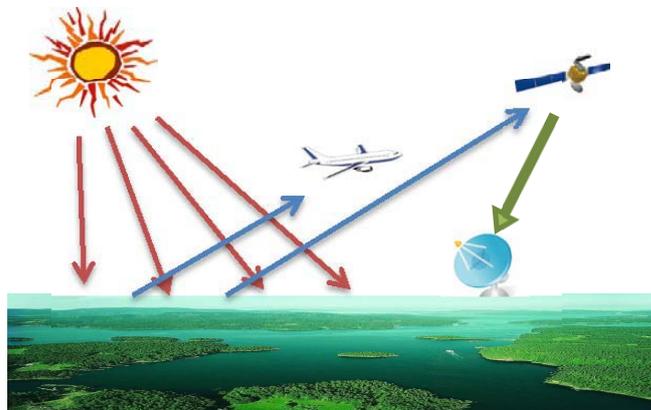
Second-order nonlinear optical imaging of chiral crystals (SONICC) is an emerging technique for crystal imaging and characterization<sup>86</sup> SONICC images can be used to determine the presence or absence of protein crystals through both manual inspection and automated analysis.

Real-Time, Subwavelength Terahertz Imaging is based on the electro-optic (EO) imaging combined with the brightness of recently developed intense THz sources. It has been reported that permits the imaging of subwavelength-size samples without compromising spatial resolution or acquisition time<sup>87</sup>.

M. A. Fadel et al. have reported the extraction of pure spectral signatures and corresponding chemical maps from electron paramagnetic resonance (EPR) imaging data sets<sup>88</sup>.

### **Remote sensing hyperspectral imaging in environment monitoring**

Remote sensing is the technique for acquisition information from an object or phenomenon without making physical contact with the object<sup>89</sup>. In the narrow sense, remote sensing is defined as the measurement of object properties on the earth's surface using data acquired from aircraft and satellites. It is necessary to rely on the propagated signals like optical, acoustical or microwave<sup>90</sup>. The acquisition of image data uses cameras and other sensors.



**Figure 2- 1** Illustration of remote sensing

Figure 2- 1 is an example for illustration of remote sensing. Reflected sunlight is the common source of radiation measured by satellite or airplane. Normally, the sensor includes photography, infrared spectroscopy or radiometers. In this Thesis, the interest is on data obtained by hyperspectral imaging spectroscopy.

Remote sensing data provides either discrete point measurements or a profile of measurements along a flight path. In this work, we are mostly interested in data acquired along a two dimensional spatial grid, which is an image. Remote sensing systems, like those installed on satellites, provide information for monitoring short-term and long-term changes on earth, and possible impacts from human activities on a particular area<sup>91</sup>. Sensor data provide instrument views of the physical objects by recoding electromagnetic radiation emitted or reflected from them and surrounding landscape. Remote sensing images can provide information about position, size, and interrelationships between objects. By their nature, spectral signature profiles can be recognized and used for description of the objects.

Hyperspectral remote sensing is a useful technology in many scientific and technological fields and it has been proposed also in the environment area from years ago. Since 1960s, multispectral imagery has been used as data source for environment observational remote sensing from airborne and satellite systems<sup>92</sup>. Initially, only three to six spectral bands were used in a single observation from the visible and near-infrared regions of the spectrum. Recently, advances in sensor technology have overcome this limitation of the data collecting systems, with the development of new hyperspectral sensor technologies. It is now possible to apply this technique to acquire land surface information, using high resolution spectra. Remote sensing can play a unique and essential role because of its ability to acquire synoptic information at different time and space scales.

The first Landsat Multispectral Scanner System (MSS) was launched in 1972<sup>93</sup>. The concept of imaging spectroscopy originated in the 1980's at NASA's Jet Propulsion Laboratory, when a revolution in remote sensing began with the introduction of the Airborne Visible Infra-Red Imaging Spectrometer. In this Thesis, this type of remote sensing is an important topic and it is the object for evaluation of the performance of the MCR-ALS method on hyperspectral imaging (Chapter3 and papers<sup>94,95</sup>).

At present, remote sensing spectral information can be provided by several hyperspectral sensors such as Hyperion, Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Compact Airborne Spectrographic Imager (CASI), Airborne Imaging Spectrometer for Applications (AISA) and HyMap (from HyVista, Castle Hill, Australia).

**Table 2-1** Source of remote sensing hyperspectral imaging data and their reference webpage.

Source of data	Reference webpage
Specim, Spectral Imaging, Ltd (Finland)	<a href="http://www.specim.fi/products/aisa-airborne-hyperspectral-systems/aisa-series.html">http://www.specim.fi/products/aisa-airborne-hyperspectral-systems/aisa-series.html</a>
National Aeronautics and Space Administration (NASA, USA)	<a href="http://aviris.jpl.nasa.gov/">http://aviris.jpl.nasa.gov/</a>
ITRES (Canada)	<a href="http://www.itres.com/">http://www.itres.com/</a>
National Aeronautics and Space Administration (NASA, USA)	<a href="http://eo1.gsfc.nasa.gov/Technology/Hyperion.html">http://eo1.gsfc.nasa.gov/Technology/Hyperion.html</a>
HyVista-Integrated Spectronics Pty Ltd (Australia)	<a href="http://www.hyvista.com/">http://www.hyvista.com/</a>

Spatial resolution used in remote sensing is a critical aspect because it determines the level of accuracy in the classification of objects on the ground by using a minimum amount of data. Low spatial resolution can hardly discriminate objects on the ground resulting in lower classification accuracy.

In the past half century, hyperspectral imaging has been successfully applied on remote sensing for environmental assessment and monitoring (urban growth<sup>96</sup>, hazardous waste<sup>97</sup>), agriculture (crop condition<sup>98</sup>, yield prediction<sup>99</sup>, soil erosion<sup>100</sup>), nonrenewable resource exploration (minerals<sup>101</sup>, oil<sup>102</sup>, natural gas<sup>103</sup>), renewable natural resources (wetlands<sup>104</sup>, soils<sup>105</sup>, forests<sup>105</sup>, oceans<sup>106</sup>), meteorology(atmosphere dynamics<sup>107</sup>, weather prediction<sup>108</sup>), mapping<sup>105</sup> (topography, land use, civil engineering), and military surveillance<sup>109</sup>. Traditional approaches for remote sensing imaging were focusing on the discovery of images shape. The primary interest was in the spatial relationships among features of the ground, which followed naturally from the similarity between the aerial or satellite images and the cartographic map. The common goal was the creation of a map.

As mentioned above, hyperspectral remote sensors acquire images across narrow contiguous spectral bands, mainly including the visible, near-infrared and mid-infrared portions of the electromagnetic spectrum<sup>110</sup>. Typically, hyperspectral sensors measure the reflected spectrum at wavelengths between 350 and 2500 nm using 150–300 contiguous bands with 5 to 10nm bandwidths<sup>111</sup>. Recent scanners support even higher spectral resolutions in the sub-nanometer range. Although hyperspectral remote sensing has a relatively short history compared with other types of remote sensing such as aerial photographs, hyperspectral sensors have been very effective for mapping the spatial extent of native and non-native species across all types of communities and ecosystems<sup>112</sup>. Recent advances in materials and optics have allowed the development of smaller, more stable, accurately calibrated imaging spectrometers that can quantify properties of the investigated land surface using absorbing and scattering characteristics.. Hyperspectral imaging is becoming more widely available from government and commercial sources; thus, it is increasingly feasible to use data from imaging spectroscopy for environmental research purposes<sup>111</sup>.

Previous studies have shown that from remote measurements, direct identification of surface materials can be obtained. Remote sensing instruments can provide detailed information on the mineralogy and geochemistry of the rock types comprising the Earth's surface. Based on that, remote sensing has been used for decades to map rocks, mineral assemblages and weathering characteristics<sup>8, 113, 114</sup>.

In environmental monitoring analysis, for example, for air protection, people require the continuous monitoring of the emissions produced by coal or oil power plants, municipal and hazardous waste incinerators cement plants, as well as many other industrial source pollutions. These monitoring studies are usually performed using extractive sampling systems coupled with spectroscopy techniques. Environmental monitoring describes the processes, and activities taken place in the environment. Hyperspectral imaging can be used for environment impact assessment as well as in many other circumstances of human activities assessment.

The application of hyperspectral imaging to natural resources, vegetation and surface water is an important field in environmental monitoring studies<sup>115</sup>. The spectral signatures of vegetation are influenced by the presence of pigments (mainly chlorophyll-a, chlorophyll-b, xanthophyll and carotenoids), and by the physical structure and water content of leaves<sup>116</sup>. Hyperspectral analysis of surface water bodies (lakes, rivers and coastal waters with riparian zones) provides information on key water quality parameters and on its trophic state.

The enhanced capability of hyperspectral technologies in environmental monitoring using remote sensing allows managers to take correct decisions with the necessary detail and in

an efficient time frame. With the fast development of space borne and airborne sensors, better hyperspectral data are more frequently and conveniently received. As a result, data storage and proper management, and effective information extraction are becoming key challenges to hyperspectral remote sensing science and technology. Especially, automatic or semi-automatic information extraction is expected to be growing in hyperspectral investigations and applications.

Some recent studies have used hyperspectral remote sensing to assess current spatial distribution and future dispersion of invasive plants at local, regional and global scales. With a hyperspectral sensor, many narrow bands can capture a range of absorption features including leaf or biochemical constituents such as carotenoids, water, nitrogen, cellulose and lignin<sup>117</sup>. As leaves and plant species vary in the concentration of their biochemical constituents, the reflectance spectra vary as well<sup>118</sup>. Kate S. He et al. have reported hyperspectral remote sensing results for detecting, mapping and predicting the spatial spread of invasive species<sup>119</sup>. Herold, M. evaluated how spectral resolution of high-spatial resolution optical remote sensing data influences detailed mapping of urban land cover. A comprehensive regional spectral library and low altitude data from the AVIRIS were used to characterize the spectral properties of urban land cover<sup>120</sup>.

At the Laboratory of Hydraulics of DICEA-Sapienza University of Rome, an effective methodology for hyperspectral monitoring has been developed<sup>115</sup>. It is based on the use of two innovative experimental devices for acquiring hyperspectral images, one based on the use of tunable interference filters, the other on the use of spectrometers<sup>121, 122</sup>. In L. G. Olmanson's work, airborne hyperspectral remote sensing was applied to assess the spatial distribution of water quality features in large rivers, the Mississippi River and its tributaries in Minnesota. All these results show that hyperspectral imaging can be used to distinguish and map key variables under complex conditions, particularly to separate and map inorganic suspended sediments independently of chlorophyll levels<sup>123</sup>.

F Salem reported a case study about a prototype of oil spill leaks in Patuxent River in Maryland, and the associated image analysis for detecting oil spills using hyperspectral imagery and the effect on soil, water, wetland, a vegetation contaminated by oil spill<sup>123</sup>. F Salem et al. in another work, reported hyperspectral allowed precise identification of grass stress and soil (oil contaminated wetland) damaged by oil polluted water<sup>124</sup>.

Space borne and airborne images, are the main source for getting real-time data. In the event of an oil spill, this information can be retrieved in short time to help authorities to plan the quickest route to the spill and formulate an effective environmental protection plan. That could be a way to reduce damages.

In chapter 3 of this Thesis, remote sensing hyperspectral imaging (AVIRIS data sets from NASA) data has been analyzed and different physical objects, such as buildings, vegetation, soil, water, etc. were properly resolved.

### **Hyperspectral imaging in food analysis**

More and more food species are supplied by food industry nowadays. Food can be a very complex product for consumption. It can have a plant or an animal origin, and it consists of carbohydrates, fats, proteins, vitamins or minerals. Assessment of food quality parameters like meat, vegetable has always been a big concern in all processes of the food industry because consumers are always demanding high quality of food raw materials and products. Analysis of food products in industry is a complex task.

Food quality is an important requirement, which can usually be defined in terms of consumer appreciation of texture and flavor, and of food safety, which includes health implications from both compositional and microbiological properties<sup>125</sup>. Food quality assurance plays an important role, because it is directly related to the consumer health. In the food industry, the producers need to be able to guarantee the quality of the product they produce<sup>126</sup>. When consumers choose food, they always take care about its visual appearance, textural patterns, geometrical features and color. Food quality assessment can use methods like sensory analysis, nutrition analysis, bacteriological and chemical analysis. These methods give judgments based on taste, healthiness, convenience, appropriate packaging, environmental friendliness, and so on<sup>127</sup>. For example, people can use some chemical properties for evaluation, such as water holding capacity, freshness, content in proteins, and vitamins<sup>128</sup>.

The other important aspect to consider especially concerned by producers, distributors and consumers is product authenticity and authentication. Food adulteration has been practiced since thousands years ago, but it has become more sophisticated in the recent years. Food ingredients most likely to be targets for adulteration include those which are of expensive and/or undergo a number of processing steps before they can appear in the market. All these frequent situations require the use of appropriate methods and technologies for food analysis.

To analyze a particular food sample, there are different possibilities, using different techniques, depending on the information that the analyzer wishes to obtain. Conventional methods of food analysis rely on a subjective visual judgment and on the use

of laboratory chemical tests. In addition, traditional grading routines and quality evaluation methods are time consuming, destructive and associated with inconsistency and variability due to human subjectivity<sup>129</sup>. Therefore, evaluation of food quality in recent food processing lines requires the use of new analytical instrumentation that is fast, specific, robust, and durable enough for the harsh environments in food processing plants and overcome all disadvantages of traditional methodology<sup>130</sup>. This instrumentation also has to be cost effective to reflect their competitiveness in food and agriculture markets.

Food industry is currently undergoing dramatic changes in applying the most advanced technological innovations, and it has gained acceptance and respect in handling, quality control and assurance, packaging and distribution. Many different methods for measuring food quality are available, which are based on different principles, procedures or instruments. Over the past few years, a number of methods have been developed to measure food quality.

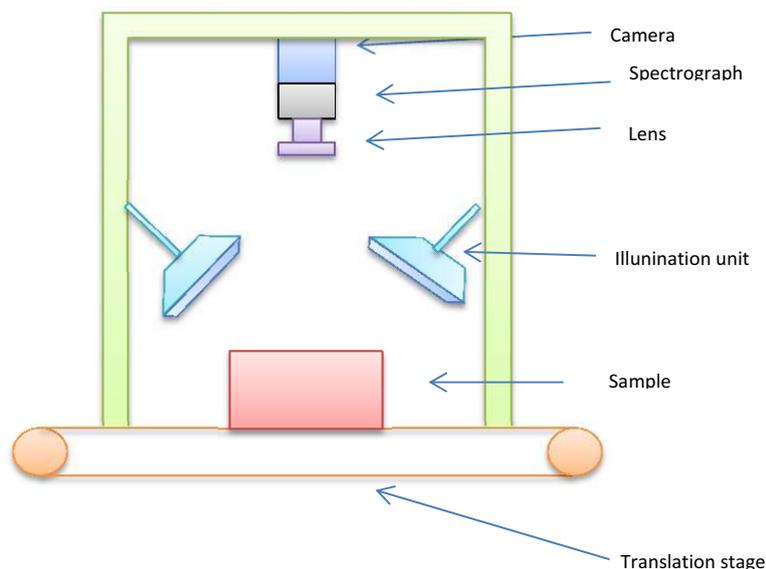
Analytical techniques most commonly used for food analysis are the following:

- **Spectroscopic techniques (IR, NIR, Raman, UV-VIS, fluorescence):** They study the interaction between radiation and molecules. Molecular vibrations give spectral signatures, which characterize food composition and may be considered as fingerprints of food samples. They can be used in image analysis.
- **Chromatography (GC, LC):** They are applied to the detection and quantitation of particular analytes of interest in food samples. It is based on the interaction of analytes (in mixtures) with a stationary phase when they travel with different speeds in a mobile phase. Chromatographic methods can be used in conjunction with different detectors. Usually, UV absorption molecular spectroscopy (via a diode array detector, DAD) or mass spectroscopy (MS), can be used as detectors.
- **Mass spectroscopy (MS, MS/MS):** Mass spectroscopy can be used not only as a detector for Chromatography, but also it may be used efficiently for direct detection and quantitation of elements and other constituents of food samples and for structure analysis.
- **Nuclear magnetic resonance (NMR):** It can be applied to a wide range of liquid and solid samples to elucidate the structure of organic molecules and constituents, and to perform their semiquantitation in food samples. It can be used in imaging analysis of food samples.
- **Thermal Analysis:** foods can be submitted to variations in temperature during production, transport, storage, preparation or consumption. Temperature changes cause alteration of the physical and chemical properties of the food products. Thermal analysis is the measurement of changes with temperature in mass, density, heat capacity and other physical properties of foods.

- **Rheology analysis:** measures flow or changes in shape of food materials when a force is applied.

In the food industry, quality evaluation and control are still performed in many complicated ways. In the traditional way, the analysis work is tedious, laborious, costly and time consuming. Sometimes human errors and inconsistency may happen. This is the reason why there is a great interest to work on hyperspectral imaging systems for evaluation of food quality. Nowadays, the analysis of agricultural products is also based on automation or semi-automation of measurements, and therefore there is also a high demand for application of hyperspectral imaging.

Hyperspectral imaging is a fast growing area in food analysis, which expands and improves the capabilities of traditional spectral analysis. It has the flexibility to tackle all type of samples, whatever their size. For example, it is able to deal with microscopic particles, with a single kernel or with the whole sample (for example, with a fruit or with a piece of meat) in order to study the distribution of a wide range of chemical compounds. Research on application of hyperspectral imaging in food analysis was already starting in 1990s<sup>2</sup>. With the fast development of imaging and computer technologies, hardware of hyperspectral imaging applied to food industry updates at a rather rapid speed, and the cost of hardware decreases correspondingly.



**Figure 2- 2** Scheme of a hyperspectral imaging instrument

As it can be seen from Figure 2- 2, hyperspectral imaging system for food analysis normally consists of the following parts: a light dispersion device and an imaging unite to function as an eye; a decision-making component such as computer and software. The light provided by the light source interacts with the food samples, and its environment. Light interaction with physical and chemical features of the sample will be dispersed and projected onto the detector.

The illumination unit provides light to the sample. This part makes a significant impact on the performance and reliability of the system. Tungsten halogen lamps, which add halogen elements (F, Cl, Br, I) into quartz tubes, are widely used as illuminators in hyperspectral imaging systems<sup>131, 132</sup>. Tungsten-halogen lamps have higher luminous efficiency compared to normal light bulbs. Furthermore, halogen frequency can guarantee sustained glow and long lifetime, which is four times more than normal light bulbs. In addition, other kinds of durable lamps such as HgAr lamps or LED source have been applied as well<sup>133</sup>.

The sensor is the most important part of the hyperspectral imaging systems, which helps in generating a spectrum for each point on the scanned line. With the development of sensor technology, many different types of sensors have been developed, which can acquire images in different spatial resolution, temporal resolution and spectral resolution.

CCD (Charged Couple Device) or CMOS (Complementary metal–oxide–semiconductor) are two widely used image sensors that have been developed rapidly in recent years. The role of CMOS Image Sensors since their birth around the 1960s has been changing a lot. Unlike the past, current CMOS Image Sensors are becoming competitive with regard to CCD technology. They offer many advantages with respect to CCD, such as lower power consumption, lower voltage operation, on-chip functionality and lower cost. Nevertheless, they are still too noisy and less sensitive than CCDs<sup>134</sup>.

The computers' memory, hard disk capacity as well as the processor makes data storage, and data processing enhanced a lot. In the near future, hyperspectral hardware, in particular, for optical devices, will be more intelligent due to the increasing development of computer technologies, which facilitate operational control automatic and intelligent.

Sometimes, for online monitoring, the light is transmitted through optical fibers towards a line light reflector. Nowadays, many kinds of fiber optic sensors are developed such as Bragg grating optical sensors<sup>135</sup>, fiber optic Febry–Perot temperature sensors<sup>136</sup> and intensity modulated sensors.

Discrimination and classification for constituents detection have been the key quality control stages in food industry<sup>2</sup>. Hyperspectral imaging has already shown a great

capability to quantify and control quality parameters in food industry with high precision<sup>137</sup>.

In fruits and vegetables, hyperspectral imaging has been successfully applied to the detection of contaminants, bruises and chilling injury in apples<sup>138</sup>, to the estimation of quality parameters in strawberry<sup>139</sup> and cucumbers<sup>140</sup>. The results of these previous works in fruits and vegetables emphasized the fact that hyperspectral imaging techniques have a notable performance for estimating physical and chemical attributes such as firmness, presence of bruises, dry matter, soluble solids content, pH, and sugar contents in addition to displaying their spatial distribution<sup>141</sup>. This enables early sorting of products and thereby they improved significantly quality management operations.

Meat tenderness is the primary determinant for consumer satisfaction which affects his decision to purchase certain sort of meat. Hyperspectral imaging was applied to meat products such as decal and tumor<sup>142</sup> detection in chicken carcass<sup>143</sup>, pork quality classification<sup>144</sup>, tenderness assessment of beef<sup>145</sup>, fish quality evaluation<sup>3</sup>, lamb muscles<sup>146</sup>. Both reflectance and scattering modes have been used. Researchers should be careful when using scattering mode. Scatter effects in the spectra resulting from physical sample variations may pose more or less severe analytical problems<sup>147</sup>.

There are few publications in utilizing hyperspectral imaging systems for quality evaluation of dairy products. Commercially produced cheeses containing varying amounts of protein, fat, and carbohydrate have been tested using a hyperspectral imaging system<sup>148</sup>. Similarly, Gowen et al. acquired a hyperspectral images of high and low fat cheese slices using a pushbroom hyperspectral imaging system<sup>148</sup>.

The various applications outlined above have shown the benefits of hyperspectral imaging technique for sample characterization and chemical species distribution. Hyperspectral imaging has proven to be a very valuable tool for a wide range of applications in food analysis. It provides attractive analytical solutions to meeting consumer demands for food quality and safety.

Right now, in most of the cases, hyperspectral imaging systems are still not directly implemented in online systems for quality evaluation because of the long time needed for image acquisition and subsequent data treatment. Anyway, hyperspectral imaging technologies can be a very useful as a research tool for determining key wavelengths with the help of chemometric methods. When only a reduced number of key wavelengths are used, the technology can be implemented in a real-time multispectral imaging system. These optimum wavelengths should not only contain the physical/chemical information, but also maintain the successive decimation and classification efficiency.

Chapter 4 of this Thesis shows the application of Raman and Infrared hyperspectral imaging to the analysis of the constituents of commercial chocolate samples.

## 2.2 Chemometric methods

This part introduces the chemometric methods used in this Thesis and their application to the analysis of hyperspectral images.

Chemometrics was first introduced in early 1970s as computers became increasingly used in analytical chemistry experiments and other scientific fields<sup>149</sup>. The international chemometrics society was established by Svante Wold and Bruce R. Kowalski in 1974<sup>149, 150</sup>. Since then, chemometrics has been developing steadily, and now it is widely spread to different fields of chemistry, especially analytical chemistry. According to the International Chemometrics Society, chemometrics can be defined as the chemical discipline that uses the theory and methods from mathematics, statistics, computer science and other related disciplines to optimize the procedure of chemical measurement, and to extract as much as possible chemical information from chemical data<sup>151</sup>.

Chemometrics is used for the treatment of chemical data, for the evaluation of instrumental signals, for the extraction of useful information, and for decision making in analytical chemistry<sup>152</sup>. Well-known approaches for multivariate calibration, chemical resolution, and pattern recognition have been developed in analytical studies<sup>153</sup>. Chemometrics now has two aspects, 1) the development of new theoretical principles and algorithms for manipulating chemical data, and 2) the applications of the chemometrics techniques to different disciplines of chemistry such as environmental chemistry, food chemistry, agricultural, medicinal, or chemical engineering, and more recently to the analytical omics (genomics, transcriptomics, metabonomics, proteomics) field etc.

During years, chemical laboratories introduced many new instrumental analytical techniques in all kinds of fields, in particular in many instrumental methods like those based on spectroscopy (UV-VIS, IR, NIR, fluorescence, etc.) and those called hyphenated, such as High Performance Liquid Chromatography with a Diode Array Detector (HPLC-DAD)<sup>154</sup>, Gas Chromatography combined with a Mass Spectroscopic detector(GC-MS)<sup>155</sup>, Gas Chromatography with an Infrared spectroscopic detector(GC-IR)<sup>156</sup>, High-Performance Liquid Chromatography with a Mass Spectroscopic detector(HPLC-MS)<sup>157</sup>, or Capillary Electrophoresis with a Diode Array Detector(CE-DAD)<sup>158</sup> among other. These hyphenated instruments can give more precise, accurate, sensitive results and allow for the qualitative and quantitative determination of multiple compounds even at trace levels. All the data produced by these instruments, like spectra, chromatograms, voltammograms, kinetic curves, titration curves, and other type of responses can be stored as data vectors or

matrices<sup>159</sup>. Univariate deterministic methods cannot solve complex chemical and physical systems that appear in nature or in the laboratory. Mathematical tools, such as classification models, calibration models, or regression models can be applied to extract the information that analytical signals contain and solve the problems being investigated. Alternatively, multivariate data analysis methods can be chosen to solve these problems, and in particular, chemometric methods should be considered for this purpose. Chemometric methods allow for the analysis of high dimensional complex datasets obtained from modern analytical instruments.

The use of chemometric tools in hyperspectral image analysis is crucial to take advantage of the full experimental measurement. Hyperspectral images provide a tremendous amount of information stored in large data sets. At each pixel of the image, there is one spectrum at multiple wavelengths<sup>160</sup>. Hyperspectral images usually include redundant information. Like for other spectroscopic methods, hyperspectral data also suffer from the problem of multi-collinearity. The characterization of multicomponent systems by hyperspectral imaging requires to cope with big datasets obtained in experimental analysis<sup>161</sup>. A typical hyperspectral imaging includes thousands of spectra, and each spectrum can contain a large number of spectral bands. This information can vary due to the composition, structure, position and orientation of the species. The huge amount of data from the two directions of the sample surface and from the spectral direction poses considerable computational challenges. To fully use the information of hyperspectral datasets, different data analysis methods such as classification, resolution or segmentation are needed.

In the late 1980s, chemometrics and other communities began to analyze and use hyperspectral images on a pixel basis, taking benefit of all the work already developed in conventional spectrometric data, and giving rise to the concept of Multivariate Image Analysis (MIA)<sup>34</sup>. In this approach, the image cube is unfolded into a data matrix on which multivariate exploration, regression or resolution techniques are then applied. By using these methods to hyperspectral images, the following aspects should be considered:

- Hyperspectral image compression. It is usually necessary to reduce the memory requirements by removing redundant information present in the images.
- Image analysis, including classification and resolution.
- Speeding up the process of collection of hyperspectral imaging data.
- Increasing the resolution of the hyperspectral imaging.

This thesis is mainly focused on the resolution of the pure spectra (signatures) and distribution map of the constituents present in hyperspectral images. In addition, some pretreatment methods are discussed to improve the results of the analysis. They are introduced first.

### **Data pretreatment chemometric methods**

Data pretreatment methods usually try to reduce large baseline variations, physical effects (like scattering), detector drifts and non-linearities, dimensionality, collinearity and noise levels from observed spectra<sup>162</sup>. Usually spectra obtained by hyperspectral imaging have the problem of noisy signatures, baseline drift, and surface scattering. Moreover, sometimes they have the problem of having different constituents at different scales. Many reasons can cause these problems. Different spectroscopic techniques are influenced by different factors, for example, light scattering in NIR, or fluorescence background in Raman spectroscopy, when samples are irradiated. Correction of optical errors may be induced by lenses and atmospheric disturbances. Scattering effects are often encountered when measuring diffuse reflectance objects in hyperspectral imaging.

Spectral preprocessing techniques should be considered a key step for the successful analysis of hyperspectral data. Hyperspectral imaging preprocessing improves the visual quality of images and makes the image analysis steps more reliable by reducing noise, and easing the edge detection. To deal with the problems above mentioned, different methods have been proposed for different aims.

### **Dead pixels and spikes**

In hyperspectral images, the presences of dead/bad pixels having zero or maximum signal values, and/or of non-informative background, can be caused by several sources. Most of the measuring systems are based on diode array detectors or tunable filters. The dysfunction of one of the diodes in the detector array may generate dead pixels (missing or zero values), unexpected spectral readings (extreme values) or spiked points in one specific wavelength. Dead pixels can distort multivariate models. Many of the routines for multivariate data analysis (e.g. PCA) can handle a limited amount of missing values. Once the dead pixels have been located, the best choice is replacing them by interpolation with neighbor pixels, due to their evident connection in structure and correlation<sup>163</sup>.

Spikes can be defined as a sudden and sharp signal rise followed by a sharp signal decline in the spectrum intensity<sup>164</sup>. Spikes can appear due to an abnormal behavior of the

detector, imperfections of electronic circuits or environmental conditions, like cosmic ray events when instruments like Charge-Coupled Device detectors (CCDs) are used for Raman imaging<sup>165</sup>. Approaches based on the derivatives of the signal, median and Median-Modified Wiener Filters (MF and MMWF, respectively)<sup>166</sup>, or Wavelet Transforms(WT)<sup>165</sup> have been proposed to remove spikes.

The geometry of the samples in the acquisition of the images plays an essential role in the selection of the Regions of Interest (ROI). Hyperspectral techniques usually acquire square images. If the sample does not cover all the scanned area, the area left outside the sample must be eliminated, since this area is usually composed by highly noisy spectra. By the selection of ROI, only the desired area will be analyzed. Selecting a threshold value obtained from the scores of a previous PCA model, background removal can be performed as another common practice.

### **Baseline correction**

Pre-treatments that can correct spectra for offset and changes in the baseline include de-trending, derivatives and Asymmetric Least Square baseline correction (AsLS).

De-trending: The trend of the baseline is estimated by least squares, adjusting each spectrum with a first or second order polynomial and then subtracting this adjusted spectrum from the original to obtain the corrected spectrum. Higher order polynomials can be used, but then there is the associated risk of removing relevant information.

Derivatives: First order derivatives may reduce spectral offset; additionally, the second derivative may remove a uniform slope from the baseline.

Asymmetric least squares (AsLS): This method calculates complex baseline shapes by adjusting two parameters: the asymmetry parameter which is related to the position of the baseline, and the smoothness parameter related to the flexibility in the shape of the baseline.

The AsLS method has been proposed by H. C. Eilers<sup>167</sup>. It reduces the possible baseline and background contributions, especially when it changes over time. It estimates a baseline/background contribution by minimizing the penalized least squares function based on the Whittaker smoother<sup>168</sup>.

Assume that  $m$  spectra  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$  are in column vectors, and that  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m$  are their corresponding baselines to be estimated. In order to fit a smooth series  $\mathbf{z}$  to  $\mathbf{s}$  the following optimization is performed :

$$\mathbf{z} = \arg \min_{\mathbf{z}} \{ \sum_i w_i (\mathbf{s}_i - \mathbf{z}_i)^2 + \mu \sum_i (\Delta^2 \mathbf{z}_i)^2 \} \quad \text{Equation 2- 1}$$

where  $\mathbf{s}_i$  is the considered spectrum,

$$\Delta^2 \mathbf{z}_i = (\mathbf{z}_i - \mathbf{z}_{i-1}) - (\mathbf{z}_{i-1} - \mathbf{z}_{i-2}) = \mathbf{z}_i - 2\mathbf{z}_{i-1} + \mathbf{z}_{i-2}, \quad \text{Equation 2- 2}$$

Weights  $w_i$  are chosen asymmetrically:  $w_i = p$  if  $\mathbf{s}_i > \mathbf{z}_i$ , otherwise  $w_i = 1 - p$ . The values of  $w$  should be set to 1 where  $\mathbf{s}_i$  is observed or allowed to influence  $\mathbf{z}$ . In all other places, the values of  $w$  were set to 0,  $\mathbf{z}$  values are smoothly interpolated).  $\mu$  is a regularization parameter, which sets the weight of the second term and acts as roughness penalty, the larger  $\mu$ , the smoother  $\mathbf{z}$  will be. Experience shows that starting from  $w \equiv 1$  and iterating between the two computations of  $w_i = p$  if  $\mathbf{s}_i > \mathbf{z}_i$ , otherwise  $w_i = 1 - p$ , leads to a solution in about 10 iterations. With  $p$  near to zero and a rather large  $\mu$ ,  $\mathbf{z}$  will tend to follow the valleys of  $\mathbf{s}$ . In the work of analysis chocolate using Raman and Infrared hyperspectral imaging in Chapter 4, the parameters  $p$  and  $\mu$  are set to 0.001 and  $1 \times 10^7$  respectively.

### Scattering effects correction

When analyzing solid and heterogeneous samples, they may exhibit sample-to-sample variability using spectroscopic instrumentation. These variations come from the changes in optical path length due to physical differences between samples, due to particle size and shape, sample packing, and sample surface. It may result in the multiplicative light scattering effect masking the spectral variations relating to the differences between the chemical compounds within a sample. Multiplicative scattering effects present in spectra are commonly corrected by the following methods:

**Scaling:** Multiplicative effects in spectra are suppressed by dividing each spectrum by some attribute such as its mean, median, minimum, maximum or norm.

**Multiplicative Scatter Correction (MSC):** The best line fitting each spectrum to a reference spectrum is calculated, and each value in the spectrum is then adjusted by subtracting the offset and dividing by the slope of the line of best fit; commonly, the reference spectrum is the mean spectrum of the data set.

MSC is a general signal treatment method which has been applied for eliminating the effect of scattering produced by physical effects, and also as a preprocessing to linearize spectral data. Multivariate linear statistical regression methods can give better predictions when applied to MSC corrected data. It corrects for the scatter level of a group of sample spectra to the level of a preselected spectrum, which usually is the sample average spectrum<sup>169</sup>. Each spectrum is fitted to this average spectrum by least squares:

$$\mathbf{x}_i = \mathbf{a}_i + \mathbf{b}_i \bar{\mathbf{x}}_j + \mathbf{e}_i \quad \text{Equation 2- 3}$$

where  $\mathbf{x}_i$  is an individual spectrum,  $\bar{\mathbf{x}}_j$  the mean spectrum of the group, and  $\mathbf{e}_i$  the residual spectrum, which ideally represents the chemical information of the spectrum related to the specific concentration changes of the different constituents of this particular sample. The corrected spectrum  $\mathbf{x}_{i, \text{MSC}}$  is calculated using the fitted constants  $\mathbf{a}_i$  and  $\mathbf{b}_i$ :

$$\mathbf{x}_{i, \text{MSC}} = (\mathbf{x}_i - \mathbf{a}_i) / \mathbf{b}_i \quad \text{Equation 2- 4}$$

In this thesis, in the Chapter 4 about the analysis of chocolate samples using Raman and Infrared hyperspectral imaging, the preselected spectrum for MSC pretreatment was the average of the raw spectra.

Baseline drift always blurs or even swamps signals and ruin analytical results. Correct baseline drift will help performing further data analysis. A typical baseline correction based on linear models or on more complex mathematical functions, can be performed. Other strategies, such as working with derivative spectra instead of with the raw spectra, can help to eliminate instrumental variations such as offsets or linear baselines that are unrelated to the chemical composition of the image.

Extended Multiplicative Signal Correction (EMSC)<sup>170</sup> is an extension of MSC that can be used to remove additional unwanted interferences in the signal, correct for the square of the reference spectrum or first or second order polynomial of the spectra. Therefore this method cannot only correct the spectra for multiplicative effects but can also simultaneously correct the baseline of the spectra and suppress the effect of known interfering compounds. This method outperforms MSC in some cases<sup>171</sup>.

Standard Normal Variate (SNV): Offset and multiplicative effects are removed from each individual spectrum by mean centering and scaling by its own standard deviation. This method has been applied to hyperspectral imaging, and has been demonstrated to improve performance of PLS models developed by using raw spectra and other pretreatments<sup>172</sup>.

Robust Normal Variate (RNV): Selection of a predefined percentile, and calculation of the standard deviation of the values lower than the selected percentile have been proposed as improvements of the SNV method to increase its robustness against extreme values and to solve the closure problem<sup>173</sup>. Selection of the proper percentile may be decided by testing different values.

Noise to signal ratio may be increased by multiplicative correction, especially when the spectrum is divided by a very small attribute. This can overcome benefits gained by reduction or removal of the multiplicative effects. On the other hand, model performance

may be unimproved if the component being modeled (response variable) is related to the scattering of the sample.

### Smoothing

In recent years, many possible methods have been designed to de-noise spectroscopic datasets, and they can be also applied to hyperspectral images. Some of these methods rely on smoothing procedures, whether by averaging or by fitting neighboring spectral channels to a polynomial function, or also based on mathematical signal filtering procedures.

Savitzky-Golay (SG) smoothing<sup>174, 175</sup> is a very popular technique, which uses convolution arrays derived from the polynomial coefficients of least squares fit formulas<sup>176</sup>. The convolution can be understood as a weighted moving average filter with weights from a polynomial of a certain degree. Instead of simply using the averaging technique, the Savitzky-Golay filter employs the regression fitting capacity to improve the smoothing results. It has been used in this work to remove large noise contributions of Raman signals.

Selection of the window size of the Savitzky-Golay filter is important. When the window size increases, the smoothing effect becomes more significant, but the signal resolution decreases. When Savitzky-Golay smoothing is applied, the span of the moving average and the smoothing polynomial degree has to be set.

### Kramers-Kronig transformation

Infrared active photons, both longitudinal and transverse, are mostly determined from reflection spectra in the infrared spectral range. These spectra do not reveal the photon frequencies directly. Kramers-Kronig spectra pretreatment is a powerful tool enabling a calculation of absorption spectra in case only reflectance spectra can be measured.

The complex reflectance spectrum is mathematically decomposed into two separate spectra- the extinction coefficient and refractive index spectrum. The extinction coefficient spectrum can be then used to calculate the absorption spectrum.

The real ( $n$  = refractive index) and imaginary ( $k$  = extinction) parts of the complex index of refraction are calculated from the reflectance spectrum using the following formulas:

$$n(\nu) = \frac{1-R(\nu)}{1+R(\nu)-2\sqrt{R(\nu)}\cos(\theta(\nu))} \quad \text{Equation 2- 5}$$

$$k(\nu) = \frac{-2\sqrt{R(\nu)}\cos(\theta(\nu))}{1+R(\nu)-2\sqrt{R(\nu)}\cos(\theta(\nu))} \quad \text{Equation 2- 6}$$

where  $R$  is the reflectance,  $n$  is wavenumber,  $\theta$  is phase shift angle of the sample.

For a given wavenumber, the phase shift is calculated using the equation:

$$\theta(v_m) = \frac{2v_m}{\pi} \int_0^{\infty} \frac{\ln\sqrt{R(v)}dv}{v^2 - v_m^2} \quad \text{Equation 2- 7}$$

The Kramers-Kronig transform algorithm assumes that the reflectance spectra are measured at incidence angles close to zero.

In this thesis, in the work about Raman hyperspectral imaging analysis of chocolate constituents, Multiplicative Scatter Correction (MSC)<sup>177, 178</sup>, Asymmetric Least Squares (ASLS)<sup>179, 180</sup> and Savitzky-Golay smoothing<sup>176</sup> were used as signal pretreatment methods to eliminate undesired light scattering, instrumental and background effects. In the third Chapter, infrared hyperspectral imaging used the Kramers-Kronig transform.

## Chemometric data analysis methods

### Bilinear method

Depending on the structure presenting data to analyze, it may apply different types of models and chemometric methods. Bilinear chemometric methods can be used in hyperspectral imaging data analysis. To describe the bilinear decomposition of a two way dataset, the following expression is used:

$$x_{ij} = \sum_{n=1}^N c_{in}s_{nj} + e_{ij} \quad \text{Equation 2- 8}$$

and the reproduction of data can be expressed in way of the matrix:

$$\mathbf{X} = \mathbf{CS}^T + \mathbf{E} \quad \text{Equation 2- 9}$$

where  $x_{ij}$  represents the  $ij^{\text{th}}$  element in the two-way dataset  $\mathbf{X}$  ( $i=1, \dots, I$ , and  $j=1, \dots, J$ ).  $c_{in}$ ,  $s_{nj}$ , and  $e_{ij}$  are the elements of  $\mathbf{C}$ ,  $\mathbf{S}^T$  and  $\mathbf{E}$ .  $N$  ( $n=1, \dots, N$ ) is the number of considered components, (chemical species contributing to the signal);  $I$  is the number of rows (for spectroscopic data,  $I$  is the number of samples spectra/profiles) in the data matrix  $\mathbf{X}$ ; and  $J$  is the number of columns (for spectroscopic data,  $J$  is the number of bands) in the data matrix  $\mathbf{X}$ .  $\mathbf{C}$  is the matrix describing how the contributions of the  $N$  species change in the  $I$  different rows of the data matrix (concentration profiles).  $\mathbf{S}^T$  is the matrix describing how

the responses of these  $I$  species change in the  $J$  columns of the data matrix (pure spectral profiles).  $E$  is the residual matrix with the data variance unexplained by  $CS^T$ , like experimental error and uncertainties.

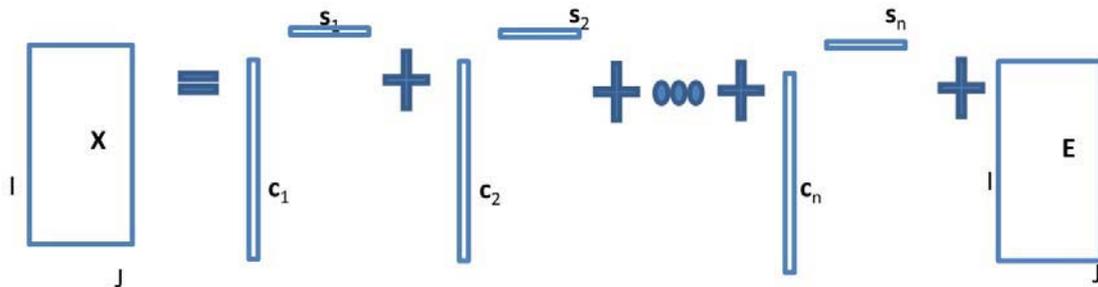


Figure 2- 3 Decomposition of a matrix set according to a bilinear model.

Principal component analysis (PCA), independent component analysis (ICA), multivariate curve resolution (MCR) etc. are bilinear methods used in the analysis of hyperspectral imaging.

### Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a well-known linear dimension reduction procedure that has been proved to be optimal in the sense of the mean squared error. PCA provides efficient abstract bilinear model decomposition to describe the variation of a data set. PCA is a technique that condenses all the spectral information into a few number of components (Principal Components, PC) or latent variables. Once the decomposition is performed, each score vector can be refolded to show the relative distribution map for each component<sup>181</sup>. It gives an abstract decomposition of experimental data, which maximizes the explained variance under the constraint of orthonormality of the components. PCA decomposes the measurement matrix  $X$  into the scores  $T$  and loadings  $P$  factor matrices, that is:

$$X = TP^T + E \quad \text{Equation 2- 10}$$

The aim of the method is to maximize the explained variance in the data with a minimum number of components. Score and loading factor matrices represent a concise summary of the original data that in most cases can aid in interpreting the underlying data variance sources. The scores and loading must be interpreted together and have little meaning alone. However, due to the applied constraints during the PCA bilinear decomposition (orthogonality, normalization and maximum variance), score ( $T$ ) and loading ( $P^T$ ) profiles

are not providing the profiles of the true variance sources, but a linear combination of them fulfilling the applied constraints. True variance source profiles for instance do not fulfill the requirement of orthogonality and they are overlapped, and in many chemical and physical systems profiles should be non-negative.

In this thesis, PCA is used as a data compression noise filtering method and to investigate the number of variance sources, as a first step of the other image resolution methods. Scores and loadings obtained by PCA facilitate also the interpretation of the behavior and description of the sources of data variance, but it does not resolve adequately their true natural profiles.

Although PCA is a widely used approach for feature extraction and data reduction, PCA suffers from high computational cost, large memory requirement and low efficacy in dealing with large dimensional datasets such as Hyperspectral Imaging (HSI). J. Zabalza et al. proposed Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging<sup>182</sup>.

In Chapter 6, bilinear methods of image resolution were assessed, and the comparison between MCR-ALS, PCA, and other related methods were shown.

### **Independent Component Analysis (ICA)**

Independent component analysis (ICA) is a statistical method for finding underlying factors or components from multivariate data, which can be proposed to solve the blind source separation problem<sup>183</sup>. ICA is used to find a mathematical transformation of the data into a linear combination of statistically independent components. The difference of ICA to other methods, is that it looks for components that are both statistically independent and non-Gaussian<sup>184</sup>. Bilinear decomposition ICA constraints are more severe than PCA constraints, since statistical independence is a stronger condition than the linear independence (orthogonality constraints) assumed by PCA (Orthogonality is a special case of linear independence)<sup>185</sup>.

ICA assumes that the data is a linear combination of non-Gaussian, mutually independent latent variables in an unknown mixing matrix. ICA reveals hidden independent sources and their relative contributions. The aim of ICA is the decomposition of multivariate signals into statistically independent component contributions with minimum loss of information. Given a set of observations represented by a data matrix  $\mathbf{X}$  (I,J), having I measurement vectors, the ICA bilinear model can be written as following:

$$\mathbf{X} = \mathbf{AS} + \mathbf{E} \quad \text{Equation 2- 11}$$

where  $\mathbf{S}$  (N,J) is the N dimensional vector of independent sources and  $\mathbf{A}$  (I,N) is the unknown basis matrix.  $\mathbf{E}$  (I,J) is the error matrix. Then, ICA searches for the linear transformation of the data  $\mathbf{W}$ , such that the projected variables:

$$\mathbf{WX} = \mathbf{S} \quad \text{Equation 2- 12}$$

are as independent as possible. It has been shown that the model is completely identical if the sources are statistically independent and at least N-1 of them are non-Gaussian. If the sources are Gaussian, the ICA transformation can be estimated up to an orthogonal transformation. Estimation of mixing and un-mixing matrices can be done maximizing diverse objective functions, among the non gaussianity of the sources and the likelihood of the sample.

Principal Component Analysis (PCA), Independent Component Analysis (ICA) can be considered to share the same overall objective: the decomposition of a data matrix into a reduced number of components.

ICA is an increasingly popular method to resolve complex data sets, like hyperspectral image data, into component signatures (spectra) and their distribution maps. Unfortunately, the pre-requisite of statistical independence severely limits the application of ICA. W. Windig and M. R. Keenan proposed a pre-processing method to extend the use of ICA. For a certain class of data, increasing the sparsity of a data set, increases the independence of the components, and thus enables the successful application of ICA. The sparsity can be increased by simply adding zeros to the data set or by applying a Haar-wavelet transform<sup>186</sup>.

In Chapter 6 comparisons between ICA and other methods like PCA or MCR-ALS (below) are shown for some specific cases.

### **Vertex Component Analysis (VCA)**

VCA (Vertex Component Analysis) is an unsupervised method based on the geometry of a convex data set which considers that the pure image constituents are located at the vertices of a simplex. It is a method that is well known in remote sensing area for endmember (pure component) extraction. For example, in a mixture defined by three constituents, the simplex boundaries built by VCA are a triangle, whose vertices correspond to the estimations of the pure constituents or endmembers. VCA assumes that, for every constituent, exists one or more pixels in the image where only this constituent exists (pure pixels). If this condition is fulfilled, VCA will find these vertex pixels and therefore the corresponding pure components. To find out these pure pixels and the spectral signatures associated to them, VCA iteratively projects the data onto a direction

which is orthogonal to the subspace spanned by the endmembers already estimated. The new endmember signatures are located at the extreme of the new projection. This iteration procedure goes on until all the endmember signatures are estimated. A more detailed description of the method can be found in<sup>187</sup>.

### **Minimum Volume Simplex Analysis (MVSA)**

MVSA (Minimum Volume Simplex Analysis), also considers that the underlying mixing model is bilinear, i.e. that the mixed hyperspectral vectors are a linear combination of the signatures (spectra) of pure components. MVSA is a method that finds the pure components (end members) in hyperspectral image by fitting a minimum volume simplex to the hyperspectral data, under some constraint such as for every pixel no less than zero (non-negativity constraint) abundance fractions and of sum is equal to one (closure). MVSA does not use a least squares approach, but a Sequential Quadratic Programming (SQP) method, based on a quasi-Newton non-linear optimization procedure under linear constraints. In order to prevent local minima and to achieve a faster speed, VCA is usually applied as the initial step to provide an inflated version of the initial simplex<sup>188</sup>. More details about the MVSA can be seen in references and the original paper<sup>189</sup>.

Geometrical-based data analysis approaches are based into two main categories, Pure Pixel (PP) or Minimum Volume (MV). Pixel Purity Index (PPI)<sup>190, 191</sup>, Iterative Error Analysis (IEA)<sup>192</sup>, Vertex Component Analysis (VCA) and Simplex Growing Algorithm (SGA)<sup>193</sup> are pure pixel based algorithms. The Minimum Volume Simplex Analysis (MVSA)<sup>189</sup>, the Minimum Volume Enclosing Simplex (MVES)<sup>194</sup> and the Simplex Identification via variable Splitting And Augmented Lagrangian (SISAL)<sup>195</sup> are minimum volume approaches which fit a minimum volume simplex to the hyperspectral data. In this thesis, VCA and MVSA methods were compared with MCR-ALS for the resolution of remote sensing hyperspectral images in Chapter 3.

### **Other recently developed image resolution methods**

When considering the problem of hyperspectral image resolution, most of the literature in the image processing areas relies on the widely used Linear Mixing Model (LMM). However, this model may be not valid and other nonlinear models can be considered. When there are strong multi-scattering effects or intimate interactions, several significant contributions have been proposed to overcome the limitations of linear model.

Altmann, Y. etc. have reported a Bayesian algorithm to estimate the parameters involved in the model yielding an unsupervised nonlinear unmixing algorithm<sup>196</sup>. This group also studied a linear Radial Basis Function Network (RBFN) for unmixing hyperspectral images<sup>197</sup>. They proposed to estimate the model abundances using a linear combination of radial basis functions whose weights are estimated using training samples. Yoann Altmann et al. reported a nonlinear method using Gaussian processes. The first step is using the Bayesian estimation of the abundance vectors for all the image pixels and the use of a nonlinear function to relate the abundance vectors to the experimental observations. The pure components (endmembers) are subsequently estimated using a Gaussian process regression<sup>198</sup>. Jie Chen etc. formulated a new kernel-based paradigm that relies on the assumption that the mixing mechanism can be described by a linear mixture of pure component (endmember) spectra, with additive nonlinear fluctuations defined in Hilbert space<sup>199</sup>.

### Multivariate Curve Resolution Alternating Least Squares (MCR-ALS)

Multivariate Curve Resolution (MCR) has been chosen in this thesis as the main strategy for the resolution of hyperspectral images. MCR is defined as a group of techniques, which help to resolve mixtures by determining the number of constituents, their response profiles (spectra), and their estimated concentrations, when no prior information is available about the nature and composition of them<sup>200</sup>.

The basic theory of MCR is that the bilinear model is valid for the investigated spectroscopic system (i.e. the extended multicomponent multiwavelength Lambert Beer's law). For a mixture with component  $i = 1, 2, \dots, N$ :

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T = \sum_{i=1}^N \mathbf{c}_i \mathbf{s}_i^T \quad \text{Equation 2-13}$$

Here  $N$  is the number of absorbing components coexisting in the system, and the  $\mathbf{c}_i$  and  $\mathbf{s}_i$  ( $i=1, 2, \dots, N$ ) are the pure concentration profiles and spectra respectively. The problem to be solved is from the measurements matrix  $\mathbf{D}$ , we need to determine:

- The number of absorbing chemical components is  $N$ ;
- The spectrum of each chemical component  $\mathbf{s}_i$  ( $i=1, 2, \dots, N$ );
- The concentration profile of each chemical component  $\mathbf{c}_i$  ( $i=1, 2, \dots, N$ ).

MCR methods are widespread and powerful methodologies for the analysis and modeling of multivariate data in many different application fields. MCR methods have been successfully used to analyze data coming from UV-vis spectroscopy<sup>201</sup>, infrared spectroscopy<sup>202</sup>, chromatography<sup>203</sup>, mass spectroscopy<sup>204</sup>, nuclear magnetic resonance<sup>205</sup>, hyperspectral imaging<sup>206</sup>, voltammetry<sup>207</sup>, microarray<sup>208</sup>, LC-MS metabolomics<sup>209</sup>. MCR is at present well developed both from a theoretical and the applied point of view, and it is still expanding its range of applications.

The wide applicability of MCR can be related to the versatility of the design and implementation of data-specific constraints. The adequate formulation and application of different constraints based on the chemical and physical features of the component

profiles (spectra and concentration when analyzing hyperspectral imaging data) to be resolved are the reason of its success.

MCR-ALS has developed very fast during the last years. An updated version of the graphical user-friendly interface related to the Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) algorithm has been published<sup>210</sup>. It includes recently published advances of this algorithm, including the implementation of additional constraints, such as kinetic hard-modeling and calibration, as well as constraints linked to model structure for multiset and multi-way data analysis, and the possibility to use fully or partially multilinear models to describe the data set.

Recently, the performance and validation of the extension of MCR-ALS with the quadrilinear constraint in the analysis of noisy datasets was studied. Results suggested that MCR-ALS with the quadrilinear constraint can be used efficiently for the analysis of four-way quadrilinear environmental datasets<sup>211</sup>. A Non-Linear Multivariate Curve Resolution Alternating Least Squares (NL-MCR-ALS) method has been proposed and applied for the calibration of spectral data that includes the so-called saturated peaks, which are flattened because of samples with ultrahigh absorbance. In spite of serious violations of the Lambert–Beer law, the NL-MCR-ALS prediction results were quite satisfactory, and the accuracy achieved is better than using other competing methods<sup>212</sup>. In another example, the photodegradation of carbofuran deposited on TiO<sub>2</sub> film under UV light exposure was analyzed using a Hard–Soft Multivariate Curve Resolution-Alternating Least Squares (HS-MCR-ALS)<sup>213</sup>.

MCR-ALS can be extended to the analysis of multi-set data (via column- and row-wise augmented data matrices), and application of multilinear/multiset constraints to profiles to be resolved. MCR-ALS has been used rather extensively also to the analysis of environmental monitoring data tables, for example, to explore the possible interaction between O<sub>3</sub> and NO<sub>x</sub> pollution patterns<sup>214</sup>. MCR-ALS has been also applied to the resolution of concentration and spectral profiles of the azo-dyes photodegradation reactions by UV–visible-DAD and <sup>1</sup>H-NMR spectroscopy data fusion<sup>215</sup>. M.J. Culzoni et al. have developed a liquid chromatographic method with fast scan fluorimetric detection with MCR-ALS in the simultaneous determination of five marker pteridines in urine samples<sup>216</sup>. Keshav Kumar et al. have shown that the synchronous fluorescence spectra of the mixture constituents at various wavelength offsets using Total Synchronous Fluorescence Spectroscopy (TSFS) data set can be extracted using MCR-ALS<sup>217</sup>.

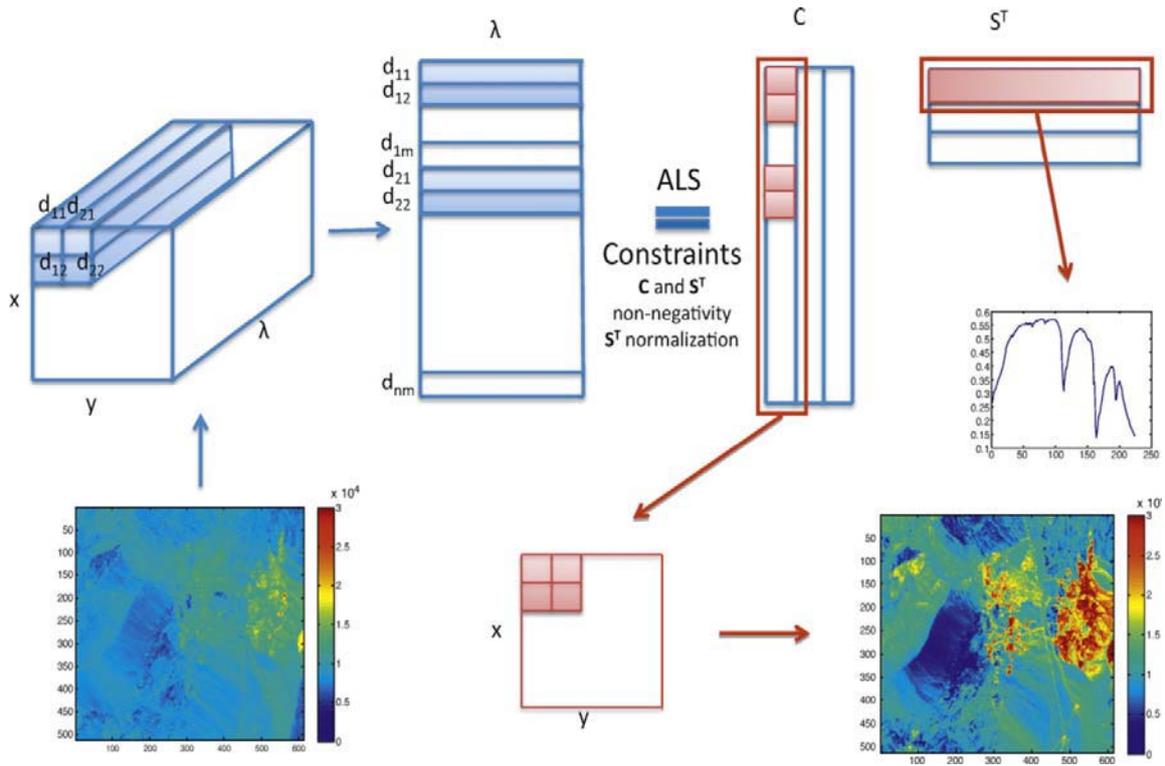
Although second- and higher order data sets are the natural type of data to be analyzed with MCR-ALS, recently this method has been also extended for the analysis of first-order

data using a correlation constraint<sup>218</sup>. In this case, MCR-ALS can be used for first-order multivariate calibration, and the estimations of analyte concentration are correct.

MCR-ALS has been recently applied to the identification of the different paramagnetic centers from complex mixtures of them analyzed by Electron Paramagnetic Resonance (EPR) spectroscopy<sup>219</sup>. In another study, the analytical data obtained from a chronoamperometric experiment for the analysis of a weak host–guest complexation system was also analyzed by MCR-ALS. The data matrices collected at different potentials were simultaneously analyzed by MCR-ALS to recover the thermodynamic parameters (e.g., stoichiometric ratio and formation constant) of cyclodextrin–dopamine complex<sup>220</sup>.

MCR-ALS has been applied in biology, food and medicine industry in recent years. MCR-ALS has been proposed to investigate biological pathways in metabolic networks. By using this method, different constraints can be included in the model, and the same source of variability can be present in different pathways, which is reasonable from a biological standpoint<sup>221</sup>. Silvia Grassi et al. have used MCR-ALS to model milk lactic acid fermentation processes. MCR-ALS applied to FT-NIR spectroscopy can be used as a control system which can also be implemented in-line, as a reliable management method for fermentation process monitoring and to estimate coagulation profiles, no matter what are the operative conditions adopted for the process<sup>222</sup>. Maha A Hegazy et al. have used MCR-ALS to estimate and quantify spectrally overlapped Vitamin B<sub>1</sub>, Vitamin B<sub>6</sub>, benfotiamine and diclofenac sodium quaternary mixtures<sup>222, 223</sup>.

Based on the previous experience and advantages of the MCR-ALS method, in this Thesis, this method has been extended to the resolution of hyperspectral imaging data sets. The goal of multivariate image analysis is to identify and estimate the distribution of the chemical constituents on the scanned sample. Hyperspectral imaging systems coupled to MCR-ALS have already been applied for the resolution of different type of problems<sup>95, 224</sup>.



**Figure 2- 4** Bilinear matrix decomposition of a hyperspectral image using the MCR-ALS method. Resolution of spectra (signatures) and of 2D image concentrations of the pure components.

In hyperspectral imaging, the measured spectroscopic intensity data in matrix  $\mathbf{D}(x, y, \lambda)$  is a function of two variables: pixel position  $x, y$ , and spectral wavelength  $\lambda$ . The pixel spectra are organized as a data matrix  $\mathbf{D}$ , which is decomposed into two factor matrices,  $\mathbf{C}$  and  $\mathbf{S}^T$ , containing the profiles related to the pure component contributions in the image, according to the expression:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{Equation 2- 14}$$

where  $\mathbf{C}$  is the concentration matrix giving the composition of the different resolved components at every pixel (at the position  $(x, y)$ ) and  $\mathbf{S}(\lambda)$  the spectral intensity of these components at wavelength  $\lambda$ . Imaging spectral data can be obtained from different techniques, such as Raman, Infrared, Near Infrared, fluorescence, among other. The same principles can be applied for all of them.

In MCR-ALS, the resolution is accomplished using an Alternating Least Squares (ALS) algorithm, which is initialized using an initial estimation of the spectral or concentration profiles for each intervening species. When it is applied to hyperspectral images resolution, MCR-ALS works following the next basic steps:

- (1) Determination of the number of constituents in the raw image (**D**).
- (2) Generation of initial estimates ( $\mathbf{S}^T$ ) based on the selection of the purest pixel spectra (SIMPLISMA).
- (3) Given **D** and  $\mathbf{S}^T$ , calculation of **C** under constraints.
- (4) Given **D** and **C**, calculation of  $\mathbf{S}^T$ , under constraints.
- (5) Reproduction of **D** from the product of **C** and  $\mathbf{S}^T$ .
- (6) Go to step 3 until convergence is achieved.
- (7) Recovering spectral information from  $\mathbf{S}^T$  and identification of image constituents.
- (8) Recovering spatial information from refolded **C** matrix, and mapping of the image constituents.

The number of image constituents can be either known beforehand based on the knowledge of the samples analyzed, or be determined by PCA or SVD on the whole image. The alternating optimization should always start by using the original preprocessed measurement, **D**, and an initial guess of either the **C** or the  $\mathbf{S}^T$  matrices. Typically, in images, the initial estimate is a  $\mathbf{S}^T$  matrix formed by pixel spectra picked up from the image according to previous knowledge (from pixels in areas of interest) or as a result of applying chemometric tools for purest pixel selection, such as SIMPLISMA (Simple-to-use Interactive Self-modeling Mixture Analysis)<sup>225</sup>.

During the iterative optimization process (steps 4 and 5), the unconstrained least-squares solution for the concentration profiles can be calculated by the following equation:

$$\mathbf{C} = \mathbf{D}(\mathbf{S}^T)^+, \text{ under constraints} \quad \text{Equation 2- 15}$$

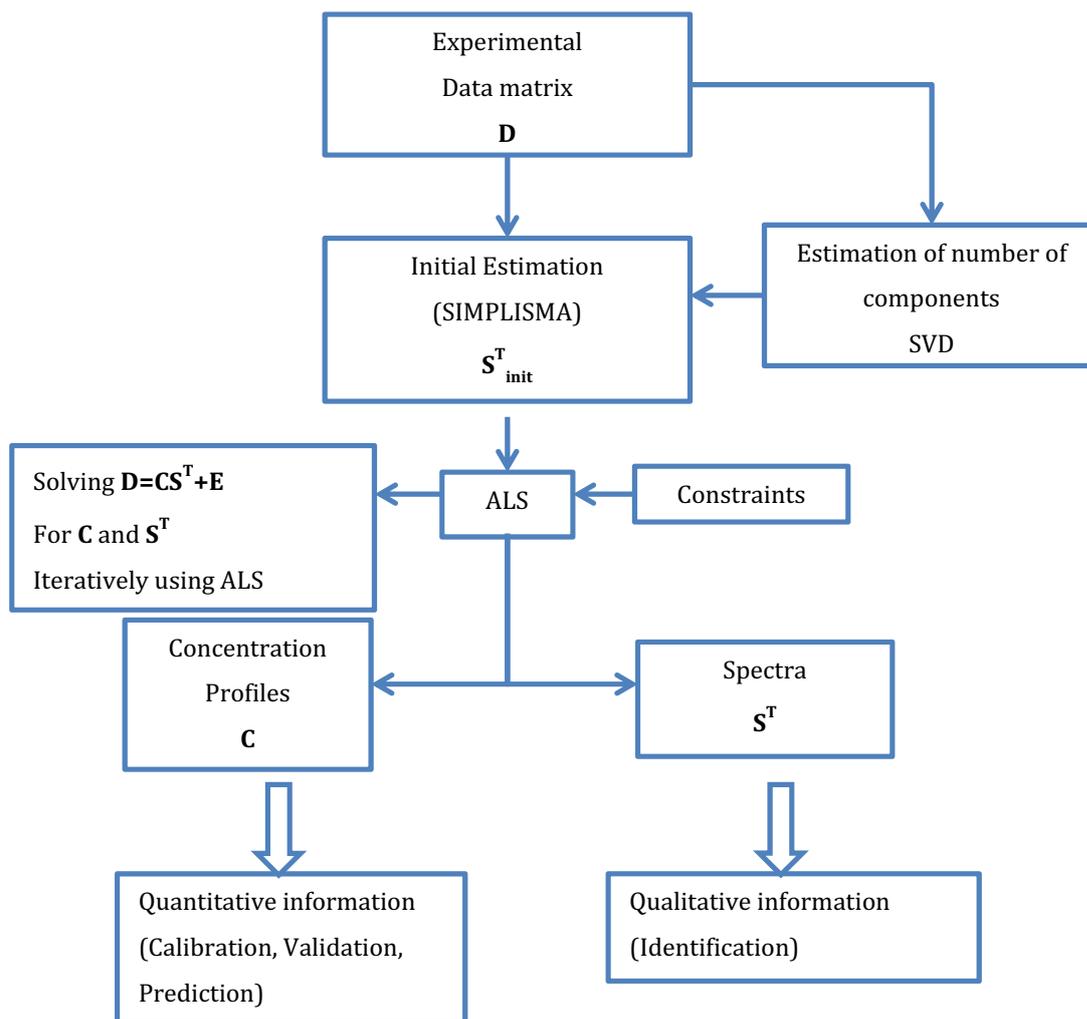
where  $(\mathbf{S}^T)^+$  is the pseudoinverse of the spectral matrix  $\mathbf{S}^T$ , and it is equal to  $[\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}]$  when  $\mathbf{S}^T$  is full rank. The new estimation of the **C** matrix can be then used to recalculate by least squares a new estimation of the  $\mathbf{S}^T$  matrix:

$$\mathbf{S}^T = \mathbf{C}^+\mathbf{D}, \text{ under constraints} \quad \text{Equation 2- 16}$$

where  $\mathbf{C}^+$  is the pseudoinverse of **C**,  $\mathbf{C}^+ = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$  when **C** is full rank. Each of these two equations, are solved under constraints. Constraints are defined as chemical or mathematical properties that should hold the resolved profiles. For example, constraints like non-negativity which is based on chemical or physical assumptions on the data can be applied during the iteration<sup>226</sup>. Similarly, constraints like local rank for selectivity

constraints which are not necessary in relation to the chemistry or the physics of the system but related to mathematical properties can be applied. Constraints are used to have profiles with shapes that are physically meaningful and to approximate the solutions to the true ones (see below). The proper application of constraints is the more important aspect in the application of the MCR-ALS method (see below for a more detailed description of constraints).

At each iteration step, a new estimation of the  $\mathbf{C}$  and  $\mathbf{S}^T$  matrices is obtained and used as a new estimation for the next iteration. Thus the alternating least squares process continues until convergence is reached (There is no significant variation among the fitting results nor profiles in consecutive cycles.). If the applied constraints are fulfilled by the data, the final solutions are close or equal to the real solution. In order to avoid scale indeterminacies during the ALS optimization, usually it is necessary to apply normalization either to  $\mathbf{C}$  or to  $\mathbf{S}^T$  profiles. In this thesis, usually, the pure spectra of the resolved components have been normalized<sup>225</sup>. Equal vector length / area normalizations are selected in this Thesis for hyperspectral imaging resolution (Loadings  $\mathbf{S}$  are divided by the square root of the sum of the squares of all the elements of the considered spectra profile, so that all components have the same spectral contribution, and the quantitative information is contained in  $\mathbf{C}$  matrix).



**Figure 2- 5** Processes of MCR-ALS for experimental data resolution

$\mathbf{C}$  and  $\mathbf{S}^T$  are optimized iteratively using an Alternating Least Squares (ALS) algorithm until convergence is reached. External application of constraints during ALS iterations of the has some advantages: it has the flexibility to cope with many kinds of data structures and chemical problems and the ability to accommodate external information in the resolution process. Both soft and hard constraints can be applied during the ALS optimization process.

Constraints are essential to drive the iterative resolution process to chemically meaningful solutions and to decrease the rotational ambiguity inherent to the multivariate curve resolution results (see below). The following constraints can be considered:

**Non-negativity constraints:** Non-negativity concentration constraint is a general constraint used in curve resolution methods.<sup>227</sup>. Physical concentrations can only be positive or zero ( $\mathbf{C} \geq 0$ ), and in many spectroscopies, for example Raman, Infrared applied in this Thesis,

spectral values also can be only positive or zero ( $\mathbf{S}^T \geq 0$ ). In all the cases of this Thesis, non-negativity concentration constraints and non-negativity spectra constraints have been applied for hyperspectral imaging resolution. There are different ways to apply this constraint, such as, direct replacement of negative values to zero, or using more rigorous iterative methods based on non-negative least squares (eg. Non-negative Least Squares, (NNLS)<sup>228</sup>, and Fast Non-Negative Least Squares (FNNLS)<sup>229</sup>). In this Thesis, FNNLS have been applied for non-negative constraint.

Unimodality constraints: In a wide variety of situations, concentration profiles have a unimodal shape, for example they only have a maximum. This is the case for instance for chromatographic elution profiles. In contrast, spectral profiles are usually not unimodal in the general case. Some other types of instrumental signals may also be unimodal, as in electrochemistry. Whenever the shape of the profiles is unimodal, unimodality is a useful additional constraint. In this case, the algorithm intends to avoid the formation of secondary maxima. First the highest maximum is detected and then all the departures from the unimodal condition are constrained, discarding left and right maxima. This constraint is not suitable for the cases of hyperspectral imaging resolution in this thesis. See reference<sup>230</sup> for more detailed descriptions of the unimodality constraint.

Closure constraints: This case is frequently encountered for instance in reaction-based systems, where a mass balance equation is obeyed by the concentration profiles of the species present in the system. With this constraint, the sum of the concentrations or fractions of all of the species involved in the reaction or mixture (the suitable elements in each row of the C matrix) is forced to be equal to a constant value (the total concentration or fractions) at each stage in the reaction or in the mixture. The closure constraint is an example of an equality constraint. This constraint is not suitable for the cases of hyperspectral imaging resolution in this Thesis. See reference<sup>230</sup> for more detailed descriptions of the closure constraint.

### **Selectivity/local rank constraints**

Selectivity/local rank constraints use the information related to the presence or absence of certain components in particular pixels. Selectivity and local rank constraints are among the most powerful constraints to decrease rotational ambiguities (see below) in curve resolution studies<sup>231</sup>.

The rank of a data matrix is equal to the number of linearly independent rows or columns<sup>232</sup> of the considered data matrix. Assuming concentration profiles and spectra to be linearly independent and noiseless, the rank of the data equals the number of pure components. This number can be estimated from Singular Value Decomposition (SVD) as the number of singular values which are larger than a threshold value associated to noise

contributions. In general, it is practical to define a 'chemical rank' counting only the number of singular values that are greater than a cut-off determined by the noise level. This implies that only major sources of variance are considered, those related with the physical phenomena under study, for instance the number of chemical species in a chemical reaction or the number of eluted components in a chromatographic separation.

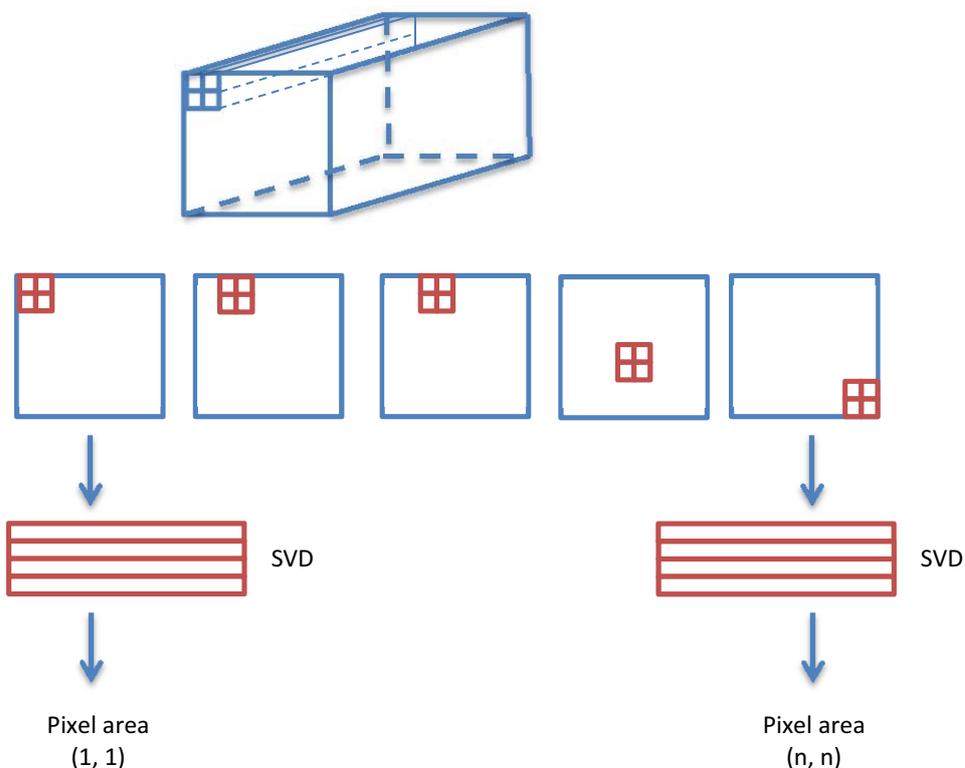
When analyzing a complex chemical imaging system, the 'chemical rank' evaluated by SVD equals to the number of chemical component in the sample<sup>233</sup>. Local rank detects the rank in a small area of the image. From chemical local rank information, it is possible to infer the presence or absence of particular constituents in the image<sup>234</sup>. Local rank constraints use this rank information for improve the MCR resolution results.

Local rank analysis methods have been used in the study of evolving processes, where the concentration profiles of the different components evolve smoothly and, often following a sequential pattern<sup>235</sup>. Using gradually growing size windows, imitating the stepwise progress of a process, the evolution of the number of significant components indicates the number of different components emerging and decaying. In sequential processes, the proper use of this information can provide concentration windows. For example, it can indicate where a particular component is present, and give approximate concentration profiles for the different components. Evolving Factor Analysis (EFA)<sup>236</sup> was the principal method used for local rank analysis of chemically evolving processes. Modified versions of this method were oriented to confirm the sequential evolution of components in unknown processes, to improve the setting of concentration windows and to overcome problems linked to the analysis of rank-deficient processes.

Local rank analysis methods are based on the use of fixed size moving windows covering the whole data set. The results obtained by the application of local rank analysis methods inform about the local complexity of the data set. They give information about how many components overlap in the different data set regions enclosed by the selected windows. They can be used to locate selective regions in the data set, i.e., windows with only one component, which are critical to achieve resolution without ambiguities<sup>237</sup>.

To adapt these methods to the hyperspectral imaging data structures, a modified FSMW-EFA methodology was developed, called Fixed Size Image Window–Evolving Factor Analysis (FSIW-EFA)<sup>234</sup>(See figure 2-6). It has been adapted to the structure of image datasets with spatial information and it provides the information to be used as MCR-ALS constraints. Hyperspectral imaging scans material surface pixel by pixel, covering the whole surface of the object. Local rank maps describe the pixels where local rank constraints can be potentially applied. The absence of a particular component can be used as a constraint in pixels where the chemical rank is lower than the total number of compounds in the image. From a rank analysis of the appropriate data windows

(submatrices) it is possible to find out the number of components (constituents) present in the corresponding pixel.



**Figure 2- 6** How fixed size moving windows are used to detect local ranks

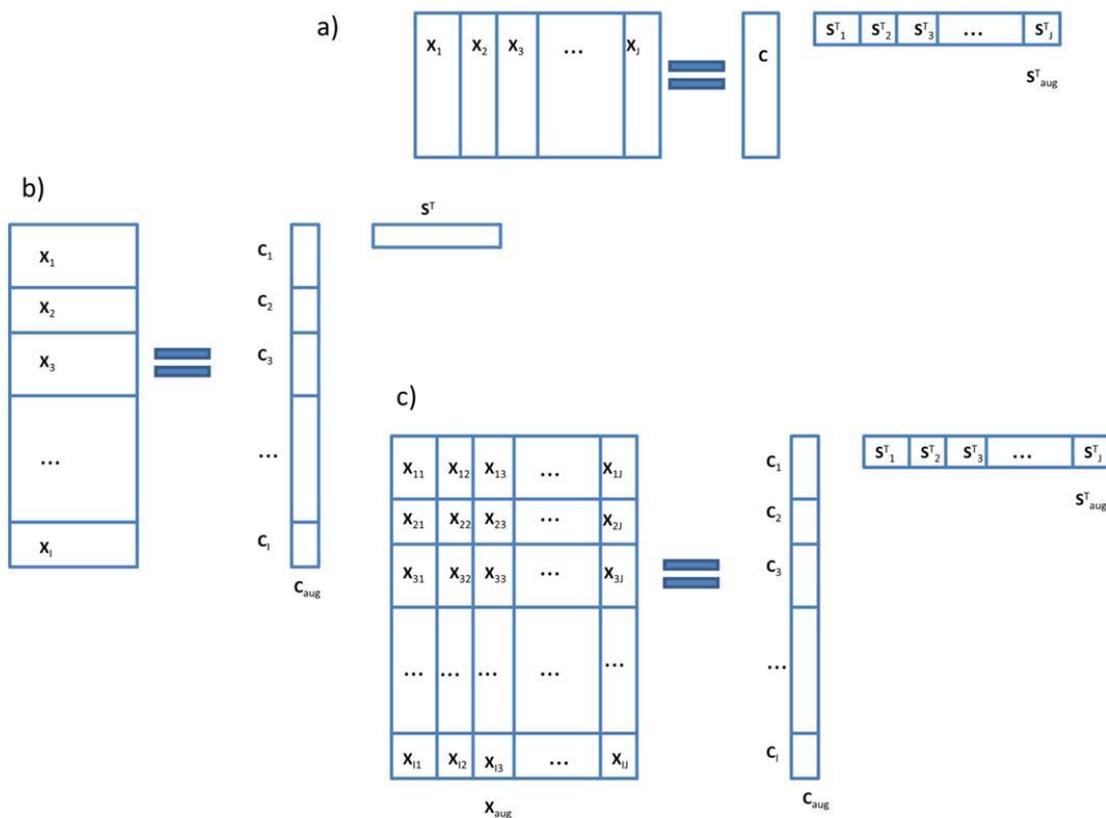
Identification of the compounds absent in these pixels requires the use of additional information. For example, comparison of the pixel spectrum with known pure spectra, when available or with the purest selected spectra of the image.

In Chapter 3, when analyzing remote sensing image datasets, USGS spectral library was used for spectral identification. Also, if available, the spectra from external samples or for reference spectra can be used during the image resolution. In Chapter 4, an example of application of local rank constraints is shown for chocolate analysis. In food industry, it is usual to have some previous knowledge about the possible chemical constituents of food products and it is relatively easy to have their pure spectra.

### **MCR-ALS applied to the simultaneous analysis of multiple data matrices**

MCR-ALS can be extended to the analysis of multiset and multiway data structures. A single data matrix can be row-wise, column-wise, or row and column-wise augmented to

form a multiset data structure when other matrices have either the same variables are in their rows, in their columns, or in both. In these cases they can be appended in the appropriate direction (see Figure 2- 7)<sup>238</sup>.



**Figure 2- 7** Extension of the MCR model to multiset data analysis arranged in augmented data matrices. a) MCR-ALS of a row-wise augmented data matrix, b) MCR-ALS of a column-wise augmented data matrix, c) MCR-ALS of a column- and row-wise augmented data matrix.

To have a new meaningful data structure, all individual data matrices (slices) in this type of data set should share some information with the other appended matrices; otherwise it does not make any sense to build such a new data arrangement. Different ways of building row-wise, column-wise, and column- and row-wise augmented matrices are shown in Figure 2- 7.

Different data augmentation arrangements are possible and they may be written in a more concise way. For instance, in the work of Chapter 4, experimental data of chocolate samples and reference data share the same pure spectra, and column-wise augmented matrices were used (where reference spectra of pure component were added at the bottom of the experiment data matrix to use selectivity constraints). Column-wise

augmented matrices,  $\mathbf{X}_{\text{aug}}$ , may be written (in MATLAB notation, The Mathworks, Inc.<sup>239</sup>) as  $[\mathbf{X}_1; \mathbf{X}_2; \mathbf{X}_3; \dots; \mathbf{X}_I]$ , where the semicolon ‘;’ notation is used to indicate that the different data matrices  $\mathbf{X}_I$ ,  $i = 1, \dots, I$ , are column-wise appended, one on top of each other, keeping the same number of columns. In this data arrangement, the different data matrices are supposed to share their column vector space. Row-wise augmented data matrices are written as  $[\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots, \mathbf{X}_J]$ ,  $j = 1, \dots, J$  where the semicolon ‘,’ notation is used to indicate that the different data matrices are set one beside the other. These different matrices are supposed to share their row vector (concentration) space. When Raman, Infrared, Near Infrared or other hyperspectral imaging analyzed a similar sample with similar pixel locations, row-wise augmented matrices can be proposed.

Column- and row-wise augmented data matrix augmentation also can be proposed in case of the different individual data matrices share their row- and column- spaces. Bilinear modelling applied to these systems are shown in Figure 2- 7c and in Equation 2- 17:

$$\begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \mathbf{X}_{13} & \dots & \mathbf{X}_{1J} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & \mathbf{X}_{23} & \dots & \mathbf{X}_{2J} \\ \mathbf{X}_{31} & \mathbf{X}_{32} & \mathbf{X}_{33} & \dots & \mathbf{X}_{3J} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{X}_{I1} & \mathbf{X}_{I2} & \mathbf{X}_{I2} & \dots & \mathbf{X}_{IJ} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \\ \vdots \\ \mathbf{C}_I \end{bmatrix} [\mathbf{S}_1^T \quad \mathbf{S}_2^T \quad \mathbf{S}_3^T \quad \dots \quad \mathbf{S}_J^T] + \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} & \mathbf{E}_{13} & \dots & \mathbf{E}_{1J} \\ \mathbf{E}_{21} & \mathbf{E}_{22} & \mathbf{E}_{23} & \dots & \mathbf{E}_{2J} \\ \mathbf{E}_{31} & \mathbf{E}_{32} & \mathbf{E}_{33} & \dots & \mathbf{E}_{3J} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{E}_{I1} & \mathbf{E}_{I2} & \mathbf{E}_{I2} & \dots & \mathbf{E}_{IJ} \end{bmatrix}$$

$$= \mathbf{C}_{\text{aug}} \mathbf{S}_{\text{aug}}^T + \mathbf{E}_{\text{aug}} \quad \text{Equation 2- 17}$$

In Equation 2- 17, a row-wise augmented spectral matrix  $[\mathbf{S}_1^T, \mathbf{S}_2^T, \mathbf{S}_3^T, \dots, \mathbf{S}_J^T]$  is obtained to describe the different spectra in the different individual data matrices (i.e., different spectroscopies), and a column-wise augmented concentration matrix  $[\mathbf{C}_1; \mathbf{C}_2; \mathbf{C}_3; \dots; \mathbf{C}_I]$  is obtained to describe the concentration changes in the different individual data matrices. The  $\mathbf{X}_{IJ}$  matrices in this data arrangement are assumed to share their two vector spaces, column or row space. Thus,  $\mathbf{X}_I$  matrices, i.e., matrices with identical index I, share their row (concentration) space, whereas  $\mathbf{X}_J$  matrices, i.e., matrices with identical index J, share the column (spectral) space.

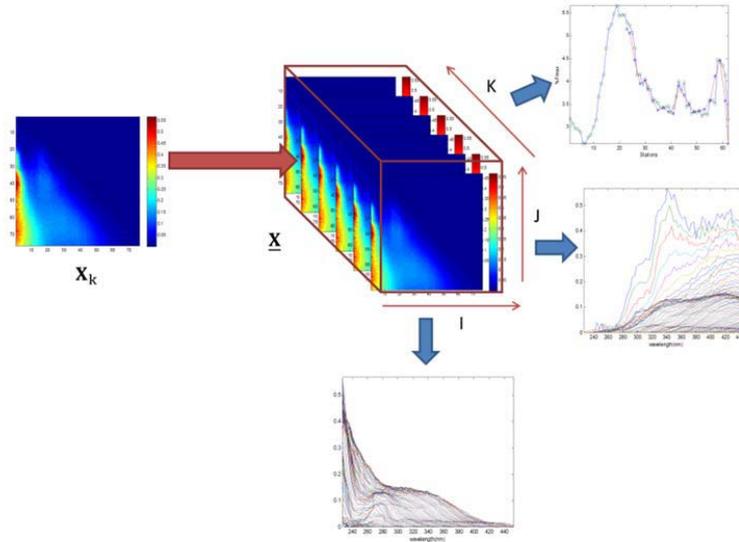
The first outstanding advantage of this data matrix augmentation is that the good features for resolution presented by one or some of the included data matrices will have a positive effect on the resolution of the most complex ones. For instance, introducing selectivity or favorable local rank conditions into this type of data sets is generally easy and possible by appending matrices coming from simpler processes or by appending data matrices from the pure response of one or few of the analytes (like standard sample of sucrose, lactose,

butter and whey which were used in the work introduced in chapter 4) present in a much more complex mixture data matrix.

### **MCR-ALS applied to multiway data using multilinear models**

Multivariate Curve Resolution (MCR) methods can be extended also to the analysis of multi-way data. According to Sanchez and Kowalski, the analytical data generated by instrumental techniques can be classified as follows: zero-order data (e.g. a univariate or single data per analyte), first-order data (e.g. a data vector per sample), second order data (e.g. a data matrix per sample)<sup>240</sup>. Correspondingly, zero-, first- and second-order calibration methods have been developed for extracting chemical information from each type of data. The trilinear/PARAFAC<sup>241</sup> model are gaining widespread acceptance among the analytical chemistry area. A range of analytical instrumentation is available that enables high dimensionality data to be resolved and explained. Chemometric techniques for treating second order data are well developed at present. The second order methods have the so-called “second-order advantage”. For example, they can predict the concentration of the analytes of interest even in the presence of unknown interferes. This also enables several analytes to be determined simultaneously. MCR has been extended to the study of second order data arranged in three-way data cubes<sup>242</sup>. As stated before, MCR is based on the bilinear decomposition given by Equation 2- 17. Although MCR-ALS was initially designed for the analysis of individual and augmented two-way data matrices under the assumption of a bilinear model, MCR-ALS was extended very soon to the investigation of three-way data with the trilinearity which forces the fulfillment of the trilinear model also. Analysis of a three-way data set using a trilinear model generates three sets of component profiles with equal number of components in the three modes.

In the multiway data analysis field, the more established method of analysis is PARAFAC. PARAFAC is a generalization of PCA to higher order/multiway data arrays. Three-way data analyzed by PARAFAC are characterized by several sets of variables that are measured in a crossed fashion and arranged in a data cube. Chemical examples of three-way data are for instance Emission Excitation Matrix (EEM) fluorescence spectra for analysis DOM in several samples (see Figure2-8) river Ter (Details in Chapter 6).



**Figure 2-8** Decomposition of the data cube of EEM data for DOM analysis giving the profiles corresponding to three data directions using a trilinear model.

PARAFAC decomposes the data cube in three factor matrices described the three data modes/way/directions<sup>243</sup>. In the case of EEM dataset  $\underline{\mathbf{X}}$  is a data cube  $I \times J \times K$ , where  $K$  is the number of the samples, and  $I$  and  $J$  are the number of wavenumbers of the emission and excitation spectra. The trilinear model used by PARAFAC can be expressed for each data slice of the cube,  $\mathbf{X}_k$  as:

$$\mathbf{X}_k = \mathbf{C}\mathbf{Z}_k\mathbf{S}^T + \mathbf{E}_k \quad \text{Equation 2- 18}$$

where every/slice data matrix,  $\mathbf{X}_k$ , is decomposed  $\mathbf{C}(I,N)$ ,  $\mathbf{S}(J,N)$ , and  $\mathbf{Z}_k(N,N)$ .  $\mathbf{C}$  is the matrix of loadings in the first mode/way/direction and  $\mathbf{S}^T$  is the matrix of loadings in the second mode/way/direction. And the factor matrix  $\mathbf{Z}_k$  of loadings in the third direction or mode is a diagonal matrix giving the relative amounts of every component in each considered data matrix  $\mathbf{X}_k$ .  $\mathbf{E}_k$  is the residual term for slice  $\mathbf{X}_k$ .

As described by the matrix notation in Equation 2-18,  $\mathbf{Z}_k$  changes for the different matrices  $\mathbf{X}_k$ .  $\mathbf{C}$  and  $\mathbf{S}^T$  matrices are the same (unique, invariant) in all simultaneously analyzed matrices  $\mathbf{X}_k$ .

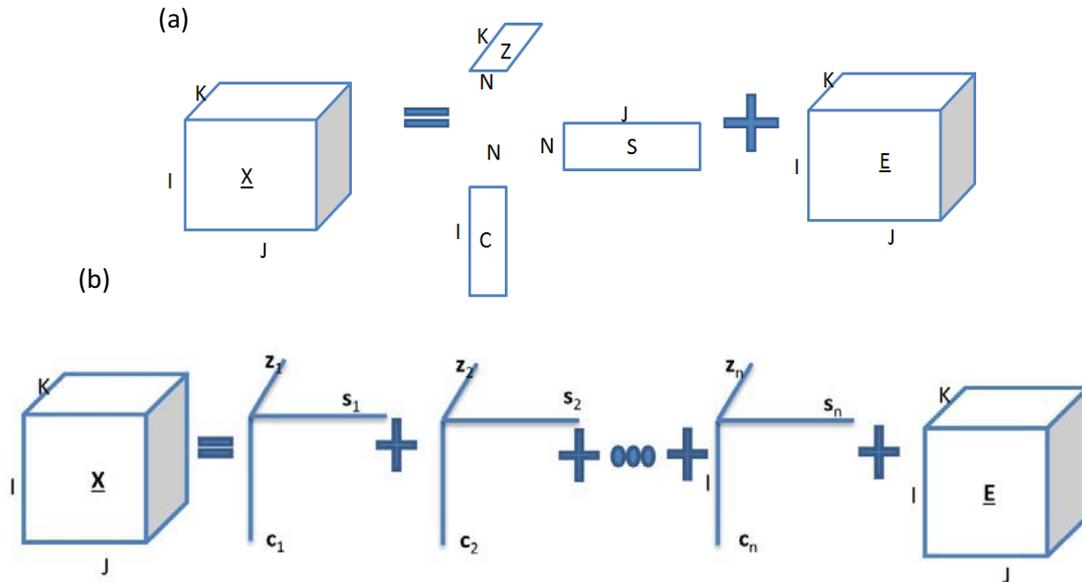
This is a strong constraint that enforces trilinear decompositions to give unique solutions for the three-factor matrices (apart from scale and trivial permutation rotation ambiguities), and avoids the presence of rotational ambiguities associated with lower structured bilinear models.

The trilinear model can also be in an element-wise way as:

$$x_{ijk} = \sum_{n=1}^N c_{in} s_{jn} z_{kn} + e_{ijk} \quad \text{Equation 2-19}$$

where  $x_{ijk}$  represents the  $ijk^{\text{th}}$  element in the three-way dataset it means the measured data ( $i=1, \dots, I, j=1, \dots, J, \text{ and } k=1, \dots, K$ ),  $N$  is the number of components (rank) common to the three modes ( $n=1, \dots, N$ ),

$c_{in}$ ,  $z_{kn}$  and  $s_{jn}$  are the elements in  $\mathbf{C}(I, N)$ ,  $\mathbf{Z}(K, N)$  and  $\mathbf{S}(J, N)$  (as shown in Figure 2-9) used to obtain the  $x_{ijk}$  element.  $e_{ijk}$  is the residual term (part of the data not explained by the model).

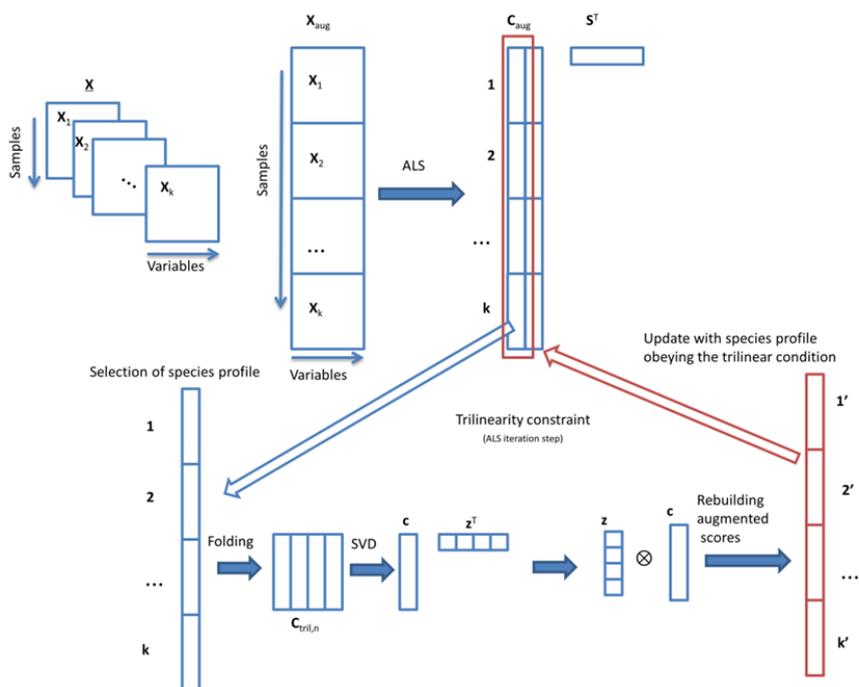


**Figure 2-9** Decomposition of a three-way data set according to a trilinear model. (a) Expressed as decomposition in loading matrices and (b) expressed as the sum of pure component triads.

The data cube  $\underline{\mathbf{X}}$  can be also written as the sum of the individual contributions of the  $N$  components (Figure 2-9). Each component is defined by a trio of profiles,  $c_n$ ,  $z_n$  and  $s_n$  corresponding to the three data directions. The estimated models of PARAFAC are compact in a mathematical sense and rather simple to interpret. In PARAFAC rotation problem is solved without the ambiguities associated to bilinear models (see below). PARAFAC provides unique solutions; when the inner structure of the data set matches perfectly the underlying model the solution provided by PARAFAC is the correct one (the

true one). In Chapter 6, PARAFAC is compared with MCR-ALS with the trilinearity constraint that is described immediately below.

In MCR-ALS, the trilinear model described above can be implemented as a constraint during ALS for each resolved component separately. This is achieved algorithmically by a constraint during the ALS optimization and it has been described in detail elsewhere<sup>244</sup>. The trilinear constraint forces profiles of the same component in the different data matrices for a particular direction to be the same (to have the same shape) and only change in scale/intensity.



**Figure 2- 10** Implementation of the trilinear constraint in the MCR-ALS (see text for details of the procedure).

The first step in the MCR-ALS method is the determination of the number of components. Next, as shown in Figure 2- 10, initial estimates of  $C_{aug}$  and  $S^T$  as the following steps:

- 1 First it is needed to reshape the three-way data set  $\underline{X}$  as a two-way one  $\underline{X}_{aug}$ , leave  $S^T$  as the spectral from which will obtain the pure spectral as the initial data for the iteration calculation of ALS.
- 2 Estimation of the purest (more different) spectra from the raw data set (SIMPLISMA)<sup>225</sup>.

3 Given  $\mathbf{D}_{\text{aug}}$  and initial  $\mathbf{S}^T$  estimation of  $\mathbf{C}_{\text{aug}}$  under non-negative and trilinearity constraints (trilinearity is introduced below).  $\mathbf{C}_{\text{aug}} = \mathbf{D}_{\text{aug}}(\mathbf{S}^T)^+$ ,  $(\mathbf{S}^T)^+$  is the pseudoinverse of the spectral matrix  $\mathbf{S}^T$ , and it is equal to  $[\mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}]$  when  $\mathbf{S}^T$  is full rank. The new estimation of the  $\mathbf{C}_{\text{aug}}$  matrix can be then used to recalculate by least squares a new estimation of the  $\mathbf{S}^T$  matrix:

4 Given  $\mathbf{D}_{\text{aug}}$  and  $\mathbf{C}_{\text{aug}}$  estimation of  $\mathbf{S}^T$  under non-negative and normalization constraints.  $\mathbf{S}^T = \mathbf{C}_{\text{aug}}^+ \mathbf{D}$ , under constraints, where  $\mathbf{C}_{\text{aug}}^+$  is the pseudoinverse of  $\mathbf{C}_{\text{aug}}$ ,  $\mathbf{C}_{\text{aug}}^+ = (\mathbf{C}_{\text{aug}}^T \mathbf{C}_{\text{aug}})^{-1} \mathbf{C}_{\text{aug}}^T$  when  $\mathbf{C}_{\text{aug}}$  is full rank.

5 Reproduction of  $\mathbf{D}_{\text{aug}}$  from  $\mathbf{C}_{\text{aug}}$  and  $\mathbf{S}^T$ .

6 Back to step 3, until convergence is achieved. To end the optimization process, maximum numbers of iterations and the convergence criterion have been set.

7 Decompose  $\mathbf{C}_{\text{aug}}$  in  $\mathbf{Z}$  and  $\mathbf{C}$ .

In Figure 2- 10, when the trilinear constraint is applied, the concentration profiles of the same component in the different sub matrices  $\mathbf{C}_k$  in  $\mathbf{C}_{\text{aug}}$  are forced to have the same shape during iterations of the ALS optimization. To do this, in step 3, a one-component concentration profile matrix,  $\mathbf{C}_{\text{tril},n}$  is built with  $I$  rows (number of rows of each data matrix) and  $K$  columns (number of matrices simultaneously analyzed). For a full trilinear model, every component has a  $\mathbf{C}_{\text{tril},n}$  matrix ( $n=1, \dots, N$ ), which is approximated by their related one-component bilinear decomposition (using for instance PCA or singular value decomposition (SVD)).  $\mathbf{c}$  is the one column vector, which contains the common (average) concentration profile of the component  $n$  in the different sub matrices in  $\mathbf{C}_{\text{aug}}$ , and  $\mathbf{z}^T$  is a row vector with the relative amounts of this concentration profile.  $\mathbf{C}_{\text{tril},n} = \mathbf{c}\mathbf{z}^T$  (See figure 2-10). After this, the full concentration  $\mathbf{C}_{\text{aug}}$  augmented matrix is rebuilt and updated by replacing their initial profiles by the  $\mathbf{C}_{\text{tril},n}$  profiles, obtained from the different individual.

At the end of the MCR-ALS optimization, only the recovered spectral information in  $\mathbf{S}^T$  can be straight forwardly matched with one of the three PARAFAC modes. The matrix  $\mathbf{C}_{\text{aug}}$  contains implicitly the information related to matrices  $\mathbf{C}$  and  $\mathbf{Z}$  in the other two PARAFAC modes. MCR-ALS has complete freedom in the correspondence of profiles among different components and modes, in different way than PARAFAC. It models also situations where two or more components have common profiles in one of the modes, in what is called interaction models<sup>200</sup> or situations where one component should be described with more than one profile in one of the modes.

In the work of using EEM to analysis DOM in water system, MCR-ALS method with the trilinearity constraint is proposed. More details are explained in Chapter 6.

**Uncertainties in MCR results**

Bilinear model decompositions differ in the way how the data decomposition is performed. PCA and ICA apply very strong constraints which produce unique solutions, but their solutions may not have direct physical meaning. MCR methods use softer constraints but with more physical meaning, like non-negativity, unimodality, closure or local rank and selectivity (see previous section 2.2). These constraints are following the known physical and chemical (natural) features of the profiles of the constituents of the analyzed samples.<sup>200</sup> However profiles recovered by MCR methods may be affected by intensity (or scale ) and rotation ambiguities<sup>245, 246</sup>.

MCR model of Equation 2- 20 can also be written as:

$$\mathbf{D} = \mathbf{C}\mathbf{T}\mathbf{T}^{-1}\mathbf{S}^T = \mathbf{C}^*\mathbf{S}^{*T} \quad \text{Equation 2- 20}$$

where T stands for any transformation matrix and:

$$\mathbf{C}^* = \mathbf{C}\mathbf{T} \text{ and } \mathbf{S}^{*T} = \mathbf{T}^{-1}\mathbf{S}^T \quad \text{Equation 2- 21}$$

This is the mathematical formulation of the rotational ambiguity, which means that the same lack of fit can be obtained in the description of the data set  $\mathbf{D}$  using different sets of profiles ( $\mathbf{C}^*$  and  $\mathbf{S}^{*T}$ ), shaped differently from the true ones ( $\mathbf{C}$  and  $\mathbf{S}^T$ ).

Another source of ambiguity different from the rotation ambiguity is the scale/intensity ambiguity, which can be expressed as:

$$\mathbf{D} = \sum_{i=1}^n \left( \frac{1}{k_i} \mathbf{c}_i \right) (k_i \mathbf{s}_i^T) \quad \text{Equation 2- 22}$$

The resolved profiles ( $\mathbf{c}_i$ ,  $\mathbf{s}_i$ ) for each pure component can present profiles with the same shape but with one of them  $k_i$  times smaller and the other  $k_i$  times larger than expected, without changing the final result. This is the reason why in MCR, if no additional scale information (like mass balance or closure), a normalization constraint is usually applied to one of the two set of profiles of every component, either to the concentration (like closure) or to the spectra (normalization to equal length/area/intensity)

The extent of rotation ambiguities in MCR results can be significantly decreased or even suppressed by the use of constraints. The more constrained a system is, the fewer the number of possible solutions (profile combinations) can fulfill the required data patterns with the same fit. The use of constraints forces results to be closer to the true profiles and fit optimally the data set  $\mathbf{D}$  (only constraints that are fulfilled by the data should be applied). The presence of rotation ambiguities in MCR results has been compelled

somewhat the wide spread use of MCR methods. Recently there have been some attempts to evaluate the extent of rotation ambiguities in MCR solutions<sup>246-250</sup>. Since rotation ambiguity is component and profile-dependent and, within the same data set, we may find components or profiles with low ambiguity and others that have it in a larger extent. Due to the presence of rotation ambiguities, instead of unique solutions, a range of feasible solutions that fit the data giving the same data-fitting values and fulfill the same applied set of constraints can be obtained in all resolution methods. In these cases, there is no one better solution than others, because they are undistinguishable considering the previous knowledge of the system. MCR results from hyperspectral image analysis can show the presence of rotation and intensity ambiguities.

### **Calculation of Rotation ambiguities and Area of Feasible Solutions (AFS)**

Lawton and Sylvestre proposed a first algorithm for determining the AFS of two-component systems under non-negative constraints in spectra and concentration profiles in early 1980s<sup>227</sup>. In 1985 Borgen et al.<sup>251</sup> developed an analytical solution of the feasible ranges of the pure component spectra from mixture spectra and extended it for three-component systems using a tangent and simplex rotation algorithms. Low dimensional representation of this continuum of solutions is also called the Area of Feasible Solutions (AFS)<sup>252, 253</sup>.

Henry and Kim proposed a procedure to calculate the outer boundary of the region of feasible solutions by using linear programming methods and determine all its vertices. These methods require the use of non-negativity and other physical constraints, to restrict feasible solutions to single points (unique solutions) and not to regions (non-unique solutions)<sup>254</sup>. Leger and Wentzell developed a dynamic Monte Carlo SMCR (Self Modelling Curve Resolution) method<sup>255</sup> which seeks to define the boundaries of feasible pure component profiles. The algorithm employs a directed Monte Carlo approach to search for valid solutions with high efficiency. The parameters for the search (direction, step size) are set through bootstrap estimates of the geometry of the solution space. However this method suffers from large instabilities in the calculations.

Rajkó and István<sup>256</sup> revised Borgen's study. They considered that the normalized feasible solutions are embraced in a (N-1)-dimensional simplex (N is the number of components) with the vertices being the N-normalized pure profiles. Computational geometry tools were used instead of the linear programming tools used to draw Borgen plots of three-component systems. The Borgen plot (BP) method used for these calculations is based on SVD/PCA and it is highly sensitive to noise.

H. Abdollahi, M. Maeder and R. Tauler have proposed a systematic grid search method for all feasible solutions and the results have been displayed in appropriate mesh and contour plots which reveal their boundaries<sup>249</sup>. In 2013, A. Golshan, M. Maeder and H. Abdollahi have developed the method using the simplex volume of feasible solutions for determination and visualization of rotational ambiguity of four-components mixture<sup>257</sup>.

Recently, Sawall et al. suggested a fast and accurate algorithm to find AFS in two and three-component systems using the inflation of polygons as a searching method<sup>258</sup>. This procedure starts with an initial triangle located in a topologically connected subset of the AFS, and an automatic extrusion algorithm is then used to form a sequence of growing polygons that approximate the AFS from the interior. For a three-component system ( $s = 3$ ), the AFS is a two-dimensional geometric region. For a numerical computation of the AFS two methods have been developed: the triangle boundary-enclosure scheme and the polygon inflation method. FAC-PACK is a MATLAB toolbox for the computation of non-negative multi-component factorizations and for the numerical approximation of the Area of Feasible Solutions (AFS) using the inflation polygon algorithm<sup>258</sup>. Effective tools included in FAC-PACK are the reduction of the AFS by a factor-locking approach and the use of the complementarity (duality) theory. This software comes with a Graphical User Interface (GUI) in MATLAB. The software implements the polygon inflation algorithm and the generalized Borgen plots. Non-negative Matrix Factorization (NMF) is used as the starting point for the polygon inflation algorithm in FAC-PACK. It supplies two or three points within the AFS depending on the number of components of the system. From these points an initial polygon can be constructed, which is a first approximation of the AFS. The initial polygon is inflated to the AFS by means of an adaptive algorithm. This algorithm allows computing all three segments of an AFS separately.

### **MCR-BANDS**

Gemperline<sup>255</sup> presented years ago an alternative method which can be applied to any multivariate data set, irrespective of the level of overlap and of the number of components. This method is the basis of the MCR-BANDS method used in this Thesis for the evaluation of the extent of the rotation ambiguities. MCR-BANDS method has been proposed by R. Tauler<sup>255</sup> and it estimates the extent of rotation ambiguities producing a range of feasible solutions based on a fast maximization and minimization of the relative signal component contribution of each component (SCCF). MCR-BANDS method has no limitation to a number of components and uses the same constraints as those applied to

find the MCR solution. It gives a simple evaluation of the extent of rotation ambiguity from the difference of the value of the optimization function giving maximum and minimum values.

MCR-BANDS evaluates for each component the minimum and maximum the relative signal contribution function (SCCF) <sup>246</sup>. This method can be used to evaluate the extent of rotation ambiguities associate to a particular MCR solution and under a set of constraints. The method is based on the maximization and minimization of the signal component contribution function, SCCF, which is defined for component n as:

$$\mathbf{SCCF}_n = \frac{\|\mathbf{c}_n \mathbf{s}_n^T\|}{\|\mathbf{C} \mathbf{S}^T\|} \quad \text{Equation 2- 23}$$

$\mathbf{SCCF}_n$  is a scalar value between 0 and 1, which gives the relative signal contribution of a certain component in relation to the whole signal due to the mixture of N components ( $n=1\dots N$ ).  $\mathbf{C}$  and  $\mathbf{S}^T$  are as defined above;  $\mathbf{c}_n$  and  $\mathbf{s}_n$  are the nth column and row of  $\mathbf{C}$  and  $\mathbf{S}$  matrices respectively.  $\|\cdot\|$  is the Frobenious norm  $\|\mathbf{C} \mathbf{S}^T\|$  gives the signal contribution of all the components present in the whole image and  $\|\mathbf{c}_n \mathbf{s}_n^T\|$  only the contribution of component n <sup>259</sup>.

For every component, SCCF is maximized and minimized under the set of considered constraints and equal data fitting. When the maximum and minimum values of SCCF of a particular component N are practically equal, it means that for this component there is practically no ambiguity remaining. On the contrary when this difference is large and close to one (SCCF is scaled between 0 and 1), then there is a large amount of ambiguity. See references <sup>246</sup> for more details about how this procedure is implemented and works. MCR-BANDS performs the minimization and maximization of SCCF using a general sequential quadratic programming procedure to solve the non-linear optimization of SCCF under non-linear constraints <sup>246, 259, 260</sup>. The amount of ambiguity resting for a particular solution can be evaluated from the differences between the maximum and minimum values of the SCCF function <sup>259</sup>.

The general idea of MCR-BANDS is finding the boundary of SCCF function that provides the maximum and the minimum relative contribution of the considered component signal to the overall signal measured. The maximum and minimum boundaries of this function should correspond to profiles that fulfill the constraints and should reproduce the data set with the optimal fit. This approach is valid for data sets with unlimited number of components. The following equations can be written:

$$\mathbf{D} = \mathbf{C}_{\text{inic}} \mathbf{S}_{\text{inic}}^T = \mathbf{C}_{\text{inic}} \mathbf{T}_{\text{min}} \mathbf{T}_{\text{min}}^{-1} \mathbf{S}_{\text{inic}}^T = \mathbf{C}_{\text{min},n} \mathbf{S}_{\text{min},n}^T = \mathbf{C}_{\text{inic}} \mathbf{T}_{\text{max},n} \mathbf{T}_{\text{max},n}^{-1} \mathbf{S}_{\text{inic}}^T = \mathbf{C}_{\text{max},n} \mathbf{S}_{\text{max},n}^T$$

Equation 2- 24

It is clear that if no constraints are considered, there is an infinite number of possible solutions of Equation 2- 24 for any non-singular matrix  $\mathbf{T}$ . It is usually possible to reduce considerably this infinite number of possible solutions by means of constraints derived from the physical nature and previous knowledge of the problem under study.

The boundaries will be related to specific rotation matrices  $\mathbf{T}$  for every component  $n$ , which are called  $\mathbf{T}_{\text{max},n}$  and  $\mathbf{T}_{\text{min},n}$ . Consider a particular set of solutions fulfilling the constraints defined by the problem,  $\mathbf{C}_{\text{inic}}$  and  $\mathbf{S}_{\text{inic}}^T$ . Feasible solutions  $\mathbf{C}_{\text{inic}}$  and  $\mathbf{S}_{\text{inic}}^T$  can be initially postulated or found using the MCR-ALS under a set of preselected constraints. The maximum band boundaries  $\mathbf{C}_{\text{max},n}$  (for the concentration profiles) and  $\mathbf{S}_{\text{max},n}^T$  (for the spectral profiles) and the minimum band boundaries  $\mathbf{C}_{\text{min},n}$  (for the concentration profiles) and  $\mathbf{S}_{\text{min},n}^T$  (for the spectral profiles) can be defined by the Equation 2- 24. The goal of the method described here is to find a way to calculate these values of  $\mathbf{T}_{\text{max},n}$  and  $\mathbf{T}_{\text{min},n}$  which define the maximum and minimum boundaries of the SCCF function which define feasible solutions of the component profiles under a set of constraints defined for a particular data set.

MCR-BANDS has been applied to the evaluation of the extent of rotation ambiguities associated to MCR-ALS solutions obtained in the analysis of hyperspectral images (see Chapters 5 of this Thesis). In the general case of the analysis of complex multi-component hyperspectral images with soft constraints like non-negativity, unique solutions are not guaranteed and rotation ambiguities may be present. Because of these ambiguities, a set of feasible solutions, fulfilling the applied constraints (e.g. non-negativity constraints) fitting equally well the analyzed data are possible.

MCR-BANDS code can be downloaded from the <http://www.mcrals.info/>



## **Chapter 3**

# **Application of MCR-ALS on remote sensing hyperspectral imaging**



## Introduction

In this chapter, different simulated datasets and other data sets (free AVIRIS, Airborne Visible/Infrared Imaging Spectrometer data sets) from NASA (The National Aeronautics and Space Administration) were analyzed by means of MCR-ALS.

The simulated data sets were prepared using pure spectra (signatures) from USGS (The U.S. Geological Survey) library. This spectra library includes samples of minerals, rocks, soils, physically reconstructed, as well as mathematically computed mixture spectra, and spectra from plants, vegetation communities, microorganisms, and man-made materials. The spectra collected from this dataset were assembled in this Thesis with the purpose of simulating the use of spectral features from remote sensing detection of these and similar materials<sup>261</sup>.

AVIRIS is a proven instrument in the realm of Earth Remote Sensing. It is a unique optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (bands) with wavelengths from 400 to 2500 nanometers. AVIRIS has been installed on four aircraft platforms: NASA's ER-2 jet, Twin Otter International's turboprop, Scaled Composites' Proteus, and NASA's WB-57. The ER-2 jet flies at approximately 20 km above sea level, at about 730 km/hr. The Twin Otter aircraft flies are acquired at 4 km above ground level at 130 km/hr. AVIRIS has flown North America, Europe, portions of South America, and Argentina. The dataset we used in this Thesis was obtained by NASA's ER-2 jet.

The main objective of the AVIRIS project is to identify, measure, and monitor constituents of the Earth's surface and atmosphere based on molecular absorption and particle scattering signatures. Research with AVIRIS data is predominantly focused on understanding processes related to the global environmental and climate change.

The first dataset we have used from AVIRIS is a hyperspectral image of the geographical region of Cuprite, Nevada, USA, obtained in 1997<sup>262</sup>. In this case, the analysis was limited to a 614 × 512 pixels sub-image. The Cuprite area is an arid area with a small amount of vegetation cover, and it has an excellent rock exposure, comprising alteration zones characterized by the occurrence of key indicator minerals. In this area, the use of hyperspectral remote sensing of minerals is well documented. Many minerals can be identified from airborne images, and their correlation with the presence of valuable minerals has been established. NASA has catalogued several of the possible minerals present in this area, as well as their spectral signatures. This spectroscopic image has been widely used for remote sensing experiments and for development and application of imaging data analysis methods. Goetz and Srivastava (1985) were the first to use hyperspectral data for mineral mapping at Cuprite<sup>263</sup>.

Dataset 2 is extracted from the hyperspectral image of a naval airport in San Diego, California, collected by AVIRIS sensor. In this case, the analysis was limited to a  $200 \times 200$  pixels sub-image. This image has 195 useful spectral bands.

Dataset 3 was acquired over Moffett Field, CA, in 1997 by the AVIRIS sensor. In this case, the analysis was limited to a  $400 \times 400$  pixels sub-image, and this image has 199 useful spectral bands. The investigated area is mainly composed of water, building, soil, and vegetation.

Methods for analyzing hyperspectral data were discussed in Chapter 2 of this Thesis (see Chapter 2, section 2.2). Many methods have been proposed for remote sensing hyperspectral imaging, such as the Pixel Purity Index (PPI)<sup>114</sup>, the Vertex Component Analysis (VCA)<sup>264</sup>, the Minimum Volume Simplex Analysis (MVSA)<sup>188</sup>, and others<sup>265</sup>. We propose in this Thesis the use of MCR-ALS for the resolution of remote sensing hyperspectral imaging datasets.

The way MCR-ALS has been used to resolve hyperspectral imaging data obtained from remote sensing was described in Chapter 2 of this Thesis. Two research papers have been written in the frame of this Thesis describing the application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging, emphasizing the use of local rank-based spatial information and of constraints for the improvement of remote sensing hyperspectral imaging resolution. Remote sensing has its own characteristics and differences compared to ordinary hyperspectral imaging. Normally, in remote sensing hyperspectral imaging is possible to find out some purest pixels that can be used for the application of local rank constraint. In this work, a new procedure based on the use of correlation coefficients with known reference spectra was proposed to detect pure pixels. Details about how local rank constraints are applied have been described in the Part 2 of this Thesis and in the methods section of the two papers:

**The two published papers of this chapter are:**

**Paper 1:** Zhang, X.; Tauler, R., Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. *Analytica chimica acta*. 2013, 762, 25-38.

**Paper 2:** Zhang, X.; Juan, A.; Tauler, R., Local rank-based spatial information for improvement of remote sensing hyperspectral imaging resolution. Submitted to *Talanta*.



Contents lists available at SciVerse ScienceDirect

Analytica Chimica Acta

journal homepage: [www.elsevier.com/locate/aca](http://www.elsevier.com/locate/aca)

## Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging<sup>☆</sup>

Xin Zhang, Romà Tauler<sup>\*</sup>

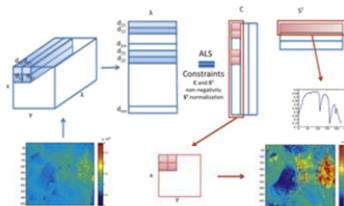
IDAEA-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain

### HIGHLIGHTS

- ▶ MCR-ALS is successfully applied to remote sensing hyperspectral images.
- ▶ Pure spectra and relative concentrations of image constituents were obtained.
- ▶ MCA-ALS results are favorably compared to results obtained by MVSA and VCA methods.
- ▶ Physical constraints were implemented to decrease the rotational ambiguities.
- ▶ MCR-BANDS is used to evaluate the presence and extent of rotational ambiguities.

### GRAPHICAL ABSTRACT

This article summarizes the use of the MCR-ALS method as a powerful tool for the resolution of hyperspectral images on their constituents. Non-negativity, spectral normalization and local rank constraints are used to get physical meaningful resolution results and decrease rotation ambiguities. MCR-BANDS method is used to evaluate the presence and extend of rotational ambiguities.



### ARTICLE INFO

**Article history:**  
 Received 17 September 2012  
 Received in revised form  
 14 November 2012  
 Accepted 26 November 2012  
 Available online 3 December 2012

**Keywords:**  
 Multivariate Curve Resolution  
 Remote sensing hyperspectral imaging  
 MCR-ALS  
 MVSA  
 VCA  
 Endmember

### ABSTRACT

The application of the MCR-ALS method is demonstrated on two simulated remote sensing spectroscopic images and on one experimental reference remote sensing spectroscopic image obtained by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). By application of MCR-ALS, the spectra signatures of the pure constituents present in the image and their concentration distribution at a pixel level are estimated. Results obtained by MCR-ALS are compared to those obtained by other methods frequently used in the remote sensing spectroscopic imaging field like VCA and MVSA. In the case of the analysis of the experimental data set, the resolved pure spectra signatures were compared to reference spectra from USGS library for their identification. In all cases, results were also evaluated for the presence of rotational ambiguities using the MCR-BANDS method. The obtained results confirmed that the MCR-ALS method can be successfully used for remote sensing hyperspectral image resolution purposes. However, the amount of rotation ambiguity still present in the solutions obtained by this and other resolution methods (like VCA or MVSA) can still be large and it should be evaluated with care, trying to reduce its effects by selecting the more appropriate constraints. Only in this way it is possible to increase the reliability of the solutions provided by these methods and decrease the uncertainties associated to their use.

© 2012 Elsevier B.V. All rights reserved.

### 1. Introduction

Hyperspectral imaging is a breakthrough in remote sensing technology [1], which can be obtained by Raman, infrared and fluorescence spectroscopies [2], and it is a useful methodology that can be applied for analytical purposes to agriculture [3], biology [4], environmental [5] and other earth science fields [6]. It has emerged as a very important field in recent years [7]. Remote sensing is the

<sup>☆</sup> Paper presented at the XIII Conference on Chemometrics in Analytical Chemistry (CAC 2012), Budapest, Hungary, 25–29 June, 2012.

<sup>\*</sup> Corresponding author. Tel.: +34 93 4006140; fax: +34 93 2045904.  
 E-mail address: [Roma.Tauler@idaea.csic.es](mailto:Roma.Tauler@idaea.csic.es) (R. Tauler).

acquisition of information about an object or phenomenon, without making physical contact with the object. In modern usage, the term generally refers to the use of aerial sensor technologies to detect and classify objects on Earth. Remote sensing hyperspectral imaging is a combination of both technologies, hyperspectral imaging and remote sensing, which provides spectral and spatial analytical information about a particular geographical area. Based on the resolved pure spectra of the constituents of the image and on their constituents distributions, additional information can be derived, such as the possible identification of these constituents from available libraries and their relative quantitation within the image. One of the implications of remote sensing hyperspectral imaging is that at every pixel of the measured image, a mixture of several spectral signatures of different materials is generally present. Resolution of the hyperspectral image is required to extract these signatures (pure spectra) and to figure out how is the distribution of the corresponding image constituents.

Hyperspectral remote sensing includes large data sets generally composed of about 100–200 spectral bands of relatively narrow bandwidths for about 5–10 nm. Hyperspectral images are represented in the form of data cubes. The spatial information is collected in the  $X$ – $Y$  plane, and the spectral information is represented in the  $Z$ -direction. The analysis of these multidimensional datasets requires sensitive detectors, fast computers, and large data storage capacities, potentially exceeding hundreds of megabytes. PCA (principal component analysis) [8] allows processing spectroscopic image datasets by reducing their dimensions without a significant loss of relevant information. PCA is a useful tool to estimate the number of more significant components (constituents) of the image and to remove noise and non-informative parts of it. However, because of the maximum variance criterion and of applied orthogonal constraints, PCA do not provide the true signatures (spectra) and relative concentration profiles of the image constituents directly.

The concept of endmember has been proposed, to refer to the pure constituent spectra present in the image pixels. Diverse mathematical methods based on convex geometry and subspace projection like the PPI (Pixel Purity Index) [9],  $N$ -FINDER ( $N$ -finder) [10], and VCA (Vertex Component Analysis) [11] methods have been proposed and applied. PPI projects the vector of pixels on unit vectors selected at random directions, and counts for the number of times the value of each projected pixel reaches an extreme value (this extreme value either can be a maximum or a minimum projected value). The endmembers are identified as those pixels with the highest scores. The  $N$ -FINDER method is based on the fact that in  $N$  spectral dimensions, the  $N$ -volume contained by a simplex formed by the purest pixels is larger than any other volume formed from any other combination of pixels. It begins with an initial simplex composed by a set of random pixels, then iteratively the simplex volume is increased inside the data set until the simplex with the smallest volume containing all data pixels is found. VCA projects the data to the identified orthogonal subspace in an interactive way, and finds the endmembers by repeated iteration. The endmember are at the vertices of the simplex. In all cases, these methods find the most likely pure pixels by an approximate method, but they do not estimate the smallest simplex (or the convex hull) directly. All these methods are based on the assumption that pure pixels do exist in the measured dataset, which of course may not be the general case for most of the natural systems and situations. More recently MVSA (minimum volume simplex analysis) [12] which is based on geometrical image analysis [13] and AMEE (Automated Morphological Endmember Extraction) [14] which is based on morphological image analysis have been reported too. The MVSA method has been proposed to cope with the situation where no pure pixel exists in the measured image. It is also based on the concept of the minimum simplex volume estimation,

unmixing the image by fitting a minimum volume simplex to the data, constraining the abundance fractions to be positive and belong to the most probable simplex. It is a fast method, but, in its present implementation, it does not guarantee non-negative spectra, which would not have physical nor chemical sense.

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) has been proposed and extensively used to resolve multiple pure responses and concentrations of the components present in unknown mixtures [15]. It has been applied to analyze multicomponent chemical systems like chemical reactions [16], industrial processes [17], chromatographic coelution problems [18], spectroscopic mixtures [19], environmental monitoring data [20], and it can be applied to many other type of mixture analysis cases. MCR-ALS has been reported also to be a useful method for the resolution and segmentation of hyperspectral biomedical and other type of hyperspectral imaging [21–25]. In MCR-ALS, the measured analytical signals are assumed to follow a generalized bilinear additive model (like the extension of Beer's law in absorption spectroscopy [26]). The contribution of each component to the measured signal depends on its concentration and on its own spectral sensitivity response (pure spectrum). MCR-ALS can also be applied to obtain quantitative information and it provides physical and chemical meaningful solutions. This is accomplished because in MCR-ALS, any type of constraints can be easily applied to the sought solutions, like non-negativity [27,28], unimodality [29,30], local rank [31,32], and trilinearity [29,33].

In this work, three different spectroscopic images have been analyzed to test the MCR-ALS method and to compare its results with those obtained by VCA and MVSA methods. Two of these images are simulated data sets and the third one is an experimental remote sensing hyperspectral airborne image from the Cuprite area in Nevada (USA). Finally, the MCR-BANDS method [34] is applied to evaluate the amount of rotation ambiguity associated to the image resolved pure spectra and concentration profiles under the constraints imposed during their estimation.

## 2. Methods

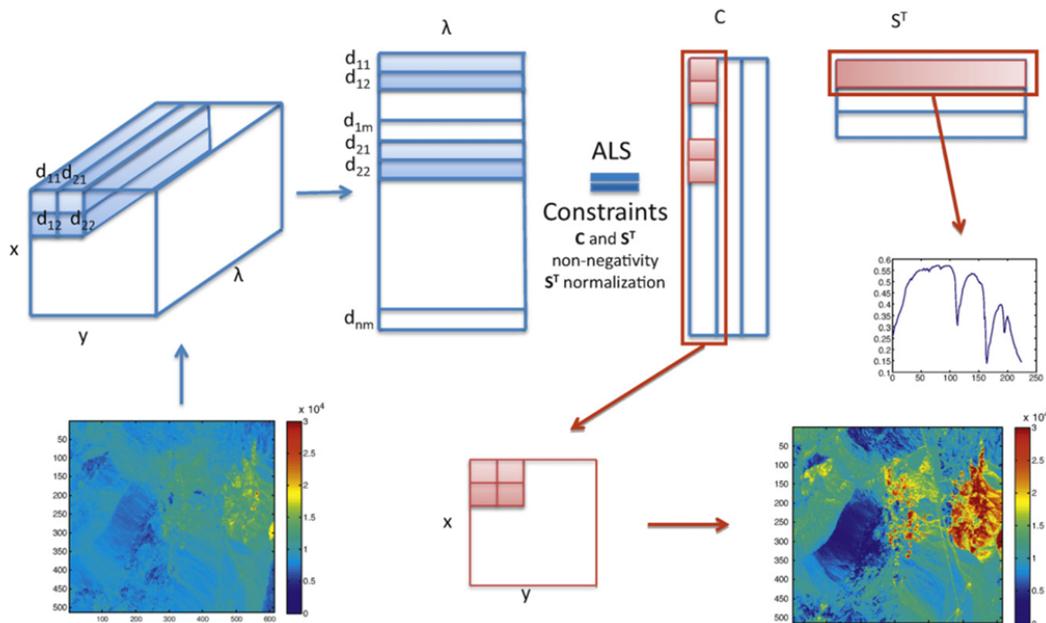
### 2.1. MCR-ALS method

The MCR-ALS method is used to decompose the hyperspectral image into the signatures or pure spectra of the image constituents and into their concentration (relative amounts) on the image (distribution map). MCR-ALS is based on a bilinear model which assumes that the observed spectra are a linear combination of the spectra of the pure components in the system [21,26]. This model can be written in matrix form as:

$$\mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad (1)$$

where  $\mathbf{D}$  is the reshaped image data matrix.  $\mathbf{C}$  is the matrix of the relative amounts or concentrations,  $\mathbf{S}^T$  is the pure spectra.  $\mathbf{E}$  is the matrix associated to noise or experimental error.

In hyperspectral imaging using visible/infrared spectroscopies, the intensity or absorption of the radiation in every pixel should not be negative and neither should be negative the concentrations of the different constituents of the image (physical constraints). Thus, during the alternating least squares procedure previously described, non-negative constraints are applied on both, on the pure spectra and on the image concentrations. Moreover, to avoid scale indeterminacies and stabilize the ALS iterative optimization, spectra matrix  $\mathbf{S}^T$  is normalized. Instead of spectra normalization, an alternative possible constraint to be used to scale adequately the obtained solutions is a mass balance or closure condition on  $\mathbf{C}$  matrix, i.e., that concentrations or relative amounts of the constituents on the pixels sum equal to a constant value. This is for



**Fig. 1.** Bilinear matrix decomposition of a hyperspectral image using the MCR-ALS method. Resolution of spectra (signatures) and of 2D image concentrations of the pure components. In the left bottom corner, the raw USGS Airborne Visible Infrared Imaging Spectrometer (AVIRIS) remote sensing spectroscopic image experimental data set [42] analyzed in this work. Spectroscopic image obtained. In the right bottom corner, 2D image concentrations for one of the components resolved by MCR-ALS.

instance one of the constraints used by the MVSA method (see below). At the end of the ALS optimization procedure, the finally obtained **C** matrix can be reshaped again as shown in Fig. 1, to recover the 2D image (distribution map) of the concentrations (relative amounts) of every component at every pixel of the image plane.

Concentration profiles from hyperspectral images are significantly different from those get usually in chemical process analysis which tend to show continuous smooth shapes. Usually used constraints, like unimodality, closure or hard-modeling resulted not to be adequate in spectroscopic imaging. On the other hand, the use of local rank constraints [29,35] can be extremely helpful to improve the reliability of MCR-ALS solutions. When local rank constraints are applied, the chemical rank (equal to mathematical rank in absence noise) of a small number of contiguous pixel windows is estimated and used to improve resolution. From local rank information, it is possible to infer where in the image the possible constituents are or are not present. Fixed Size Moving Window Evolving Factor Analysis can be applied for this purpose [35], to detect in which pixel a particular component is present or is missing. In this work, however, the different image constituents have very similar, quasi collinear, spectra profiles or signatures, being therefore very difficult to detect and find out what pixels can have a particular component missed.

To evaluate the quality of the data fitting finally achieved after application of the MCR-ALS procedure, the percentage of lack of fit (*lof*) (Eq. (2)) and the percentage of explained variance ( $R^2$ ) (Eq. (3)) are calculated according to the two following equations:

$$lof (\%) = 100 \times \sqrt{\frac{\sum_{ij} d_{ij} - \hat{d}_{ij}}{\sum_{ij} d_{ij}^2}} \quad (2)$$

where  $d_{ij}$  is the element of the hyperspectral image data matrix **D**, and  $\hat{d}_{ij}$  is the corresponding element of this data matrix recalculated by ALS,  $\hat{\mathbf{D}}$ . This lack of fit value gives a measure of the fit quality in relative terms with the same units as the measured data, and comparable with experimental relative error estimations.

For the explained variances,  $R^2$  is calculated as:

$$R^2 = 100 \times \left( 1 - \frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2} \right) \quad (3)$$

where  $e_{ij}$  are the elements of the **E** matrix and  $d_{ij}$  are the elements of the raw data set **D**.

Additionally, the correlation ( $r^2$ ) between a particular resolved profile and the true profile (in simulations the true profiles are known) and the vector angles between them can be evaluated using Eqs. (4) and (5),

$$r^2 = \frac{\mathbf{xy}^T}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (4)$$

$$\text{Angle} = \frac{180}{\pi} \times \arccos \left( \frac{\mathbf{xy}^T}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \quad (5)$$

where **x** is the vector of resolved profiles and **y** is the vector of true profiles.

All these parameters can be also easily calculated for the solutions obtained by the other methods tested in this work, like VCA and MVSA (see below).

MCR-ALS algorithm code and GUI for MATLAB (The Mathworks Inc., MA, USA) is freely available from the home page of MCR at <http://www.mcrals.info/>.

## 2.2. VCA method

VCA is an unsupervised method based on the geometry of a convex data set which considers that the pure image constituents are located at the vertices of a simplex. For example, in a mixture defined by three constituents, the simplex boundaries are a triangle, whose vertices correspond to the pure constituents or endmembers. VCA assumes that, for every constituent, exists one or more pixels in the image where only this constituent exists (pure pixels). To find out these pure pixels and the spectral signatures associated to them, VCA iteratively projects the data onto a direction which is orthogonal to the subspace spanned by the endmembers already estimated. The new endmember signatures are located at the extreme of the new projection. This iteration procedure goes on until all the endmember signatures are estimated. A more detailed description of the method can be found in [11].

## 2.3. MVSA method

MVSA also considers that the underlying mixing model is linear, i.e., that the mixed hyperspectral vectors are a linear combination of the signatures (spectra) of the pure components. MVSA resolves the hyperspectral image by fitting a minimum volume simplex to the hyperspectral data, under constraints (for every pixel) of no less than zero abundance fractions and sum equal to one [12]. Differently to MCR-ALS, MVSA does not use a least squares approach, but a sequential quadratic programming (SQP) method, based on a quasi-Newton non-linear optimization procedure under linear constraints. In order to prevent local minima and to achieve a faster speed, VCA is usually applied as initial step to provide an inflated version of the initial simplex [12]. More details about the MVSA can be seen in Refs. [12,13].

MVSA and VCA MATLAB codes can be downloaded from: <http://www.lx.it.pt/~bioucas/code.htm>.

## 2.4. MCR-BANDS method

In the general case of complex multi-component hyperspectral images, solving Eq. (1) under a set of constraints like non-negativity, does not guarantee unique solutions, nor, that the achieved solutions are the true correct ones. Solutions obtained by all bilinear modeling methods like MCR-ALS (but also by MVSA or VCA) are often not the true correct ones because of the presence of rotational and intensity ambiguities, which are ubiquitously present in all factor analysis bilinear decompositions of data matrices. Because of these ambiguities, a set of feasible solutions, fulfilling the applied constraints (e.g., non-negativity constraints) fitting equally well the analyzed data are possible. In these cases, there is no better solution than others, because they are undistinguishable considering the previous knowledge of the system. The true solution is one among all the solutions fitting the data the same and fulfilling also the applied constraints. For instance, starting with different initial estimations in MCR-ALS (when only non-negativity constraints are applied) can give different final solutions, all equivalent from a data fitting point of view and fulfilling the applied constraints. Some methods like MVSA, when applied to the same data set converge always to the same final unique solution, but this does not guarantee that this solution be the true physical one, and as it will be shown in the results section, it may even happen that this final solution is not feasible because it does not fulfill the known constraints of the system, like spectra non-negativity. In order to evaluate the extent of rotation ambiguities associated to a particular MCR solution, the MCR-BANDS procedure has been proposed [34]. This method can be used to evaluate the extent of rotation ambiguities associate to a particular MCR solution and under a set of constraints. In this paper it is applied to

evaluate the extent of rotation ambiguities associated to MCR-ALS solutions, and as extension, also to VCA and MVSA solutions. The method is based on the maximization and minimization of the signal component contribution function, SCCF, which is defined for component  $n$  as:

$$SCCF_n = \frac{\|\mathbf{c}_n \mathbf{s}_n^T\|}{\|\mathbf{C} \mathbf{S}^T\|} \quad (6)$$

$SCCF_n$  is a scalar value between 0 and 1, which gives the relative signal contribution of a certain component in relation to the whole signal due to the mixture of  $N$  components ( $n = 1, \dots, N$ ).  $\mathbf{C}$  and  $\mathbf{S}^T$  are as defined above;  $\mathbf{c}_n$  and  $\mathbf{s}_n$  are the  $n$ th column and row of  $\mathbf{C}$  and  $\mathbf{S}$  matrices respectively.  $\|\cdot\|$  is the Frobenius norm  $\|\mathbf{C} \mathbf{S}^T\|$  gives the signal contribution of all the components present in the whole image and  $\|\mathbf{c}_n \mathbf{s}_n^T\|$  only the contribution of component  $n$  [36]. For every component, SSCF is maximized and minimized under the set of considered constraints and equal data fitting. When the maximum and minimum values of SSCF of a particular component  $n$  are practically equal, it means that for this component there is practically no ambiguity remaining. On the contrary when this difference is large and close to one (SCCF is scaled between 0 and 1), then there is a large amount of ambiguity. See Ref. [34] for more details about how this procedure is implemented and works. MCR-BANDS performs the minimization and maximization of SSCF using a general sequential quadratic programming procedure to solve the non-linear optimization of SSCF under non-linear constraints [34,36,37]. The amount of ambiguity resting for a particular solution can be evaluated from the differences between the maximum and minimum values of the SSCF function [36].

As said above, the MCR-BANDS method allows for the estimation of the amount of rotation ambiguity associated to a particular MCR solution under a particular set of constraints. In the case of spectroscopic images, non-negativity, spectra normalization and local rank constraints can be usually applied. Local rank constraint can be applied to one or more components, and in the extreme case to all the components of the investigated system. Local rank constraints depend on the previously information or estimation about those pixels where one or more components are absent or are present. Assessing for the extent of the feasible solutions region in the analysis and resolution of hyperspectral images, as in any other general mixture analysis problems, is a mathematically ill-posed problem which requires the use of state of the art non-linear optimization with non-linear constraints techniques [38] such as the one proposed in the MCR-BANDS procedure which requires the use *fmincon* MATLAB function of the *Optimization Toolbox* of MATLAB, (The Mathworks Inc., MA, USA).

MCR-BANDS code can be downloaded from the <http://www.mcrals.info/>.

## 3. Data

In this work, to evaluate the use of MCR-ALS, VCA and MVSA in the resolution of remote sensing hyperspectral imaging, two simulated datasets and one real airborne spectroscopic image dataset have been used. Simulated datasets included two examples, with different type of pixel relative amounts of the constituents (concentrations), one following a Dirichlet distribution and another one showing regular patterns between adjacent pixels.

### 3.1. Simulated spectroscopic image in data set 1

First simulated data set consists of a mixture of three spectral signatures with 224 bands extracted from the 1997 USGS Digital Spectral Library [39,40]. The whole data set includes 2000 pixels, where the concentration of each component in each pixel was

obtained following a Dirichlet distribution (see below). The whole size of the data set is 2000 pixels and 224 wavelengths, ( $\mathbf{D}$  matrix has dimensions  $2000 \times 224$ ). The spectroscopic image of this simulated dataset was built according to the bilinear model given in Eq. (1), where  $\mathbf{D}$  is the spectroscopic image dataset;  $\mathbf{S}^1$  is the three pure spectral signatures matrix from the spectral library.  $\mathbf{C}$  (concentration matrix), is generated first using the gamma  $\Gamma$  random function (*gamrand* MATLAB m function), and then the following equation:

$$\mathbf{C} = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}, \quad \alpha = (\alpha_1, \dots, \alpha_K) \quad (7)$$

where  $\mathbf{C}$  is concentration matrix.

The simulated data using this equations have a random pattern and they did not have any pure pixel, i.e., more than one constituent is present in all pixels, with abundance fractions (relative amounts or concentrations) not larger than 0.8. The sum of these abundance fractions was constant for all pixels and equal to one. This data image is very well suited to compare results and performances of MCR-ALS and MVSA methods, since data concentrations do agree with MVSA applied constraints.

Homocedastic noise,  $\mathbf{N}$ , (of zero-mean and constant variance equal to 0.01%) was added to simulated data. In the way this image was created, there were some pixels for one of the components at low concentrations. For instance for component 1, component 2 and component 3 there were some pixels where these components were at very low concentration values. In these pixels one of the image constituents can be considered to be absent. This information can be used to decrease rotation ambiguities when is implemented as a constraint in MCR-ALS and MCR-BANDS methods.

### 3.2. Simulated spectroscopic image 2 in data set 2

The second data set simulates a hyperspectral image with a regular pattern distribution of three different chemical components. The three spectral signatures ( $\mathbf{S}^1$ ) were extracted from the 1997 USGS Digital Spectral Library. The dataset has  $33 \times 33$  pixels and the spectra have 224 wavenumbers. This image shows different scenes for each component with a certain shape (see Fig. 3). Eq. (1) is used to calculate the data matrix,  $\mathbf{D}$ , having the spectroscopic image for data set 2; concentrations,  $\mathbf{C}$ , or relative amounts of the constituents in adjacent pixels follow a certain trend which made the image more natural. These concentrations follow the relationship given in the equations:

$$\mathbf{C}_1 = \frac{\sin \sqrt{x_{ij}^2 + y_{ij}^2}}{\sqrt{x_{ij}^2 + y_{ij}^2}} \quad (8)$$

$$\mathbf{C}_2 = x e^{(-x_{ij}^2 - y_{ij}^2)} \quad (9)$$

$$\mathbf{C}_3 = x e^{(1/2)(-x_{ij}^2 - y_{ij}^2)} \quad (10)$$

where  $x$  and  $y$  values were generated using the meshgrid MATLAB function. They give the regular positions of the pixels in the two directions of the  $x$ - $y$  image plane. In Appendix A, MATLAB codes for generation of the concentration matrix,  $\mathbf{C}$ , of the 3 components are given. In the way this image was created, there were some pixels with local rank properties. For component 1, component 2 and component 3, there were pixels where only the concentration of one of the three constituents was at high concentrations, and the other two remaining were at much lower concentration. This information can be used to eliminate rotation ambiguities when is implemented as a constraint in MCR-ALS and MCR-BANDS methods. As in previous example, experimental noise was added to the

data (white Gaussian noise with zero-mean and constant variance equal to 0.01%).

### 3.3. Experimental remote sensing spectroscopic image data set 3

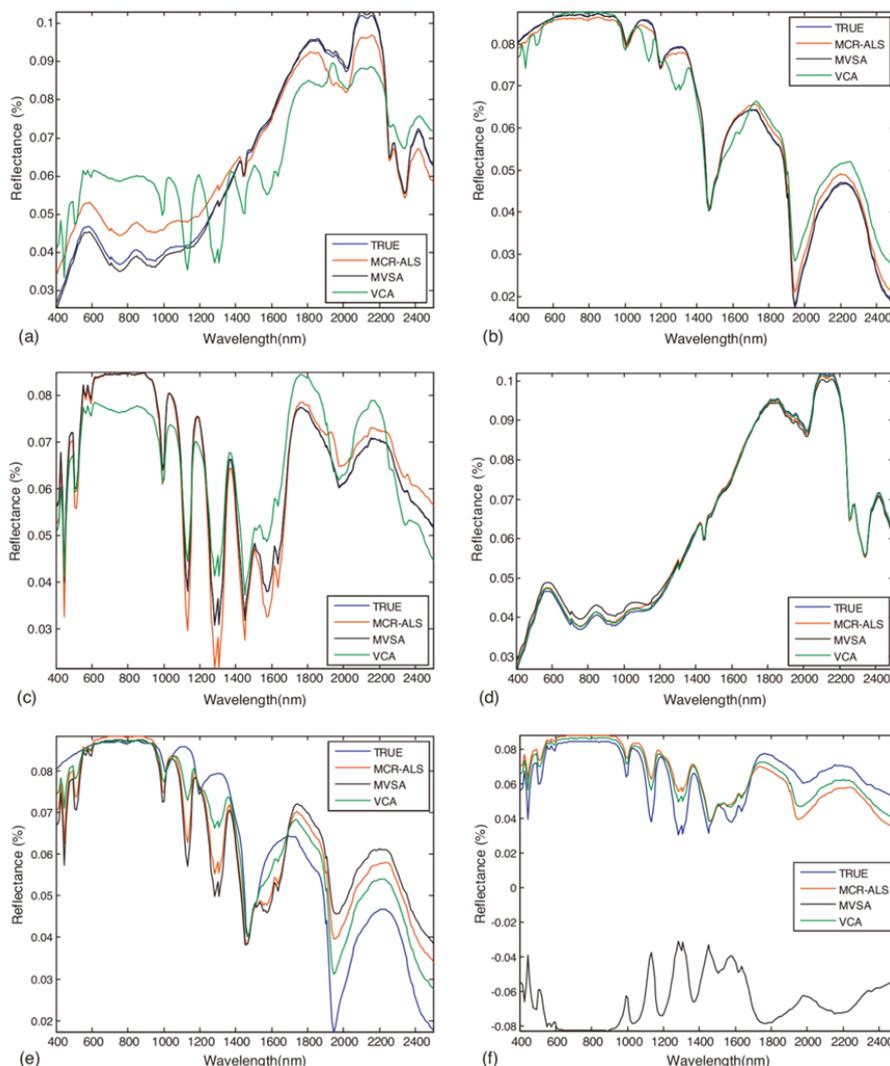
Additionally, resolution and quantification of a real experimental hyperspectral image was also performed using the three tested methods, MCR-ALS, MVSA and VCA. This is a hyperspectral image of the geographical region of Cuprite, Nevada, USA, obtained in 1997 [41]. It has a 17 m per pixel footprint with 224 spectral bands over the 400–2500 nm vis-NIR spectral region. The whole size of the data set is  $614 \times 512 \times 224$ . These images were obtained by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [42], AVIRIS, on NASA's ER-2 jet aircraft platform. It has approximately 10 nm spectral resolution and 20 m spatial resolution. The Cuprite area is an arid area with a small amount of vegetation cover, and it has an excellent rock exposure, comprising alteration zones characterized by the occurrence of key indicator minerals. In this area, the use of hyperspectral remote sensing of minerals is well documented. Many minerals can be identified from airborne images, and their relation to the presence of valuable minerals has been established. NASA has cataloged several of the possible minerals present in this area as well as their spectral signatures. This spectroscopic image has been widely used for remote sensing experiments and methods. This image is shown in the right corner of Fig. 1.

## 4. Results

### 4.1. Simulated spectroscopic image in data set 1

Before analyzing this data set with MCR-ALS, initial estimates of the three spectral signatures were first obtained by SIMPLISMA. SVD confirmed the presence of 3 components, since three larger singular values were clearly distinguished from the rest which were explaining data noise. Non-negativity constraints were applied to both spectra and concentration profiles of the pure components, and pure spectra were normalized.

MCR-ALS resolved spectra were compared with those used for the data simulation. In Fig. 2 above, the comparison of MCR-ALS, MVSA, VCA resolved spectra obtained in the analysis of the simulated data set 1 are compared with the true spectra used for the simulation (from USGS reference library). Blue lines are true profiles, red lines are MCR-ALS resolved spectra, black lines are MVSA resolved spectra, green lines are VCA resolved spectra. In order to perform such a comparison, spectra resolved by MVSA and VCA were renormalized in the same way as for MCR-ALS, to have their Frobenius norm equal to one. This is necessary due to the fact that MVSA and VCA do not apply such normalization as a constraint (total sum equal to one of the abundance fractions is used instead). As it was explained above, in this particular case, the simulated data set 1 was built according to this closure condition. In Fig. 2a–c, results for the first, second and third component are given respectively. For the first component, MVSA resolved spectra (black lines) were the closest to the true ones (blue lines), MCR-ALS spectra (red lines) were also very similar and VCA (green lines) gave mixed spectral bands. For the second and third, the same situation is observed. In Table 1, the comparison between the results (for both, spectra and concentration profiles) obtained by the three different methods (MCR-ALS, VCA and MVSA) is expressed numerically using correlation coefficients, ( $r^2$ ) and angle values, between true and resolved pure spectra and concentration profiles. MVSA gave the closest to one  $r^2$  values and the smaller angles for the three components. MCR-ALS gave also very good results and results of VCA were the worst ones. VCA results were the less accurate because there were



**Fig. 2.** (Upper part) Comparison between the spectra used for the data simulation (blue, from USGS reference library), with those obtained by MCR-ALS (red), MVSA (black) and VCA (green) in the analysis of the simulated data set 1 for the first (a), second (b), and third (c) image constituents. (Bottom part) Comparison between the spectra used for the data simulation (blue, from USGS reference library), with those obtained by MCR-ALS (red), MVSA (black) and VCA (green) in the analysis of the simulated data set 1 for the first (d), second (e), and third (f) image constituents. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

no pure signatures or spectra (there were no pixels with a single component) in this simulated dataset.

In Table 2 model fitting results expressed as lack of fit (*lof*) and explained variance ( $R^2$ ) values (Eqs. (2) and (3)) are given for the whole model including the three resolved components, by MCR-ALS, MVSA and VCA. MCR-ALS and MVSA lack of fit for the full model were in both cases of only 0.99%, which are significantly lower (better) than the lack of fit value for VCA, which was 4.43%. In terms of explained data variances,  $R^2$ , the three methods gave more similar results (MCR-ALS and MVSA explained 99.99% of the data variance and VCA 99.8%). Explained variances by the individual components given in Table 2 are rather high for the three components and they

sum more than when they were considered together in the full model (up to 148%), which reflects the extremely high overlapped variance among them. From all these results, it is concluded that the results obtained by MVSA and MCR-ALS in the analysis of data set 1 were rather similar and close to the true ones. Small differences observed between solutions obtained by MCR-ALS and MVSA compared to true ones are due to possible rotation ambiguities when only non-negativity constraints were applied. MVSA and MCR-ALS solutions are equivalent from data fitting point of view (*lof* of 0.99% in Table 2) and they both fulfill the known constraints of the system (non-negativity). See below in the MCR-BANDS section, a deeper discussion about rotation ambiguities and how to decrease them.

**Table 1**  
Correlation coefficients ( $r^2$ , Eq. (4)) and angles (Eq. (5)) between true profiles and component profiles obtained by MCR-ALS, MVSA and VCA in the analysis of the two simulated data sets.

	Spectra <sup>a</sup>			Concentration <sup>b</sup>		
	MCR-ALS	MVSA	VCA	MCR-ALS	MVSA	VCA
<i>Data set 1</i>						
$r^2$	0.9970	0.9998	0.9810	0.9905	0.9999	0.9783
	0.9997	0.9999	0.9975	0.9920	0.9999	0.9944
	0.9988	0.9999	0.9955	0.9942	0.9999	0.9255
Angle	4.40	0.87	11.18	7.90	0.39	11.96
	1.37	0.81	4.07	7.25	0.39	6.05
	2.84	0.10	5.47	6.15	0.68	22.26
<i>Data set 2</i>						
$r^2$	0.9999	0.9997	1.0000	0.9910	0.9999	1.0000
	0.9882	0.9810	0.9958	0.9906	0.9967	0.9936
	0.9845	-0.9998	0.9925	0.9083	0.1781	0.5932
Angle	0.63	1.40	0.36	7.69	0.69	0.06
	8.81	11.19	5.25	7.86	4.68	6.50
	10.09	178.74	7.04	24.73	79.74	53.62

<sup>a</sup> Spectra profiles used for data simulation were obtained from USGS reference library [41].

<sup>b</sup> Concentration profiles used for data simulation were obtained using Eqs. (7)–(10) (simulated data sets 1 and 2) and the MALAB codes given in Appendix A (simulated data set 2).

#### 4.2. Simulated spectroscopic image in data set 2

MCR-ALS was applied to this second data set using non-negative and spectra normalization constraints as in the previous case. In Fig. 2 bottom, spectra profiles resolved by MCR-ALS, MVSA and VCA are given, and compared with the true spectra profiles used for the data stimulation. Red lines are MCR-ALS profiles, black lines are MVSA profiles and green lines are VCA ones. According to lower part of Table 1, first component was properly resolved by MCR-ALS, MVSA and VCA, with  $r^2$  values very close to one in the three cases. VCA, in this case, got an excellent recovery of this spectrum, because there were pixels near to pure for this component in the raw dataset. For the second component, resolved spectra by all of the three methods showed some distortion [43]. The three methods mixed some of the bands from the other components. This is a consequence of the presence of rotation ambiguity not solved when only non-negativity constraints were applied (see Section 2.4) and of the absence of pure pixels. For the third component, it is worth to pay attention that MVSA failed and gave a wrong spectrum profile with a similar symmetrical shape to the true profile, but with unreasonably negative values. The optimization problem solved by MVSA minimized the simplex volume giving negative values for the spectrum of this component. In its current implementation, MVSA method could not apply non-negative constraints to spectra profiles and therefore the solution can give negative values, which has no physical meaning. Although MVSA provides a unique solution to the mixture analysis problem, this does not guarantee that this solution is the true one and in fact, in some circumstances, like this one, the solutions provided by MVSA are not feasible because they do not fulfill the non-negativity constraints of the system. The reason why negative spectra are produced by MVSA is due to the fact that data set 2 does not conform to the closure (sum to one) constraint implemented in this method. Whereas the total amounts of the constituents in different pixels were the same in data set 1, in data set 2, these total amounts were changing from pixel to pixel,

with very different total spectra intensities. Therefore, whereas the equal sum to one constraint implemented by MVSA to all pixels is right for data set 1, it is not right for data set 2 and the recovered spectra will be distorted in this case, giving negative values.

This is in clear contrast to MCR-ALS solutions, in which the solutions are always feasible solutions for the applied constraints, although they are not unique and they may depend on initial estimates in case of high rotation ambiguity. Closure constraints were not applied in MCR-ALS simply because this is not general case for spectroscopic images. The fact that MCR-ALS solutions are not unique should not be considered a limitation of this method but of the insufficient applied constraints to provide uniqueness to the bilinear decomposition given in Eq. (1). When MCR-ALS solutions are unique (rotation ambiguity totally removed), they would then correspond to the true one. The amount of rotation ambiguity still present in MCR-ALS solutions can be checked using an independent method like the MCR-BANDS method (see below). The art of MCR-ALS is therefore, to choose the appropriate constraints to restrict the number of possible solutions.

MVSA and VCA methods did not estimate the relative concentrations of each component at every pixel (concentrations). Possible known information about the concentrations, like local rank or selectivity is therefore not used explicitly and it cannot be used as a constraint during the profiles optimization. Relative concentrations should be then obtained in a separate step at the end of the optimization by an ordinary least squares method, for instance using the equation,  $\mathbf{C} = \mathbf{D}(\mathbf{S}^T)^+$ , where  $(\mathbf{S}^T)^+$  is the pseudoinverse [44] of the estimated spectral pure signatures of the components. Again, this calculation can produce negative values for the estimated  $\mathbf{C}$  concentrations unless a non-negative least squares method is used [45]. In contrast, in MCR-ALS, both  $\mathbf{C}$  and  $\mathbf{S}^T$  profiles, are alternative obtained under the required constraints by projection in the two subspaces of the raw data matrix  $\mathbf{D}$ , column and row and subspaces and using non-negative least squares projections [45]. For the first component in this data set 2, and according to  $r^2$  and angle

**Table 2**  
MCR-ALS, MVSA, VCA data fitting results for the analysis of the two simulated data sets.

	Data set 1			Data set 2		
	MCR-ALS	MVSA	VCA	MCR-ALS	MVSA	VCA
Full model lof <sup>a</sup>	0.99	0.99	4.43	0.16	0.75	0.61
Full model $R^2$ <sup>b</sup>	99.99	99.99	99.8	99.99	99.99	99.99

<sup>a</sup> Percentage of lack of fit (lof, Eq. (2)) considering only one single component and the three components (full model) together.

<sup>b</sup> Percentage of explained variances ( $R^2$ , Eq. (3)) considering only one single component and the three components (full model) together.

values, either MCR-ALS, MVSA or VCA provided accurate estimations of the pure component spectra and concentrations (see lower part of Table 1). For the second component, MVSA and VCA  $r^2$  values were a little better (closer to one) than those of MCR-ALS, probably in agreement with the existence of pure pixels for this component as discussed above. For the third component however, the situation resulted to be completely different especially for MVSA, as a consequence as said before, of the negative spectra values present in its spectrum, giving a negative  $r^2$  value and a large angle (higher than  $90^\circ$ ). See also the values for the corresponding wrong values for the concentration profile calculated for this third component resolved by MVSA. Table 2 illustrates that for this data set 2, the three methods gave very similar explained variances and lack of fit values for the full model considering simultaneously the three image constituents. Since the resolved profiles showed appreciable differences (lower part of Fig. 2 and of Table 1), this confirms again the presence of rotation ambiguities in the solutions (see below in Section 2.4). As for data example 1, the variances explained by the individual constituents were high and overlapped since their sum highly exceeded 100%.

In the simulated data set 2, adjacent pixels had common patterns, as it may happen in real spectroscopic images from natural systems in many circumstances. In this example, relative concentrations of a constituent were higher at the center of a particular region of the image and then they decreased smoothly at the edges of this region. In Fig. 3, the 2D reconstructed images of the concentration values at a pixel level in the  $x$ - $y$  plane (distribution maps) resolved using the three methods, MCR-ALS, MVSA and VCA, are compared separately and to the true 2D simulated images for each of the three image constituents. Relative concentrations for the first resolved pure component (Fig. 3, left side), resulted to be almost perfect using any of the three tested methods. For the second component (Fig. 3, center), the shape of the MCR-ALS reconstructed image did not changed much either. The color of the image obtained was only a little different in the red part, with slightly higher intensity values for this component than for the true one (possible rotation ambiguity). On the other side, second component concentrations calculated for MVSA and VCA gave some negative image values in the deep blue region. This was much more severe and worse for the third component obtained by any of these two methods (Fig. 3, right side). As a consequence of this, the relative image intensities obtained for these two (second and third) components by MVSA and VCA were highly distorted and wrong compared to the true ones (Fig. 3, upper part). MCR-ALS provided in all cases, much better and reasonable reconstructed 2D images of the three pure component concentrations for this data set. It is concluded therefore that MCR-ALS provided more accurate estimations of pure spectra and concentrations than MVSA and VCA in the analysis of this second data set.

#### 4.3. Results of experimental data set

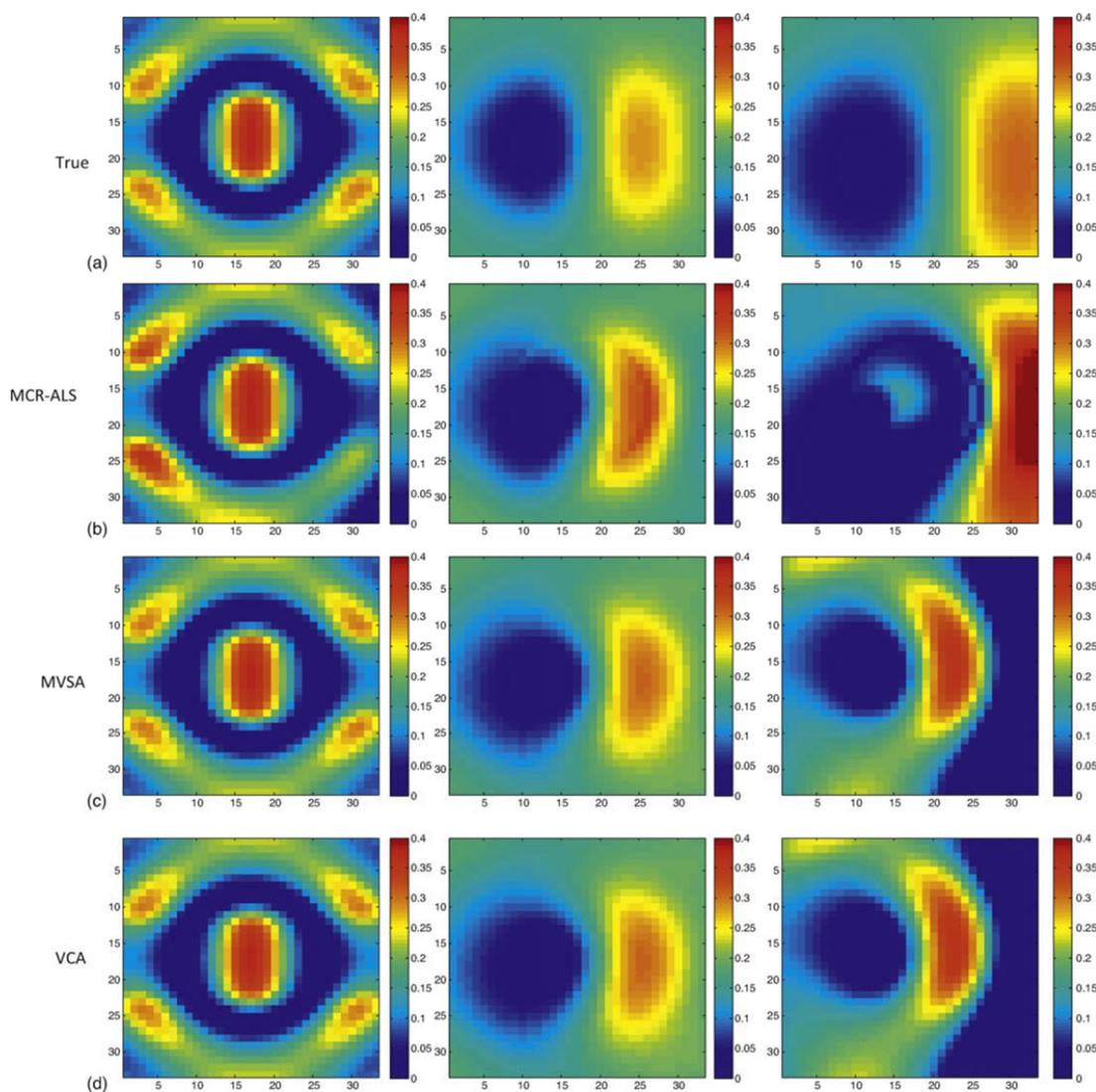
In the lower left corner of Fig. 1, the original raw remote sensing spectroscopic image was shown. As mentioned before, MCR-ALS method requires an initial estimation of the number of components present in the data matrix  $\mathbf{D}$  and an initial estimation of one of the two factor matrices,  $\mathbf{C}$  or  $\mathbf{S}^T$ . As a preliminary estimation of the number of components, PCA indicated the possible presence of up to 4–5 constituents on the analyzed spectroscopic image. Initial spectra estimates were then obtained from the measured spectra in the 'purest' pixels, using the same procedure as in SIMPLISMA [46]. After that, MCR-ALS optimization was carried out under non-negativity constraints in spectra and concentration profiles, and resolved spectra were normalized.

Fig. 4a–e shows the comparison between the pure spectra (signatures) resolved by MCR-ALS (red), and those resolved by

MVSA (black), VCA (green) and also with reference (blue) best matching USGS library spectra for the four resolved components giving a larger contribution to the total data variance. With five components, the total explained variances ( $R^2$  value in Eq. (3)) were already very high (99.98%). Table 3 shows the correlation coefficients,  $r^2$ , between MCR-ALS pure spectra and the corresponding spectra from the USGS library. As it can be seen in this table, these values were close to one. Individual constituents were postulated to be Alunite, Andradite, Montmorillonite, Nontronite and Kaolinite, which individually explain data variances of 27.19%, 38.58%, 7.72%, 16.59% and 6.10% respectively. The major spectral bands agree well with those from the spectral library, although relative intensities showed some differences. In Table 3,  $r^2$  (correlation coefficients) values between MCR-ALS resolved pure spectra and reference library ones for Andradite, Montmorillonite and Nontronite were very close to 1 and their angles between them were rather small, proving that they match with the proposed ones rather well. For Kaolinite and Alunite, correlation values were not so good, with  $r^2$  values of 0.9414 and 0.9296 respectively. Their inclusion in the proposed model of five components did not have much effect on the final resolution and fit of the experimental dataset, and their presence is therefore more questionable. The relative concentration contributions of Montmorillonite and Kaolinite resulted to be rather low, therefore they are closer to the contribution of the background signal level, and they should not be as reliable as the other major ones. In previous works from the literature [11,47], a larger number of spectra signatures (endmembers) were proposed for this image, explaining low variances in many circumstances. Including a too large number of image constituents or endmembers can produce unstable and inaccurate models and model over fitting. As shown also in Fig. 4, VCA and MVSA spectra for the last two components, four and five, resulted to be rather different to those obtained by MCR-ALS. Comparing them with those from the spectral library, gave other possible candidates. However, again, MVSA spectra estimates produced negative values of these two minor components, and they were again unfeasible.

In Fig. 5, 2D images of the relative concentrations of the five components at each pixel in the  $x$ - $y$  plane estimated by MCR-ALS are given. To better illustrate the relative contributions of the different components, the same intensity scales have been used for all resolved components. The first constituent assigned as Andradite (Fig. 5a), is present a wide area of the image and in the right part has a higher concentration. The second constituent is Alunite (Fig. 5e), and it is located in the right side of the image. The third constituent is Montmorillonite (Fig. 5c). The fourth constituent is assigned to Kadinite, it is at the right side of the image too, and it does not contain Alunite (Fig. 5d). The fifth constituent is Nontronite, which is encountered mainly in the lower part of the image (show in Fig. 5e). Kadinite is at lower concentration than Alunite and Nontronite.

From the 2D image contribution of the different components, conclusions about the mineral distribution over the investigated geographical area can be drawn. It is first concluded that Alunite mineral is concentrated in the more east part of the studied area, together with Kaolinite, which is also in the same area, being almost complementary to each other. Previous results of Clard and Swayze [48,49] for the location of these two components in the same image were similar to ours, although the contribution ratios of these two components were a little different. Our results also showed that Andradite can be distributed in a wide area of the geographical area under investigation and that it has higher concentrations at the northeast part of it, with a contribution higher than for the other constituents (see relative scales and colors in the different images of Fig. 5). Whereas Clard and Swayze did not conclude the same about the possible presence of Andradite in their paper, Chan et al. on the other hand, did obtain a similar result [47], with a similar relative

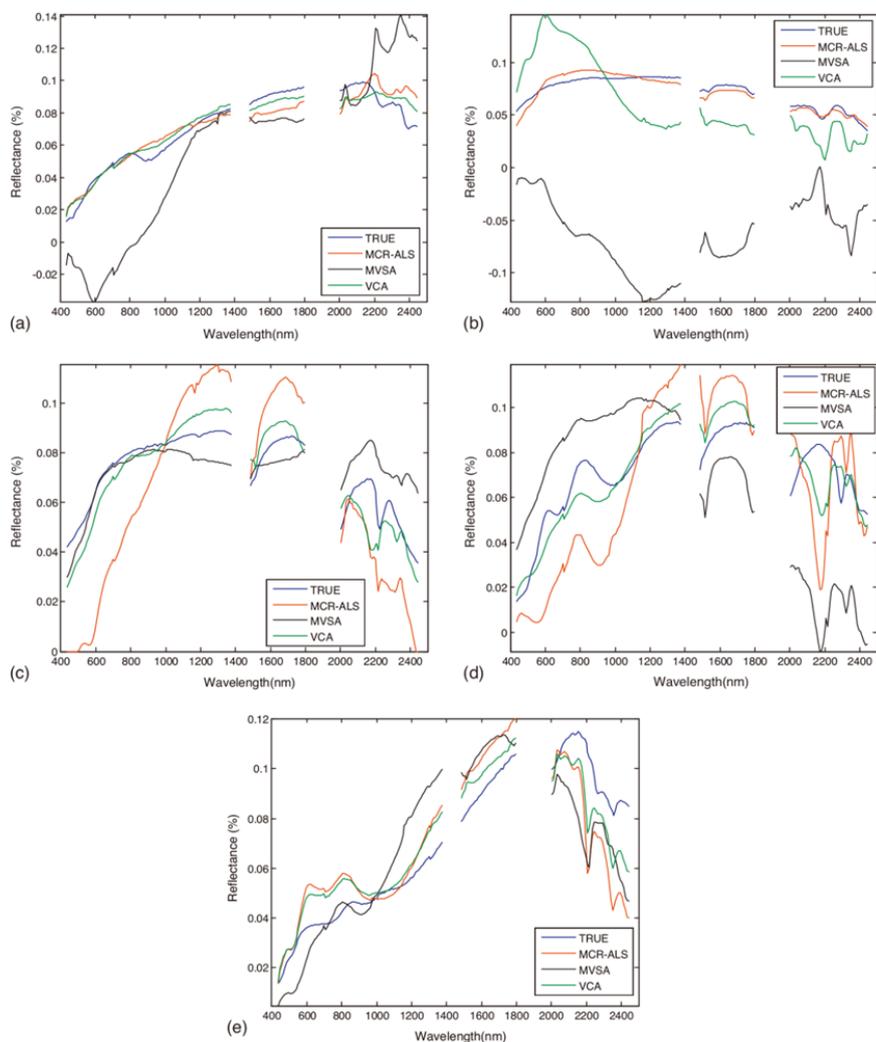


**Fig. 3.** 2D reconstructed images of the concentration values at a pixel level in the  $x$ - $y$  plane resolved using the three methods. True (a), MCR-ALS (b), MVSA (c) and VCA (d) concentrations for each of the three image constituents.

**Table 3**

Correlation coefficients ( $r^2$ , Eq. (4)) and angles (Eq. (5)) between USGS library spectra and those obtained by MCR-ALS, VCA and MVSA in the analysis of the experimental data set.

	Andradite	Alunite	Montmorillonite	Kaolinite	Nontronite
MCR-ALS					
$r^2$	0.9974	0.9414	0.9739	0.9296	0.9932
Angle	4.09	19.70	13.11	21.63	6.69
MVSA					
$r^2$	0.9669	Not matching	0.8786	0.8709	0.9882
Angle	14.78	Not matching	28.53	29.44	8.82
VCA					
$r^2$	0.989	Not matching	0.9909	0.9969	0.9933
Angle	8.49	Not matching	7.75	4.48	6.63



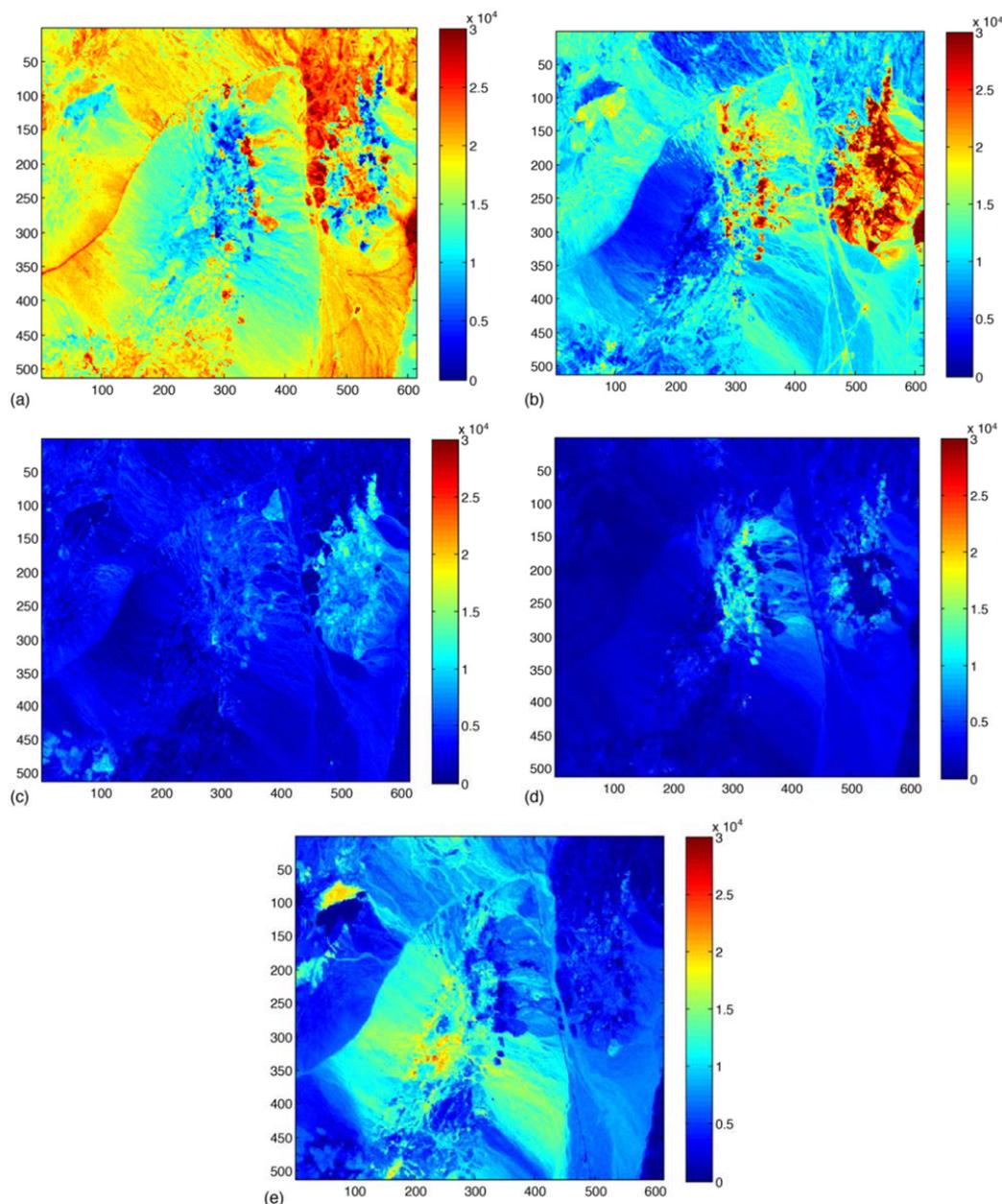
**Fig. 4.** Comparison between library spectra (blue, from USGS reference library), with those obtained by MCR-ALS (red), MVSA (black) and VCA (green) in the analysis of the experimental data set. Proposed image constituents are: (a) Andradite, (b) Alunite, (c) Montmorillonite, (d) Kaolinite, and (e) Nontronite. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

concentration distribution over the investigated area. Montmorillonite would be concentrated in almost the same area as Alunite and Kaolinite but with a rather lower concentration and only in some pixels is at a little bit larger concentrations. Nontronite distributes mostly in the southwest part of the geographical area under study, and also in a particular spot in the northwest part, where it has a relatively high concentration. The present and distribution of these two components are in agreement with those obtained by Clard and Swayze [48,49]. In the analysis of the spectroscopic image under study in this work, most of the previous studies have concentrated on the resolution of the chemical contributions to the measured signal and compared the resolved spectra profiles to the reference USGS spectral library [47,50]. In the case of remote sensing hyperspectral imaging, the presence of light scattering and other physical contributions can be important and should be considered

[51,52]. However, in the particular case of data set 3 obtained from Cuprite, Nevada of USA, the investigated region is an arid area with a very small amount of vegetation cover and nearly no water contributions, and only soil and rock minerals should be the major spectral variance sources. Therefore, physical contributions to the spectroscopic signal should be small.

#### 4.4. MCR-BANDS results

MCR-BANDS results are shown in Table 4. SCRF values for MCR-ALS, VCA and MVSA solutions are given in this table. In all performed analysis, on simulated and experimental datasets, when only non-negativity constraints (apart from spectra normalization) were used, maximum and minimum SCRF values differed considerably and their differences were rather large. This shows the presence of



**Fig. 5.** 2D reconstructed images of the concentration values at a pixel level in the  $x$ - $y$  plane resolved using MCR-ALS for the proposed image constituents: (a) Andradite, (b) Alunite, (c) Montmorillonite, (d) Kaolinite, and (e) Nontronite.

a considerably amount of rotation ambiguity in the solutions, even after the application of these constraints. On the other hand, actual SCCF values for MCR-ALS, are within the range of maximum and minimum the max and min SCCF values obtained by MCR-BANDS with non-negative and normalization constraints. In Table 4, SSCF values are evaluated considering the ambiguity associated to

MCR-ALS resolved profiles. SCCF values for the true profiles (also shown in Table 4) in the case of the simulated data sets were similar to the ones of MCR-ALS. In the first simulated data set, MVSA SCCF values are excellent and coincide with the true ones. This is a consequence of the matching conditions of this data set for an optimal application of the MVSA method, specially because of the simulated

**Table 4**

Signal component contribution function (Eq. (6)) SSCF obtained by MCR-ALS, VCA, MVSA and MCR BANDS [34,36] maximum and minimum values.

	MAX SCCF <sup>a</sup>	MAX SCCF <sup>b</sup>	MAX SCCF <sup>c</sup>	MCR	VCA	MVSA	True SCCF	MIN SCCF <sup>c</sup>	MIN SCCF <sup>b</sup>	MIN SCCF <sup>a</sup>
Dataset 1										
1st component	0.62	0.52	0.52	0.52	0.58	0.48	0.48	0.52	0.17	0.15
2nd component	0.64	0.50	0.30	0.30	0.37	0.26	0.27	0.30	0.16	0.14
3rd component	0.64	0.39	0.39	0.39	0.46	0.48	0.47	0.39	0.44	0.24
Dataset 2										
1st component	0.84	0.64	0.64	0.64	0.70	0.73	0.69	0.64	0.60	0.02
2nd component	0.61	0.28	0.28	0.28	0.38	0.41	0.18	0.28	0.27	0.08
3rd component	0.62	0.31	0.25	0.25	0.09	0.07	0.25	0.25	0.25	0.05
Experiment dataset										
1st component	0.69	–	–	0.09	0.28	0.25	–	–	–	0.04
2nd component	0.66	–	–	0.32	0.30	0.19	–	–	–	0.12
3rd component	0.67	–	–	0.42	0.56	0.19	–	–	–	0.08
4th component	0.72	–	–	0.21	0.24	0.41	–	–	–	0.05
5th component	0.57	–	–	0.09	0.50	0.46	–	–	–	0.09

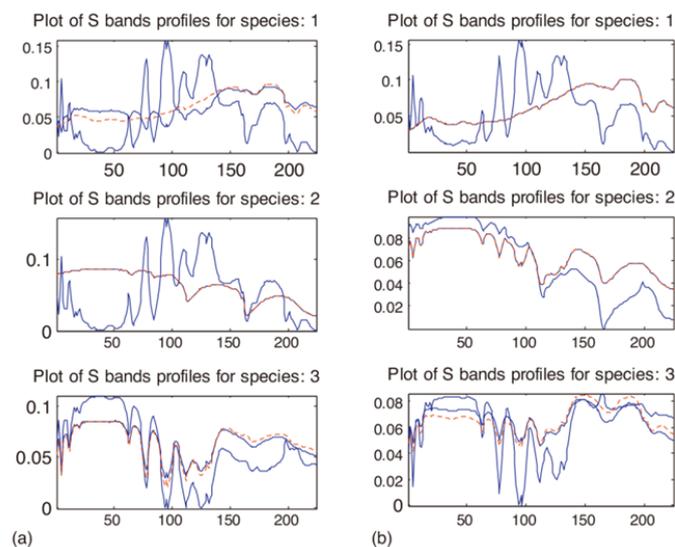
<sup>a</sup> Using only spectra and concentration non-negativity and spectra normalization constraints.<sup>b</sup> Using the same constraints as in 1 plus local rank constraint for the first component.<sup>c</sup> Using the same constraints as in 1 plus local rank constraint for the all three components.

concentration profiles following a random Dirichlet distribution which is exactly the assumption about relative abundances in the development of the MVSA method.

In Table 4, SCCF values are also given for cases where information about local rank is available. For instance, for simulated data sets 1 and 2, SCCF values are given for the cases where local rank information is available for one or for all three components. In the first case, pixels where the first component is known to be at low concentrations were forced to fulfill this condition. In case only one constituent is present in one or more pixels, the pure spectrum or spectral signature of this constituent is totally determined (apart from noise) and there should not be rotation ambiguity for its spectrum profile. As it is deduced from the results shown in the table, differences between min and max SCCF values become smaller, being both closer to the ones obtained by MCR-ALS and to the true ones. In case of using such a constraint for all the components, i.e., setting concentration values of these three components very low in some pixels and allowing for full selectivity of every

component, differences in extreme SCCF max and min values can become negligible and profiles coincide with the true ones and with those obtained by MCR-ALS. In these cases rotation ambiguities are eliminated and MCR-ALS solutions are unique and equal to the true one.

In Fig. 6, MCR-ALS resolved spectra together with maximum and minimum SCCF spectra for every component in the simulation of experimental data sets 1 and 2 are given. Again, it is clear that when only non-negativity and normalization constraints are applied, a wide set of possible pure spectra can be obtained for the same spectroscopic image, with the same data fit. This means that in all these cases, to get better, less ambiguous and more reliable estimates will require the use of local rank and selectivity information about the system. From Fig. 6, MCR-ALS (with SIMPLISMA initialization), and also VCA and MVSA, tend to choose a solution which is close to one of the two SCCF extreme values. This is probably due to the fact that these boundaries corresponds to the spectra in pure (or nearly pure component) pixels. Maximum or minimum



**Fig. 6.** MCR-BANDS spectra profiles obtained for every component in the analysis of simulated data set 1 (a) and data set 2 (b). In blue, max and min SCCF profiles and in dashed red MCR-ALS/true profiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

SCCF MCR-BANDS solutions are similar or equal to these spectra (signatures) obtained at the purest pixels. If in these pixels only one component is present (selectivity), then its spectrum is one SCCF extreme value and also will be one of the solutions provided by the different methods, MCR-ALS, MVSA and VCA. A deeper study of this fact is needed and some work is pursued in this direction.

In simulated data set 1, when local rank constraints were applied to all of the components, unique profiles were obtained which was a little different to the true one. This is probably due to the fact that pixels constrained according to local information were not totally pure ones and other components were still present at low concentrations. In simulated data set 2, the first component has a very low concentration, close to zero, and as a consequence of this, a practically unique profile, very similar to the true one was obtained. These results confirm that when there is total selectivity in a pixel (only one component present) or, what is the same, when all the other components have a very low concentration in these pixels, the spectra profiles resolved by MCR-ALS will be close to the true ones, even if MCR-BANDS indicate that there is still ambiguity. A unique result will be obtained if the constraint applied on all components, and this is a situation like MVSA. Only when all of those constraints do agree with the real situation, the unique true result is obtained.

The experimental dataset is a rather complex data set. Most of the components present in the image have similar chemical groups and their spectra profiles are highly collinear. Due to the large size of the image and computer limitations, MCR-BANDS could only be applied to a reduced section of the experimental dataset (one every eight pixels). The SCCF values of MCR-ALS results were correctly between the maximum and minimum MCR-BANDS values (see Table 3). These maximum and minimum SCCF values for the five components resulted to be rather different, for example, the maximum and minimum SCCF values for the first component are between 0.04 and 0.70 (from a range between 0 and 1). Therefore results of MCR-BANDS confirmed that in this case there is still a large amount of ambiguity in the spectra profiles resolved by MCR-ALS. Finding regions of reduced local rank and implementing them as a constraint during MCR-ALS and MCR-BANDS will reduce considerably the presence of rotation ambiguities.

Moreover, concentration changes in hyperspectral image at different pixels are not continuous, like in evolving chemical processes. In evolving chemical processes is relatively easy to get and applied the local rank information from evolving factor analysis methods [53]. However, in the case of the hyperspectral images investigated in this work, this local rank information is rather difficult to find out and discover if some pixels have one or more components absent. Further work is also planned in this direction, to develop new local rank detection techniques to allow local for the identification of image pixels where certain components are absent.

## 5. Conclusions

In this work, MCR-ALS has been shown to be a powerful tool for the resolution of hyperspectral imaging on their constituents. Results obtained by MCR-ALS are in general similar to those obtained by MVSA and VCA methods, except for the cases where the later produce spectra and concentration profiles with negative values, which are not feasible from a physical point of view and according to the desired constraints of the sought solutions. In the three studies reported here in this work, MCR-BANDS results suggest that the extent of rotation ambiguity associated to the MCR-ALS resolved profiles can be rather high and that the correct solutions can only be guaranteed if additional constraints are applied, such as those providing information about the local rank

properties of the image, i.e., about the presence or absence of the different components in the image pixels. Only in this way, the reliability and quality assessment of the proposed solutions can be better guaranteed. If this is not possible, the extent of remaining rotation ambiguities can be reported using a method like the MCR-BANDS method proposed in this work.

## Appendix A. Simulation of concentration values for simulated data set 1

### First component:

```
[X,Y]=meshgrid(-8:.5:8);
R=sqrt(X.^2+Y.^2)+eps;
C1=sin(R)/R;
surf(C1,'DisplayName','C1');figure(gcf)
C1=C1+0.2173;
C1=C1*100;
```

### Second component:

```
[X,Y]=meshgrid(-2:.1:2);
R=X.*exp(-X.^2-Y.^2);
surf(R,'DisplayName','R');figure(gcf)
C2=R(4:36,4:36);
min(min(C2))
C2=C2+0.4289;
C2=C2*10;
```

### Third component:

```
[X,Y]=meshgrid(-2:.1:2);
C3=X.*exp(0.5*(-X.^2-Y.^2));
surf(C3,'DisplayName','C3');figure(gcf)
C3=C3(4:36,4:36);
C3=C3+0.6089;
C3=C3*10;
```

X and Y give the coordinates of the pixels in the two axes of the 2D image, C1, C2 and C3 are the concentrations of three components respectively.

## References

- [1] J. Rogan, D. Chen, *Progr. Plann.* 61 (2004) 301.
- [2] C. Fischer, I. Kakoulli, *Rev. Conserv.* 7 (2006) 3.
- [3] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, *Trends Food Sci. Technol.* 18 (2007) 590.
- [4] R.A. Schultz, T. Nielsen, J.R. Zavaleta, R. Ruch, R. Wyatt, H.R. Garner, *Cytometry* 43 (2001) 239.
- [5] A.F.H. Goetz, B. Curtiss, *Field Anal. Chem. Technol.* 1 (1996) 67.
- [6] F.D. van der Meer, H.M.A. van der Werff, F.J.A. van Ruitenbeek, C.A. Hecker, W.H. Bakker, M.F. Noomen, M. van der Meijde, E.J.M. Carranza, J.B.d. Smeth, T. Woldai, *Int. J. Appl. Earth Observ. Geoinform.* 14 (2012) 112.
- [7] T. Blaschke, *ISPRS J. Photogram. Rem. Sens.* 65 (2010) 2.
- [8] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* 2 (1987) 37.
- [9] J.W. Boardman, *Summaries 4th Annual JPL Airborne Geoscience Workshop*, 1993, p. 11.
- [10] M.E. Winter, in: M.R. Descour, S.S. Shen (Eds.), *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Society of Photo-Optical Instrumentation Engineers, United States, 1999, p. 266.
- [11] J.M.P. Nascimento, J.M.B. Dias, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, 2005, p. 898.
- [12] J. Li, J.M. Bioucas-Dias, *Geoscience and Remote Sensing Symposium*, 2008. IGARSS 2008, IEEE International, IEEE, 2008, p. III250.
- [13] J.M. Bioucas-Dias, A. Plaza, *Geoscience and Remote Sensing Symposium (IGARSS)*, 2011 IEEE International, IEEE, 2011, p. 1135.
- [14] A. Plaza, P. Martínez, R. Pérez, J. Plaza, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, 2002, p. 2025.
- [15] T. Azzouz, R. Tauler, *Talanta* 74 (2008) 1201.
- [16] M. Garrido, F. Rius, M. Larrechi, *Anal. Bioanal. Chem.* 390 (2008) 2059.
- [17] R. Tauler, B. Kowalski, S. Fleming, *Anal. Chem.* 65 (1993) 2040.

- [18] E. Peré-Trepas, S. Lacorte, R. Tauler, *J. Chromatogr. A* 1096 (2005) 111.
- [19] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridors, *J. Chemom.* 12 (1998) 183.
- [20] R. Tauler, S. Lacorte, M. Guillaumon, R. Cespedes, P. Viana, D. Barceló, *Environ. Toxicol. Chem.* 23 (2004) 565.
- [21] S. Piqueras, L. Duponchel, R. Tauler, A. De Juan, *Anal. Chim. Acta* 705 (2011) 182.
- [22] A. de Juan, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler, *Infrared and Raman Spectroscopic Imaging*, Wiley-VCH Verlag, GmbH & Co. KGaA, 2009, p. 65.
- [23] A. de Juan, M. Maeder, T. Hancewicz, R. Tauler, *J. Chemom.* 22 (2008) 291.
- [24] J.M. Amigo, J. Cruz, M. Bautista, S. Maspocho, J. Coello, M. Blanco, *TrAC Trends Anal. Chem.* 27 (2008) 696.
- [25] S. Piqueras, J. Burger, R. Tauler, A. de Juan, *Chemom. Intell. Lab. Syst.* (2011), <http://dx.doi.org/10.1016/j.chemolab.2011.12.004>.
- [26] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, *Chemom. Intell. Lab. Syst.* 76 (2005) 101.
- [27] R. Bro, S. De Jong, *J. Chemom.* 11 (1997) 393.
- [28] D.J. Leggett, *Anal. Chem.* 49 (1977) 276.
- [29] R. Tauler, A. Smilde, B. Kowalski, *J. Chemom.* 9 (1995) 31.
- [30] A. de Juan, Y. Vander Heyden, R. Tauler, D.L. Massart, *Anal. Chim. Acta* 346 (1997) 307.
- [31] A.d. Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *TrAC Trends Anal. Chem.* 23 (2004) 70.
- [32] A. de Juan, S. Navea, J. Diewok, R. Tauler, *Chemom. Intell. Lab. Syst.* 70 (2004) 11.
- [33] R. Tauler, I. Marqués, E. Casassas, *J. Chemom.* 12 (1998) 55.
- [34] J. Jaumot, R. Tauler, *Chemom. Intell. Lab. Syst.* 103 (2010) 96.
- [35] A. de Juan, M. Maeder, T. Hancewicz, R. Tauler, *Chemom. Intell. Lab. Syst.* 77 (2005) 64.
- [36] R. Tauler, *J. Chemom.* 15 (2001) 627.
- [37] A. de Juan, E. Casassas, R. Tauler, *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Ltd., 2006.
- [38] R. Fletcher, *Practical Methods of Optimization*, vol. 1, Wiley, 1987.
- [39] R.N. Clark, G. Survey, *USGS Digital Spectral Library splib06a*, US Geological Survey, Denver, CO, 2007.
- [40] R.N. Clark, G. Survey, *The US Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 μm*, US Geological Survey, Denver, CO, 1993.
- [41] NASA, Jet Propulsion Laboratory, [http://aviris.jpl.nasa.gov/data/free\\_data.html](http://aviris.jpl.nasa.gov/data/free_data.html), 1997.
- [42] G. Vane, R.O. Green, T.G. Chrien, H.T. Enmark, E.G. Hansen, W.M. Porter, *Remote Sens. Environ.* 44 (1993) 127.
- [43] R.b. Rajkó, *Anal. Chem.* 82 (2010) 8750.
- [44] G. Golub, W. Kahan, *J. Soc. Ind. Appl. Math. Ser. B: Numer. Anal.* (1965) 205.
- [45] C.L. Lawson, R.J. Hanson, *Solving Least Squares Problems*, Society for Industrial Mathematics, 1995.
- [46] W. Windig, J. Guilment, *Anal. Chem.* 63 (1991) 1425.
- [47] T.H. Chan, C.Y. Chi, Y.M. Huang, W.K. Ma, *IEEE Transactions on Signal Processing*, vol. 57, 2009, p. 4418.
- [48] R.N. Clark, G.A. Swayze, vol. NASA, no. 19980201628, NASA, United States, 1996, p. 49.
- [49] R.N. Clark, G.A. Swayze, K.E. Livo, R.F. Kokaly, S.J. Sutley, J.B. Dalton, R.R. McDougal, C.A. Gent, *J. Geophys. Res.* 108 (2003) 5.
- [50] D. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, J. Feng, *Remote Sens. Environ.* 110 (2007) 287.
- [51] T. Isaksson, T. Naes, *Appl. Spectrosc.* 42 (1988) 1273.
- [52] A.S. El-Hagrasy, H.R. Morris, F. D'amico, R.A. Lodder, J.K. Drennen, *J. Pharm. Sci.* 90 (2001) 1298.
- [53] M. Maeder, *Anal. Chem.* 59 (1987) 527.

## Local rank-based spatial information for improvement of remote sensing hyperspectral imaging resolution

Xin Zhang<sup>1</sup>      Anna de Juan<sup>2</sup>      Romà Tauler\*<sup>1</sup>

<sup>1</sup> IDAEA-CSIC. Jordi Girona 18, Barcelona, 08034, SPAIN, Roma.Tauler@idaea.csic.es

<sup>2</sup> Chemometrics group. Dept. Analytical Chemistry. Universitat de Barcelona. Diagonal 645, Barcelona, 08028, SPAIN.

Corresponding author: Romà Tauler

### Abstract:

This paper shows the effect of using local rank and selectivity constraints based on spatial information of spectroscopic images to increase the performance of Multivariate Curve Resolution (MCR) methods and to decrease the ambiguity of final results. Fixed Size Image Window-Evolving Factor Analysis (FSIW-EFA) is applied to discover which pixels are more suitable for the application of local rank constraints. An automated method to help in setting appropriate threshold values for the application of FSIW-EFA, based on global and local use of Singular Value Decomposition (SVD) is proposed. Additional use of correlation coefficients between selected reference spectra and pixel spectra of the image is shown to provide an alternative way for the application of the selectivity constraint in spectroscopic images for the first time. This alternative method resulted to be satisfactory when pure pixels exist.

**Keywords:** MCR, local rank constraints, rotation ambiguity, hyperspectral imaging, remote sensing

## 1 Introduction

Multivariate curve resolution (MCR) methods have many advantages for the analysis of remote sensing hyperspectral images and have been already applied in this field[1-4]. MCR methods can recover pure spectral (related to chemical features) and spatial (related to the physical shape or distribution map) information about the spectroscopic image constituents[5, 6]. MCR methods are based on a bilinear model. Hyperspectral images are usually a mixture of multiple spectra constituents and their resolution under particular constraints, like non-negativity, does not guarantee unique solutions because of rotation ambiguities inherent to bilinear model decompositions[7]. This implies that apart from the sought solutions under the required constraints, other solutions are possible by linear combination of them.

In remote sensing studies, distribution maps of the image constituents resolved by MCR are important information to recover because they may include information about mineral location, forest distribution, air pollution sources and other environmental elements. Chemical characterization of image constituents by remote sensing hyperspectral imaging can be considered at global and local scales. In remote sensing hyperspectral images, experimental data are represented in the form of a data cube, in which the spatial information is given in the X-Y directions and the spectral information is given in Z direction. MCR-ALS (multivariate curve resolution- alternating least squares) is applied to the unfolded image dataset sized  $(X \times Y, Z)$ , where X and Y are the number of pixels in the X and Y directions and Z the number of spectral channels. Detecting the details about the distribution of chemical constituents over the scanned surface in remote sensing hyperspectral imaging gives information about the features of the investigated objects.

The chemical rank (equal to the mathematical rank in the absence of noise, i.e. the number of image constituent distinct to noise in the tested area of the image) can be estimated in windows formed by a small number of contiguous pixels and this information can be used afterwards to improve the resolution of the constituents of an image. Local rank constraints have been proposed to use this information gathered from the local rank analyses of a moving window along the image surface. This information helps to know which constituents (chemical species) are present or absent in a certain pixel area.[8-11]. Local rank information may include regions of total selectivity where pure spectral information allows the identification of single particular constituents and also includes information about regions where some particular constituents are absent in some pixels. To detect and to locate these regions can be of great help to enhance the resolution and to improve the reliability of the resolution process. The local rank information can also help to describe the properties of remote sensing images, making the image sharper, and making the objects of the image be seen more clearly. Local rank information may be very useful to describe appropriately the presence and properties of the different objects on the image of the scanned surface. Local rank information can be used as a constraint in MCR methods to reduce rotation ambiguity significantly and to obtain unique resolution results in some particular cases. As it has been shown in previous works[12, 13], methods based on Evolving Factor Analysis (EFA[14]) can be applied for this purpose and have been adapted to hyperspectral image specificities.

Since in most of remote sensing systems, image constituents use to have very similar, quasi collinear spectra profiles or signatures, it can be rather difficult to estimate the local rank and to find out how many components each pixel has and which pixels can have a particular component present or absent. FSIW-EFA (Fixed Size Image Window-Evolving Factor Analysis) is applied to figure out which pixels are available in the remote sensing hyperspectral images for the application of local rank constraints and to improve MCR-ALS resolution[12, 15]. In this work, the use of local rank-based spatial information is shown to improve the resolution of remote sensing hyperspectral images. Suggestions are proposed to set adequate threshold values to improve the accuracy of the estimations of local rank in local pixel areas. The proposed FSIW-EFA method is then tested on one specially prepared simulated spectroscopic image dataset and on two real remote sensing spectroscopic images. A new method based on the calculation of correlation coefficients between selected reference spectra and pixel spectra of the image is proposed to help in the detection of selective pixels and to encode this information as a constraint. .

## 2 Methods and Data sets

### 2.1 Method

In previous papers, the application of MCR-ALS using local rank constraints[5, 12, 16] was already introduced. In this paper, the method is extended for remote sensing hyperspectral imaging resolution. The application of local rank constraints needs developing first a local rank map to decide which pixels have low complexity, i.e., a rank lower than the total number of constituents in the image, and, therefore, can be potentially constrained. Thus, before starting the ALS constrained optimization steps[17], raw remote sensing image data are submitted to Fixed Size Image Window-Evolving Factor Analysis[12] to look for pixels where local rank information can be applied. The details of this procedure are shown in the following section.

#### 2.1.1 Obtaining Image local rank Information

FSIW-EFA (Fixed Size Image Window-Evolving Factor Analysis) is a method for local rank analysis which was designed to take into account the spatial structure of images[12]. It performs multiple local singular value decompositions (SVD) moving small windows along the full spatial surface of the raw image data set. These small windows comprise each particular pixel and a limited number of neighboring pixels. The number of pixels in every moving window should exceed (at least in one unit) the total number of possible components in the whole image dataset, which can be initially estimated from SVD of the whole image data.

Remote sensing hyperspectral imaging always has considerable amounts of noise. SVD can be used to estimate the levels of noise. The first step of FSIW-EFA is using SVD to explore the spatial structure of the image. When  $n$  components are chosen for resolution, the first  $n$  components are considered to explain enough information of the raw data. The rest of components have lower singular values and are considered to describe mostly noise contributions.

From the singular values obtained in the local rank analyses, a figure can be derived (local rank map) that has the same spatial structure as the raw image. Values on each pixel in the figure represent the number of overlapping contributions to the signal in this position. A threshold value is initially set to mark the limit between signal and noise singular values. Pixels that have higher singular values than the threshold value will be considered to have useful information about the chemical constituents (not noise), and pixels with singular values lower than this threshold value will be considered to contain only noise. In pixels with chemical information, the number of singular values higher than the threshold indicates the number of overlapping signal contributions in that pixel area. There are different ways to set threshold values and to obtain complete and partial image local rank maps of the image. When only one threshold value is set, a complete local rank map is obtained giving local information about the whole image complexity, including information about how many constituents overlap in each pixel area. Partial local rank maps are derived using a threshold band instead of a threshold value. The threshold band marks the lowest and highest boundaries for sensible threshold values. The only pixels shown in the partial local rank map are those whose rank is invariant within the threshold boundaries.

Selection of adequate threshold band values is critical for the calculation of the partial local rank map. When the threshold values are not proper, subsequent steps linked to set local rank constraints will mislead the model and produce worse instead of better image resolutions. Although visual inspection of suitably displayed singular value plots can generally provide adequate threshold values<sup>15</sup>, a proposal for automatic threshold band selection is given in this work. Thus, if resolution requires  $n$  components, the minimum threshold boundary can be set as equal to the maximum singular value of  $n+1^{\text{th}}$  component obtained from all SVD local rank analyses. The maximum threshold boundary adopted will be the maximum singular value of the  $n^{\text{th}}$  component from all local rank analyses performed along the pixel areas of the image surface. Note that if a value higher than this one was chosen for the maximum threshold boundary, there could not be any pixel having a number of overlapping components equal to  $n$ , which could conflict with the global SVD results. This is a restrictive, but practical threshold band to use to set local rank constraint since values with invariant rank value among lower and higher threshold boundaries can only be those having a rank lower than the total number of components of the system.

Local rank maps show one rank value per pixel of the image. Absent constituents in a particular pixel can be identified using the correlation coefficient between the raw pixel spectrum and preselected reference spectra, i.e. obtained from databases, from purest pixels by SIMPLISMA [18], or from a preliminary MCR analysis of the image<sup>15</sup>. In this paper the latter option was chosen. Those components with the lowest correlation coefficients with the pixel spectrum are the ones likely to be absent in this pixel.

Pixels where some of the image components are absent can be used for the application of local rank constraints. In practice, however, it may be difficult to detect which components are absent in a particular pixel, especially when the samples for analysis are complex mixtures or when they have many components with very similar spectra (signature). FSIW-EFA[12] needs to be used in this situation.

Moreover, when the image for analysis has pure pixels (selective), a different and simple strategy based on the use of correlation coefficients for selective pixel detection can be proposed. Correlation coefficients can be used to describe the similarity between raw measured spectra at each pixel and known pure spectra. When the correlation coefficient is equal or very close to 1, it means that on that particular pixel a single component is dominant and the other components are absent at very low contributions. It is possible then to use this information for selectivity constraint. When this strategy is applied to look for pure pixels, it is necessary first to have reference spectra for some of the components. In remote sensing hyperspectral imaging, this is usually difficult. In these situations, spectra obtained from purest pixels by purest variable selection methods such as SIMPLISMA [18] or from a preliminary MCR analysis of the image using only non-negativity constraints can be used for this purpose, although then, they will not correspond to pure signatures (spectra) of chemical constituents, but to unknown mixtures of them. Therefore in these situations, MCR results will have this limitation and they should be interpreted accordingly.

### 2.1.2 MCR steps for image resolution

In this work, the multivariate curve resolution alternating least squares (MCR-ALS) method [17] was applied to resolve remote sensing hyperspectral images. MCR-ALS is proposed to get the pure spectra (signatures) and distribution maps of the chemical constituents of spectroscopic images [5, 19]. MCR-ALS is based on a bilinear model (see Equation 1 below), which assumes that the observed spectra are a linear combination of the spectra of the pure components of the investigated system [17].

$$\mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad \text{Equation 1}$$

Here  $\mathbf{D}$  contains the spectra of the raw remote sensing images, which are unfolded in a data matrix (from the raw data cube),  $\mathbf{S}^T$  has the pure spectra of the image constituents and  $\mathbf{C}$  contains their related concentration profiles on the scanned image pixels.  $\mathbf{C}$  matrix can be refolded to give the 2D distribution maps of the contributions in relative amounts of every component at every pixel on the image surface.

In the general case, MCR solutions are not unique and they have a certain amount of ambiguity. Although the application of Multivariate Curve Resolution methods does not require any previous knowledge, the incorporation of any preliminary information, either chemical or mathematical, previously obtained from the exploratory analysis of the spectroscopic data can influence positively the resolution of the image and decrease mathematical rotation ambiguities. During the ALS optimization process, the inclusion of information can be achieved by the application of constraints, which encode mathematical or chemical properties that pure component profiles should obey. MCR-ALS optimization requires giving initial estimates of either  $\mathbf{S}^T$  or  $\mathbf{C}$  profiles in Equation 1. Typically, spectral estimates can be provided from pure variable detection methods [17, 20]. For hyperspectral images, the most applied constraints are non-negativity in the

concentration ( $\mathbf{C}$ ) and spectral ( $\mathbf{S}^T$ ) profiles. Local rank and selectivity constraints are applied to the pixels in  $\mathbf{C}$  and are an efficient way to reduce rotation ambiguities in MCR-ALS results [8, 10, 21].

The application of local rank constraints in hyperspectral imaging resolution has aspects different to those present in other spectroscopic situations. Global chemical rank (the total number of distinguishable chemical constituents in the hyperspectral image) is initially estimated by SVD, which provides an efficient model to describe the data variance. In order to apply local rank constraints, local rank information should be provided (see method section 2.1.1). In the local rank map, pixels where a single component is present can be used to implement the selectivity constraint. Pixels with some absent constituents can also be constrained with local rank constraints. In spectroscopic remote sensing, only pixels with very well defined information in terms of rank estimation and identification of absent components will be constrained.

In this work, the MCR-BANDS method was applied to evaluate the effect of the application of local rank constraints on the extent of rotation ambiguities associated with MCR-ALS results. For this purpose a dataset specially prepared is used. Details of the MCR-BANDS method can be found elsewhere [7]. MCR-BANDS method is based on the maximization and minimization of the signal component contribution function, SCCF, which is defined for a particular component as:

$$SCCF_n = \frac{\|\mathbf{c}_n \mathbf{s}_n^T\|}{\|\mathbf{C} \mathbf{S}^T\|} \quad \text{Equation 2}$$

SCCF is a scalar value between 0 and 1, which gives the relative signal contribution of a certain  $n$  component in relation to the whole signal due to the mixture of all components. This function is maximized and minimized for all the possible linear combinations of the MCR solutions under the same constraints applied in the MCR-ALS analyses. All these solutions will fit equally well the data and obey the constraints used during their estimation, but they will be different. More details are given elsewhere [7, 22]. When maximum and minimum SCCF values are coincident for a particular component, it means that there is no rotation ambiguity associated with its resolution.

MCR-ALS and MCRBANDS algorithm code and GUI for MATLAB (The MathWorks Inc., MA, US) is freely available from the home page of MCR at <http://www.mcrals.info/>.

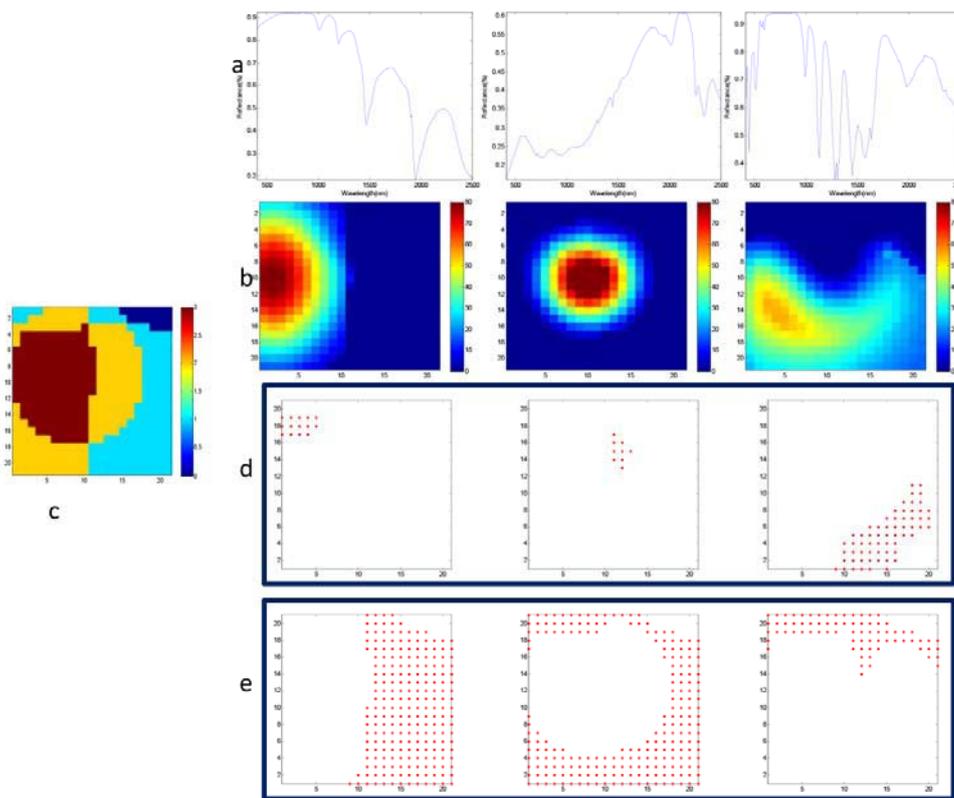
## 2.2 Data

In this work, the use of local rank constraints on MCR-ALS in the resolution of remote sensing hyperspectral imaging is performed using one dataset specially prepared for this purpose and two real spectroscopic image datasets.

### 2.2.1 Simulated spectroscopic image data set 1

Figure 1a and 1b shows the distribution maps and spectral signatures of diverse complexity used to generate the simulated hyperspectral image. The resulting dataset has 21×21 pixels and the

spectra have 224 wavenumbers. This data set gives the data matrix  $\mathbf{D}$ , which was built using a bilinear model (Equation 1) using spectral signatures ( $\mathbf{S}^T$ ) extracted from the 1997 USGS Digital spectral Library of minerals [23]. This library includes reflectance spectra on minerals, vegetation, and miscellaneous stuff and their spectra are used as references for material identification in remote sensing images. Concentration  $\mathbf{C}$ , giving the relative amounts of image constituents in adjacent pixels was created using the set of distribution maps simulating scenarios of different spatial complexity shown in Figure 1b. As said above, the image data were reconstructed according to bilinear model using equation 1, where  $\mathbf{E}$  is the noise added to the data which was calculated according to a white Gaussian noise distribution with zero-mean and constant standard deviation equal to 1% of maximum signal intensity (homoscedastic noise).



**Figure 1.** Analysis of dataset 1 a) Pure spectra of simulated dataset 1. b) Simulated distribution maps. c) Local rank map of dataset 1 obtained by FSIW-EFA. d) Selective pixels from the image of dataset 1 selected by the correlation coefficient method. e) Pixels where a certain component is absent from the image of dataset 1, selected by the FSIW-EFA method, suitable for the application of local rank constraints during MCR-ALS analysis.

### 2.2.2 Experimental remote sensing spectroscopic image datasets 2 and 3

To illustrate the use of local rank constraints in the analysis of experimental remote sensing imaging, hyperspectral data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) were used. AVIRIS measures the solar reflected spectrum from 400 to 2500 nm through 224 contiguous spectral channels at 10-nm intervals across the spectrum[24] were used. Two scenes of the free standard datasets were downloaded from the AVIRIS website (<http://aviris.jpl.nasa.gov/data/index.html>). Experiment dataset 2 is extracted from the hyperspectral image of a naval airport in San Diego, California, collected by AVIRIS (Airborne Visible and Infrared Imaging Spectrometer) sensor. In this case the analysis was limited to a  $200 \times 200$  pixels sub-image. This image has 195 useful spectral bands. Experiment dataset 3 was acquired over Moffett Field, CA, in 1997 by the AVIRIS Spectrometer. In this case the analysis was limited to a  $400 \times 400$  pixels sub-image and this image has 199 useful spectral bands. The investigated area is mainly composed of water, building, soil, and vegetation.

## 3 Results

### 3.1 MCR-ALS of a simulated hyperspectral dataset (dataset 1)

The dataset 1 is used to illustrate and validate the application of local rank constraints in MCR-ALS method. Since the noise level added to this dataset is relatively low (1%), it will be relatively easy to detect the presence of the different constituents in the image pixels. As mentioned above, this data set contains three constituents that were mixed in the simulated image. Non-negativity on both spectral and concentration directions and normalization on spectral direction were used as constraint for image resolution. MCR-ALS results were compared for this dataset with and without local rank constraints.

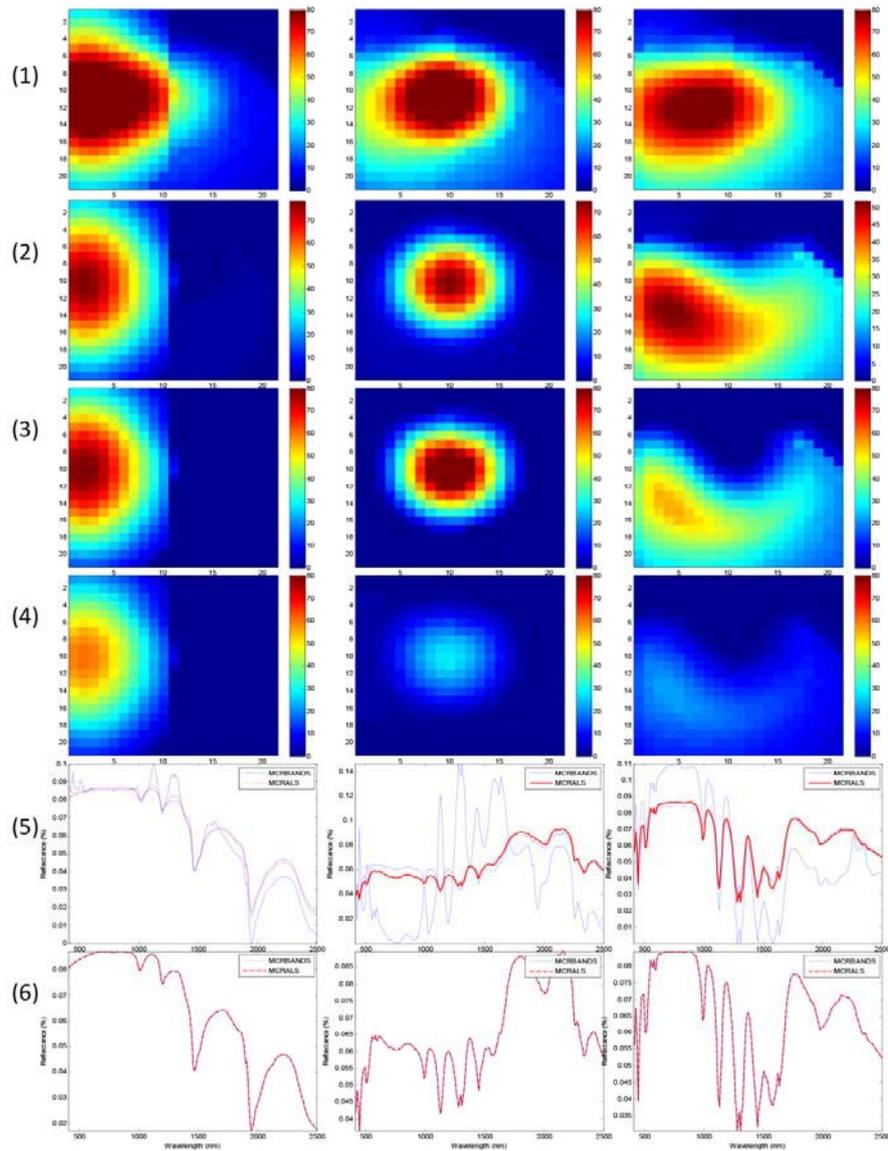
Local rank constraints can be set using the correlation coefficient method (to look for selective pixels) or by the method based on FSIW-EFA results (for both selectivity and local rank conditions). For an easy assessment of the results obtained by these two methods, the distribution maps of the components are shown (Figure 1b) and a local rank map of the image (Figure 1c) and The local rank map reproduced well the true overlapping pattern of the different components in the image and it allows for the proper application of local rank constraints. In the red-brown image section three components were overlapping, in the yellow section two components were estimated, in the cyan section only one component is estimated (selectivity region), and in the dark blue section of the image no component was detected (only noise region).

Since pure pixels (pixels having mostly only one single constituent apart from noise) exist in the image, they can be used as a constraint if their position in the image is previously detected. This situation can be encountered in geology, food, pharmacy, environment, etc. where image regions with pure pixels can be present. In simulated dataset 1, there were some pixels where only one component appeared. Detection of these pure pixels was performed checking for the correlation coefficients between possible known pure spectra (signatures obtained from MCR results with

only non-negativity constraints) and the spectra measured at each pixel. Those pixels with highest correlation coefficients can be used for the application of selectivity constraints. In Figure 1d, pixels with high correlation coefficient, close to 1, between pixel spectrum and reference pure spectrum, are illustrated as red points. These pixels are considered to have a single component. Therefore, the contribution of the other components on these pixels can be set to 0 (or better to a very small number) and use them for the application of the selectivity constraint during the ALS optimization. When selectivity constraints were used for these single component pixels, the least squares optimization converged faster, for example, in the dataset 1, resolution with and without the application of the local rank constraint needed 16 and 46 ALS iterations respectively. Comparison with the left figure shows that pixels picked up by the correlation coefficient method were mostly the same as those in the cyan regions of the local rank map obtained by FSIW-EFA, where only one component was detected.

For a better illustration of the procedure for pixel selection and appropriate application of local rank constraints, FSIW-EFA results are also given and compared with those obtained using the correlation procedure previously mentioned. For the application of the local rank constraint there should be a pixel with one or more components absent in the hyperspectral image. Figure 1c gives the FSIW-EFA estimated local rank map of the image. Figure 1e shows that the results from FSIW-EFA for the application of the local rank constraint were correctly estimated (note that in this figure, red dots mean absence of the related compound in Figure 1b in the marked pixel and, therefore, local rank constraints would set a 0, or very small value, for these components in the adequate pixels).

Both methods, the one based on the correlation coefficients and the one based on FSIW-EFA method resulted to be complementarily useful to find out selective pixel image regions, where a single chemical constituent is present at significant contributions. Equally important information, detected by FSIW-EFA, is to know the image regions where a particular component is absent, since this information can also be used as a local rank constraint during MCR-ALS. As a consequence of this, more accurate pure spectra and contribution profiles were obtained (Figure 2).

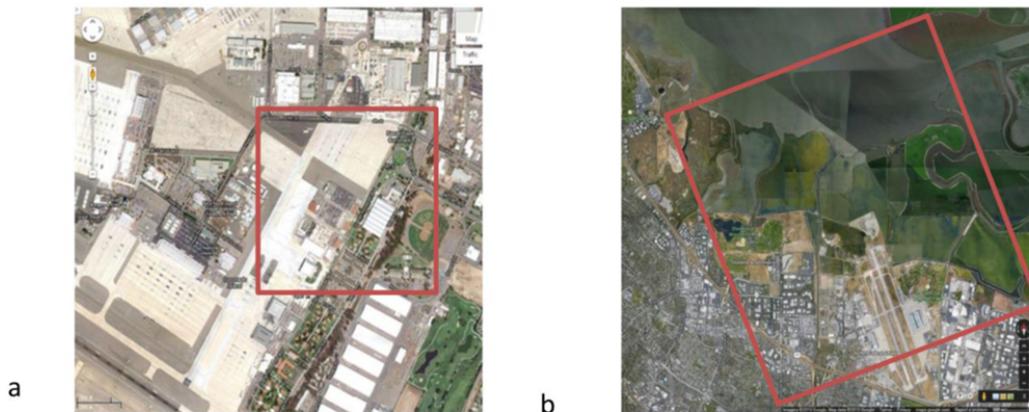


**Figure 2.** Extent of rotation ambiguities estimated by the MCR-BANDS method in the MCR-ALS analysis of dataset 1, with and without application of local rank constraints. Upper plots represent maps from maximum SCCF boundary, resolved MCR-ALS maps and maps from minimum SCCF boundary. Lower plots display the related spectra for the maximum and minimum values of SCCF (in blue), and the resolved spectrum obtained in the MCR-ALS solution (in red). (1) MCRBANDS<sub>MAX</sub>. (2) MCR-ALS with non-negativity constraint and spectra normalization. (3) MCR-ALS with non-negativity, local rank constraint and spectra normalization. (4) MCRBANDS<sub>MIN</sub>. (5) MCR-ALS with non-negativity constraint and spectra normalization. (6) MCR-ALS with non-negativity, local rank constraint and spectra normalization.

Results shown in Figure 2 include also the evaluation of the extent of rotation ambiguities obtained by MCR-BANDS in the case of the analysis of dataset 1. MCR-BANDS results for the case of only considering non-negativity and normalization constraints are shown. MCR-ALS and MCR-BANDS results with the application of local rank constraints are given. Results are given separately for every component. In the upper part, subfigures show, distribution maps for the maximum signal component contribution function (SCCF) optimized by MCR-BANDS, for the solution obtained by MCR-ALS (with and without local rank constraint) and for minimum of SCCF, respectively. In the lower part, the corresponding spectra for the maximum and minimum values of SCCF are also given in blue, whereas the red line is the resolved spectrum obtained in the MCR-ALS solution. When local rank constraints were not applied (upper line), resolution results were still ambiguous, with a large amount of rotation ambiguity (although the recovered MCR-ALS solution looks similar to the simulated maps and spectra). In contrast, in Figure (lower line), MCR-ALS solution and both MCR-BANDS solutions at maximum and minimum of SCCF are coincident, showing that rotation ambiguities were practically eliminated when local rank constraints were applied and solutions were unique.

### 3.2 MCR-ALS of the hyperspectral image of the San Diego airport (dataset 2)

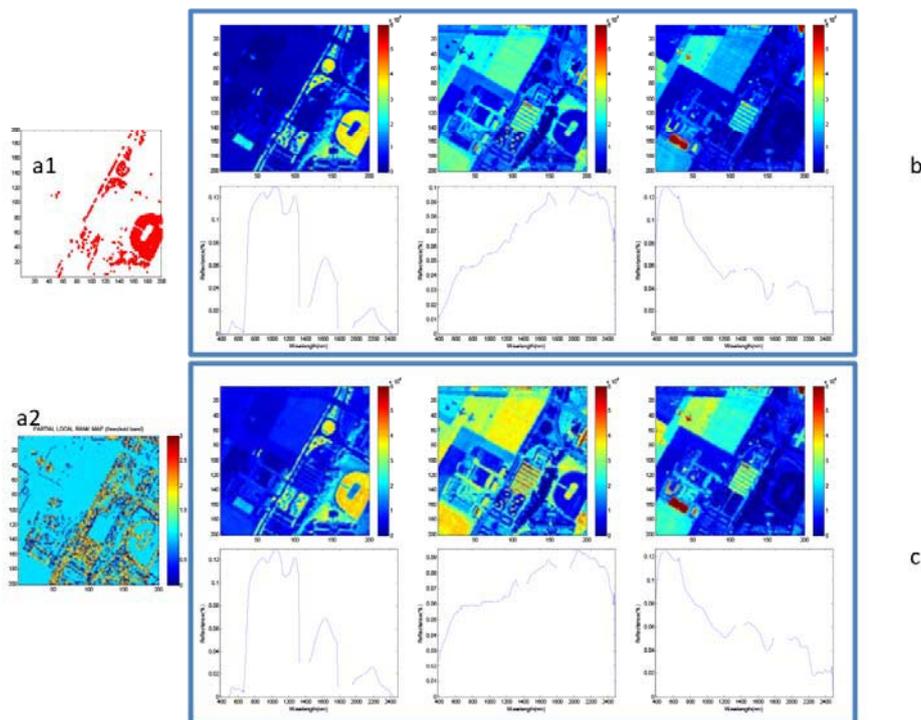
For dataset 2 from the image of the airport in San Diego, California, a small region was chosen from the raw image (Figure 3a) to focus on the area where airplanes were located. Results with and without local rank constraints were compared.



**Figure 3.** Google earth images. a) Photography (NOT spectroscopic image) of San Diego airport, California, USA (latitude:32.707, longitude:-117.206) corresponding to dataset 2. b) Photography of Moffet field, California, USA. (latitude:37.576, longitude:-122.202) corresponding to dataset 3.

Similarly to the previous analysis of dataset 1, in the analysis of remote sensing dataset 2, MCR-ALS was first used with only non-negativity and normalization constraints. Based on SVD, three

components were chosen for initial MCR-ALS analysis. Correlation coefficients were obtained by comparison of the pure spectra resolved by MCR-ALS (MCR with only non-negativity and normalization constraints in the first step) with spectra in each pixel of the raw image. Pixels with high correlation coefficients were chosen to build a masking local rank matrix to be used as a constraint (Figure 4a1). In Figure 4a1, areas with only the first component are shown in red (It was difficult to find clear selective regions for other components by the correlation coefficient method and, therefore, these components were not considered). In the selective pixels for the first component, contributions of the second and third component were set to very low values as local rank constraints, trying to improve the resolution of the image.



**Figure 4.** MCR-ALS analysis of dataset 2, San Diego (California, USA) airport spectroscopic image. a1) Pixels selected by the correlation coefficient method a2) Partial local rank map obtained by the FSIW-EFA method. b) Pure spectra (signatures) and constituent distribution maps estimated by MCR-ALS with non-negativity constraints. c) Pure spectra (signatures) and constituent distribution maps estimated by MCR-ALS with non-negativity and local rank constraint.

FSIW-EFA was also applied to estimate the local rank map of the image. Considering the difficulties in the estimation of an appropriate threshold value, the partial local rank alternative method (see Method section) was applied. First maximum and minimum threshold values should be found. In this case, the minimum threshold value was chosen as the maximum singular value of 4<sup>th</sup> singular

value in all SVD local rank analyses. This contributes to remove the effect of noise, as it was mentioned in method session. The maximum threshold value used the maximum singular value of the 3<sup>rd</sup> component in all SVD local rank analyses (see section 2.1.1 When FSIW-EFA was applied to this image, a partial local rank map (see method section) was obtained (Figure 4a2). In Figure 4a1, pure pixels corresponding to component 1 are given. They are recognized as rank one pixels in Figure 4a2. Boundary regions among different kinds of elements appear in yellow, showing the presence of two components overlapping. Dark blue zones in this case are pixels for which the rank is not well defined and that would not be constrained according to the normal FSIW-EFA procedure.

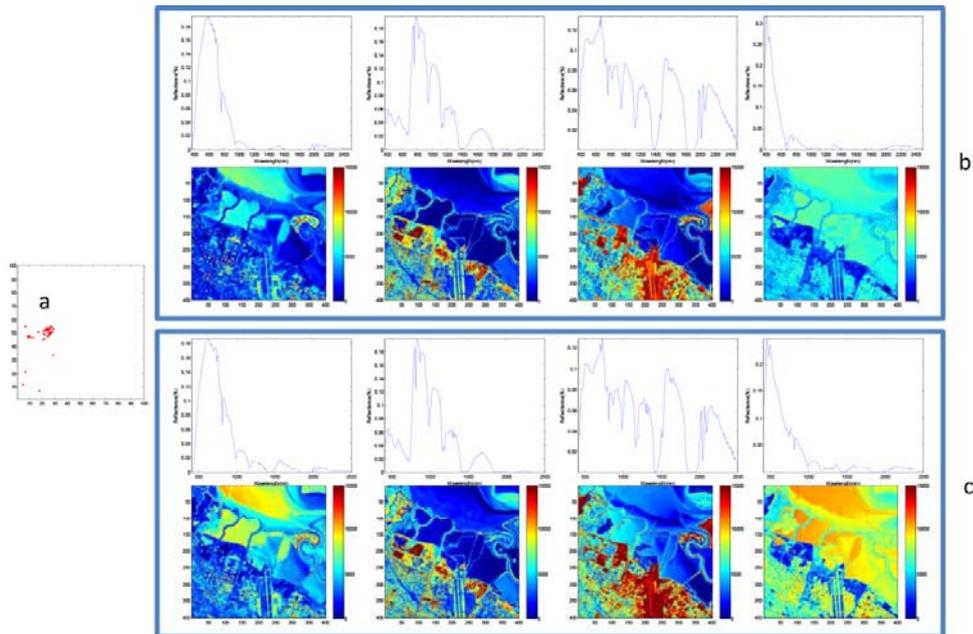
Pure spectra (signatures) resolved by the MCR-ALS method explained the most characteristic bands for the three resolved components. First, second and third MCR-ALS components corresponded to image constituents such as green plants (or vegetation), hard ground (including parking apron) and metallic materials (roofs, and more importantly, the airplanes). From the image of the airport, airplane signature has some similar patterns to the hard soil ground one. Hard ground materials in the airport have normally some kinds of concrete, it is reasonable that these materials give a certain level of contribution like metallic signatures of the airplanes

Folded distribution maps of the three components resolved by MCR give their corresponding images. Figures 4b and 4c show the results obtained by MCR-ALS without and with the application of local rank constraints, respectively. Intense red colors mean that in this image region the corresponding components have very high contributions, and dark blue means that they are at low contributions.

Figures 4c show that, after the application of selectivity constraint to the first component, many objects become generally better defined. This is the case for the vegetation in the first component, the plane shadow in the second component or for the plan and the shape of the roof in the third component. From Google maps[25], the position of this air station was obtained, and its corresponding image was downloaded. Distribution of the different image components did agree with results obtained from MCR-ALS on the corresponding hyperspectral image.

### **3.3 MCR-ALS of the hyperspectral image of the Moffet field (dataset 3)**

MCR-ALS analysis of the hyperspectral image obtained over Moffett Field, CA, in 1997 was also used to test the use of the correlation coefficient method for application of the local rank constraints. The area chosen in this case includes green plants in water and fields, hard soil ground, buildings and a big lake. Pixels with higher correlation coefficient were chosen to build a mask data matrix for the application of local rank constraints (Figure 5a). Figure 6a illustrated the area where only one image constituent (corresponding to the second MCR-ALS component) was detected. In this image region, pixels contributions of the 1<sup>th</sup>, 3<sup>th</sup> and 4<sup>th</sup> components were set to a very low value in these pixels. No clear selective pixels were obtained for other components and, therefore, selectivity was not applied.



**Figure 5.** MCR-ALS analysis of dataset 3, Moffet field (California, USA) spectroscopic image. a) Pixels selected by the correlation coefficient method. b) Pure spectra (signatures) and constituent distribution maps estimated by MCR-ALS with non-negativity constraints. c) Pure spectra (signatures) and constituent distribution maps estimated by MCR-ALS with non-negativity and local rank constraint.

Pure spectra resolved by MCR-ALS explained most characteristic bands for the four major image constituents. Resolved pure spectra and distribution maps of the 1<sup>th</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> image constituents were considered to be green plants in water, soil, hard ground (or building) and water respectively. Comparing Figures 5c and 5d shows that MCR applied with local rank constraint show clearer image details in the water areas, in the building areas and in the soil fields. From Google maps, the position of this geographical area was obtained, and its corresponding image was downloaded. Compared to the image from Google maps (Figure 3b), distribution of the different components did agree with results obtained from MCR-ALS of the corresponding hyperspectral image. Remarkably, resolution results obtained for each component by MCR-ALS from this hyperspectral image give a very detailed description and details of the contribution distribution of the different image constituents, which is therefore one of the main advantages of this type of analysis, apart from the resolution also of the pure spectra (signatures) of these constituents, which can be extremely useful for their identification.

MCR-BANDS optimized the maximum and minimum of the signal component contribution function (SCCF) for MCR results with non-negativity and, selectivity constraints of data2 and data3. In these two cases, rotation ambiguities were practically eliminated when local rank constraints were applied and solutions were unique.

## **4 Conclusions**

Appropriate use of local rank and selectivity constraints can improve significantly the quality of the pure spectra (signatures) and of the constituent distribution maps resolved by the MCR-ALS analysis of hyperspectral images in remote sensing studies. As a conclusion, the application of the FSIW-EFA method should be preferred for the general situation where only one or more components are absent in certain pixels. It doesn't matter whether there is complete selectivity (only one component) or not. Eigenvalues obtained from SVD can be used to set threshold values and to get the information about local rank will be applied during the ALS resolution of the remote sensing hyperspectral image. In contrast, the correlation coefficient method proposed in this work is a very convenient and efficient way to look for pixels where only one component exists. These pixels can be then used advantageously for a simpler application of selectivity constraints.

In the analysis of hyperspectral images obtained from the San Diego airport and the Moffet field, MCR-ALS results showed clearly the presence of the different objects of the image, like the airplanes, the green plant areas, lakes, the hard ground field and other related objects. By appropriate use of local rank constraints, objects in the rebuilt image were defined more sharply (and more easily recognized).

## **5 Acknowledgements:**

Authors acknowledge research grant CTQ2012-3816 from Ministerio de Economía y competitividad, Spain. Xin Zhang thanks the grant from China Scholarship Committee (2011637007).

## 4 References

- [1] J.A. Richards, J. Richards, Remote sensing digital image analysis, Springer 1999.
- [2] T. Selige, J. Böhner, U. Schmidhalter, High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures, *Geoderma*, 136 (2006) 235-244.
- [3] R.A. Schowengerdt, Remote sensing: models and methods for image processing, Academic press 2006.
- [4] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, J.F. Huntington, ICE: A statistical approach to identifying endmembers in hyperspectral images: Learning from Earth's Shapes and Colors, *IEEE transactions on Geoscience and Remote Sensing*, 42 (2004) 2085-2095.
- [5] S. Piqueras, L. Duponchel, R. Tauler, A. De Juan, Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares, *Anal. Chim. Acta*, 705 (2011) 182-192.
- [6] M. Berman, A. Phatak, R. Lagerstrom, B.R. Wood, ICE: a new method for the multivariate curve resolution of hyperspectral images, *J. Chemom.*, 23 (2009) 101-116.
- [7] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, *Chemometrics Intellig. Lab. Syst.*, 103 (2010) 96-107.
- [8] A. De Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution, *Anal. Chim. Acta*, 500 (2003) 195-210.
- [9] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three - way data analysis and ambiguity in multivariate curve resolution, *J. Chemom.*, 9 (1995) 31-58.
- [10] P. Geladi, S. Wold, Local principal component models, rank maps and contextuality for curve resolution and multi-way calibration inference, *Chemometrics Intellig. Lab. Syst.*, 2 (1987) 273-281.
- [11] R. Tauler, Multivariate curve resolution applied to second order data, *Chemometrics Intellig. Lab. Syst.*, 30 (1995) 133-146.
- [12] A. de Juan, M. Maeder, T. Hancewicz, R. Tauler, Local rank analysis for exploratory spectroscopic image analysis. Fixed size image window-evolving factor analysis, *Chemometrics Intellig. Lab. Syst.*, 77 (2005) 64-74.
- [13] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis, *TrAC, Trends Anal. Chem.*, 23 (2004) 70-79.
- [14] H. Keller, D. Massart, Evolving factor analysis, *Chemometrics Intellig. Lab. Syst.*, 12 (1991) 209-224.
- [15] A. de Juan, M. Maeder, T. Hancewicz, R. Tauler, Use of local rank-based spatial information for resolution of spectroscopic images, *J. Chemom.*, 22 (2008) 291-298.
- [16] X. Zhang, R. Tauler, Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging, *Anal. Chim. Acta*, 762 (2013) 25-38.

- [17] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometrics Intellig. Lab. Syst.*, 76 (2005) 101-110.
- [18] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, *Anal. Chem.*, 63 (1991) 1425-1432.
- [19] A. de Juan, M. Maeder, T. Hanczewicz, L. Duponchel, R. Tauler, *Chemometric Tools for Image Analysis, Infrared and Raman Spectroscopic Imaging*, Wiley-VCH Verlag GmbH & Co. KGaA2009, pp. 65-109.
- [20] S. Piqueras, L. Duponchel, R. Tauler, A. De Juan, Resolution and segmentation of hyperspectral biomedical images by Multivariate Curve Resolution-Alternating Least Squares, *Anal. Chim. Acta*, 705 (2011) 182-192.
- [21] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three - way data analysis and ambiguity in multivariate curve resolution, *J. Chemom.*, 9 (2005) 31-58.
- [22] R. Tauler, B. Kowalski, S. Fleming, Multivariate curve resolution applied to spectral data from multiple runs of an industrial process, *Anal. Chem.*, 65 (1993) 2040-2047.
- [23] R.N. Clark, G. Survey, USGS digital spectral library splib06a, US Geological Survey Denver, CO2007.
- [24] R.O. Green, M.L. Eastwood, C.M. Sarture, T.G. Chrien, M. Aronsson, B.J. Chippendale, J.A. Faust, B.E. Pavri, C.J. Chovit, M. Solis, Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS), *Remote Sens. Environ.*, 65 (1998) 227-248.
- [25] Google Earth Google Inc., USA, 2012.

## Discussion

The first paper<sup>266</sup> shows how MCR-ALS allowed the resolution of pure spectra and concentration distribution of the constituents present in the analyzed remote sensing hyperspectral imaging simulated datasets and in the remote sensing datasets obtained from NASA, AVIRIS (Airborne Visible/Infrared Imaging Spectrometer).

Effective use of remote sensor data requires the analysis and interpretation of the acquired data and interprets the recovered information to address practical problems, such as the position of landfills or searching for mineral deposits. Different objects are discriminated according to their spectra resolved by MCR-ALS.

In this Thesis, MVSA and VCA endmember (pure component) methods were compared with MCR-ALS in the analysis of hyperspectral remote sensing data. Table 1 in paper<sup>266</sup> shows how similar are the component true profiles and MCR-ALS, MVSA and VCA resolved profiles, in the analysis of two different simulated data sets. Table 4 in paper<sup>266</sup> gives the signal component contribution function SCCF obtained by MCR-ALS, VCA, MVSA and MCR-BANDS (maximum and minimum values) methods. These results showed that MVSA and VCA are not as good as MCR-ALS for hyperspectral imaging resolution. MVSA produced negative spectra values in some circumstances, which are not physically meaningful, whereas MCR-ALS with non-negative constraints always obtained results with a clear chemical and physical meaning.

MCR-BANDS method has been proposed to ascertain the extent of rotation ambiguity associated to MCR-ALS resolved spectra profiles. In the situation of remote sensing hyperspectral imaging, since the data have high level of noise, and some constituents are highly collinear (overlapped), the extent of rotation ambiguity of MCR-ALS (or of any other MCR method) resolved profiles can be rather high.

Local rank information can be used as a constraint in MCR-ALS methods to reduce rotation ambiguity significantly and to obtain unique resolution results in some particular cases (see Chapter 2 section 2.2 and papers<sup>266 94</sup>). From the resolved results for the airport of San Diego dataset, and for the Moffett Field (California) data set, shown in the other paper of this chapter<sup>94</sup>, it is concluded that appropriate use of local rank and of selectivity constraints can improve significantly the quality of the pure spectra (signatures) and of the constituent distribution maps resolved by MCR-ALS in the analysis of hyperspectral remote sensing images

To apply local rank constraints in MCR-ALS, the FSIW-EFA method (see Paper<sup>234</sup>, and Chapter 2 of section 2.2) is preferred for the general situation, when one or more components are absent in certain pixels. It doesn't matter whether there is complete

selectivity (only one component) or not. When FSIW-EFA is applied singular values obtained from SVD are used to set the threshold values for local rank analysis. This information have been used during the ALS resolution of the hyperspectral images in the work of this chapter<sup>94</sup>.

In contrast, the correlation coefficient method<sup>94</sup> can be proposed as a as an alternative and efficient way to look for pixels where only a single constituent or component exists. These pixels can then be used advantageously for the application of selectivity constraints (pixels with only one component present).

For the analysis of Cuprite, Nevada dataset (the first dataset analyzed in this chapter and paper<sup>267</sup>), several individual constituents were preliminary identified as Alunite, Andradite, Montmorillonite, Nontronite and Kaolinite. In previous studies<sup>194, 268, 269</sup>, apart from VCA and MVSA methods mentioned in this chapter and the paper<sup>52</sup>, the Automated Morphological Endmember Extraction (AMEE) method, the Minimum Volume Enclosing Simplex (MVES) method and the Simplex Identification via variable Splitting and augmented Lagrangian (SISAL) method, were also applied by the people working on remote sensing to the analysis of this similar dataset in previously published papers. Mainly, these methods were used to obtain the pure spectra (endmembers, signatures) of the different objects or constituents present in the image. In contrast, the application of MCR-ALS provided both, the resolution of the pure spectra (signatures) and the estimation of the concentration distribution maps at the same time. This information can be used supplementary to take decisions about mineral distribution in mining studies. Independent Component Analysis (ICA) also has been used for the resolution on hyperspectral imaging of exactly the same geographical area<sup>270</sup>(see Chapter 2 Section 2.2 for the description of this method). In the paper<sup>267</sup>, it is shown that ICA can provide unique results but that they may not be the true solutions. In contrast, using local rank constraints in MCR is a way to provide unique resolutions which are coincident with the true physically meaningful ones. In Chapter 5 a more detailed comparison between ICA and MCR is given.

The second experimental data set example (hyperspectral images from San Diego, California and Moffett Field, CA, in 1997) resulted easier to analyze because it was possible to find pixels with some components absent and suitable for the application of local rank constraints during MCR-ALS. In the analysis of hyperspectral images obtained from the San Diego airport and of the Moffet field, MCR-ALS results showed clearly the presence of different objects in the image, like airplanes, green plant areas, lakes, hard ground field and other related objects. By appropriate use of local rank constraints, objects in the rebuilt images were defined more sharply (and more easily recognized).

In other studies<sup>271, 272</sup> Anomaly detection (AD) or target detection (TD) were applied to hyperspectral imaging data analysis of the airport in San Diego, California (the second dataset used in this chapter and paper<sup>94</sup>). Their results could identify the presence of airplanes in the airport directly. K.C. Tiwari etc. compared different methods with ICA for the detection of these airplane targets. These models performed differently depending on the spectral variability<sup>273</sup>. Inverse Euclidean distance based super-resolution mapping method was reported to make better detection of those airplanes<sup>274</sup>. From the results obtained by us in the application of MCR-ALS to the same hyperspectral images, airplanes could be detected successfully from the resolved distribution map. Other objects with very similar signatures to those of the airplane were also located in the distribution map obtained by MCR-ALS.

Physical objects, such as buildings, vegetation, soil, water, etc.(Moffett Field, CA dataset in this chapter and paper<sup>94</sup>) studied in paper<sup>94</sup> are usually investigated in remote sensing applications such as geology, forestry, soil science, geography, and urban planning<sup>275</sup>. Resolution of distribution maps involves to locate where these particular objects are, and to identify the spectral features of them. For example, rivers, geologic structures and vegetation may be mapped for environmental applications. Airports, troop convoys or missile sites are for special interest for military purposes. The analysis of the objects located at a surface of the ground can be examined from the MCR-ALS resolved distribution maps. This technique would allow taking decisions from the detailed information contained in these images.

## **Chapter 4**

# **Application of hyperspectral imaging combined with chemometrics on food analysis**



## Introduction

Spectroscopic analysis alone cannot provide information about the spatial distribution of food constituents. Hyperspectral imaging integrates spectroscopic and imaging techniques, providing both spectral and spatial information simultaneously to cope with the increasing demand for safe foods. Hyperspectral imaging techniques are also applied for visual evaluation of food quality<sup>2</sup>.

The use of hyperspectral imaging technique is increasing owing to its high speed, simplicity and safety, as well as its capacity to measure multiple attributes simultaneously without the need of slow sample preparation. NIR, IR and Raman hyperspectral imaging are the most successful techniques used in food industry. They enable the qualitative and quantitative assessment of chemical and physical features.

PCA, Neural network classification, fuzzy logic, or SIMCA chemometric methods have been applied for food sample classification and analysis using hyperspectral imaging data<sup>276</sup>. These methods can provide information about texture, color, shape, or size features of food samples. More details of chemometric methods applied in hyperspectral imaging can be found in Chapter 2 of this Thesis.

In this Thesis, we propose the use of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) for the hyperspectral imaging analysis of chocolate samples as a case study of food analysis, and to recover the spectra of its constituents and their corresponding distribution maps on the analyzed chocolate samples.

Chocolate is a well appreciated food which is made of a complex emulsion of different constituents. It is a semi-solid suspension of fine solid particles containing sugar, butter and cocoa<sup>277</sup>. Primary chocolate categories are dark, milk and white. They differ in the content of cocoa solid, milk fat and cocoa butter<sup>278</sup>. Different kinds of chocolate have varying proportions of carbohydrates, fat and proteins. Chocolate manufacturing processes differ due to variation in national consumer preferences and company practices.

The flavor of a piece of chocolate can be affected by the type and amount of ingredients present, as well as by the processing techniques used<sup>279</sup>. In processing ordinary chocolate, it is conventional to grind a mixture of chocolate liquor and sugar to the desired degree of fineness using ordinary roll refiners, thereafter, to incorporate the desired amount of cocoa butter. The chocolate mixture and cocoa butter is then placed in a suitable device capable of agitating or working the mixture. Because of the need of continued agitation which usually involves further grinding, the sugar and skim milk solids become completely coated with fat<sup>280</sup>.

Particle size distribution and ingredient composition influence perception of primary taste and oral volatiles release of chocolate<sup>281, 282</sup>. Research<sup>279</sup> also supports that differences in the flavor of milk chocolate tend to arise from processing changes rather than from differences in ingredients. Highly skilled chocolate manufacturers are able to manipulate flavors in the chocolate<sup>279</sup>. The use of hyperspectral imaging to analyze the particulars of chocolate constituents can help to monitor the quality of chocolate, and it can reveal the existing relation between chocolate taste and product processing.

The applicability of Raman micro-spectroscopy mapping (imaging) to the analysis of both white and milk chocolate samples was reported in reference<sup>283</sup>. In this report, they failed to analyze milk chocolate because of the strong fluorescence signal from cocoa solids. Infrared hyperspectral imaging has lower spatial resolution than Raman but it does not suffer from the effect of fluorescence, and it has faster scanning speeds because it does not require so much focus at the sample surface. Moreover, infrared hyperspectral imaging has many common features to Raman, which also can be used for spatial and spectral information detection.

For Raman spectra acquisition, a HORIBA JobinYvon Raman microscope (HORIBA, Ltd.) was used using a 532 nm radiation laser. For Infrared spectra acquisition and high-resolution mapping, the Nicolet iN10 MX Infrared Imaging Microscope with MCT (Mercury cadmium telluride) detector was used. The instruments used in the experimental determinations are shown in Figure 4-1.



**Figure 4-1** Hyperspectral image collection instruments used in the experimental analysis of chocolate samples

Hyperspectral imaging instruments for research are developing very fast. For example, Bruker Corporation developed a new Raman system: SENTERRA Dispersive Raman Microscope, which can scan at multiple wavelengths, (1064 nm, 785 nm, 633 nm, 532 nm

and 488 nm), with high confocal depth profiling performance. It has an open architecture version for the study of large size samples (e.g. in art works) with high lateral resolution for the investigation of living cells, etc. All of these features have been improved a lot in comparison with the instrument we have applied in the experimental work of this Thesis. Collecting the dataset of the same sample with similar pixel numbers needed 24 hours with the old instruments, while the new instruments only need about 30min. For FT-IR, Bruker Corporation has also developed two new series of instruments, named LUMOS and HYPERION. LUMOS is a fully automated stand-alone FT-IR microscope. HYPERION is an instrument with highest measuring sensitivity and better spatial resolution. Now, these instruments can only work at laboratory for research. But it is no doubt that in the near future more hyperspectral instruments will be developed suitable for online and inline industry applications.

To develop the analysis of chocolate samples using hyperspectral imaging, Raman and Infrared hyperspectral imaging were used combined with MCR-ALS method. The following research paper has been written and published on *Applied Spectroscopy*.

**The published paper in this chapter**

Zhang, X.; Juan, A.; Tauler, R., Multivariate Curve Resolution applied to hyperspectral imaging analysis of chocolate samples. *Applied Spectroscopy*. 2015, 69(8).

## Multivariate Curve Resolution applied to hyperspectral imaging analysis of chocolate samples

Xin Zhang<sup>1</sup>, Anna de Juan<sup>2</sup>, Romà Tauler<sup>1\*</sup>

<sup>1</sup>IDAEA-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain

<sup>2</sup>Chemometrics group, Department of Analytical Chemistry, University of Barcelona, Barcelona 08028, Spain

### Abstract:

This paper shows the application of Raman and Infrared hyperspectral imaging combined with Multivariate Curve Resolution (MCR) to the analysis of the constituents of commercial chocolate samples. The combination of different spectral data pretreatment methods allowed decreasing the high fluorescent Raman signal contribution of whey in the investigated chocolate samples. Using equality constraints during MCR analysis, estimations of the pure spectra of the chocolate sample constituents were improved, as well as their relative contributions and their spatial distribution on the analyzed samples. In addition, unknown constituents could be also resolved. White chocolate constituents resolved from Raman hyperspectral image indicate that at macro scale, sucrose, lactose, fat and whey constituents were intermixed in particles. Infrared hyperspectral imaging did not suffer from fluorescence and could be applied for white and milk chocolate. As a conclusion of this study, micro-hyperspectral imaging coupled to the MCR method is confirmed to be an appropriate tool for the direct analysis of the constituents of chocolate samples, and by extension, it is proposed for the analysis of other mixture constituents in commercial food samples.

**Keywords:** Raman, Infrared, Hyperspectral image, MCR-ALS, chocolate, chemometric

### Introduction

The application of hyperspectral imaging is important in the simultaneous analysis of multiple constituents in complex food, pharmaceutical, agricultural<sup>1</sup> and environment samples<sup>2</sup> from different origin. For example, it has been applied in the homogenous analysis of food samples<sup>3</sup> or pharmaceutical tablets<sup>4</sup>. In addition, it represents a huge increase of the speed of experimental measurements and analyses, and therefore it represents a significant reduction of analysis cost. Hyperspectral images result from spectroscopic readings of hundreds of contiguous spectral channels at each spatial position (pixel) of the target sample under study. Each pixel in a hyperspectral image contains one multichannel (multi-wavelength) spectrum of the specific location. In the analysis of a sample, hyperspectral imaging provides a data block with three dimensions, two spatial dimensions ( $x$  and  $y$ ) and one spectral dimension ( $\lambda$ ) (see Figure 1). Hyperspectral imaging has been shown to be a very suitable approach for the qualitative and quantitative analysis of complex food products (like chocolate) where different ingredients can be mixed with different heterogeneity levels and particle sizes<sup>5</sup>.

Raman hyperspectral has been traditionally used for the analysis of the constituents of food and agriculture products. However, it is mostly only used for single spectral measurements at fixed locations of unaltered samples<sup>6</sup> or for measurement of average constituent after samples mixing and dissolution. Using this spectroscopic method, the configuration and relative amounts of carbon double bonds ( $C = C$ ) of sample constituents, the nature and amount of proteins and carbohydrates, and the chemical composition of many food products can be investigated directly<sup>7</sup>. Moreover, Raman hyperspectral imaging has a very good spatial resolution, like near field Raman imaging which can reach nanoscale resolution in some particular cases<sup>8-10</sup>. Although, in the past, Raman imaging was considered to be a slow technology, recently, with the advent of new detection strategies, hyperspectral images acquisition speed has increased significantly<sup>11-13</sup>. Most probably, in the near future, acquisition speed of Raman images will be suitable for process analysis in industry and its use will be standardized. When hyperspectral imaging is applied to research and development (R & D) activities, for instance in food analysis or medicine, online speed is unnecessary, and the present Raman market equipment's technique already meet analytical requirements. Infrared hyperspectral imaging has many common features to Raman, it does not suffer from the effect of fluorescence, and it has lower spatial resolution and has faster scanning speeds than Raman because it does not require so much focus on the sample surface. Although at present, infrared hyperspectral imaging has good resolution properties, it suffers from the interference signal from water. Infrared hyperspectral imaging needs an appropriate strategy to cope with this problem.

Due to the huge size and complexity of the spectral data obtained by hyperspectral imaging, chemometric methods, and in particular Multivariate Curve Resolution methods, are a good choice to extract information about the nature, amount and location of food constituents in samples from the information included in the raw measured images. MCR-ALS (Multivariate Curve Resolution Alternative Least Square) is a resolution method oriented to recover the underlying concentration profiles (which can be folded back into spatial distribution maps), as well as the pure spectra profiles (signatures) of the constituents of the analyzed samples (see

Figure 1). To perform the resolution of the concentration and spectra profiles, MCR-ALS solves a bilinear model under constraints based on known chemical and mathematical properties of the profiles to be resolved<sup>14</sup>. Hyperspectral imaging systems coupled to MCR-ALS have already been applied for the resolution of different type of problems using hyperspectral images<sup>15,16</sup>.

In the general case, results of MCR analysis are not unique and they can have an unknown amount of ambiguity<sup>17</sup>. In food industry, usually, the pure spectra of one or more components can be known from databases or experimentally<sup>18,19</sup>. In these cases, the use of this information as a constraint during MCR can help to decrease or remove the amount of MCR inherent ambiguity.

In this work, chocolate samples were chosen as a case study because like many other food and agriculture products, they are a complex multiphase mixture of particulate matter (sugar, cocoa and certain milk components) and of continuous phases (cocoa butter, milk fat and emulsifiers). Chocolate is a common product in food industry, whose analysis is time consuming and tedious. During chocolate manufacturing, refining and coating, properties such as particle size, suspension consistency and viscosity give specific textural and sensory qualities to the final product<sup>20</sup>. Control of the size of fat particle size distribution can affect the rheological and textural properties of chocolate, and it can also have impact of fat content in chocolate, whose absorption in consumers body may have health effects<sup>21</sup>. Using hyperspectral imaging, detection of constituent particles in chocolate and estimation of their size can be obtained and used for chocolate quality control. By extension, Raman hyperspectral imaging combined with data analysis chemometric methods is proposed in this work for other food and agriculture products analysis.

Hyperspectral imaging often faces the problem of noisy signals, baseline drifts, or sample surface scattering, and sometimes it also has the problem of different components at different scales. Data pretreatment usually reduces the contribution of large baseline variations, dimensionality reduction, spectral collinearity and high noise levels of the experimental spectra. This initial spectral data pretreatment is a key step for the successful analysis of hyperspectral data.

In this work, the MCR-ALS chemometric method was used to resolve the contributions and spatial distributions of the different constituents in samples of white and milk chocolate. Reference spectra of sucrose, lactose, butter and whey were available, which allowed the possibility to validate the results, and also the possibility to use them for the implementation of selectivity/local rank information as equality constraints and for further resolution improvement of the different chocolate constituents<sup>22</sup>. Finally, infrared hyperspectral imaging was also applied and the results compared with those obtained by Raman hyperspectral imaging.

## Experimental

### Materials

White and milk chocolate samples were collected from supermarket. For comparison, both Raman and Infrared hyperspectral imaging techniques were applied on the same block sample of chocolate (of the same packet).

All reagents are reference materials, sucrose (S7903, Sigma-Aldrich), lactose (PHR1025, Sigma-Aldrich), Butter (BCR-519, LGC STANDARS) and whey powder (IRMM-801, LGC STANDARS). Butter and whey were certified reference material from Institute for reference materials and measurements, European Commission Joint Research Centre (ECJRC).

### Experimental Method

HORIBA JobinYvon Raman microscope (HORIBA, Ltd.) was used for Raman spectra acquisition using 532 nm radiation laser. High-resolution mapping was carried out over a surface area of 41×41  $\mu\text{m}$ . 41 points per line and 41 lines per image were recorded. The instrument resolution is 1  $\mu\text{m}$ . Spectra were recorded every 0.2 s per point and had a spectral resolution of 1.6  $\text{cm}^{-1}$  in the spectral range of 1890-219  $\text{cm}^{-1}$ . Laser power was 15 mW.

The Nicolet iN10 MX Infrared Imaging Microscope with MCT (Mercury cadmium telluride) detector was used for Infrared spectra acquisition and high-resolution mapping over a surface area of 300×300  $\mu\text{m}$ , 30 points per line and 30 lines per image. The instrument resolution is 10  $\mu\text{m}$ . The spectral range was from 4000  $\text{cm}^{-1}$  to 675  $\text{cm}^{-1}$ , the spectral resolution was 4  $\text{cm}^{-1}$ , for every pixel average spectrum of 8 scans was considered, Acquisition of every spectrum at each pixel of the image took 1s. And an internal background was used as a reference. NIR spectral data were recorded using Thermo Scientific™ OMNIC™ Picta™ Software.

Individual Raman and Infrared pure spectra of sucrose, lactose, butter and whey, were obtained in the similar situation as the chocolate samples.

### Data Analysis

#### Data Pretreatment

##### Data Pretreatment for Raman hyperspectral imaging

Multiplicative Scatter Correction (MSC)<sup>23, 24</sup>, Asymmetric Least Squares (AsLS)<sup>25, 26</sup> and Savitzky-Golay smoothing methods<sup>27</sup> were applied to eliminate undesired light scattering, instrumental and background effects.

MSC is a general signal treatment method which has been applied for eliminating the effect of scattering physical effects, and also as a preprocessing method to linearize spectral data, in such a way that better predictions are obtained when applied to MSC corrected data. It corrects the scatter level for a group of samples spectra to the level of a preselected spectrum, which usually is the sample average spectrum<sup>28</sup>. In this work, the preselected spectrum was the average spectrum of the raw spectra. This kind of correction is particularly suited when the baseline caused by scattering among samples is similar in shape and differs only in slope and offset, although it can also improve the measured signal in cases of more irregular baselines.

AsLS is another pretreatment method proposed by P. C. Eilers<sup>29</sup>, which reduces the possible baseline and background contributions, especially when they are irregularly shaped, lower in frequency than the useful signal (as it happens when comparing fluorescence contribution and spectroscopic features in a Raman spectrum) and pixel-to-pixel variations may occur. It estimates a baseline/background contribution by minimizing a penalized least squares function based on the Whittaker smoother<sup>30</sup>. In this work, the Whittaker smoother parameters  $\rho$  and  $\mu$  were set to 0.001 and  $1 \times 10^7$ , respectively; this choice is always data set-dependent.

Savitzky-Golay (SG) is a very popular smoothing technique, which uses convolution arrays derived from the polynomial coefficients of least-squares fitting formulas<sup>27</sup>. The convolution can be understood as a weighted moving average filter with a weighting polynomial of a certain degree. It is used in this work, to remove the larger noise contributions of Raman signals. In this paper, Savitzky-Golay smoothing filter window width and smoothing polynomial degree were 5 points width and third polynomial degree respectively.

#### Data Pretreatment of Infrared hyperspectral images

Infrared active photons, both longitudinal and transverse, are mostly determined from reflectance spectra in the infrared spectral range. These spectra do not reveal photon frequencies directly<sup>31</sup>. It is a powerful tool enabling a calculation of absorption spectra in case only reflectance spectra can be measured.

The complex reflectance spectrum is mathematically decomposed into two separate spectra-extinction coefficient and refractive index spectrum. The extinction coefficient spectrum can be then used to calculate the absorption spectrum<sup>31</sup>.

The real ( $n$  = refractive index) and imaginary ( $k$  = extinction) parts of the complex index of refraction are calculated from the reflectance spectrum using the following formulas:

$$n(\nu) = \frac{1-R(\nu)}{1+R(\nu)-2\sqrt{R(\nu)}\cos(\theta(\nu))} \quad \text{Equation 1}$$

$$k(\nu) = \frac{-2\sqrt{R(\nu)}\cos(\theta(\nu))}{1+R(\nu)-2\sqrt{R(\nu)}\cos(\theta(\nu))} \quad \text{Equation 2}$$

where  $R$  is the reflectance,  $n$  is wavenumber,  $\theta$  is phase shift angle of the sample.

For a given wavenumber, the phase shift is calculated using the equation:

$$\theta(\nu_m) = \frac{2\nu_m}{\pi} \int_0^\infty \frac{\ln\sqrt{R(\nu)}d\nu}{\nu^2-\nu_m^2} \quad \text{Equation 3}$$

The Kramers-Kronig transform algorithm assumes that the reflectance spectra are measured at incidence angles close to zero.<sup>32</sup>

In this work, Kramers-Kronig transformation has been implemented as initial standard spectral data preprocessing using appropriate software (Thermo Scientific™ OMNIC™ Picta™ Software, from Thermo Fisher Scientific Inc.).

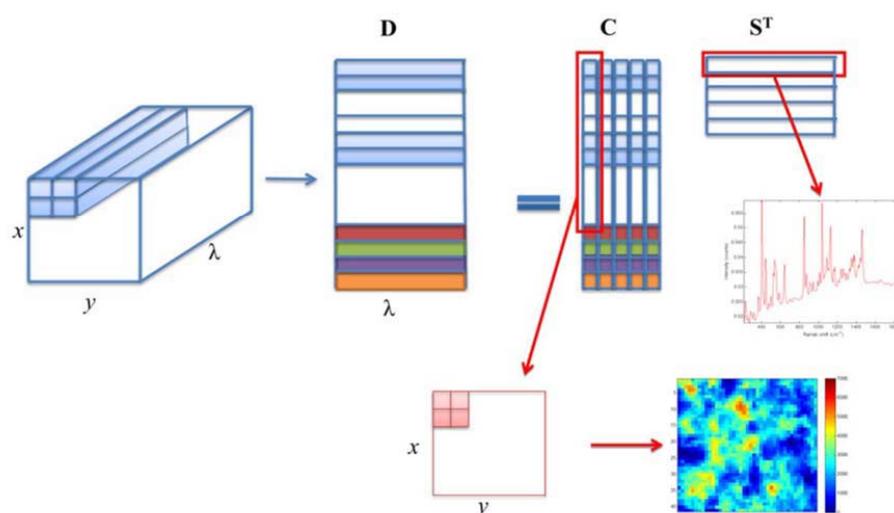
#### Resolution of Raman and Infrared hyperspectral images from chocolate samples

MCR-ALS method has been used to decompose the 2D chocolate Raman and Infrared hyperspectral images into the contributions (distribution map) and pure spectra (or

signatures) of the image constituents. MCR-ALS is based on a simple bilinear model which assumes that the constituents of the analyzed sample weighted according to their relative concentrations are expressed by the simple matrix equation as:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{Equation 4}$$

where  $\mathbf{D}$  ( $n,m$ ) is the unfolded data matrix coming from the 2D image see Figure 1, of dimensions  $n$  (equal to  $x \times y$ ) pixels by  $m$  wavenumbers. The  $n$  rows of this matrix contain the experimentally measured spectra for every pixel of the spectroscopic image.  $\mathbf{C}$  ( $n,n_c$ ) is the matrix of the relative amounts or concentrations of the  $n_c$  components in the  $n$  pixels,  $\mathbf{S}^T$  ( $n_c,m$ ) is the pure spectra (signatures) matrix associated with these  $n_c$  components present in the image at  $m$  wavenumbers.  $\mathbf{E}$  ( $n,m$ ) is the matrix associated with noise or experimental error (variance not explained by the  $n_c$  resolved components).



**Figure 1.** Scheme of MCR-ALS analysis of a Raman hyperspectral image. Resolution of the pure spectra (signatures) of the image constituents and of their 2D image distribution (mapping). The known reference spectra of sucrose (red), lactose (green), butter (purple) and whey (orange) were added at the bottom of the unfolded data matrix.

MCR-ALS working steps have been already described in detail in previous works<sup>15,33</sup>. Figure 1 illustrates how MCR-ALS was used for the resolution of the Raman hyperspectral images analyzed in this work.

The first step is to unfold the 3D data cube ( $x \times y \times \lambda$ ) as a 2D data matrix ( $n \times m$ , where  $n = x \times y$ ,  $m = \lambda$ ) ready for MCR-ALS analysis. Since the information provided in the two spatial directions is of the same type, the data should be modelled by a bilinear model. In addition, in a second step of the data analysis, some or all of the known reference spectra of sucrose, lactose, butter and whey can also be considered and added at the bottom of the unfolded data matrix ( $(n+4) \times m$ ). This last step has the aim of taking into account the possibility of using this prior information available and, consequently, improving MCR-ALS resolution by reduction of possible rotation ambiguities<sup>17</sup> in the resolved profiles. In food chemistry, this situation is rather common, and reference spectra (spectral signatures) of some of but not all of the chemical constituents of the analyzed food samples are possibly known. In this work, since chocolate is a rather well characterized material, and their main components, like sucrose, lactose, have well known spectra from libraries or from the analysis of samples only containing them at similar experimental conditions<sup>7, 34, 35</sup>, they can be used to improve MCR-ALS results. Experimental spectra of sucrose, lactose, butter and whey were obtained experimentally in this work using the same Raman and FT-IR spectroscopic systems used for the analysis of chocolate samples.

The first steps before proceeding to the ALS optimization of  $\mathbf{C}$  and  $\mathbf{S}^T$  are the estimation of the number of components in  $\mathbf{D}$  and the generation of initial estimates of  $\mathbf{C}$  and  $\mathbf{S}^T$ . The number of components can be initially estimated by methods, such as Singular Value Decomposition (SVD). For hyperspectral images, spectral initial estimates were obtained from the selection of purest (more dissimilar) spectra of the raw hyperspectral image data set  $\mathbf{D}$ <sup>36</sup>.

MCR-ALS was applied using non-negativity constraints to concentration and pure spectra profiles of the image constituents<sup>37, 38</sup>. To avoid scale indeterminacies and stabilize the ALS iterative optimization, spectral profiles in matrix  $\mathbf{S}^T$  were also normalized to unit length.

Optionally, during ALS optimization, equality constraints were also tested in the concentration profiles to fix some specific values to zero<sup>14, 33</sup>. This was applied for some of the spectra included in the analysis which correspond to the experimentally measured reference spectra of the constituents of chocolate (like sucrose, lactose, butter and whey). These spectra were appended to the image data matrix (see Figure 1), and only the concentration of them is allowed to be different to zero during the ALS optimization (zero concentration equality or selectivity constraint).

The detailed procedure describing how to apply this zero concentration equality constraint in MCR-ALS is displayed in Figure 1. The experimental reference spectra of the chocolate constituents are appended at the bottom of the  $\mathbf{D}$  image data matrix. The *cseI* masking concentration matrix variable is defined, with the same rows and columns as the  $\mathbf{C}$  matrix. All matrix elements of *cseI* with unknown concentration values are marked with the MATLAB notation 'NaN' (not a number), and those elements where the concentration of one or more components is known, will contain this known value. For instance for the case of the added reference spectrum of sucrose, the corresponding values of the concentrations of the other components are set to zero and only in the sucrose position in the *cseI* matrix appears the symbol *NaN*. It is important to mention that equality constraints applied in this way are an efficient way to provide the estimation of the reference spectra during the optimization, and

that this approach is better than if reference spectra are fixed in the  $\mathbf{S}^T$  matrix during ALS. This approach gives more flexibility for the optimal least squares final estimation of the corresponding pure spectra and its possible adaptation to small spectral changes in chocolate samples compared to reference spectra.

The quality of MCR-ALS data fitting is evaluated by the percentage of lack of fit (lof) (Equation 5) and the percentage of explained variance ( $R^2$ ) (Equation 6) calculated according to the two following equations:

$$\text{lof}(\%) = 100 \times \sqrt{\frac{\sum_{ij} (d_{ij} - \hat{d}_{ij})^2}{\sum_{ij} d_{ij}^2}} \quad \text{Equation 5}$$

$$R^2 = 100 \times \left[ 1 - \frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2} \right] \quad \text{Equation 6}$$

Where  $d_{ij}$  is the element of the hyperspectral image data matrix  $\mathbf{D}$ , and  $\hat{d}_{ij}$  is the corresponding element of this data matrix recalculated by the ALS model,  $\hat{\mathbf{D}} = \mathbf{CS}^T$ . Lacks of fit values give a measure of the fit quality in relative terms with the same units as the measured data, and comparable with experimental relative error estimations.  $e_{ij}$  are the elements of the  $\mathbf{E}$  matrix and  $d_{ij}$  are the elements of the raw data set  $\mathbf{D}$ .

Additionally, when reference spectra are available, the agreement between them and a particular resolved profile can be calculated using the correlation ( $r^2$ ) coefficient and the vector angle between them, using Equations 7 and 8,

$$r^2 = \frac{\mathbf{xy}^T}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad \text{Equation 7}$$

$$\text{angle} = \frac{180}{\pi} \times \arccos \left[ \frac{\mathbf{xy}^T}{\|\mathbf{x}\| \|\mathbf{y}\|} \right] \quad \text{Equation 8}$$

where  $\mathbf{x}$  is the vector of resolved profiles and  $\mathbf{y}$  is the vector of reference spectra vectors.

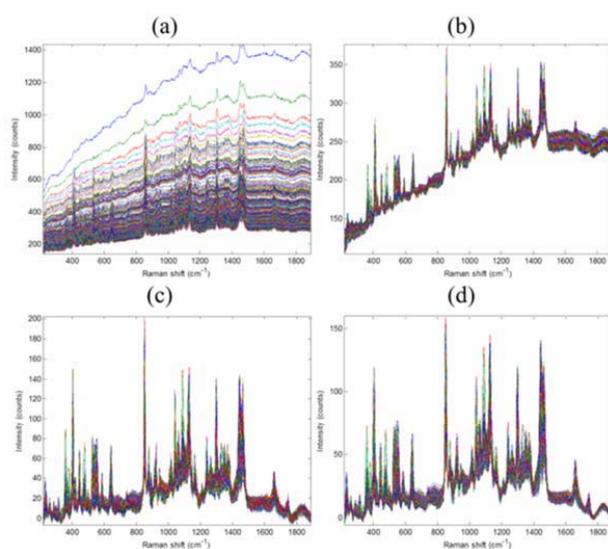
MCR-ALS algorithm code and its GUI for MATLAB (The Mathworks Inc., MA, US) is freely available from the home page of MCR at <http://www.mcrals.info/>.

## Results:

### MCR-ALS resolution of Raman hyperspectral images

In Raman hyperspectral images, milk chocolate signal was totally dominated by fluorescence, therefore only white chocolate results will be discussed in this section.

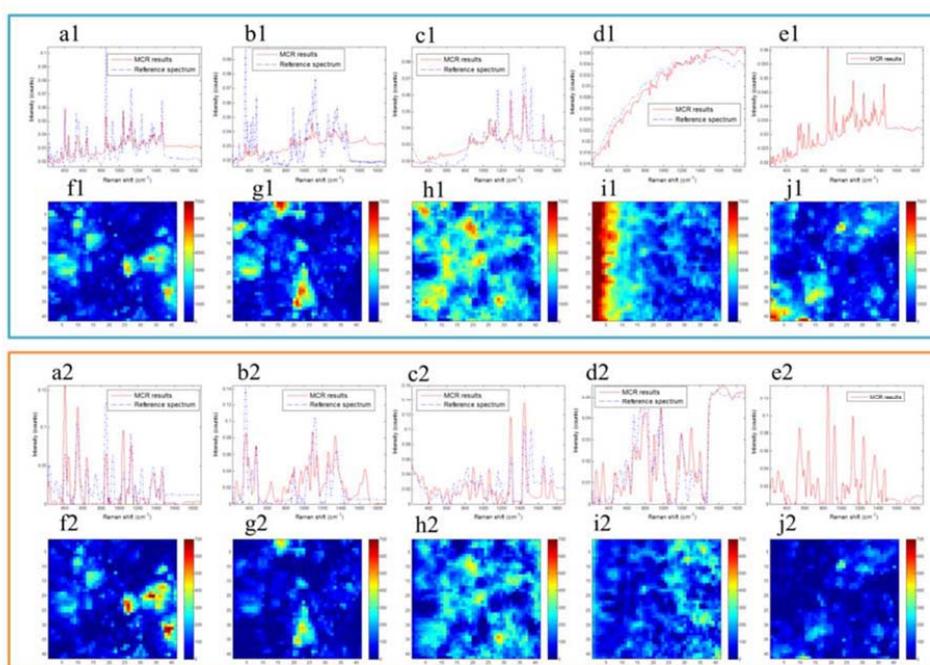
Results using different pretreatment methods on raw Raman spectra are given in Figure 2. In Figure 2a, raw (untreated) Raman spectra on chocolate image pixels are shown. High irregular baseline and noise contributions are observed in them. In Figure 2b, MSC (Multiplicative Scatter Correction, <sup>23,24</sup>) processed Raman spectra have much lower scatter contributions. In Figure 2c, previous MSC spectra were additionally processed using the AsLS procedure <sup>29,30</sup>, to correct their fluorescent background contribution. Although the fluorescent contribution was removed satisfactorily, the results still show noisy patterns. To diminish this noisy patterns, the Savitzky-Golay smoothing procedure <sup>39</sup> was applied to previously MSC-AsLS pretreated spectra, and instrumental and environmental noise contributions were significantly removed (Figure 2d).



**Figure 2.** (a) Raw Raman spectra of a chocolate sample. Spectra after (b) MSC. (c) MSC + AsLS and (d) MSC + AsLS + Savitzky-Golay pretreatment.

Tables 1 and 2 summarize MCR-ALS results with and without signal pretreatment (MSC + AsLS + Savitzky-Golay smoothing). In Table 1, explained variances and lack of fit values for the different MCR-ALS tested models are given. MCR-ALS results gave good data fitting results (Table 1) in both cases (untreated and pretreated spectra). Contributions of sucrose, lactose and butter were higher after signal pretreatment but, in contrast, the contribution of whey was significantly lower. This decrease of whey contribution is due to the fact that the resolved spectrum of this compound in the raw images included a high fluorescence contribution and this was significantly removed after signal pretreatment with the AsLS pretreatment method. Indeed, after signal pretreatment, whey resolved spectrum had a lower baseline contribution and showed more characteristic and meaningful bands. As a consequence of this whey signal contribution decrease, relative amounts of sucrose, lactose and butter increased in the results obtained with pretreated data.

Table 2 gives also the comparison between MCR-ALS resolved spectra and the corresponding reference spectra of sucrose, lactose, butter and whey. Correlation coefficients ( $r^2$ , Equation 7) and angles (Equation 8) between MCR-ALS resolved spectra and reference spectra are given for different combinations of data pretreatment methods (like MSC, AsLS and SG), and for different combination of constraints (non-negativity and equality). Best recovery values are underlined in this Table. MCR-ALS resolved spectra from chocolate sample constituents from data without and with signal pretreatments (with MSC, AsLS and Savitzky-Golay smoothing together) are given in the upper part and lower parts of Figure 3, respectively. Red lines are the MCR-ALS resolved pure spectra and blue lines are the reference spectra.



**Figure 3.** Pure spectra of the image constituents and their 2D image distribution resolved by MCR-ALS with (lower part) and without (upper part) any signal pretreatment. Reference and MCR-ALS resolved spectra are given in red and blue lines respectively. (a) sucrose, (b) lactose, (c) butter, (d) whey and (e) a fifth unknown resolved component.

As it is seen from Tables 1 and 2 and in Figure 3, application of the different signal pretreatment methods could remove a significant amount of the fluorescence contribution from the Raman signal. The high contribution shown in Figure 3i1 was removed as shown in Figure 3i2. The agreement between reference and resolved spectra was considered very satisfactory in all cases.

Equality/selectivity constraints<sup>17</sup> were also applied during MCR-ALS analysis in two different ways (see description of how this was applied in section 2.3.1). First, they were applied to three of the components (sucrose, lactose and butter) and then they were also applied to all four known components (sucrose, lactose, butter and whey). Since the reference spectrum of whey had a very high fluorescence contribution, it was difficult to distinguish its characteristic spectral bands. When equality constraints were applied (information about reference spectra was given by using *cse1* mask concentration matrix, see method section 2.3.1), resolution of the pure spectra of the constituents was better than when the equality constraints were not applied (see Tables 1 and 2). Results obtained with equality constraints applied to either three or four components gave some small differences but they did not change significantly<sup>40</sup>.

### **MCR-ALS resolution of the spatial distribution of chocolate constituents on Raman hyperspectral images**

Figure 3 also shows the MCR-ALS resolved spatial distributions (maps) of the different chocolate constituents on the resolved Raman hyperspectral image. Butter (Figures 3h1 and 3h) is rather uniformly distributed over the whole sample image. Distributions of sucrose (Figures 3f1 and 3f2), of lactose (Figures 3g1 and 3g2), and of an unknown contribution (Figures 3j1 and 3j2), were mostly clustered heterogeneously in separate particles over the whole image. Spatial distribution of sucrose, of lactose and of the fifth additional component on the image showed more or less similar particle sizes. Whey in contrast, has more uniform distribution than these sugars, and it has smaller particle sizes. Butter and whey are distributed in an interweave net of both (thicker for butter and thinner for whey), and wrapped around the other components.

The four reference spectra suffered from fluorescence contributions at different levels. Sucrose, lactose and butter have a low level of fluorescence contribution. Comparing the results before and after spectra pretreatment, resolved distribution maps of sucrose, lactose and butter were almost equal except for a small change in their relative contribution.

With no data pretreatment, reference and resolved spectra of whey showed always high levels of fluorescence and it was difficult to find its characteristic spectral bands. Spatial distribution of the whey contribution requires further discussion. Figure 3i1 gives the relative spatial distribution map of whey component when no pretreatment was applied. Due to the high fluorescence contribution of this component, it appears at high amount in the left side of this Figure, just at the beginning of the hyperspectral image scanning. This fluorescence contribution caused a significant baseline which hid completely the information about the other spectral bands and image constituents. After the appropriate data pretreatment (especially using background/baseline correction by the AsLS method), fluorescence disappeared considerably (Figures 3d2 and 3i2) and the other chemical image constituents could be better resolved.

Before and after data pretreatment, MCR-ALS resolved spectra, and the related distribution maps of sugar and butter did not suffer from the fluorescence contribution. As a conclusion, if the aim of the work is to resolve sugars and butter fat constituents in chocolate samples, and

the estimation of the distribution map of whey is not important, MCR-ALS can be applied directly to raw data, without losing relevant information and without the need of spending long time in the application and validation of signal pretreatment methods. On the other hand, when the estimation of whey contribution is important, then fluorescence contribution should be removed by application of signal pretreatment methods such as MSC, and Savitzky-Golay smoothing, and specially AsLS. This will allow for a good distribution map of this component at a pixel level in the image (Figure 3i2) and to improve the resolution of its spectra too.

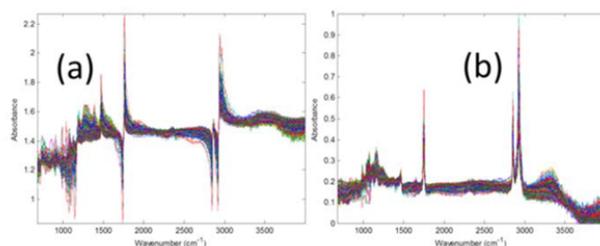
More important band frequencies can be deduced from the finally MCR-ALS resolved Raman spectra, which can be used then for their assignment to specific chemical groups. For instance, the fifth unknown component resolved by MCR has Raman bands at  $843\text{ cm}^{-1}$  (C-C),  $916\text{ cm}^{-1}$  (CO),  $1121\text{-}1131\text{ cm}^{-1}$  (COH) and  $1461\text{ cm}^{-1}$  (CHO). When compared to library spectra, this spectrum resulted to be rather similar to the Raman spectra of glucose as defined in a previous report .

#### Resolution of Infrared hyperspectral images by MCR-ALS

Differently to Raman imaging, infrared hyperspectral imaging does not suffer the impact of fluorescence and can obtain information from both milk and white chocolate.

#### Spectra of chocolate constituents on Infrared spectral images resolved by MCR-ALS

MCR-ALS resolved four components in white chocolate samples. Kramers-Kronig processing was applied during dataset exportation (Figure 4) after spectra acquisition (Thermo Scientific™ OMNIC™ Picta™ software). Table 3 gives the correlation coefficients and angles between reference spectra profiles and spectra profiles resolved by MCR-ALS. MCR-ALS fitting results are also given in the table. In all cases, either selectivity constraints were applied or not, the obtained MCR models were satisfactory. The use of selectivity constraints always improved spectral recoveries. For comparison, Table 3 also gives the results with pretreatment (MSC + AsLS + Savitzky-Golay smoothing). In the case of the infrared dataset, spectra pretreatment removed baseline but it also removed some chemical information from the raw dataset. Lack of fit (lof) increased a little amount after data pretreatment. Resolved spectra of sucrose fitted better its reference spectrum but lactose fitted worse. Therefore, in this case MSC + AsLS + Savitzky-Golay smoothing pretreatments were not applied in the case of infrared hyperspectral imaging.

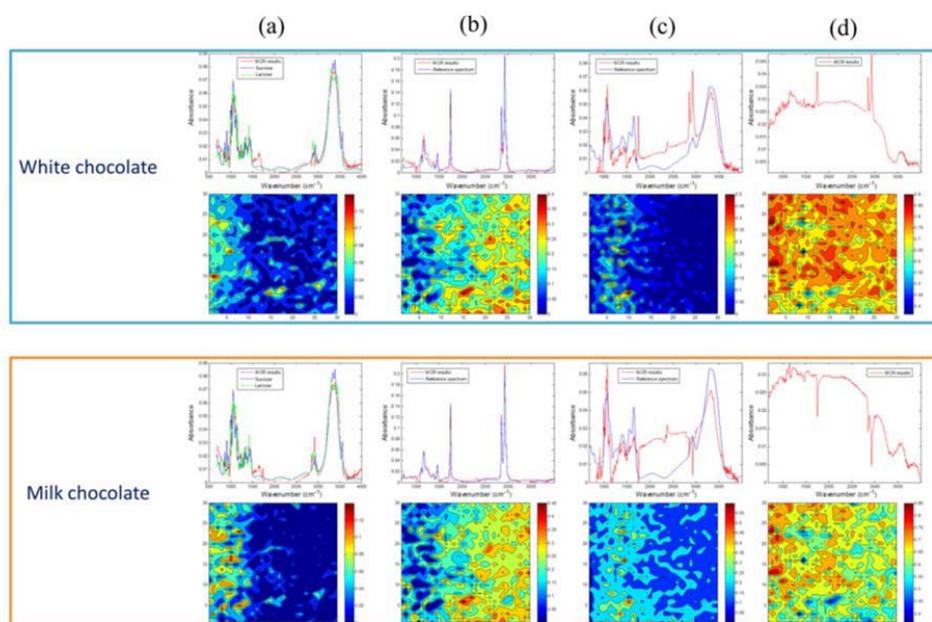


**Figure 4.** (a) Raw infrared spectra of a chocolate sample. (b) Spectra after Kramers-Kronig pretreatment.

MCR-ALS resolved spectrum for component 1 was compared with sucrose and lactose reference spectra. Since sucrose and lactose pure FT-IR spectra are very similar, MCR-ALS had an inherent difficulty distinguishing them, and their contributions were resolved in the same component. An additional fourth unknown spectrum was resolved by MCR-ALS which may correspond to another component of chocolate, like cocoa. MCR-ALS resolved spectra in milk and white chocolate were very similar. The only different spectrum was the unknown fourth component preliminary assigned to cocoa.

#### Spatial distribution of chocolate constituents on Infrared spectral images resolved by MCR-ALS

Figure 5 also shows the MCR-ALS resolved distribution map of the different components in white chocolate and milk chocolate analyzed samples by FT-IR. Sugars were mixed with butter fat in the constituent particles. Milk chocolate sample had less contribution of sugars and fat and more of the fourth component. When the figures are directly looked, all the constituents showed similar distribution patterns in milk and white chocolate samples. Comparing the results of the MCR-ALS analysis of FT-IR and Raman hyperspectral images, it is clear that the former did not provide as much resolution as Raman imaging and it could not show as many details of the particles shape. Using a better resolution infrared hyperspectral imaging instrument, distribution maps of both constituents white and milk chocolate could be obtained.



**Figure 5.** Pure spectra of the image constituents and their 2D image distribution of white (upper) and milk (lower) samples resolved by MCR-ALS. Normalized reference spectra of sucrose, butter and whey

are shown in blue, lactose is shown in green, and normalized MCR-ALS resolved spectra are given in red. (a) sugar (sucrose and/or lactose), (b) butter, (c), whey and (d) a fourth unknown resolved component.

### Conclusions

In this work, Raman and FT-IR hyperspectroscopic imaging combined with MCR-ALS with non-negativity and equality constraints have properly identified and resolved the main constituents of commercial white chocolate samples and mapped their spatial distribution on the analyzed images. Sucrose, lactose, fat, whey and an additional unknown constituent were resolved by MCR-ALS. This unknown constituent resulted to have a Raman spectrum very similar to glucose. Application of image data pretreatment methods such as Multiplicative Scatter Correction (MSC), Asymmetric Least Squares (AsLS) and Savitzky-Golay (SG) smoothing reduced significantly the presence of strong fluorescence background in Raman spectra, and resolved whey characteristic bands. From the distribution of the different constituents on the image, butter and whey were spread on the image in way of interweave (thicker in butter and thinner in whey), and wrapped around sugar particles. The other three constituents were heterogeneously distributed in separate particulates.

Application of infrared hyperspectral imaging was not affected by the strong fluorescence of some chocolate constituents, and it could be applied to the analysis of white and milk chocolate.

Results achieved in this work are representative of the possibilities offered by hyperspectral imaging analysis of food samples, and confirmed the potential use of the MCR-ALS as a complementary tool method for this type of analysis. Other similar fields of application of this combination of methods (hyperspectral imaging and MCR-ALS) include the analysis of other material surfaces, agriculture products as well as biological tissues in medicine.

### 5 Acknowledgements:

Authors acknowledge research grant CTQ2012-3816 from Ministerio de Economía y competitividad, Spain. Xin Zhang thanks the grant from China Scholarship Committee (2011637007).

## References

1. R. Lu, Y.-R. Chen. In "Hyperspectral imaging for safety inspection of food and agricultural products", Photonics East (ISAM, VVDC, IEMB), 1999; International Society for Optics and Photonics: 1999; Pp 121-133.
2. L.J. Otten III, E.W. Butler, B. Rafert, R.G. Sellar. In "Design of an airborne Fourier transform visible hyperspectral imaging system for light aircraft environmental remote sensing", SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics, 1995; International Society for Optics and Photonics: 1995; Pp 418-424.
3. Y.Z. Feng, D.W. Sun. "Application of hyperspectral imaging in food safety inspection and control: a review". *Crit. Rev. Food Sci. Nutr.* 2012. 52(11): 1039-58.
4. A. de Juan, M. Maeder, T. Hanczewicz, R. Tauler. "Local rank analysis for exploratory spectroscopic image analysis. Fixed size image window-evolving factor analysis". *Chemometrics Intellig. Lab. Syst.* 2005. 77(1): 64-74.
5. A. Gowen, C. O'Donnell, P. Cullen, G. Downey, J. Frias. "Hyperspectral imaging—an emerging process analytical tool for food quality and safety control". *Trends Food Sci. Technol.* 2007. 18(12): 590-598.
6. A.M. Herrero. "Raman spectroscopy a promising technique for quality assessment of meat and fish: A review". *Food Chem.* 2008. 107(4): 1642-1651.
7. S. Keller, T. Löchte, B. Dippel, B. Schrader. "Quality control of food with near-infrared-excited Raman spectroscopy". *Fresenius J. Anal. Chem.* 1993. 346(6-9): 863-867.
8. N. Anderson, P. Anger, A. Hartschuh, L. Novotny. "Subsurface Raman Imaging with Nanoscale Resolution". *Nano Lett.* 2006. 6(4): 744-749.
9. A. Hartschuh, E.J. Sánchez, X.S. Xie, L. Novotny. "High-Resolution Near-Field Raman Microscopy of Single-Walled Carbon Nanotubes". *Phys. Rev. Lett.* 2003. 90(9): 095503.
10. C. Jahncke, M. Paesler, H. Hallen. "Raman imaging with near - field scanning optical microscopy". *Appl. Phys. Lett.* 1995. 67(17): 2483-2485.
11. S. Schlücker, M.D. Schaeberle, S.W. Huffman, I.W. Levin. "Raman Microspectroscopy: A Comparison of Point, Line, and Wide-Field Imaging Methodologies". *Anal. Chem.* 2003. 75(16): 4312-4318.
12. B.M. Davis, A.J. Hemphill, D. Cebeci Maltaş, M.A. Zipper, P. Wang, D. Ben-Amotz. "Multivariate Hyperspectral Raman Imaging Using Compressive Detection". *Anal. Chem.* 2011. 83(13): 5086-5092.
13. L. Opilik, T. Schmid, R. Zenobi. "Modern Raman Imaging: Vibrational Spectroscopy on the Micrometer and Nanometer Scales". *Annu. Rev. Anal. Chem.* 2013. 6(1): 379-398.
14. R. Tauler, B. Kowalski, S. Fleming. "Multivariate curve resolution applied to spectral data from multiple runs of an industrial process". *Anal. Chem.* 1993. 65(15): 2040-2047.
15. S. Piqueras, L. Duponchel, R. Tauler, A. De Juan. "Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares". *Anal. Chim. Acta* 2011. 705(1): 182-192.
16. X. Zhang, R. Tauler. "Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging". *Anal. Chim. Acta* 2013. 762(0): 25-38.
17. R. Tauler, A. Smilde, B. Kowalski. "Selectivity, local rank, three - way data analysis and ambiguity in multivariate curve resolution". *J. Chemom.* 1995. 9(1): 31-58.
18. M. de Veij, P. Vandenabeele, T. De Beer, J.P. Remon, L. Moens. "Reference database of Raman spectra of pharmaceutical excipients". *J. Raman Spectrosc.* 2009. 40(3): 297-307.
19. L.G. Thygesen, M.M. Løkke, E. Micklander, S.B. Engelsen. "Vibrational microspectroscopy of food. Raman vs. FT-IR". *Trends Food Sci. Technol.* 2003. 14(1): 50-57.
20. E.O. Afoakwa, A. Paterson, M. Fowler. "Factors influencing rheological and textural qualities in chocolate – a review". *Trends Food Sci. Technol.* 2007. 18(6): 290-298.

21. N. Konar, B. Özhan, N. Artık, S. Dalabasmaz, E.S. Poyrazoglu. "Rheological and physical properties of inulin-containing milk chocolate prepared at different process conditions". *CyTA-- J. Food* 2014. 12(1): 55-64.
22. J. Jaumot, R. Gargallo, A. de Juan, R. Tauler. "A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB". *Chemom. Intell. Lab. Syst.* 2005. 76(1): 101-110.
23. T. Isaksson, T. Næs. "The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy". *Appl. Spectrosc.* 1988. 42(7): 1273-1284.
24. S. Romero-Torres, J.D. Pérez-Ramos, K.R. Morris, E.R. Grant. "Raman spectroscopic measurement of tablet-to-tablet coating variability". *J. Pharm. Biomed. Anal.* 2005. 38(2): 270-274.
25. J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, J. Tan. "Asymmetric least squares for multiple spectra baseline correction". *Anal. Chim. Acta* 2010. 683(1): 63-68.
26. W. Newey, J.L. Powell. "Asymmetric Least Squares Estimation and Testing". *Econometrica* 1987. 55(4): 819-47.
27. M.U. Bromba, H. Ziegler. "Application hints for Savitzky-Golay digital smoothing filters". *Anal. Chem.* 1981. 53(11): 1583-1586.
28. P. Geladi, D. MacDougall, H. Martens. "Linearization and scatter-correction for near-infrared reflectance spectra of meat". *Appl. Spectrosc.* 1985. 39(3): 491-500.
29. P.H. Eilers. "Parametric time warping". *Anal. Chem.* 2004. 76(2): 404-411.
30. P.H. Eilers. "A perfect smoother". *Anal. Chem.* 2003. 75(14): 3631-3636.
31. K. Ohta, H. Ishida. "Comparison among several numerical integration methods for Kramers-Kronig transformation". *Appl. Spectrosc.* 1988. 42(6): 952-957.
32. H.H. Dunken, R. Stephanowitz. "Infrarot - Reflexionsspektren von Festkörper - und Glasoberflächen". *Zeitschrift für Chemie* 1983. 23(10): 353-365.
33. J. Jaumot, R. Gargallo, A. de Juan, R. Tauler. "A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB". *Chemometrics Intellig. Lab. Syst.* 2005. 76(1): 101-110.
34. I. Larmour, K. Faulds, D. Graham. "Rapid Raman mapping for chocolate analysis". *Anal. Methods* 2010. 2(9): 1230-1232.
35. H. Dahlenborg, A. Millqvist - Fureby, B.D. Brandner, B. Bergenstahl. "Study of the porous structure of white chocolate by confocal Raman microscopy". *Eur. J. Lipid Sci. Technol.* 2012. 114(8): 919-926.
36. W. Windig, J. Guilment. "Interactive self-modeling mixture analysis". *Anal. Chem.* 1991. 63(14): 1425-1432.
37. W.H. Lawton, E.A. Sylvestre. "Self Modeling Curve Resolution". *Technometrics* 1971. 13(3): 617-633.
38. D.W. Osten, B.R. Kowalski. "Multivariate curve resolution in liquid chromatography". *Anal. Chem.* 1984. 56(6): 991-995.
39. A. Savitzky, M.J.E. Golay. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures". *Anal. Chem.* 1964. 36(8): 1627-1639.
40. S. Söderholm, Y.H. Roos, N. Meinander, M. Hotokka. "Raman spectra of fructose and glucose in the amorphous and crystalline states". *J. Raman Spectrosc.* 1999. 30(11): 1009-1018.

## Application of hyperspectral imaging combined with chemometrics on food analysis

Table 1 MCR-ALS results in the analysis of a white chocolate sample using Raman hyperspectral imaging data

Constraints		Sucrose <sup>1</sup>	Lactose <sup>1</sup>	Butter <sup>1</sup>	Whey <sup>1</sup>	Comp5 <sup>2</sup>	LoF <sup>3</sup> (lack of fit)%	R <sup>2</sup> (%) <sup>4</sup>
Raw Data No Pretreatment	Non-negativity	13.79	10.45	22.75	34.90	11.52	1.04	99.99
	Non-negativity +equality constraint on 3 components	13.65	10.78	20.59	35.11	11.68	1.08	99.99
	Non-negativity +equality constraint on 4 components	13.63	10.02	22.92	36.28	11.62	1.51	99.98
Pretreated data (MSC+AsLs+ +Savitzky Golay)	Non-negativity	14.03	14.14	34.16	12.95	13.12	6.67	99.56
	Non-negativity +equality constraint on 3 components	15.38	10.41	35.70	13.10	13.04	7.56	99.43
	Non-negativity +equality constraint on 4 components	15.38	10.41	35.70	13.10	13.04	7.56	99.43

<sup>1</sup> Percentage of explained variance (R<sup>2</sup>) considering only this component.

<sup>2</sup> Percentage of explained variance (R<sup>2</sup>) considering only the fifth unknown resolved component.

<sup>3</sup> MCR-ALS lack of fit (lof) considering the five components (full model) together.

<sup>4</sup> MCR-ALS percentage of explained variances (R<sup>2</sup>) considering the five components (full model) together.

17

Table 2 Correlation coefficients (r<sup>2</sup>, Equation 7) and angles (Equation 8) between the resolved spectra by MCR-ALS (full model) and the reference spectra in the analysis of the Raman hyperspectral image of the white chocolate sample

Applied Constraints <sup>1</sup>			Sucrose	Lactose	Butter	Whey
Raw data No pretreatment	Non-negativity	r <sup>2</sup> (angle <sup>3</sup> )	0.9590(16.46)	0.9255(22.26)	<u>0.9698(14.12)</u>	0.9989(2.65)
	Non-negativity + + equality constraints on 3 components	r <sup>2</sup> (angle)	0.9596(16.33)	0.9220(22.78)	0.9687(14.38)	0.9990(2.61)
	<b>Non-negativity + + equality constraints on 4 components</b>	r <sup>2</sup> (angle)	<b>0.9709(13.85)</b>	<b>0.9220(22.78)</b>	<b>0.9680(14.54)</b>	<b>0.9991(2.57)</b>
Pretreatment (MSC+AsLs+ +Savitzky Golay)	Non-negativity	r <sup>2</sup> (angle)	9.3612(20.59)	0.6701(47.93)	0.8972(26.21)	<u>0.9814(11.08)</u>
	Non-negativity +equality constraint on 3 components	r <sup>2</sup> (angle)	<u>0.9606(16.13)</u>	0.8431(32.43)	<u>0.9042(25.29)</u>	0.9582(16.63)
	<b>Non-negativity +equality constraint on 4 components</b>	r <sup>2</sup> (angle)	<b>0.9497(18.25)</b>	<b>0.9910(7.71)</b>	<b>0.9006(25.77)</b>	<b>0.9739(13.12)</b>

<sup>1</sup> MCR-ALS constraints

<sup>2</sup> Correlation coefficients (r<sup>2</sup>) between MCR-ALS resolved and reference spectra of pure components.

<sup>3</sup> Angles between MCR-ALS resolved and reference spectra of pure components

18

## Application of hyperspectral imaging combined with chemometrics on food analysis

Table 3 Correlation coefficients ( $r^2$ , Equation 7) and angles (Equation 8) between the resolved spectra by MCR-ALS (full model) and the reference spectra in the analysis of the Infrared hyperspectral image of the white and milk chocolate samples

	Applied Constraints <sup>1</sup>		Sugar		Butter	Whey	lof (lack of fit) <sup>4</sup> %	R <sup>2,5</sup>	
			Sucrose	Lactose					
No pretreatment	White Chocolate	Non-negativity	$r^{2,2}$ (angle <sup>3</sup> )	0.9584(17.16)	0.9580(17.26)	0.9834(10.40)	0.8260(34.30)	4.41	99.81
		Non-negativity +equality constraints	$r^3$ (angle)	0.9820(10.90)	0.9811(11.51)	0.9834(10.44)	0.8260(34.29)	5.55	99.72
	Milk chocolate	Non-negativity	$r^2$ (angle)	0.9534(17.57)	0.9594(16.38)	0.9864(9.45)	0.9175(23.44)	4.45	99.80
		Non-negativity +equality constraints	$r^2$ (angle)	0.9703(13.99)	0.9733(13.27)	0.9868(9.32)	0.8548(31.26)	4.65	99.78
With pretreatment	White Chocolate	Non-negativity	$r^2$ (angle)	0.9572(16.83)	0.8038(36.51)	0.9820 (10.90)	0.8260(34.29)	8.61	99.06
		Non-negativity +equality constraints	$r^2$ (angle)	0.9998(0.1700)	0.9011(21.67)	0.9819(10.91)	0.8260(34.29)	10.46	98.90
	Milk chocolate	Non-negativity	$r^2$ (angle)	0.9692(14.27)	0.8051(36.37)	0.9820 (10.90)	0.8260(34.29)	5.64	99.70
		Non-negativity +equality constraints	$r^2$ (angle)	0.9803(10.99)	0.9699(14.66)	0.9819(10.91)	0.8260(34.31)	5.76	99.67

<sup>1</sup> MCR-ALS constraints

<sup>2</sup> Correlation coefficients ( $r^2$ ) between MCR-ALS resolved and reference spectra of pure components.

<sup>3</sup> Angles between MCR-ALS resolved and reference spectra of pure components

<sup>4</sup> MCR-ALS lack of fit (lof) considering the four components (full model) together.

<sup>5</sup> MCR-ALS percentage of explained variances (R<sup>2</sup>) considering the four components (full model) together.

## Discussion

In this chapter, the application of Raman and infrared hyperspectral imaging combined with MCR-ALS chemometric data analysis is shown in the analysis of chocolate constituents as a case of study. In contrast to the applications of hyperspectral remote sensing in previous chapter, the hyperspectral imaging analysis of food samples, like the chocolate, has the advantage of having easily available reference spectra of the sample constituents, obtained either by independent experiments in the laboratory or from libraries of reference spectra. On the other side, food samples can have different constituents with very similar chemical structure and spectra, which makes more challenging their resolution and identification.

Reviews of hyperspectral imaging applied to food analysis are given in references <sup>276, 284, 285</sup> and in the Chapter 2 of this Thesis. In these previous review papers, chemometric methods have been applied to the area of hyperspectral image analysis, mainly for regression (linear and non-linear) and classification (supervised or unsupervised) purposes. MCR-ALS is a powerful chemometrics method which was already applied for hyperspectral imaging resolution more than ten years ago. However, it is not a popular method in the food hyperspectral imaging analysis filed. The analysis of chocolate samples from this chapter is a good example of application of MCR-ALS for food analysis.

In this chapter, different pretreatment methods such as Multiplicative Scatter Correction (MSC), Asymmetric Least Squares (AsLS) and Savitzky-Golay data smoothing have been used. In the Chapter 2 of this Thesis, these methods were described. Hyperspectral imaging analysis can be performed at ambient conditions, without requiring special sample preparation. Chemical organic molecules can produce strong fluorescence signals when Raman is used. Infrared spectroscopic technique can be applied instead when Raman suffers from fluorescence and there is a high risk of sample heating from the laser. Spectral preprocessing data pretreatment techniques are a key step for the successful analysis of hyperspectral data sets.

In the case of chocolate analysis of this work, the use of data pretreatment methods such as MSC, AsLS and Savitzky-Golay smoothing methods, either separately or in combination, can facilitate the extraction of chemical information contained from the analyzed spectroscopic image data. Specifically, the undesired strong impact of fluorescence on Raman spectra can be removed significantly.

The results of this work show that using non-negativity and local rank constraints, MCR could resolve appropriately most of the chocolate constituents of white chocolate such as sucrose, lactose, butter and whey from Raman hyperspectral imaging data. Concentration

distribution maps of these constituents were displayed separately. Compared to Raman, infrared hyperspectral images did not suffer from the effect of fluorescence, and the application of the Kramers-Kronig transformation data pretreatment was enough in this case, to facilitate the extraction of the chemical information contained in the analyzed images. Constituents of milk chocolate could be well resolved by infrared hyperspectral imaging technique, in contrast to what was happening with Raman hyperspectral imaging technique.

Resolved spectra of sucrose, lactose, butter and whey were compared with reference spectra obtained experimentally. Correlation coefficients between reference and extracted spectra were satisfactory. Physical composition of chocolate constituents could be displayed adequately. Homogeneity of chocolate samples could be validated too (see in paper<sup>52</sup>).

Results obtained in this work show that hyperspectral imaging technique can be an efficient way for chocolate quality control, and that it also can be used for chocolate product research and development. The surface topography, fat crystallization are possible obtained by using resolved concentration distribution map of Raman hyperspectral image or the improved IR hyperspectral image mentioned in Chapter 4 section 4.1 (Bruker's new IR instruments have better pixel size resolution and faster data collection capability). These parameters have a relation with texture and color of the products<sup>286</sup>. Other hyperspectral imaging analysis of important ingredients (like polyphenols, cocoa butter and alternative fats) can be tried in the future for the chocolate product quality monitor.

In the paper<sup>52</sup> of this chapter, the micro-scope imaging instrument used for the measurements had pixel resolution in  $\mu\text{m}$  scale which provides more details of the constituent and reaction information in the food samples

Raman or infrared hyperspectral imaging techniques combined with MCR-ALS as proposed in this Thesis can be extended to homogeneity analysis of other type samples in pharmaceutical tablets<sup>287</sup>, powders, or suspensions, which is important for to pharmaceutical process monitoring and quality control<sup>288</sup>.



## **Chapter 5**

### **Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods**



## Introduction

Different bilinear methods have been widely applied for hyperspectral image analysis, apart from MCR-ALS, such as Independent Component Analysis (ICA)<sup>289</sup>, Principal Component Analysis (PCA)<sup>2</sup>, and Minimum Volume Simplex Analysis (MVSA)<sup>189</sup>. The theories of these methods have been introduced in Chapter 2. In this chapter (Paper <sup>267</sup>). These methods are compared using Mutual Information (MI), Amari Index (AI) and lack of fit (lof) parameters. This comparison can help to understand the performance of these bilinear models when different constraints are applied.

In Chapter 3 and Chapter 4, hyperspectral imaging resolution results have be shown to have associated a certain amount of rotation ambiguity <sup>246</sup> when only non-negativity constraints (apart from spectra normalization) were used. MCR-BANDS method was applied to evaluate the extension of rotation ambiguity in the results obtained by MCR-ALS. Apart from MCR-BANDS, methods that attempt to calculate the whole Area of Feasible Solutions (AFS) can also be used for evaluate the effect of rotation ambiguities, and different algorithm have been proposed<sup>227, 251, 255, 256, 290</sup>.

The gridding search method was proposed by Rajko and Abdollahi. A. Golshan, et al. have developed an AFS method using a simplex grid search of the area of feasible solutions which allows the visualization of rotational ambiguity in mixtures of up to four-components<sup>249, 257</sup>.

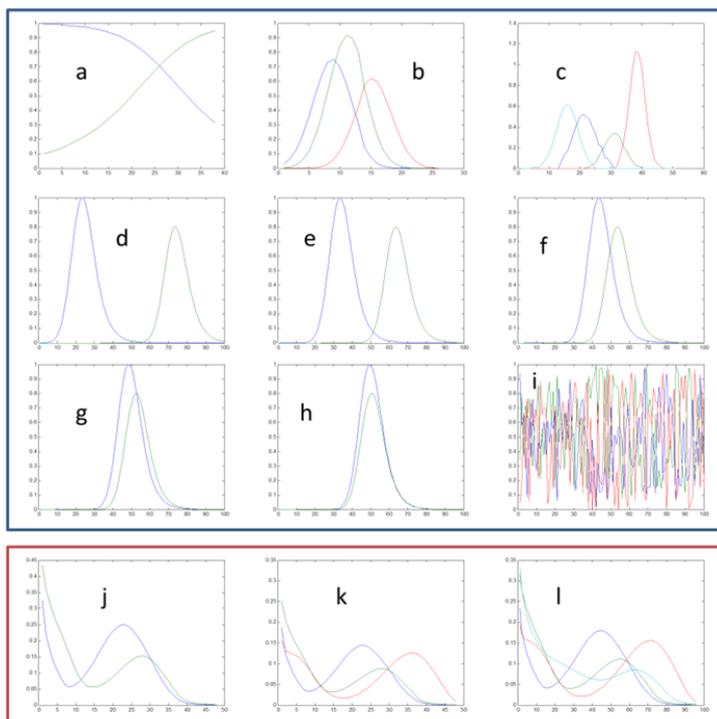
Rajkó and István<sup>256</sup> used computational geometry tools instead of the linear programming tools used to draw Borgen plots of three-component systems. They considered that the normalized feasible solutions are embraced in a (N-1)-dimensional simplex (N is the number of components) with the vertices being the N-normalized pure profiles.

R. Tauler developed the MCR-BANDS method <sup>259</sup> based on a previous idea of P.Gemperline <sup>104</sup>, for the calculation of the extension of rotation ambiguities, based on the fast maximization and minimization of a function defined by the relative Signal Component Contribution (SCCF) of each component <sup>246, 255</sup>.

Recently, Sawall et al. proposed algorithm to find AFS in two and three-component systems using the inflation of polygons as a searching method <sup>258</sup>. This procedure starts with an initial triangle located in a topologically connected subset of the AFS, and an automatic extrusion algorithm is then used to form a sequence of growing polygons that approximate the AFS from the interior. FAC-PACK is a MATLAB toolbox developed by Sawall et al. for the computation of non-negative multi-component factorizations and for the numerical approximation of the Area of Feasible Solutions (AFS) using the inflation

## Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

polygon algorithm<sup>258</sup>. In this chapter(Paper<sup>267</sup>), the results of MCR-BANDS and of FAC-PAACK method are compared.



**Figure 5-1.** Profiles used for Simulation of reaction and chromatographic data systems with two, three, and four components. (a-i) are the concentration profiles, and (j-l) are the pure spectra.

Reaction and chromatographic type of data systems with two, three, and four components were simulated for the purposed of this Chapter (Paper<sup>267</sup>). For the first data (Figure 5-1a) set, concentration profiles were obtained from a simple first order kinetic chemical reaction or process:  $X \rightarrow Y$ , with two components and spectra profiles (Figure 5-1j). For the second data set, the elution profiles (Figure 5-1b) and component spectra (Figure 5-1k) of three co-eluting components in a single HPLC-DAD chromatographic run are used. For the third data set, elution profiles (Figure 5-1c) from a signal HPLC-DAD run with four co-eluting components are used. Component spectra for them are given (Figure 5-1l).

Datasets 4-8 LC are elution profiles with different overlap (coelution) from two component systems. They have the same component spectra (Figure 5-1j), whereas the extent of chromatographic resolution decreased. The peak overlap increased makes the system more challenging for curve resolution (Figure 1 d, e, f, g, and h).

Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

---

In the last dataset, three random concentration profiles were considered for the data simulation (Figure 5-1i). Spectra of the components for this simulate dataset are given in Figure 5-1k.

**The paper in this chapter**

X. Zhang, R. Tauler. Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods. 2015.

## Measuring and comparing the resolution performance and the extend of rotation ambiguities of some bilinear modelling methods

Xin Zhang<sup>1</sup>

Roma Tauler\*<sup>1</sup>

<sup>1</sup> Jordi Girona 18, Barcelona, 08034, SPAIN

Corresponding author: Roma Tauler

Email: Roma.Tauler@idaea.csic.es

### Abstract:

Bilinear models are often used in the analysis of datasets from spectroscopy and chromatography. Whenever bilinear soft modelling approaches are applied, rotation ambiguities are ubiquitously present and they should be considered. In this work, results obtained by the application of different methods like Independent Component Analysis (ICA), Principal Component Analysis (PCA), and Minimum Volume Simplex Analysis (MVSA) are compared with those obtained by Multivariate Curve Resolution (MCR). In order to do this comparison, Mutual Information (MI), Amari index (AI) and lack of fit (lof) parameters are used for the evaluation of the different methods, and the corresponding areas or regions of feasible solutions (AFS) and their boundaries are investigated in each case. The results obtained by the MCR-BANDS method in the calculation of the extension of rotation ambiguities are discussed and compared with those obtained by the FAC-PACK method, which has been recently proposed for the estimation of the whole range of feasible solutions.

**Key words:** Rotation ambiguity, MCR, MCR-BANDS, AFS, bilinear, FAC-PACK

## 1 Introduction:

Chemometric methods provide powerful tools to analyze multi- and megavariable data from modern analytical instruments. Some of these chemometric methods, in particular Multivariate Curve Resolution (MCR) methods, have been proposed for the resolution of chemical data obtained from chromatography [1], spectroscopy [2], nuclear magnetic resonance[3], hyperspectral imaging[4], voltammetry[5], omics microarray[6] and LC-MS [7]data, among others [8, 9] etc. MCR methods are a group of methods based on the fulfillment of a bilinear model which attempt the extraction of the true underlying sources of chemical variation using a minimum amount of prior assumptions about the process under investigation. For the analysis of complex multi-component mixture systems, they offer the possibility of resolution, identification and also quantification [10] of the different components present in an unknown mixture, without needing their previous chemical and physical separation.

MCR chemometric methods have their intrinsic drawbacks, especially that they cannot assure encountering a unique solution to explain the measured experimental variation in the data, and that a range of feasible solutions may be obtained by their application. Ambiguities appear because different linear combination of the component profiles fulfilling the constraints of the system fit equally well the data[11]. Unfortunately, the presence of rotation ambiguities and of non-unique solutions decreases the reliability of MCR methods and makes their assessment more difficult. The only way to reduce the extent of rotation ambiguities and to obtain solutions closer to true ones is by the application of additional constraints (soft or hard) which imply using more knowledge about the data system, or also moving from bilinear modeling to multilinear modeling [12].

Bilinear modeling methods like Minimum Volume Simplex Analysis (MVSA), Independent Component Analysis (ICA), Principal Component Analysis (PCA) and Multivariate Curve Resolution-Function Minimization (MCR-FMIN) have already been compared in previously published papers [13, 14]. MVSA initially was developed for satellite imaging individual component (endmember) resolution, and more recently it has been also proposed in analytical chemistry[15]. PCA considers the information between the different components to be orthogonal or linearly uncorrelated[16]. ICA assumes that the components are mutually statistically independent[17]. These assumptions are statistically different (the latter is more restricted than the first), and therefore the results are different. In particular, ICA and PCA can be used for different purposes like data preprocessing, exploration, classification, regression and resolution. All these methods have been proposed for analytical chemistry purposes, and some authors have investigated whether one method is better than the other. Different from these approaches, based on statistical assumptions, Multivariate Curve Resolution methods, especially those based in Alternating Least Squares (MCR-ALS), use more natural and physically and chemically meaningful assumptions by

means of constraints, like non-negativity, unimodality, closure, selectivity, or local rank, and by means of other constraints related to the data structure (like trilinearity, or multilinearity) and find an optimum solution from a least squares fitting convergence criterium [18]. MCR-FMIN have been also proposed as a different way for multivariate curve resolution and it is based on non-linear optimization algorithms using non-linear constraints [19]. MCR-FMIN use PCA scores and loadings to define the subspace of MCR solutions and rotates them to fulfill the constraints of the system. Therefore it is also interesting to compare their solutions with those obtained by PCA and MCR-ALS.

In order to evaluate the effect of rotation ambiguities associated to a particular MCR solution and to measure its extent, Lawton and Sylvestre [20] already proposed a first algorithm for determining the area of feasible solutions (AFS) in two-component systems under the assumption of non-negative spectra and concentration profiles. Borgen et al. [21] extended Lawton and Sylvestre method to three component systems and proposed a linear programming optimization method to calculate the permitted ranges of pure component spectra using tangent and simplex rotation algorithms. Rajkó and István[22] revised Borgen's study and used computational geometry tools, to draw Borgen plots of three-component systems. Leger and Wentzell developed a dynamic Monte Carlo SMCR method[23] which seeks to define the boundaries of allowable pure component profiles. For the calculation of the whole range of feasible solutions, a systematic grid search method based on species-based Particle Swarm Optimization have been proposed for three-component systems by H. Abdollahi et al [24, 25]. A. Golshan et al. have also developed a method that finds the simplex volume containing all feasible solutions and facilitate the determination and visualization of rotational ambiguities of four-components mixture[25].

R. Tauler developed the MCR-BANDS method [26] based on a previous idea of P. Gemperline [27], for the calculation of the extension of rotation ambiguities, based on the fast maximization and minimization of a function defined by the relative Signal Component Contribution (SCCF) of each component [11, 23]. This method has no limitation for the number of components and it uses the same constraints as those applied to find out the MCR solution. It gives a simple evaluation of the extent of rotation ambiguity from the difference between the maximum and minimum values of the SCCF function. Recently, Sawall et al. suggested a fast accurate algorithm to find the AFS for two and three-component systems based on the use of a polygon inflation algorithms.[28]. FAC-PACK, is an interactive MATLAB toolbox for the computation of non-negative multi-component factorizations and for the numerical approximation of the area of feasible solutions using the inflation polygon algorithm[28].

In this work, FAC-PACK results are compared to those obtained by MCR-BANDS, and with the solutions obtained by different bilinear model methods such as PCA, ICA, MVSA, MCR-FMIN

and MCR-ALS. The aim of this work is to get a deeper understanding of MCR methods and evaluate their performance under different constraints. In addition, the extension of rotation ambiguities associated to MCR solutions is investigated by the MCR-BANDS and FAC-PACK methods. The comparison of results obtained by these two methods can help to evaluate the reliability of their results and to get a deeper understanding of their principles.

## 2 Theory

The second-order bilinear data generally can be decomposed by bilinear model based methods according to Equation (1).

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} = \mathbf{D}^* + \mathbf{E} \quad \text{Equation 1}$$

where  $\mathbf{D}$  (I,J) is the experimental data matrix corresponding to a bilinear system with I different samples and J different variables.,  $\mathbf{C}$  (I,N) is the contributions of the N components in each sample,  $\mathbf{S}$  (J,N) is the pure response matrix of the N components,  $\mathbf{E}$  (I,J) is the matrix associated to noise or experimental error. Giving the data matrix  $\mathbf{D}$ , the aim of bilinear model is to determine the two factor matrices  $\mathbf{C}$  and  $\mathbf{S}$ .

The concept of rotation ambiguities is explained using the following reasoning. For any non-singular matrix  $\mathbf{T}$  (N,N) the identity matrix  $\mathbf{I} = \mathbf{T}^{-1}\mathbf{T}$  can be inserted into Equation 1 as the following equations:

$$\mathbf{D}^* = \mathbf{C}\mathbf{T}^{-1}\mathbf{T}\mathbf{S}^T = \mathbf{C}_{\text{new}}\mathbf{S}_{\text{new}}^T \quad \text{Equation 2}$$

where

$$\mathbf{C}_{\text{new}} = \mathbf{C}\mathbf{T}^{-1} \text{ and } \mathbf{S}_{\text{new}}^T = \mathbf{T}\mathbf{S}^T \quad \text{Equation 3}$$

According to Equation 3, any rotation of factor matrices,  $\mathbf{C}$  and  $\mathbf{S}^T$ , using a non-singular  $\mathbf{T}$  matrix, will produce a new valid solution of the bilinear model. Therefore, in absence of enough constraint, an infinite number of rotations and solutions are possible.

### Principal Component Analysis (PCA)

PCA provides a mathematical and very efficient way to solve the bilinear model and perform the matrix decomposition given in Equation 1. PCA decomposes the measurement matrix  $\mathbf{D}$  into the scores  $\mathbf{C}_{\text{PCA}}$  and loadings  $\mathbf{S}_{\text{PCA}}$  orthogonal factor matrices, and a reduced number of components are selected which explain maximum data variance. The aim of the method is to maximize the explained variance in the data with a minimum number of components. Due to the applied constraints during the PCA bilinear decomposition (orthogonality of scores and loadings, normalization of loadings and maximum variance),

score and loading profiles do not resemble in general the true variance sources, but a linear combination of them fulfilling the applied constraints. True variance source profiles do not fulfill for instance the requirement of orthogonality and they are overlapped, and in many chemical and physical systems profiles should be non-negative.

#### **Independent Component Analysis (ICA)**

The aim of ICA is the decomposition of the measured multivariate signals into statistically independent component contributions with a minimum loss of information. ICA assumes that the mixing vectors in  $\mathbf{C}$  are linearly independent and that the components in  $\mathbf{S}$  are mutually statistically independent, as well as independent of noise components. This goal is equivalent to finding an unmixing matrix  $\mathbf{W}$  that satisfies:

$$\mathbf{WX} = \widehat{\mathbf{S}}^T \quad \text{Equation 4}$$

where  $\widehat{\mathbf{S}}$  is the estimation of the  $\mathbf{S}$ . The main task of ICA is to find out the unmixing matrix  $\mathbf{W}$  based on the principle that the output  $\widehat{\mathbf{S}}^T$  as independent as possible. Thus, this task turns into an optimization problem under the constraints of independency, which is generally reflected by non-gaussian profiles. MF-ICA algorithm [29] applied in this work apply non-negativity constraints to the signals.

#### **Minimum Volume Simplex Analysis (MVSA) method:**

MVSA also considers that the underlying mixing model is bilinear, i.e. that the measured spectral vectors are a linear combination of signatures (spectra) of pure components. MVSA is a method that finds the pure components (end members) by fitting the data to a minimum volume simplex, under some constraints, such as having for every pixel no less than zero abundance fractions (non-negativity constraint) and that their sum should be equal to one (closure). The MVSA method starts with an estimate of the purest spectra profiles, obtained by the Vertex Component Analysis (VCA)[30] method, which is a pure variable detection method based in an iterative algorithm. MVSA does not use a least squares approach, but a sequential quadratic programming (SQP) method, based on a quasi-Newton non-linear optimization procedure under linear constraints[31]. MVSA method provides estimations of the pure spectra  $\mathbf{S}$  of the system. Concentration profiles  $\mathbf{C}$  should be calculated by least squares subsequently.

#### **Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS)**

MCR-ALS solves Equation 1 iteratively using an ALS algorithm, which optimally fits the experimental data matrix  $\mathbf{D}$ , and resolves the 'true' pure response profiles, in concentration

$\mathbf{C}$  and pure spectra  $\mathbf{S}^T$  matrices. This optimization is carried out for a proposed number of components using initial estimates of either  $\mathbf{C}$  or  $\mathbf{S}^T$ . These initial estimates of  $\mathbf{C}$  or  $\mathbf{S}^T$  can be extracted using procedures for purest variables selection, such as SIMPLISMA[32]. During the ALS optimization, several constraints can be applied to model the shapes of  $\mathbf{C}$  or  $\mathbf{S}^T$  profiles such as non-negativity, unimodality, normalization and selectivity (local rank). Convergence is achieved when in two consecutive iterative cycles, relative differences in standard deviations of the residuals between the experimental and calculated ALS data are less than a previously selected threshold value. The use of constraints or of any other previously known property about the nature of the component profiles can decrease ambiguity significantly and in some case eliminate it totally. More details about MCR-ALS can be found in [27, 33].

#### **Multivariate Curve Resolution-objective Function Minimization (MCR-FMIN)**

MCR-FMIN is based on the minimization of an objective function defined directly from the non-fulfilment of constraints and being always in the subspace spanned by PCA solutions. In other words, an appropriate rotation of the PCA solutions is performed searching for physically meaningful solutions (like non-negative). The function can be defined as:

$$f(\mathbf{T}) = c_{\text{norm}}(\mathbf{T}) + c_{\text{non-neg}}(\mathbf{T}) + c_{\text{unimod}}(\mathbf{T}) + c_{\text{clos}}(\mathbf{T}) + c_{\text{equa}}(\mathbf{T}) + \dots \quad \text{Equation 5}$$

where  $f(\mathbf{T})$  is the objective scalar function to be minimized and  $c_{\text{norm}}(\mathbf{T})$ ,  $c_{\text{non-neg}}(\mathbf{T})$ ,  $c_{\text{unimod}}(\mathbf{T})$ ,  $c_{\text{clos}}(\mathbf{T})$ ,  $c_{\text{equa}}(\mathbf{T})$  are scalar functions for the normalization, non-negativity, unimodality, closure, and equality constraints. In this work only non-negativity and normalization constraints were used. The goal of the optimization is then to find a  $\mathbf{T}$  rotation matrix that minimizes  $f(\mathbf{T})$ . Function  $f(\mathbf{T})$  can be minimized using different optimization methods. In this work, quasi-Newton non-linear optimization method with a cubic interpolation line search was used. See reference [19]for more details.

#### **MCR-BANDS: Calculation of the extend of rotation ambiguities**

In order to evaluate the extent of rotation ambiguities associated to a particular MCR solution under a set of constraints, the MCR-BANDS procedure has been proposed[11]. In this paper it is applied to evaluate the extent of rotation ambiguities associated to MCR-ALS solutions, and as extension, also to those obtained by MCR-FMIN, PCA, ICA, and MVSA methods previously explained. In addition, MCR-BAND results are compared to those obtained by the FAC-PACK method (see below). MCR-BANDS method is based on the maximization and minimization of the signal component contribution function, SCCF, which

is defined for component n as:

$$SCCF_n = \frac{\|c_n s_n^T\|}{\|CS^T\|} \quad \text{Equation 6}$$

$SCCF_n$  is a scalar value between 0 and 1, which gives the relative signal contribution of a certain component in relation to the whole signal due to the mixture of N components ( $n=1\dots N$ ).  $C$  and  $S^T$  are as defined above;  $c_n$  and  $s_n$  are the  $n^{\text{th}}$  column and row of  $C$  and  $S$  matrices respectively.  $\| \cdot \|$  is the Frobenious norm  $\|CS^T\|$  gives the signal contribution of all the components present in the whole image and  $\|c_n s_n^T\|$  only the contribution of component n[26]. For every component, SCCF is maximized and minimized under the set of considered constraints and with equal data fitting. When the maximum and minimum values of SCCF of a particular component n are practically equal, it means that for this component, there is practically no ambiguity remaining. On the contrary when this difference is large and close to one (SCCF values are scaled between 0 and 1), then there is a large amount of ambiguity. See references[11] for more details about how this procedure is implemented and works. MCR-BANDS performs the minimization and maximization of SCCF values using a general sequential quadratic programming procedure to solve the non-linear optimization of SCCF under non-linear constraints [11, 26, 34].

#### **FAC-PACK: Evaluation of the Area of Feasible Solutions (AFS)**

The area of feasible solutions (AFS) is a subset of the two dimensional plane consisting of all pairs of solutions which represent non-negative spectra (case of the spectral AFS) or non-negative concentration profiles (case of the concentration AFS). In FAC-PACK toolbox for MATLAB, AFS is computed using a polygon inflation algorithm and its more recent implementation, the inverse polygon inflation algorithm. The polygon inflation method approximates the border of each AFS segment by a sequence of growing polygons from the interior of each AFS. An adaptive strategy for the edge selection keeps the computational costs down and guarantees a controlled quality of the boundary approximation. AFS gives different separated subsets obtained for the number of components of the system. Each subset is a segment in the visible figure. The idea of the polygon inflation algorithm[35] is to approximate the boundary of each segment of the AFS by a sequence of adaptively refined polygons. Starting with an initial triangle whose vertices are located on the boundary of an AFS segment, all the edges are subdivided and the new vertices are moved to the boundary of the AFS. Up to now it has only be shown for systems that have equal or less than three components.

### Mutual Information (MI) evaluation

To estimate the degree of independence between component profiles, Mutual Information (MI) values are used. MI values are defined as proposed in previous works as a natural measure of the mutual independence between two variables[36]. MI between two variables can be expressed using the Joint Probability density function  $p(x_1, x_2)$  and the Marginal probability density function  $p(x_1)$  and  $p(x_2)$ .

$$I(x_1, x_2) = \int dx_1 \int dx_2 p(x_1, x_2) \log \left[ \frac{p(x_1, x_2)}{p(x_1)p(x_2)} \right] \quad \text{Equation 7}$$

Krasakov et al. have proposed an efficient method or estimating MI values. More details about this algorithm are available[37].

### Concentration profiles recovery evaluation using Amari index

To estimate the reliability of the results obtained by the different methods applied here, concentration profiles were compared to the correct ones (if available) using the Amari index[13]. It shows the reliability between the compared concentration profiles, and it is defined as follows:

$$P = \frac{1}{2N} \sum_{i,j=1}^N \left( \frac{|p_{ij}|}{\max_k |p_{ik}|} + \frac{|p_{ij}|}{\max_k |p_{kj}|} \right) - 1 \quad \text{Equation 8}$$

where  $p_{ij} = (\hat{\mathbf{C}}^+ \mathbf{C})_{i,j}$ ,  $\mathbf{C}$  are the true concentration profiles and  $\hat{\mathbf{C}}^+$  are the pseudoinverse of calculated ones using the considered method. The Amari index is equal to zero when the true and estimated concentration profiles differ only in scaling or permutation of the components (in presence of only permutation and intensity ambiguities). Amari index values become lager when resolved profiles differ from the true ones and rotation ambiguities are present. Low Amari index values are desirable.

### Data fitting: calculation of lack of fit values

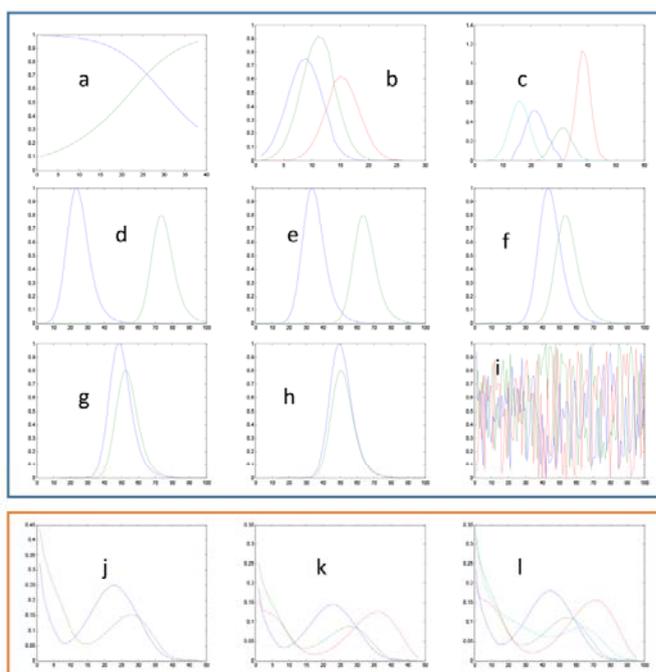
Lack of fit values are defined form the difference between input data  $\mathbf{D}$  values and their reproduced values obtained using a particular method. To evaluate the quality of the data fitting finally achieved after application of the bilinear models, the percentages of lack of fit (lof) are calculated according to the following equations:

$$\text{lof}(\%) = 100 \times \sqrt{\frac{\sum_{ij} (d_{ij} - \hat{d}_{ij})^2}{\sum_{ij} d_{ij}^2}} \quad \text{Equation 9}$$

where  $d_{ij}$  are the elements of the data matrix  $\mathbf{D}$ , and  $\hat{d}_{ij}$  are the corresponding elements recalculated by the considered method,  $\hat{\mathbf{D}}$ . Lack of fit value give a measure of the fit quality in relative terms with the same units as the measured data, which can be compared with experimental relative error estimations if known.

### 3 Datasets

In this work similar synthetic datasets to those described in previous work [13] are used. Both reaction and chromatographic type of profiles of two, three, and four components were used for the purposed of this work. The different sets of concentration and spectra profiles used as examples are given in Figure 1.



**Figure 1.** Profiles used for the simulation of the different reaction and chromatographic data systems investigated in this work, with two, three, and four components. a, b, c, d, e, f, g and i are the concentration profiles, and j, k and l are the pure spectra.

Pure spectra were normalized to unit area and relative concentration profiles were always scaled between 0 and 1 in all the cases. In this way results obtained by the different methods

were more comparable. For Data set 1 (Figure 1a) concentration profiles were obtained from a simple first order kinetic chemical reaction,  $X \rightarrow Y$ , with two components and with the spectra profiles 1 and 2 (Figure 1j).

Dataset 2 was obtained using elution profiles from a three component HPLC-DAD coeluting system (Figure 1b) and their corresponding pure spectra (Figure 1k)

For dataset 3, elution profiles (Figure 1c) from a single HPLC-DAD run with four co-eluting components were used. Component spectra for them are given in Figure 1l.

Datasets 4-8 were describe LC elution profiles of a two component system with different overlapping levels (Figure 1 d, e, f, g, h) and same component spectra (Figure 1 j).

In dataset 9, three random concentration profiles were considered (Figure 1i) with pure spectra given in Figure 1k.

### Software

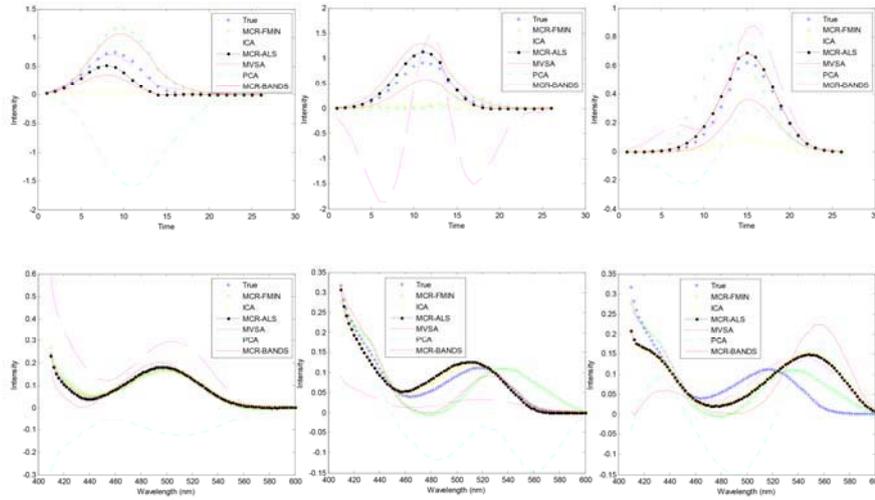
All calculations were performed in MATLAB R2013a (Mathworks Inc., Natick, MA, USA) for windows.

MF-ICA[29] methods were obtained from the ICA MATLAB toolbox V3 (ICA Toolbox Homepage (<http://isp.imm.dtu.dk/toolbox/ica>)). MVSA MATLAB codes were downloaded from: <http://www.lx.it.pt/~bioucas/code.htm>. MCR-ALS and MCR-BANDS algorithm code and GUI for MATLAB is freely available from the home page of MCR at <http://www.mcrals.info/>. FAC-PACK GUI for MATLAB is freely available from webpage <http://www.math.uni-rostock.de/FAC-PACK/>.

## 4 Results and discussion

Figure 2 shows concentration and spectral profiles resolved by the different bilinear methods applied in this work (MCR-ALS, ICA, MCR-FMIN, PCA, MVSA and SCCF  $f_{\min}$  and  $f_{\max}$  values of MCR-BANDS). All profiles obtained by MCR-ALS, ICA and SCCF  $f_{\min}$  and  $f_{\max}$  values of MCR-BANDS fulfilled non-negativity constraints. PCA used orthogonality constraints instead of non-negative constraints, and therefore negative values were encountered in both concentration and spectral profiles. MVSA method provided non-negative pure spectra too. However concentration profiles obtained by least squares from them had also negative values. MCR-FMIN gave small negative values, very near to zero (not visible in the Figure).

Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods



**Figure 2.** Concentration and spectral profiles resolved by the different bilinear methods applied in this work (MCR-ALS, ICA, MCR-FMIN, PCA, MVSA and profile corresponding to  $f_{\min}$  and  $f_{\max}$  of MCR-BANDS) in the resolution of data set 2 (concentration profiles b and spectra profiles k in Figure 1). Figure (1), (2) and (3) correspond to the first, second and third resolved component profiles: True ('\*', blue), MCR-FMIN ('+', green), ICA ('Δ', yellow), MCR-ALS ('.', black), MVSA ('— —', purple), PCA ('- -', cyan), MCR-BANDS ('—', red).

Table 1 shows MI values of all spectra profiles, as well as the Amari index (AI) for all concentration profiles and the lack of fit values obtained for the different tested methods (MCR-ALS, ICA, PCA, MCR-FMIN and MVSA) in the analysis of all datasets previously described.

Lack of fit values of the different tested methods were rather similar in general, except in the analysis of the more complex systems. Solutions obtained by the different methods were rather equivalent from a mathematical point of view and they were feasible, differing among them because of the presence of unresolved rotation ambiguities for the considered constraints.

PCA and MCR-FMIN solutions gave similar lack of fit values, because MCR-FMIN solutions are obtained by rotation of initial PCA loadings, during the optimization. PCA and MCR-FMIN gave the lowest lack of fit values, because of the tendency of these two methods to slightly over fit the raw data (some noise is embedded in the PCA solutions). This is in agreement with the general over fitting tendency of most bilinear model-based methods, which may incorporate some noise in the resolved parameters.

When the number of components is increased (from data sets 1 to 3), the values of MI for all

resolved spectra profiles were increasing. This means that the components become less independent. The lower the MI values are, the higher independence between the resolved profiles is present. MI values obtained for MF-ICA-resolved spectra should be lower (more independent) than those from obtained by the other methods. But this was not always the case and they were higher than those from FMIN, from PCA loadings and from true spectra profiles. This should be related again to the application of non-negativity constraints in MF-ICA compared to orthogonality constraints in PCA.

MCR-ALS and MCR-FMIN gave low values of the Amari index compared to the same values for ICA in some cases, which means that they recovered better the concentration profiles than MF-ICA. Results of Table 1 also show that, although the values of MI for PCA were lower than those for ICA, in many cases (bold numbers in Table 1), the Amari indices for PCA were always much worse than those for ICA and MCR, meaning that the profiles recovered by PCA were always more different to the true ones. This is obviously due to orthogonality constraints in PCA which produced negative values in the finally recovered profiles, which obviously were not present in true profiles and therefore produced worse Amari index values.

From datasets 4 to 8, overlapping of concentration profiles increased. Values of MI for resolved spectra profiles and of the Amari index for elution profiles of datasets 4-8 increased significantly when coelution also increased (less independent spectra profiles were obtained and worse recovery of the elution profiles was achieved). This means that, when concentration profiles were strongly overlapped, as in these cases, the only use of non-negativity constraints does not guarantee the correct recovery of the true profiles, because of the presence of rotational ambiguities. Interestingly MI values for ICA resulted to be lower than those from MCR. This is probably due to the MF-ICA non-negativity constraints implementation on both concentration and spectral profiles, making MF-ICA less efficient to achieve the independence condition of the sought profiles. Increasing overlapping of elution profiles affected the independence (measured by MI values) and the quality (measured by Amari Index values) of the resolved spectra profiles. Amari indices for PCA were clearly worse than those for ICA and MCR.

MI and of Amari index values show that independence and orthogonality constraints are not adequate for the appropriate resolution of pure components profiles in general in the MCR context. Highly independent or orthogonal profiles make them to differ as much as possible. However in the real situation, this condition is not fulfilled in general, profiles are neither highly independent nor orthogonal, and they may be strongly overlapped. ICA and PCA obtain good mathematical models with unique solutions, but these solutions may be rather different to the true ones.

### Evaluation of the extension of rotation ambiguities in MCR and other methods

MCR solutions have in general a certain degree of ambiguity, and this can be evaluated by methods like MCR-BANDS. MCR-BANDS is based on the optimization of the function SCCF defined by the relative signal contribution of every component in relation to all the other components in the mixture (see Equation 2). In this work, relative signal contributions of all the components measured by SCCF using the different investigated methods such as MCR-ALS, MVSA, PCA, ICA and MCR-FMIN are evaluated and compared.

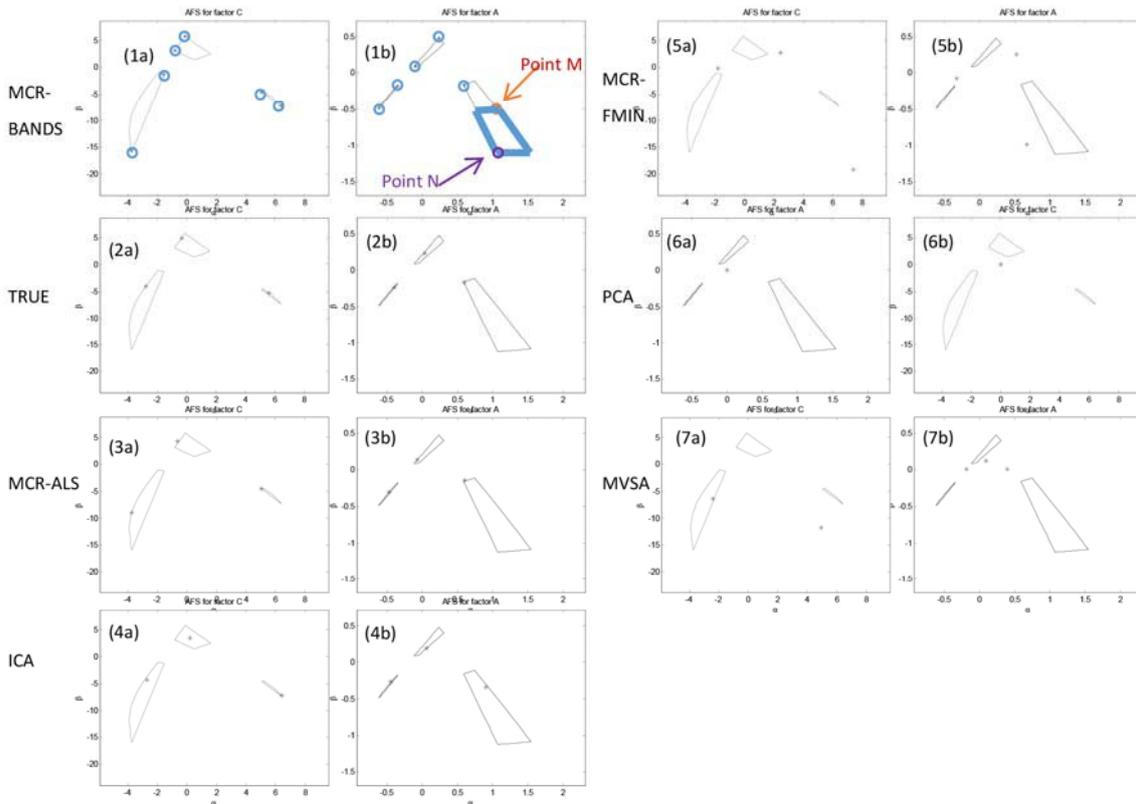
Dataset 2 (a three component system, see dataset section) was selected to study in more detail the rotation ambiguity problem. Table 2 gives SCCF values for MCR-ALS, PCA, MVSA, MCR-FMIN, ICA, and SCCF,  $f_{\min}$  and  $f_{\max}$  solutions, calculated by MCR-BANDS under non-negativity constraints, for this three component system.

SCCF  $f_{\text{MCR-ALS}}$ ,  $f_{\text{ICA}}$ , and  $f_{\text{TRUE}}$  values were all located within MCR-BANDS SCCF  $f_{\min}$  and  $f_{\max}$  range, and all of them provided non-negative profiles and fitted the data equally well. In contrast, SCCF  $f_{\text{PCA}}$  values, and some of the SCCF  $f_{\text{MVSA}}$ ,  $f_{\text{MCR-MIN}}$  values, resulted to be outside of the range between SCCF  $f_{\min}$  and  $f_{\max}$ . PCA profiles were as expected to be orthogonal and they had consequently negative values, and therefore they were out of the MCR-BANDS (non-negative) feasible range. MVSA method provided only solutions for the pure spectra but not for concentration profiles, which should be estimated by an additional least squares, and they were then always giving negative values, meaning that they did not fulfill the non-negativity requirement for the concentration profiles. As a consequence, they were also out of the MCR-BANDS range. MCR-FMIN had the problem of embedded noise from initial profiles estimated by PCA, which prevented having feasible non-negative solutions. Although these MCR-FMIN results were near to the true profiles and had low MI and low Amari indexes, their SCCF  $f_{\text{MCR-MIN}}$  values were out of range of SCCF  $f_{\min}$  and  $f_{\max}$ .

In this work, the FAC-PACK method was used to calculate and display geometrically the full range of all feasible solutions, i.e. the area of feasible solutions, called AFS, using the polygon inflation algorithm. This method provides similar results to the Borgen plot method proposed by Rajko [22] and to the gridding search method proposed by Rajko and Abdollahi [38], but its application is simpler and it is currently available (as open source application) for two- and three- component systems, under non-negative constraints. In the case of the three components system studied in this work, the AFS obtained by FAC-PACK by means of the polygon inflation algorithm are shown in Figure 3. AFS are displayed as polygons for both concentration (Figure 3 (1a)-(7a)) and spectral solutions (Figure 3 (1b)-(7b)). The FAC-PACK calculated set of polygons are composed of three isolated and topologically connected subsets corresponding to the three components of the analyzed system. Marked points in the Figure were projected for the profiles previously obtained by the different bilinear methods tested in this work (MCR-ALS, ICA, MCR-FMIN, PCA, MVSA, and SCCF  $f_{\min}$  and  $f_{\max}$

Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

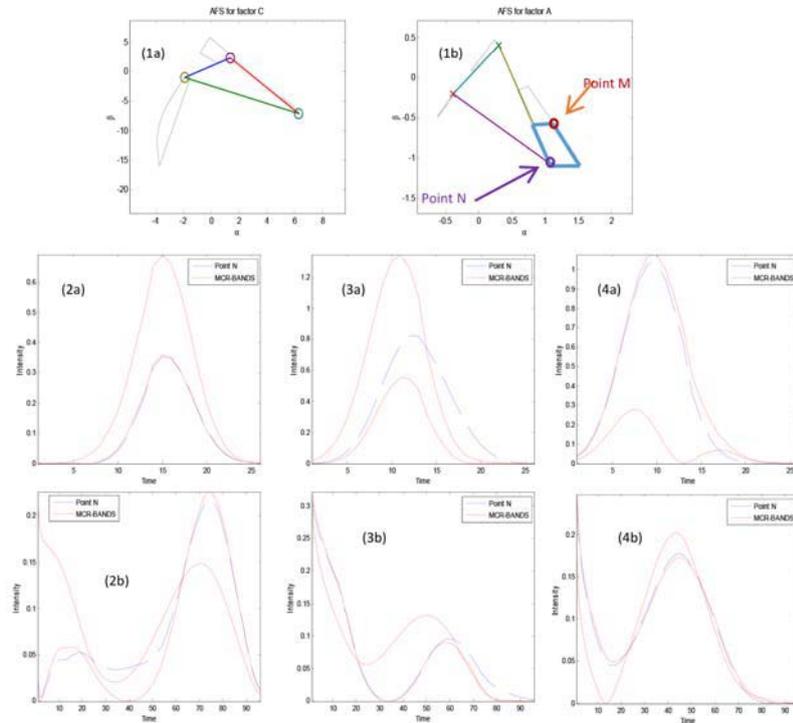
defined profiles from MCR-BANDS). As it is shown in Figure 3, MCR-BANDS profiles were always located in the border of FAC-PACK polygons, and the location of MCR-ALS, ICA and MCR-BANDS SCCF  $f_{\min}$  and  $f_{\max}$  solutions are all located inside the AFS calculated by FAC-PACK, whereas the location of some of the solutions provided by FMIN, PCA, MVSA solutions are located outside the AFS. Interestingly, SCCF  $f_{\min}$  and  $f_{\max}$  of MCR-BANDS solutions projected on this figure were usually located on one of the corners/edges of the borders (marked in blue circles) of the AFS except for one of the profiles (see point M in Figure 3) marked by a red circle and a red arrow in Figure 3.



**Figure 3.** AFS obtained by the FAC-PACK method in the analysis of data set 2. Polygons in each plot are projected area of the feasible solutions for each component calculated by FAC-PACK. Marked points (using '\*') in the figure were obtained from the projection of the profiles resolved by the different bilinear methods applied in this work (MCR-ALS, ICA, MCR-FMIN, PCA, MVSA and profile corresponding to  $f_{\min}$  and  $f_{\max}$  of MCR-BANDS). The area in blue tetragon means the area of polygon from FAC-PACK which is not between the points projected by profiles corresponding to  $f_{\min}$  and  $f_{\max}$  of MCR-BANDS. Point M (marked by a red circle and a red arrow) is the only point of SCCF  $f_{\min}$  and  $f_{\max}$  of MCR-BANDS solution projected which not located on a corner/edge of the borders of the AFS in this Figure. The other related points projected by SCCF  $f_{\min}$  and  $f_{\max}$  of MCR-BANDS solutions which located on corners/edge were marked in blue circles). Point N is the representative point in the lower corner of the polygon (marked by a purple circle) in the region located away from the point M.

Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

To further study the relationship existing between MCR-BANDS solutions and FAC-PACK feasible solutions AFS, SCCF values were evaluated in the region located away from the point M in Figure 3 (shown as a blue rectangle). For instance, the feasible solution located at point N (away from M and in the lower corner of the polygon marked by the purple circle in Figure 3-(1b)) gave spectra and concentration profiles shown as red lines in Figures 4-(2) (3) and (4). Figure 4-(1) shows the locations of these projected profiles of the three components defined by point N using the FAC-PACK method. Locations corresponding to concentration profiles are marked as 'o' and locations corresponding to spectra profiles are marked as 'x'. Figures 4 (2), (3) and (4) shows the corresponding concentration and spectral profiles of solutions at point N together with concentration and spectra profiles calculated for  $f_{\min}$  and  $f_{\max}$  SCCF values of MCR-BANDS.



**Figure 4.** Locations projected from profiles marked by point N on the polygon of dataset 2 (Locations corresponding to concentration were marked as 'o' in (1a) and spectra profiles were marked as 'x' in (1b)). Figure (2), (3) and (4) are comparison between profiles obtained for  $f_{\min}$  and  $f_{\max}$  solutions of MCR-BANDS and profiles of the three components in the analysis of data set 2 at the location marked by point N

Table 3 gives SCCF values,  $f$ , for MCR-BANDS maximum and minimum solutions ( $f_{\min}$  and  $f_{\max}$ ), as well as for the 'TRUE' and MCR-ALS solutions with non-negativity constraints, and, more

interestingly, for the feasible solutions located at point N previously described (Figure 4). It should be remarked that MCR-BANDS SCCF  $f_{\min}$  and  $f_{\max}$  values calculated either using TRUE or MCR-ALS initials with non-negativity constraints were very similar to those obtained using initials from the point located at N in Figure 4. SCCF  $f_{\text{point\_N}}$  values for the three components were all located between the range of  $f_{\min}$  and  $f_{\max}$  obtained by MCR-BANDS. This is an important result which confirms that SCCF values of the profiles in the quadrilateral blue area of Figure 3, away from point M, are also between the ranges of MCR-BANDS results. Therefore, results from MCR-BANDS and from FAC-PACK are confirmed to be in agreement, in spite of being based incompletely different approaches.

As previously shown in paper of Hamid and Tauler[24], for a two-way data system, it is possible to define AFS graphical boundaries in two dimensions to illustrate the whole set of feasible solutions. However, for higher dimensional systems (three and more components), it is not possible to illustrate graphically the whole set of feasible solutions using only the two profiles defined by SCCF  $f_{\min}$  and  $f_{\max}$  MCR-BANDS values. Since feasible solutions projections change from a line to a plane, a volume or even a more complex graphical display is needed to delimit all AFS plots. MCR-BANDS is not intended to be a method to providing visually the boundaries of the AFS, but to give an estimation of the extent of rotation ambiguity associated to a particular MCR solution. Further work is perused to confirm general ways to assess the reliability of MCR solutions, and in particular of the extent rotation ambiguities associated to them.

## Conclusions

Bilinear modelling based methods, like MCR-ALS, MCR-FMIN, ICA, PCA, or MVSA (all of them tested in this work) are appropriate methods for multivariate curve resolution. But not all of them perform in the same way. When they are applied to two-way data and appropriate local rank/selectivity conditions are not present in the data, there is no way in general to know if they provide the true solution when they are applied under similar constraints. Different resolution methods will give one of the equivalent feasible solutions for the problem under study.

Results obtained from all these methods show that, a range of feasible solutions describe and fit the data equally well, while fulfilling the bilinear model and constraints. It is not possible therefore to say that one method is better than another one. All of them might be considered as an alternative tool for the resolution of mixed spectroscopic pure signals (signatures) on certain cases. However, implementation of physical constraints, such as non-negativity and others in resolved spectra and concentration profiles is more flexible, powerful and reliable in MCR methods like MCR-ALS than in other.

The MCR-BANDS method gives an estimation of the extension of rotation ambiguities and therefore of the ranges of feasible solutions. Whereas methods like FAC-PACK provide a geometrical display of all feasible solutions. Interestingly, the extension of rotation ambiguities estimated by the MCR-BANDS method and the AFS calculated by a method like FAC-PACK are in agreement and concordant.

**Acknowledgements:**

Authors acknowledge research grant CTQ2012-3816 from Ministerio de Economía y competitividad, Spain. Xin Zhang thanks the grant from China Scholarship Committee (2011637007).

**Reference:**

- [1] V. Boeris, J.A. Arancibia, A.C. Olivieri, Determination of five pesticides in juice, fruit and vegetable samples by means of liquid chromatography combined with multivariate curve resolution, *Anal. Chim. Acta*, 814 (2014) 23-30.
- [2] M. Garrido, F. Rius, M. Larrechi, Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes, *Anal. Bioanal. Chem.*, 390 (2008) 2059-2066.
- [3] W. Windig, B. Antalek, M.J. Robbins, N. Zumbulyadis, C.E. Heckler, Applications of the direct exponential curve resolution algorithm (DECRA) to solid state nuclear magnetic resonance and mid - infrared spectra, *J. Chemom.*, 14 (2000) 213-227.
- [4] D.M. Haaland, J.A. Timlin, M.B. Sinclair, M.H. Van Benthem, M.J. Martinez, A.D. Aragon, M. Werner-Washburne, Multivariate curve resolution for hyperspectral image analysis: applications to microarray technology, in: *Biomedical Optics 2003, International Society for Optics and Photonics*, 2003, pp. 55-66.
- [5] M. Esteban, C. Arino, J. Diaz-Cruz, M. Diaz-Cruz, R. Tauler, Multivariate curve resolution with alternating least squares optimisation: a soft-modelling approach to metal complexation studies by voltammetric techniques, *TrAC, Trends Anal. Chem.*, 19 (2000) 49-61.
- [6] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne, Multivariate curve resolution of time course microarray data, *BMC Bioinformatics*, 7 (2006) 343.
- [7] I.S.n. Pérez, M.a.J. Culzoni, G.G. Siano, M.a.D.G. García, H.c.C. Goicoechea, M.a.M.n. Galera, Detection of unintended stress effects based on a metabonomic study in tomato fruits after treatment with carbofuran pesticide. Capabilities of MCR-ALS applied to LC-MS three-way data arrays, *Anal. Chem.*, 81 (2009) 8335-8346.
- [8] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: A review of advanced and tailored applications and challenges, *Anal. Chim. Acta*, 765 (2013) 28-36.
- [9] X. Zhang, R. Marcé, J. Armengol, R. Tauler, Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution, *Chemosphere*, 111 (2014) 120-128.
- [10] T. Azzouz, R. Tauler, Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples, *Talanta*, 74 (2008) 1201-1210.
- [11] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, *Chemometrics and Intelligent Laboratory Systems*, 103 (2010) 96-107.
- [12] A. Malik, R. Tauler, Performance and validation of MCR-ALS with quadrilinear constraint in the analysis of noisy datasets, *Chemometrics Intellig. Lab. Syst.*, 135 (2014) 223-234.
- [13] H. Parastar, M. Jalali-Heravi, R. Tauler, Is independent component analysis appropriate for multivariate resolution in analytical chemistry?, *TrAC, Trends Anal. Chem.*, 31 (2012) 134-143.
- [14] H. Abdollahi, R. Tauler, Uniqueness and rotation ambiguities in multivariate curve resolution methods, *Chemometrics Intellig. Lab. Syst.*, 108 (2011) 100-111.
- [15] M.B. Lopes, J.-C. Wolff, J.M. Bioucas-Dias, M.A. Figueiredo, Investigating counterfeit medicines—the near infrared chemical imaging picture, *NIR news*, 22 (2011) 10-18.
- [16] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics Intellig. Lab. Syst.*, 2 (1987) 37-52.
- [17] H. Oja, K. Nordhausen, Independent component analysis, *Encyclopedia of Environmetrics*, (2001).
- [18] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemometrics Intellig. Lab. Syst.*, 76 (2005) 101-110.
- [19] R. Tauler, Application of non-linear optimization methods to the estimation of multivariate curve resolution solutions and of their feasible band boundaries in the investigation of two chemical and environmental simulated data sets, *Anal. Chim. Acta*, 595 (2007) 289-298.

- [20] W.H. Lawton, E.A. Sylvestre, Self modeling curve resolution, *Technometrics*, 13 (1971) 617-633.
- [21] O.S. Borgen, B.R. Kowalski, An extension of the multivariate component-resolution method to three components, *Analytica Chimica Acta*, 174 (1985) 1-26.
- [22] R. Rajkó, K. István, Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry, *J. Chemom.*, 19 (2005) 448-463.
- [23] M.N. Leger, P.D. Wentzell, Dynamic Monte Carlo self-modeling curve resolution method for multicomponent mixtures, *Chemometrics Intellig. Lab. Syst.*, 62 (2002) 171-188.
- [24] H. Abdollahi, M. Maeder, R. Tauler, Calculation and Meaning of Feasible Band Boundaries in Multivariate Curve Resolution of a Two-Component System, *Anal. Chem.*, 81 (2009) 2115-2122.
- [25] A. Golshan, M. Maeder, H. Abdollahi, Determination and visualization of rotational ambiguity in four-component systems, *Anal. Chim. Acta*, 796 (2013) 20-26.
- [26] R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, *J. Chemom.*, 15 (2001) 627-646.
- [27] A. Barducci, D. Guzzi, P. Marcoionni, I. Pippi, Aerospace wetland monitoring by hyperspectral imaging sensors: A case study in the coastal zone of San Rossore Natural Park, *J. Environ. Manage.*, 90 (2009) 2278-2286.
- [28] M. Sawall, K. Neymeyr, A fast polygon inflation algorithm to compute the area of feasible solutions for three - component systems. II: Theoretical foundation, inverse polygon inflation, and FAC - PACK implementation, *J. Chemom.*, 28 (2014) 633-644.
- [29] P.A. Højen-Sørensen, O. Winther, L.K. Hansen, Mean-field approaches to independent component analysis, *Neural Comput.*, 14 (2002) 889-918.
- [30] J.M. Nascimento, J.M. Bioucas Dias, Vertex component analysis: A fast algorithm to unmix hyperspectral data, *Geoscience and Remote Sensing, IEEE Transactions on*, 43 (2005) 898-910.
- [31] J. Li, J.M. Bioucas-Dias, Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data, in: *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, IEEE, 2008*, pp. III250-III253.
- [32] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, *Anal. Chem.*, 63 (1991) 1425-1432.
- [33] R. Tauler, Multivariate curve resolution applied to second order data, *Chemometrics Intellig. Lab. Syst.*, 30 (1995) 133-146.
- [34] A. de Juan, E. Casassas, R. Tauler, *Soft Modeling of Analytical Data*, in: *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Ltd, 2006.
- [35] M. Sawall, C. Kubis, D. Selent, A. Börner, K. Neymeyr, A fast polygon inflation algorithm to compute the area of feasible solutions for three - component systems. I: concepts and applications, *J. Chemom.*, 27 (2013) 106-116.
- [36] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Networks*, 13 (2000) 411-430.
- [37] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, *Physical review E*, 69 (2004) 066138.
- [38] S. Beyramysoltan, R. Rajkó, H. Abdollahi, Investigation of the equality constraint effect on the reduction of the rotational ambiguity in three-component system using a novel grid search method, *Anal. Chim. Acta*, 791 (2013) 25-35.

## Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

Table 1. Mutual information (MI) values for spectra, Amari Index (AI) values for the resolved concentration profile, and lof values for reconstructed data sets using MCR-ALS, PCA, MVSA, MCR-FMIN, ICA methods.

Data set	lof(%) <sup>a</sup>						Mutual Information <sup>b</sup>						Amari index <sup>c</sup>				
	TRUE	MCR-ALS	PCA	MVSA	MCR-FMIN	ICA	TRUE	MCR-ALS	PCA	MVSA	MCR-FMIN	ICA	MCR-ALS	PCA	MVSA	MCR-FMIN	ICA
1	0.308	<b>0.298</b>	<b>0.298</b>	<b>0.298</b>	<b>0.298</b>	<b>0.298</b>	0.578	0.922	0.414	1.045	<b>0.326</b>	0.923	0.207	0.565	0.294	<b>0.197</b>	0.220
2	0.544	0.533	<b>0.499</b>	0.524	<b>0.499</b>	0.513	2.390	2.373	2.449	2.814	<b>2.205</b>	2.404	0.101	0.542	0.286	0.389	<b>0.076</b>
3	0.494	0.471	<b>0.463</b>	0.528	<b>0.463</b>	0.465	4.419	4.335	4.032	4.334	<b>3.683</b>	4.018	0.280	0.392	0.131	0.319	<b>0.129</b>
4	0.578	<b>0.568</b>	<b>0.568</b>	0.685	<b>0.568</b>	<b>0.568</b>	0.578	0.575	0.412	0.408	<b>0.169</b>	0.311	0.178	0.547	0.407	0.318	<b>0.157</b>
5	0.583	0.572	<b>0.571</b>	0.725	<b>0.571</b>	0.572	0.578	0.579	0.412	0.408	<b>0.271</b>	0.374	<b>0.114</b>	0.548	0.411	0.229	0.129
6	0.685	0.673	<b>0.671</b>	1.596	<b>0.671</b>	0.672	0.578	0.714	0.414	0.429	<b>0.339</b>	0.579	0.085	0.583	0.426	0.339	<b>0.002</b>
7	0.814	0.802	<b>0.798</b>	3.780	<b>0.798</b>	0.801	0.578	1.001	0.415	0.407	<b>0.557</b>	0.823	0.264	0.597	0.340	<b>0.083</b>	0.297
8	0.838	0.826	<b>0.822</b>	1.898	<b>0.822</b>	0.826	0.578	2.156	<b>0.416</b>	0.436	<b>0.416</b>	0.620	0.704	0.599	0.436	0.599	<b>0.391</b>
9	0.475	0.462	<b>0.461</b>	40.348	<b>0.461</b>	0.466	2.390	2.563	2.405	2.464	<b>2.253</b>	2.433	<b>0.080</b>	0.480	0.351	0.137	0.237

<sup>a</sup> lof values calculated according to Equation 9.

<sup>b</sup> Mutual Information (MI) values calculated according to Equation 8.

<sup>c</sup> Amari index values calculated according to Equation 7.

## Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

Table 2. Relative signal component contribution function values (SCCF),  $f^a$ , obtained by MCR-BANDS, for datasets 2.

Data set	Component	$f_{TRUE}$	$f_{MCR-ALS}$	$f_{PCA}$	$f_{MVSA}$	$f_{MCR-FMIN}$	$f_{ICA}$	$f_{max}$	$f_{min}$
1	1	0.651	0.566	<b>0.992</b>	0.551	0.641	0.621	0.675	0.301
	2	0.469	0.621	<b>0.123</b>	0.684	0.486	0.565	0.816	0.451
2	1	0.409	0.596	<b>0.982</b>	0.536	0.433	0.511	0.728	0.298
	2	0.489	0.371	<b>0.186</b>	0.358	<b>0.120</b>	0.367	0.371	0.189
	3	0.328	0.267	<b>0.041</b>	<b>0.803</b>	0.666	0.356	0.620	0.185
3	1	0.369	0.197	<b>0.944</b>	<b>0.451</b>	0.408	0.474	0.571	0.143
	2	0.226	0.658	<b>0.313</b>	<b>0.673</b>	<b>0.328</b>	0.470	0.661	0.391
	3	0.659	0.400	<b>0.097</b>	<b>0.462</b>	0.405	0.281	0.503	0.300
	4	0.421	0.376	<b>0.049</b>	<b>0.293</b>	0.312	0.236	0.433	0.145
4	1	0.625	0.624	<b>0.958</b>	<b>0.958</b>	<b>0.674</b>	0.500	0.624	0.458
	2	0.781	0.779	<b>0.287</b>	<b>0.287</b>	<b>0.573</b>	0.774	0.779	0.636
5	1	0.621	0.622	<b>0.959</b>	<b>0.959</b>	<b>0.793</b>	0.649	0.665	0.454
	2	0.776	0.775	<b>0.283</b>	<b>0.283</b>	<b>0.463</b>	0.630	0.811	0.601
6	1	0.652	0.601	<b>0.987</b>	<b>0.987</b>	0.549	0.650	0.656	0.381
	2	0.521	0.607	<b>0.163</b>	<b>0.164</b>	0.653	0.523	0.770	0.513
7	1	0.594	0.704	<b>0.998</b>	<b>1.001</b>	0.553	0.711	0.810	0.331
	2	0.475	0.385	<b>0.684</b>	<b>0.088</b>	0.518	0.361	0.763	0.257
8	1	0.582	0.736	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>0.910</b>	0.829	0.349
	2	0.465	0.323	<b>0.017</b>	<b>0.004</b>	<b>0.017</b>	<b>0.110</b>	0.669	0.205
9	1	0.327	0.344	<b>0.991</b>	0.623	0.499	0.430	0.722	0.256
	2	0.426	0.520	<b>0.114</b>	6.182	0.309	0.417	0.554	0.271
	3	0.440	0.326	<b>0.071</b>	5.195	0.425	0.330	0.564	0.232

<sup>a</sup> SCCF f value calculated for MCR-ALS, PCA, MVSA, MCR-FMIN, ICA, and maximum and minimum values of SCCF,  $f_{min}$  and  $f_{max}$ , using MCR-BANDS according to Equation 6.

## Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods

Table 3. Relative signal component contribution function values (SCCF),  $f$ , for different profiles resolved from dataset 2.

Component <sup>a</sup>	$f_{\text{Point}_N}$ <sup>b</sup>	$f_{\text{MCR-ALS}}$ <sup>c</sup>	$f_{\text{TRUE}}$ <sup>d</sup>	$f_{\text{max\_MCR}}$ <sup>e</sup>	$f_{\text{min\_MCR}}$ <sup>f</sup>	$f_{\text{max\_point}_N}$ <sup>g</sup>	$f_{\text{min\_point}_N}$ <sup>h</sup>	$f_{\text{max\_TRUE}}$ <sup>i</sup>	$f_{\text{min\_TRUE}}$ <sup>j</sup>
1	0.1835	0.2675	0.3283	0.3701	0.1814	0.3401	0.1780	0.3402	0.1780
2	0.4853	0.3701	0.4894	0.7359	0.2855	0.6821	0.3058	0.7399	0.3057
3	0.5831	0.5955	0.4088	0.6197	0.1423	0.6201	0.1603	0.6201	0.1586

<sup>a</sup> Component number

<sup>b</sup> SCCF values of profiles corresponding to point N (example feasible solution in Figure 3 and 4)

<sup>c</sup> SCCF values of MCR-ALS solutions with non-negativity constraint.

<sup>d</sup> SCCF values of pure component spectra used for simulation of the dataset.

<sup>e</sup> Maximum SCCF values obtained using MCR-BANDS from MCR resolved profiles.

<sup>f</sup> Minimum SCCF values obtained using MCR-BANDS from MCR resolved profiles

<sup>g</sup> Maximum SCCF values obtained using MCR-BANDS from point N corresponding profiles.

<sup>h</sup> Minimum SCCF values obtained using MCR-BANDS from point N corresponding profiles.

<sup>i</sup> Maximum SCCF values obtained using MCR-BANDS from pure component spectra used for simulation of the dataset.

<sup>j</sup> Minimum SCCF values obtained using MCR-BANDS from pure component spectra used for simulation of the dataset.

## Discussion

Results of Mutual Information (MI) and of Amari Index (AI) showed that resolved results with independence and orthogonality constraints usually do not match the profiles of the pure component in the experiment. Profiles independence or orthogonality makes the profiles of the components to differ as much as possible, but this does not mean that they correspond to the true profiles causing the observed data variance. In the real situation, especially in the analysis of hyperspectral imaging of this Thesis, independence or orthogonality conditions are not fulfilled by spectra profiles in general. Pure component spectral in the hyperspectral imaging are neither independent nor orthogonal, and they may be overlapped. ICA and PCA obtain good mathematical models, but these models are different from the true ones. MVSA method provided the better estimations of the pure spectra (endmembers) of the hyperspectral imaging. However concentration profiles are obtained by least squares post processing when using MVSA. And this operation was always giving negative values for them, which will make results non-realistic. Closure (sum to 1) constraint in MVSA is not an adequate constraint for the estimation of concentrations (distribution maps) since they do not match the true situation of hyperspectral imaging.

Bilinear models evaluated in this chapter can be applied as preprocessing, exploration, classification, regression and resolution tools. All of them can be considered as an alternative tool for the analysis of mixed signals on certain cases. However, implementation of physical constraints, such as non-negativity in resolved spectra and concentration profiles is more flexible and reliable in MCR methods like in MCR-ALS. This is also the reason why MCR-ALS is a more adequate method for hyperspectral imaging analysis.

MCR-BANDS delivers an approximation to the extension of rotation ambiguity. AFS represent all feasible solutions geometrically and can be depicted by a method like FAC-PACK. The results of this chapter (Paper <sup>267</sup>) have shown that the extension of rotation ambiguity calculated by MCR-BANDS and the AFS calculated by a method like FAC-PACK are concordant. This is an important conclusion for future work and applications of the different methods evaluating the rotation ambiguity extension.

## **Chapter 6**

# **Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution**



## Introduction

### Dissolved Organic Matter (DOM) analysis in surface waters

River flows include all the material flowing through the river, not only water, but also sediments, nutrients, organic debris and biomass. These flows are a key aspect in monitoring studies for the environmental quality assessment of the flowing area near to the river.<sup>291</sup> They have strong seasonal patterns and display strong changes due to natural droughts. River water quality assessment is usually based on the comparison of measured values of particular physicochemical parameters with threshold values defined in national or international directives<sup>292</sup>.

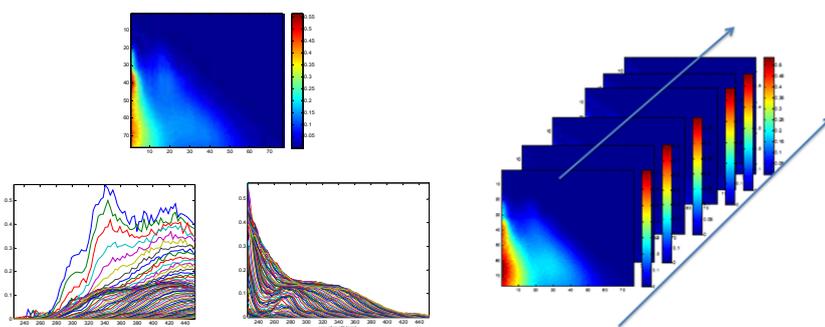
In the surface water systems, the dissolved organic matter (DOM) is a key parameter which is formed by a mixture of aromatic and aliphatic hydrocarbon structures with different attached functional groups. These functional groups have different structures with a wide range of molecular weights<sup>293</sup>. In the past, DOM in surface water systems has been considered an important element for water quality assessment. Its composition varies in time and space depending on the proximity to the sources and on the exposure to degradation processes<sup>294</sup>. In natural environments, concentration and composition changes of DOM have a potential effect on the health of ecosystems through processes like light attenuation, nutrient availability, and contamination transport<sup>295</sup>.

As the chemical nature of DOM defines its optical properties, optical measurements have been applied for tracing its variability in natural waters<sup>296</sup>. The two major DOM components that have been found to fluoresce are humic materials (giving blue fluorescence) and protein fractions (giving UV fluorescence). The most common fluorescent peaks identified in natural water included humic and fulvic-like matter with fluorescence patterns called as fluorescence A (237-260/400-500 nm) and C (300-370/400-500 nm), and protein-like material fluorescence centers, such as fluoreophores T (225-237/340-381 nm and 275/340 nm) and B (225-237/390-321 and 275-310 nm).<sup>297, 298</sup> They are respectively written as UVA, UVC, UVT and UVB for short. Additionally, there are moleculars with fluorescence patterns called UVD (soil fulvic acid, 390/509 nm), UVE (soil fulvic acid, (455/521 nm)), UVM (marine humic-like, (312/380-420 nm)) in natural environments<sup>294</sup>.

### Excitation Emission Matrix (EEM) fluorescence combined with chemometric methods for DOM analysis and water quality assessment

It is difficult to trace changes in the structure of DOM in natural water systems because of the complexity and non-conservative behavior of the composition of DOM in natural environments<sup>299, 300</sup>.

Molecular fluorescence spectroscopy is a particularly useful technique for the study of the chemical properties of fluorescent molecules. It is a highly sensitive analytical technique, allowing nondestructive direct measurements at natural environmental concentrations<sup>301</sup>. Molecular fluorescence is a very sensitive technique for the analysis of the chemical environment of fluorophors<sup>302</sup>, like those present in DOM. The excitation–emission matrix (EEM) of fluorescence generates multidimensional data structures suitable for this type of analysis. The utilization of EEM fluorescence can be used in the characterization of the properties and reactivity of DOM. EEM allows DOM analysis and monitoring in surface waters and other natural systems<sup>303</sup>. In the next Figure, EEM experimental measures of DOM in river water are shown.



**Figure 6- 1** EEM data matrix structure of the water samples obtained from Ter River.

In this Thesis the basic EEM data matrix structure for the analysis of one single water sample was composed by 76 rows (76 excitation bands) and 76 columns (76 emission bands), giving a data matrix, **D** (76, 76), which contains the fluorescence spectra of the considered water sample. Excitation and emission spectra of the compounds studied in individual samples can be arranged in data tables or data matrices. In this study only the fraction of the raw EEM that did not contain scattering bands (from 227-450nm, avoiding the region between 200 and 225nm) was selected.

As said before EEM spectra can be arranged in a data cube or three-way, three-mode data structure. Totally 122 water samples were collected at 62 stations along the Sau Reservoir system and its tributary, the Ter River. When the EEM data of 122 samples were considered, then a new data structure was obtained which can be arranged in a long column-wise data matrix **D<sub>aug</sub>** (122×76, 76) which can be arranged by setting individual

data matrices from every sample, one on the top of the other and keeping the emission wavelength axis in common.

Chemometric analysis of data sets obtained in large environmental monitoring studies of natural fresh water systems by EEM fluorescence spectra (such as those shown in previous Figure 6-1) provide a powerful and convenient way to investigate DOM source contributions of the investigated watersheds<sup>304-306</sup>.

Appropriate three-way chemometric data analysis methods, including so-called second-order calibration methods and those based on trilinear modeling<sup>200</sup> have been proposed for the analysis of EEM datasets. The high analytical potential of EEM coupled to chemometric methods has long ago been recognized. Due to the intrinsic structure of EEM data, the Parallel Factor Analysis (PARAFAC) method<sup>241</sup> based on the trilinear model is adequate and it can be successfully apply for their decomposition. PARAFAC decomposes directly the three-way data set or data cube into components using and alternative least squares method. In Chapter 2, these methods were described.

Alternatively the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS), method can be used also for the analysis of three-way data fulfilling the premises of the trilinear model<sup>200</sup>. MCR-ALS was used in this Thesis to analyze the EEM dataset (paper<sup>307</sup>).

The bilinear decomposition of the data matrix  $\mathbf{D}_{aug}$  can be described by the following equation:

$$\mathbf{D}_{aug} = \mathbf{S}_{Ex-C} \mathbf{S}_{Em}^T + \mathbf{E} \quad \text{Equation 6- 1}$$

where  $\mathbf{S}_{Em}$  (76,n) is the matrix of excitation spectra (n is the number of considered bilinear components contributing to the observed fluorescence signal),  $\mathbf{S}_{Ex}$  (Excitation Profiles) and  $\mathbf{C}$  (Contribution Profiles) are mixed in the augmented  $\mathbf{S}_{Ex-C}$  matrix (122×76, n), and  $\mathbf{E}$  is the residual matrix describing the variance not explained by the bilinear model ( $\mathbf{S}_{Ex-C} \mathbf{S}_{Em}^T$ ).

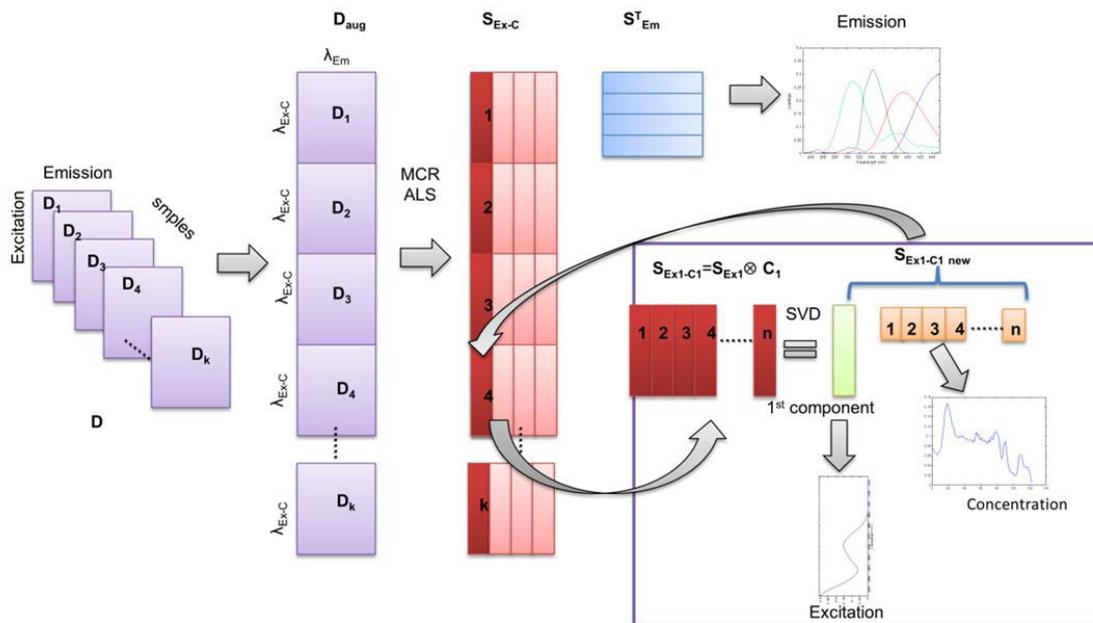
When the trilinear model<sup>308, 309</sup> holds, every individual data matrix,  $\mathbf{D}_k$ , is decomposed by the following equation

$$\mathbf{D}_k = \mathbf{S}_{Ex} \mathbf{C}_k \mathbf{S}_{Em}^T + \mathbf{E} \quad \text{Equation 6- 2}$$

where k refers to slice k and  $\mathbf{C}_k$  is a diagonal concentration matrix referring to the relative amounts of the components in slice/matrix k.  $\mathbf{S}_{Ex}$  and  $\mathbf{S}_{Em}^T$  matrices are the same for all  $\mathbf{D}_k$  simultaneously analyzed. Only the new  $\mathbf{C}_k$  diagonal matrix of dimensions (n, n), changes from sample to sample and it gives in its diagonal the relative spectral contributions of the n components to the fluorescence signal of sample matrix k.

Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

Alternating Least Squares optimization of both  $\mathbf{S}_{\text{Ex-C}}$ , estimation of  $\mathbf{S}_{\text{Em}}$  is performed iteratively under non-negative, trilinearity and normalization constraints until convergence is achieved. According to Figure 6-2, when trilinearity constraint is applied during the ALS optimization,  $\mathbf{S}_{\text{Ex-C}}$  is first decomposed by SVD for each component. Only the first singular value is considered in this decomposition. For this component, the shape of its excitation spectrum will be exactly the same for all the considered samples and that it only will change its relative intensity. After this decomposition, the full  $\mathbf{S}_{\text{Ex-C}}$  is rebuilt and updated for the next ALS iteration. Apart from forcing the shape of the excitation spectrum of the considered component to be the same for all different samples, this procedure captures the relative intensity variation of this component, which is stored in a contribution or concentration  $\mathbf{C}$  matrix (loading of the third mode), giving therefore the relative contribution of this component in the different samples.  $\mathbf{S}_{\text{Ex}}$  (Excitation spectra) and  $\mathbf{C}$  (Contribution Profiles) in  $\mathbf{S}_{\text{Ex-C}}$  matrices are finally obtained, together with  $\mathbf{S}_{\text{Em}}$  (emission spectra) matrix obtained directly during the ALS procedure (Equation 6-2).



**Figure 6-2** Decomposition of the three way excitation–emission fluorescence data set using the trilinearity constraint during the Alternating Least Squares optimization in MCR-ALS.

Compared to PARAFAC, the main advantage of MCR-ALS is that the trilinearity constraint can be applied independently on each component, in contrast to PARAFAC where all resolved components should fulfill the sought trilinear condition.

### The dataset investigated in this work

The study area investigated in this work was the middle Ter River flow, which supplied 55% of the raw water for drinking purposes to Barcelona city and its surroundings. With a 3010 km<sup>2</sup> surface basin and 208 km long, it originates in the middle of the Catalan more oriental range of the Pyrenees (at 2400 m high) and it passes through Girona city, and leads to the Mediterranean Sea<sup>310</sup>, see Figure 6-3<sup>311</sup> and Figure 6-4 (From google map). Most of the Ter river basin is characterized by a medium population density; Girona and Vic are the most populous cities. In terms of the use of the river water, 53% of the total water is assigned to agricultural irrigation and golf courses, 24% to domestic use, and the rest is for industrial (17%) and livestock farmer (7%) use<sup>312</sup>. This river receives the direct impact of some metallurgic, pulp mill, textile and tannery industries<sup>313</sup>. As a result in the river have found alkylphenols, alkylphenol ethoxylates, aliphatic and aromatic hydrocarbons, fatty acids, plasticizers, alkylbenzenes and pesticides, the discharge of urban and industrial waters, and of agricultural runoff<sup>314</sup>.



Figure 6- 3 The location of Ter River basin in Catalunya, Spain.

Main source and transformation processes of the DOM in Ter river system were characterized and located, in middle course of Ter river (which is a rather populated area) is described, several kilometers before entering its main reservoirs in Susqueda and

Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

Sau (see Figure 6-4), during its travel throughout the lentic ambient until reaching these two dams. Results obtained in this study support the hypothesis that the different fractions of DOM are differentially processed along the river-reservoir system<sup>307</sup>.



**Figure 6- 4** The upper part is the terrain map of Ter river basins and the lower part is the main reservoirs in Susqueda and Sau (The figure was obtained from online Google map).

**The published paper in this chapter**

ZHANG, Xin., Marcé, R., Armengol, J., & Tauler, R. Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution. *Chemosphere*, 2014, 111, 120-128



## Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution



Xin Zhang<sup>a</sup>, Rafael Marcé<sup>b,c</sup>, Joan Armengol<sup>c</sup>, Romà Tauler<sup>a,\*</sup>

<sup>a</sup> IDAEA-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain

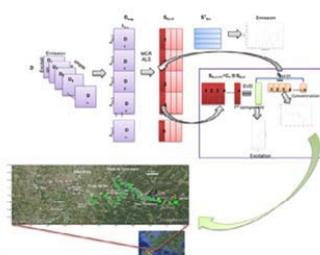
<sup>b</sup> Catalan Institute for Water Research (ICRA), Emili Grahit 101, 17003 Girona, Spain

<sup>c</sup> Department of Ecology, University of Barcelona, Diagonal 645, 08028 Barcelona, Spain

### HIGHLIGHTS

- MCR-ALS and PARAFAC provided similar solutions with chemical significance.
- Resolved emission and excitation spectra identified major DOM river constituents.
- Geographical distribution of DOM river contributions is displayed.
- Correlation between human activities and environmental impact are discussed.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

#### Article history:

Received 10 December 2013

Received in revised form 25 February 2014

Accepted 1 March 2014

Available online 19 April 2014

Handling Editor: Klaus Kümmerer

#### Keywords:

DOM

River-reservoir

Excitation and emission matrix fluorescence

MCR-ALS

PARAFAC

Fresh water system

### ABSTRACT

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method with the trilinearity constraint is proposed for the analysis of excitation–emission fluorescence data from Dissolved Organic Matter (DOM) in fresh water natural systems, and the results obtained are compared with those obtained with PARAFAC. The effects of different number of components and constraints on the stability of the proposed models are compared. MCR-ALS is shown to be an effective way to characterize and resolve DOM sources in natural fresh water systems from EEM data, with good correlation with experimentally measured DOM concentration values. MATLAB georeferenced mapping is used to illustrate the geographical distribution of resolved DOM contributions. MCR-ALS resolved EEM spectra are used to recognize the corresponding chemical groups assigned to possible DOM sources. Relationships between human activities and the environmental situation of the river system are discussed from these possible DOM sources.

© 2014 Elsevier Ltd. All rights reserved.

### 1. Introduction

Dissolved Organic Matter (DOM) consists of a complex mixture of organic molecules that vary greatly in molecular weight, like polysaccharide, polypeptide proteins, and fulvic acids (Leenheer and Croué, 2003; Stedmon and Markager, 2005a). In a fresh water

system, organic carbon processing is impacted by downstream export, and biogeochemical transformations. In rivers, lakes, and man-made reservoirs, DOM sources are usually categorized as allochthonous and autochthonous, material derived from outside the ecosystem and biota within the system respectively. Organic matter obtained from a water system of tributaries can be effectively related with the carbon and oxygen cycling in a receiving lake or man-made reservoir. Moreover, accumulation of microbial biomass and activity in the riverine section of reservoirs make this

\* Corresponding author. Tel.: +34 93 400 61 40. Fax: +34 93 204 59 04  
E-mail address: [Roma.Tauler@idaea.csic.es](mailto:Roma.Tauler@idaea.csic.es) (R. Tauler).

# Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

X. Zhang et al./Chemosphere 111 (2014) 120–128

121

section important in terms of allochthonous organic matter processing (Weissenberger et al., 2010). Reservoir ecosystems are open systems, and reservoir dynamics are influenced by river inflows (Imberger and Ivey, 1993). The relative high water load in reservoirs defines these systems as important sites for organic carbon processing at local and global scales also (Weissenberger et al., 2010).

To trace the dissolved organic matter, fluorescence measurements have been proposed to characterize and trace DOM dynamics (Coble et al., 1990; Coble, 1996; Zsolnay et al., 1999; Yan et al., 2000; Stedmon and Markager, 2005b). Excitation Emission Matrix (EEM) spectroscopy produces fluorescence (emission) spectra at many different excitation wavelengths providing an overall view of different features existing within the selected spectral range (Coble, 1996; Her et al., 2003; Yamashita and Tanoue, 2003). EEM data are usually assumed to conform adequately to a three-way trilinear model system, which can be resolved in excitation and emission spectra and the corresponding relative concentrations of the fluorescent components present in the analyzed samples.

As the chemical nature of DOM defines its optical properties, optical measurements have been applied for tracing its variability in natural waters (Stedmon and Markager, 2001). The optically active fraction of DOM is called colored dissolved organic matter (CDOM) and it can be used as a tracer for the dynamics and characteristics of the total DOM pool. CDOM can also give fluorescence when excited by light in the UV and blue VIS region of the spectrum. The two major DOM components that have been described to fluoresce are humic material (giving blue fluorescence) and protein fractions (giving UV fluorescence). The most commonly found fluorescent bands identified in natural water included, humic and fulvic-like material fluorescence centers, as fluorophores A and C, and protein-like material fluorescence centers as fluorophores T and B (Lochmueller and Saavedra, 1986; Coble, 1996). It has been proposed that the combination of EEM data and chemometrics data analysis methods can provide a very powerful tool in the assessment of DOM dynamics (Yamashita et al., 2008) in fresh water systems. The PARAFAC (Bro, 1997) method has already been effectively used in the analysis and modeling of multi-way data systems to extract both qualitative and quantitative information from three way EEM data systems. Multivariate Curve Resolution Alternating Least Squares, MCR-ALS (Jaumot et al., 2005) is another chemometrics method that has been proposed as a powerful data processing technique for the analysis of experimental data arranged in multiway and multi-set data. MCR-ALS decomposes a data matrix using a bilinear model which produces two factor matrices of reduced sizes, which provide interpretable information in physical and chemical terms about the investigated system. MCR-ALS method has been extended to the analysis of three way and multi-way data sets (Brown et al., 2009). When applied to three-way data as in EEM spectroscopy, results obtained by MCR-ALS with the trilinearity constraint give similar results than other trilinear methods like PARAFAC (Tauler et al., 1995a; Alier et al., 2011).

This paper investigates the potential use of MCR-ALS method with the trilinearity constraint for the resolution of main DOM source from EEM monitoring data in a complex fresh water system and the results obtained are compared with those also obtained by PARAFAC. The investigated system consisted in multiple water samples collected along the Ter River-Sau Reservoir system in Catalonia (Northwest Spain, Fig. 1), which is a river-reservoir system located in a watershed under strong human impact. The constituents of DOM resolved by the proposed chemometric data analysis methods in this study are discussed in terms of the possible chemical organic matter constituents responsible of the excitation and emission spectra and of their distribution along the river section under study and their impact.

## 2. Methods

### 2.1. Samples

#### 2.1.1. Samples collection and DOM experimental determination

Water samples were collected at 62 stations along the Sau Reservoir system and its tributary, the Ter River (see Fig. 1), starting at straight-line distance of 9 km upstream from the reservoir until down the Sau dam. Water sampling was approximately every 400 m along the river-reservoir system. It is a canyon-shaped reservoir area, in which mostly one single tributary (Ter river) is responsible for most part of the total river flow. Most part of the total inflow has only this single tributary, which facilitates the investigation of the established chemical gradients. The exact positions of sampling stations have been recorded with a global positioning system (GPS). Samples were collected within two days in October 2009. During this time, the river water enters with an insertion depth very close to half the maximum depth of the reservoir conforming a layer of water to contrasting physical and chemical proprieties which are discernible with simple in situ measures (water temperature, conductivity, and dissolved oxygen). From stations 23–62 (see map in Fig. 1), one sample from the river layer was samples, and one surface sample was also collected every two stations. Along the Ter River, there are two towns (Manlleu and Roda de Ter, see map in Fig. 1) located upstream from the reservoir system and that pour their organic pollutants into the river system, and two waste water treatment plants (Vic WWTP and Roda de Ter WWTP in the map) which affect also the organic matter content of the investigated water system and analyzed samples. The whole area is a highly humanized watershed and it is therefore, a good example to investigate relationships between natural DOM and anthropogenic DOM sources in a river-reservoir system.

In addition, in order to compare and validate results obtained from the analysis of excitation/emission data by MCR-ALS and PARAFAC, DOC concentration values of every analyzed sample were experimentally determined in the laboratory by combustion in a Shimadzu TOC-5000 Analyzer after filtration through a pre-combusted GF/F filter and acidification to remove inorganic carbon. These values are given in Fig. 1.

#### 2.1.2. Excitation emission fluorimetric determinations

EEM spectra were recorded in a Hitachi F-7000 fluorescence spectrophotometer. Spectra were collected using a Rhodamine B solution (Wako Pure Chemical Industries, Ltd.) and the light diffuser and filter set. Bandwidths for excitation and emission were set at 5 nm, and both excitation and emission were in the spectral range of 200–600 nm, at 3 nm increments. Scanning speed was at 12,000 nm min<sup>-1</sup> and power at 700 V. EEM spectra were corrected for inner filter effects with the corresponding absorbance spectra and converted to Raman Units (RU) using the area under the Mili-Q water Raman scatter peak at excitation 350 nm. Raman scatter was removed subtracting the EEM spectrum for Mili-Q water, while Rayleigh scatter effects were removed using a 3D interpolation approach Kowalczyk et al., 2005.

### 2.2. Chemometric methods

The basic excitation–emission matrix (EEM) data structure for the analysis of one single water sample was composed by 76 rows (76 excitation wavelengths) and 76 columns (76 emission wavelengths), **D** (76, 76), which contains the fluorescence spectra of the considered water sample. In this study only the fraction of EEM data which did not contain scatter bands (from 227 to 450 nm) was selected and the region between 200 and 225 nm

# Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

122

X. Zhang et al./Chemosphere 111 (2014) 120–128

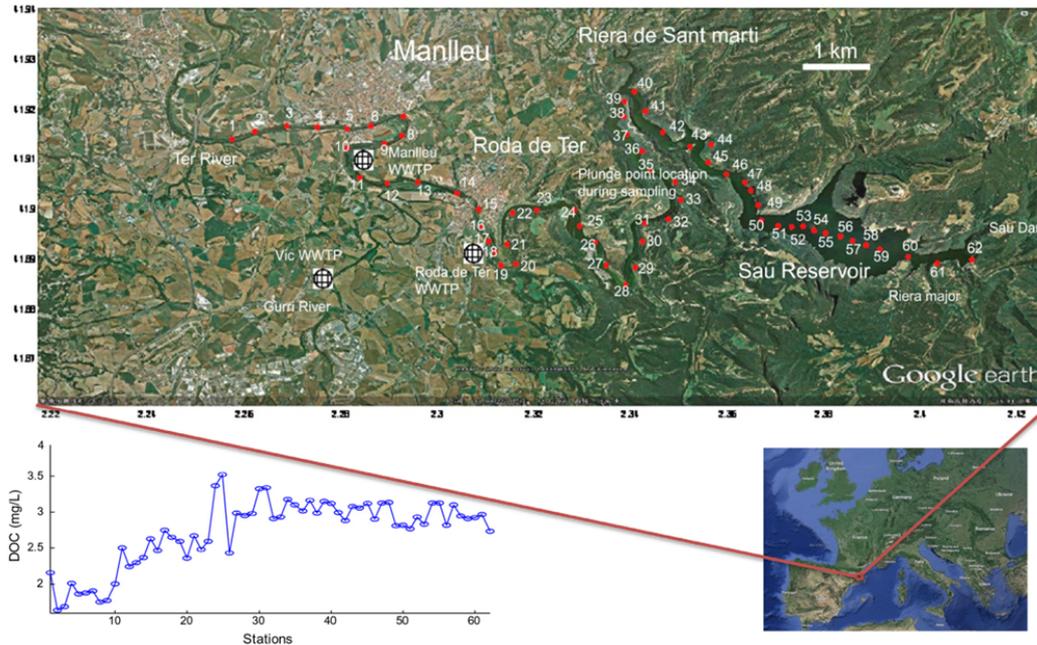


Fig. 1. Mapping, location and DOC concentrations of the different sampling sites of the Ter river-Sau reservoir system.

was eliminated. The bilinear decomposition of the data matrix  $\mathbf{D}$  can be described by the following equation:

$$\mathbf{D} = \mathbf{S}_{\text{Ex}} \mathbf{S}_{\text{Em}}^T + \mathbf{E} \quad (1)$$

where  $\mathbf{S}_{\text{Em}}$  ( $76, nc$ ) is the matrix of excitation spectra ( $nc$  is the number of considered bilinear components contributing to the observed fluorescence signal) and  $\mathbf{S}_{\text{Ex}}$  ( $76, nc$ ) is the matrix of emission spectra, and  $\mathbf{E}$  is the residual matrix describing the variance not explained by the bilinear model ( $\mathbf{S}_{\text{Em}} \mathbf{S}_{\text{Ex}}^T$ ). The unique bilinear decomposition of matrix  $\mathbf{D}$  into  $\mathbf{S}_{\text{Ex}}$  and  $\mathbf{S}_{\text{Em}}$  is not assured if only one sample is analyzed (analysis of a single data matrix). However when multiple samples of the same system are simultaneously analyzed and the trilinear model is applied, both emission and excitation spectra ( $\mathbf{S}_{\text{Ex}}$  and  $\mathbf{S}_{\text{Em}}$ ) can be recovered univocally (Tauler et al., 1995).

When the EEM data of 122 samples are considered, a new data matrix  $\mathbf{D}_{\text{aug}}$  ( $122 \times 76, 76$ ) can be arranged column-wise, i.e. by setting individual data matrices from every sample one on the top of the other and keeping the emission wavelength axis in common.

$$\mathbf{D}_{\text{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \dots \\ \mathbf{D}_i \\ \dots \\ \mathbf{D}_{122} \end{bmatrix} = [\mathbf{D}_1; \mathbf{D}_2; \dots; \mathbf{D}_i; \dots; \mathbf{D}_{122}] \quad (2)$$

where the symbol ‘;’ is used to indicate the column-wise matrix augmentation (matrix concatenation in MATLAB notation).

When the trilinear model (de Juan et al., 1998; de Juan and Tauler, 2001) holds, every individual data matrix,  $\mathbf{D}_i$ , is decomposed by the following equation:

$$\mathbf{D}_i = \mathbf{S}_{\text{Ex}} \mathbf{C}_i \mathbf{S}_{\text{Em}}^T + \mathbf{E}_{\text{aug}} \quad (3)$$

where  $\mathbf{S}_{\text{Ex}}$  and  $\mathbf{S}_{\text{Em}}^T$  matrices are the same for all  $\mathbf{D}_i$  simultaneously analyzed. Only the new  $\mathbf{C}_i$  diagonal matrix of dimensions ( $nc, nc$ ), changes from sample to sample and it gives in its diagonal the relative spectral contributions of the  $nc$  components to the fluorescence signal of sample matrix  $i$ . The trilinear model and constraint is indeed fulfilled by fluorescence data, since excitation and emission spectra of the same chemical component should be the same whatever is the sample analyzed where is present. This important property of EEM data is very advantageous to give unique solutions of Eq. (3) for all simultaneously analyzed  $\mathbf{D}_i$  matrices. In this work, MCR-ALS has been adapted to the simultaneous analysis of the  $\mathbf{D}_i$  EEM data matrices measured for all 122 monitoring samples. The trilinear condition is achieved in MCR-ALS algorithmically during the ALS optimization and it has been described elsewhere (Tauler et al., 1998) and summarized in Fig. 2. The data set is first arranged in a column-wise augmented data matrix where the emission mode is the common mode of augmentation of all individual excitation–emission data matrices, one per analyzed river sample. Then the number of components is estimated, and initial estimates of  $\mathbf{S}_{\text{Em}}^T$  are obtained using the purest rows of the augmented data matrix. Alternating Least Squares optimization of both  $\mathbf{S}_{\text{Ex}-c}$  and  $\mathbf{S}_{\text{Em}}^T$  is performed iteratively under non-negative, trilinearity and normalization constraints until convergence is achieved.

$\mathbf{S}_{\text{Ex}}$  (Excitation Profiles) and  $\mathbf{C}$  (Contribution Profiles) are mixed in the augmented  $\mathbf{S}_{\text{Ex}-c}$  matrix. In Fig. 2, it is shown how these two matrices of  $\mathbf{S}_{\text{Ex}-c}$  and  $\mathbf{C}$  profiles can be finally recovered using the procedure used also for the implementation of the trilinear constraint in MCR-ALS. According to Fig. 2, when trilinearity constraint is applied during the ALS optimization,  $\mathbf{S}_{\text{Ex}-c}$  is first decomposed by SVD for each component. Only the first singular value is considered in this decomposition, implying that for this component, the shape of its excitation spectrum is exactly the same for all the considered

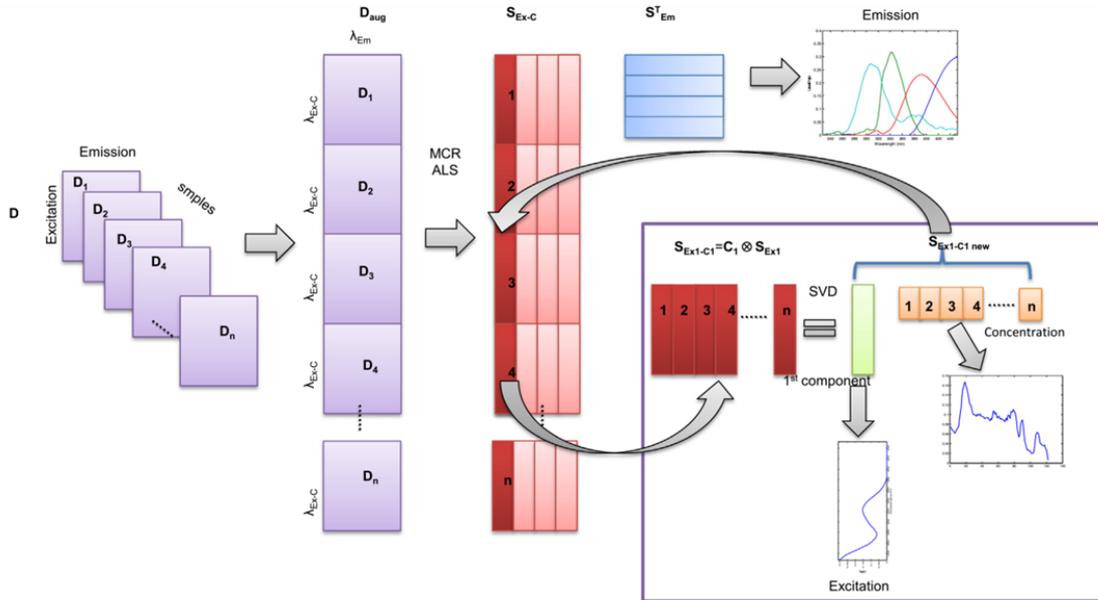


Fig. 2. Decomposition of the three way excitation–emission fluorescence data set using the trilinearity constraint during the Alternating Least Squares optimization in MCR-ALS.

samples and that it will only change its relative intensity according to concentration of this component. After this decomposition, the full  $S_{Exc-C}$  is rebuilt and updated for the next ALS iteration. Therefore, apart from forcing the shape of the excitation spectrum of the considered component to be the same for all different samples, this procedure captures the relative intensity variation of this component, which is stored in a contribution or concentration  $C$  matrix (loading of the third mode), giving the relative contribution of this component in the different samples.

The parallel factor analysis (PARAFAC) is based on the trilinear model and it decomposes directly the three-way data set or data cube into components using an alternative least squares method (Bro, 1997). In PARAFAC, it is also possible to use several constraints such as non-negativity to excitation, emission and concentration profiles. Core consistency diagnostic (CORCONDIA) is applied to estimate the number of components and to check for the appropriateness of the proposed model. PARAFAC can be considered a constrained version of the more general method Tucker3 with a super-identity core matrix. The core consistency diagnostic indicates how well the model is in concern with the distribution of superdiagonal and off-superdiagonal elements of the Tucker3 core. If the PARAFAC model is correct, then it is expected that superdiagonal elements will be close to one and the off-diagonal elements close to zero. More details can be found from literature (Bro and Kiers, 2003). The core consistency diagnostic (CORCONDIA) is defined as:

$$CORCONDIA = 100 \times \left( 1 - \frac{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (g_{def} - t_{def})^2}{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F t_{def}^2} \right) \quad (4)$$

where  $g_{def}$  is the calculated element of the core using the PARAFAC model, defined by dimensions  $(d \times e \times f)$ ;  $t_{def}$  the element of a binary array with zeros in all elements and ones in the superdiagonal

(the expected Tucker3 core) and  $F$  is the number of factors in the model.

To evaluate the quality of the results achieved by the application of MCR-ALS and PARAFAC the percentage of lack of fit (ALS *lof*) and the percentage of explained variance ( $R^2$ ) were calculated according to the following equations:

$$lof (\%) = 100 \times \sqrt{\frac{\sum_{ij} (d_{ij} - \hat{d}_{ij})^2}{\sum_{ij} d_{ij}^2}} \quad (5)$$

where  $d_{ij}$  is the element of the experimental data matrix  $D$ , and  $\hat{d}_{ij}$  is the corresponding element calculated by ALS. The lack of fit gives a measure of the fit quality in relative terms with the same units as the measured data, and comparable with experimental relative error estimations.

$$R^2 = 100 \times \left( 1 - \frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2} \right) \quad (6)$$

where  $e_{ij}$  are the elements of the  $E$  matrix and  $d_{ij}$  are the elements of the raw data set  $D$ .

MCR-ALS algorithm code of user-friendly GUI for MATLAB (The Mathworks Inc., MA, US) (2013) is available from the home page of MCR at <http://www.mcrals.info/>. PARAFAC analysis has been performed using the version implemented in PLS-toolbox (v.5.8).

Component contribution profiles (matrices  $C_i$ ) resolved by MCR-ALS can be plotted on maps obtained by Google Maps (2012) and Mapping Toolbox<sup>29</sup> under MATLAB computer and visualization environment. Using GPS values of river and reservoir water samples recorded during the monitoring campaign, the map of the fresh water system was obtained from Google

# Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

124

X. Zhang et al./Chemosphere 111 (2014) 120–128

map as a figure file (in tiff format). Contributions (matrices  $C_i$ ) of three different MCR-ALS resolved components were then plotted on the map according to their intensity and GPS value. Different components have different contributions which were plotted in the appropriated position of the map with a size proportional to the value. Color of the spots is also related to their intensity values, green means a high intensity value and red means a low intensity value.

## 3. Results and discussion

### 3.1. MCR-ALS and PARAFAC resolution of main DOM contribution sources over the investigated region using EMM spectral data

Experimental EEM data of dimensions  $122 \times 76 \times 76$ , corresponding to the analysis of the whole set of 122 water river/reservoir samples (at 76 excitation and 76 emission wavelengths) was analyzed with MCR-ALS and PARAFAC methods. SVD (single value decomposition) of the corresponding column- and row-wise augmented data matrices gave rather similar results for the first 6 components, with singular values equal to **117.34, 19.76, 7.17, 3.84, 2.51, 1.75** in the first case and **117.22, 20.67, 7.29, 2.93, 2.20, 1.33** in the second case. These results already indicate that the trilinear model assumption (de Juan and Tauler, 2001) is well justified at least for the first three components (Tauler et al., 1995).

In Fig. 3, pure emission and excitation fluorescence spectra resolved by MCR-ALS (red) with the trilinearity constraint and PARAFAC (blue) are given for different number of components (from three to six components). When only 3 components were considered MCR-ALS and PARAFAC recovered emission and excitation profiles were compared. When 4 components were considered, the two sets of emission and excitation spectra and the contribution profiles resolved by MCR-ALS and PARAFAC were also rather similar, except for the fourth component. When even more components were proposed, the trilinear model did not produce consistent results and the results obtained by both methods, MCR-ALS and PARAFAC resulted to be rather different. Table 1 shows the comparison of the fitting results obtained by MCR-ALS and PARAFAC when 3, 4, 5 and 6 components were considered. When 5 components were chosen, the shapes of the emission spectra obtained for the third and fifth components were significantly different for the two methods. Although data fitting values measured by  $R^2$  and  $lof$  (Table 1) obtained by the two methods were similar, resolved profiles obtained by both methods resulted to be rather different and results were less stable. This is even more obvious, when 6 components were considered and results can be then considered to be unstable and inconsistent. Choosing a too large number of components allowed explaining more data variance, as shown in Table 1 in  $lof$  values, but experimental noise can be then propagated and mixed with resolved profiles giving results that were not reliable nor stable anymore.

Core consistency diagnostic is also a good way to check for multilinear data structure and data complexity. In this work, core consistencies of PARAFAC model with 3 components was 86.4%. This means that when 3 components were chosen, the trilinear model was rather accurate. When 4, 5 and 6 components were chosen, core consistency values are  $-10.8\%$ ,  $-65.0\%$ , and  $-2824.6\%$  respectively. When negative values appeared, PARAFAC model results are unstable, and too much noise is probably included in the model (over fitting). From all these results, it is concluded that the more reliable model would be the one considering only three components. Selection of additional number of components is not safe enough from available data and the interpretation of the resolved components can be considered rather speculative.

### 3.2. EEM spectral properties and distribution of the main resolved DOM sources

Either excitation or emission spectra (Fig. 3) resolved by PARAFAC or by MCR-ALS (with the trilinear constraint) revealed the main fluorescence characteristics of DOM fluorophores present in the investigated fresh water system.

In the left part of Fig. 4, concentration changes of the three major possible DOM sources resolved by MCR-ALS are shown along the section of Ter river map (obtained from Google Maps application (2012)). The 62 sampling river stations investigated in this work were properly georeferenced and their latitude and longitude position marked on the river map. The three component profiles were plotted for the first component in Fig. 4a; for the second component in Fig. 4b; and for the third component in Fig. 4c. Contributions of each component were displayed on the map by circles whose areas were proportional to their intensity value, i.e. larger circles mean higher contributions and the opposite for smaller circles and lower contributions. Right part of Fig. 4 shows the corresponding lines describing the contribution changes calculated as explained in the method section for each of these three MCR-ALS resolved components. Blue lines indicate contribution profiles from samples obtained from middle depth river water samples and red lines indicate contribution profiles obtained from surface river waters samples.

The results obtained by MCR agree well with those previously reported in the literature for similar DOM sources. First MCR-ALS resolved component was similar to the one found in previous studies (Coble et al., 1998; Sharma and Schulman, 1999; Moran et al., 2000; Stedmon and Markager, 2001). Second and third MCR-ALS resolved components were also similar to those found in previous studies (Coble, 1996) (Mopper and Schultz, 1993; Determann et al., 1994; Coble, 1996; Determann et al., 1996; Yamashita and Tanoue, 2003).

Next, detailed description of possible chemical structures of the three MCR-ALS resolved components and of their concentration distribution details are given.

#### 3.2.1. First resolved DOM contribution

The first MCR-ALS resolved component explains 39.19% of the experimental variance and it shows two excitation band maxima, one at 355 nm and another at wavelengths below 240 nm, with a broad emission band near 460 nm. These spectral values are characteristic of the so called UVC type of humic matter fraction, as it has been previously reported by Stedmon and Markager (2001). At these wavelengths, low molecular weight aromatic organic matter is considered to be the main responsible of the observed spectral features, dominating the fluorescence of wastewater DOM. Deep waters and waters with greater terrestrial input have a greater potential for photo degradation than surface waters (Hudson et al., 2007). Humic matters like UVC are found to be more likely to be degrade (Moran et al., 2000). The fact that this component shows emission and excitation at longer wavelength (460 nm) than the other two components suggests that it contains more conjugated fluorescent aromatic molecules (Sharma and Schulman, 1999). Excitation at long wavelengths suggests that the chemical functional groups responsible for this fluorescence (fluorophores) are aromatic in nature or that they contain multiple overlapped unsaturated functional groups (Coble et al., 1998). Furthermore, similarities are observed between these resolved fluorescence spectra and those from hydroxy-, methoxy-, or naphthol chemical groups (Senesi, 1990). In summary, this resolved component is likely to represent the high molecular weight fraction derived from soluble terrestrial humic matter, i.e. soluble DOM fulvic acids present in most of natural water environments (Stedmon and Markager, 2005b).

# Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

X. Zhang et al./Chemosphere 111 (2014) 120–128

125

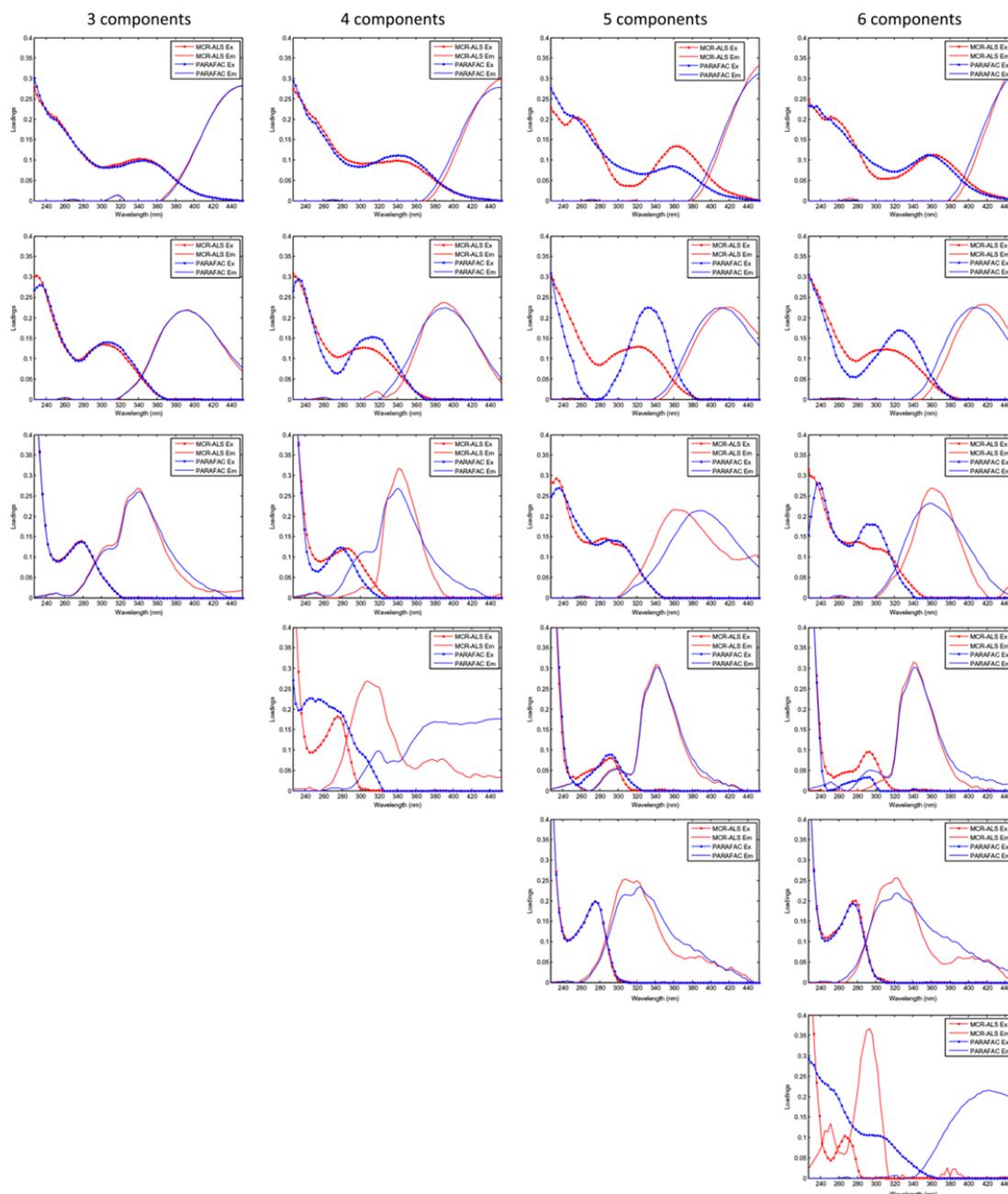


Fig. 3. MCR-ALS (red) and PARAFAC (blue) resolved emission ( $S_{Em}^i$ ) and excitation ( $S_{Ex,c}$ ) pure spectra profiles for different number of components. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Distribution of this first UVC component (Fig. 4) has a maximum contribution at the 5th station, and then a decrease of it in the next three stations. All these stations are down river flow, close to Manlieu town (see Fig. 4), a village which although has a med-

ium size population (more than 20,000 persons) and some industrial activity, does not produce a significant degradation of the water quality. From this point going down along the river, UVC increases their relative contribution, with humic acid fluorophore

# Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

126

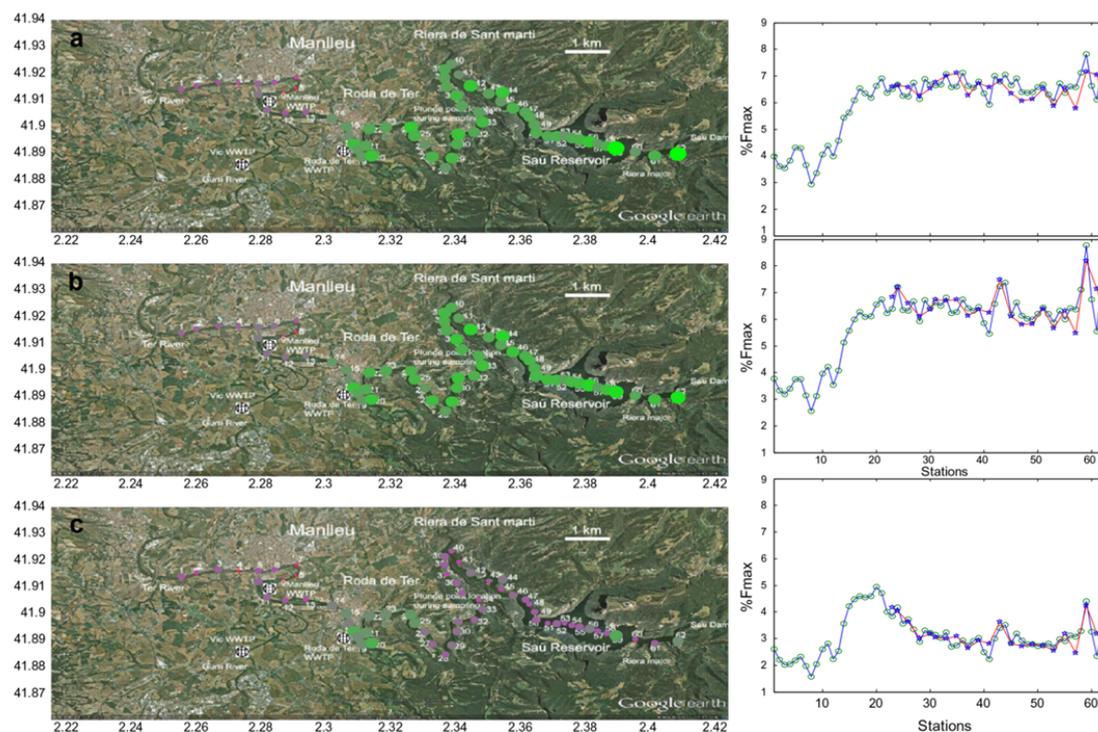
X. Zhang et al./Chemosphere 111 (2014) 120–128

**Table 1**  
Individual explained variances ( $R^2$ ) by MCR-ALS with trilinearity constrain and by PARAFAC when different number of components were considered.

		3 Components		4 Components		5 Components		6 Components	
		MCR-ALS	PARAFAC	MCR-ALS	PARAFAC	MCR-ALS	PARAFAC	MCR-ALS	PARAFAC
Component 1	em	60.6	62.5	64.0	68.5	57.2	63.2	62.4	79.7
	ex	63.2	60.9	59.1	53.1	67.3	60.1	61.1	36.5
Component 2	em	68.6	66.1	71.5	71.6	77.1	74.6	76.6	76.3
	ex	52.9	56.3	48.9	48.7	40.6	44.3	41.4	41.8
Component 3	em	84.8	86.4	88.1	84.2	88.8	86.8	87.6	87.1
	ex	28.1	25.3	22.4	29.1	21.1	24.7	23.2	24.1
Component 4	em			93.4	88.1	91.8	93.2	91.6	91.3
	ex			12.8	22.4	15.7	13.2	16.1	16.6
Component 5	em					95.2	93.6	94.9	94.1
	ex					9.4	12.5	10.0	11.5
Component 6	em							97.9	80.1
	ex							4.1	35.8
$lof^a$		6.4	6.0	5.7	5.1	4.9	4.3	5.0	3.9
$R^{2,b}$		99.6	99.6	99.7	99.7	99.8	99.8	99.8	99.9

<sup>a</sup> Percentage of lack of fit ( $lof$ ) considering all components (full model) together.

<sup>b</sup> Percentage of explained variances ( $R^2$ ) considering all components (full model) together.



**Fig. 4.** Mapping of contributions of the three main components resolved by MCR-ALS on the fresh water system under study. Fmax is the sum of the fluorescence maximum score for each component

groups increasing their presence too. From station 8th down, the contributions of this component increases sharply, until reaching their maximum at 17th station, where a WWTP (Roda de Ter) is located. Indeed, after the 18th station, the contribution of UVC has some little fluctuations. Only at the place near to Sau Dam there are some peaks which are affected by a new river tributary (Riera major).

### 3.2.2. Second resolved DOM contribution

A second MCR-ALS resolved component explains 33.73% of the total experimental data variance and it shows two excitation band maxima, one at 310 and another below 240 nm, with an emission very broad band with its maximum at 395 nm. These features are characteristic to those observed for the so called UVA fluorescence (see [Stedmon et al., 2003](#)). Its presence is attributed to the decom-

# Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

X. Zhang et al./Chemosphere 111 (2014) 120–128

127

position of organic matter generated within the river stream It can be associated to fulvic acid fluorophore groups present in most of natural environments (Stedmon and Markager, 2005a). Hydroxybenzoic, and salicylic chromophore are considered to be good models for this component (Senesi, 1990).

The contribution trend of UVA is very similar to UVC. The different part is that it goes up slowly after 40th station and that a new relative maximum contribution appears at the 44th station. The reason for this might be explained because at the 40th station, a tributary river (Riera de Sant Marti) inflows into the Ter river, and it introduces a new amount of organic matter into the river, increasing the contribution of UVA which reach their peak maximum at this station.

### 3.2.3. Third resolved DOM contribution

Fluorescence properties of the third MCR-ALS resolved component explained a lower amount of experimental data variance, only 12.27%, and it shows so called UVT spectral characteristics (Coble, 1996), with excitation bands around 280 nm and a single emission spectral band centered at 355 nm. These spectral properties resemble these of protein-like (tryptophan-like) fraction sources which induce fluorescence in different aquatic environments (Mopper and Schultz, 1993; Coble, 1996; Yamashita and Tanoue, 2003). It is from anthropogenic origin and it is also correlated to organic matter from industrial effluents and from untreated sewage. This type of chemical groups (protein and tryptophan-like) are produced by biological activity in surface waters derived from autochthonous DOM processes (Determann et al., 1994; Determann et al., 1996).

The contribution of UVT (see Fig. 4c) reflects a different DOM source which can be related mostly to organic matter caused by human (population) activities. Its contribution drops down slowly from the 1th station to the 8th station, and then it grows up sharply after the 8th station, i.e. with the increasing population activity (in Manlleu town). After this point, an increase of this contribution takes place when the river goes across this town and it decreases then significantly after the river flows out of the town. From stations 19–23, this contribution keeps being rather high (close to Roda de Ter WWTP). After station 24th, the rather slow water flowing makes this UVT contribution to have enough time for settlement and/or degradation and it decreases rapidly until station 41th. At stations 43th and 44th, the tributary river (Riera de Sant Marti) produced new changes in the contribution profile of this DOM source with a second peak maximum. And at station 48th this contribution reaches a new minimum which does not change then until station 55, where a new increase is produced as a consequence of a new river tributary (Riera del Balà). From these results, it may be concluded, that the DOM source explained by UVT is related mostly with human-activity, which has a variable contribution along the investigated water system.

### 3.2.4. Correlation of MCR-ALS resolved DOM contributions with independent experimental DOM total concentrations

The relative contributions calculated from MCR-ALS resolved profiles were compared with known values of DOC concentrations in every sample (obtained by the method introduced in 2.1.1). Correlation coefficients between 1st (UVC), 2nd (UVA), 3rd (UVT), 1st + 2nd, and 1st + 2nd + 3rd components were 0.79, 0.75, 0.23, 0.77 and 0.69 respectively. By *F*-test (Jamshidian et al., 2007), *p* values of these correlations were below 0.05, except for the 3rd component where a *p* value of 0.073 was obtained. By *t*-test (McDonald, 2009), these *p* values were also lower than 0.05 except for 3rd component with a *p* value of 0.079. Therefore, it is concluded that 1st and 2nd MCR-ALS resolved components correlate well with DOM, whereas 3rd component not. The explanation of this is because, as mentioned before, UVC and UVA resolved com-

ponents are linked to humic-like substances, stemming from the chemical and biological activity of microorganisms, whereas UVT is considered to be organic matter of anthropogenic origin, (Gone et al., 2009). These two type of matters did not show similar trends in the river system. Compared to UVC and UVA, UVT is at a lower concentrations, and it did not have significant effect on total DOC concentrations, although it provided different spectroscopic trends compared to natural (non-anthropogenic) DOC. So UVC and UVA showed a significant correlation with the DOC analyzed by DOC analyzer, whereas UVT did not.

In summary, fluorescence properties of the three main DOM components found in this study do agree with those encountered in previous studies on other freshwater and wastewater environmental studies (Coble, 1996; Chen et al., 2003; Cumberland and Baker, 2007), and also correlates well with independently experimentally estimated DOM concentration values. Resolved excitation and emission spectra showed characteristic patterns which correspond to organic chemical groups like hydroxy-benzoic (like in salicylic acid),  $\beta$ -naphthol, aromatic nitrogen rings, tryptophan, amino acid and proteins.

## 4. Conclusions

Large environmental monitoring studies of natural fresh water systems by excitation–emission matrix fluorescence spectra provide a powerful and convenient way to investigate DOM source contributions in watersheds. In this work three major DOM fluorescence contributions were resolved including a first fluorescence source assigned to a UVC humic type of fraction, a second fluorescence source assigned to a UVA humic/fulvic type of fraction, and a third fluorescence source assigned to a mixed UVT, protein-, and tryptophan-like type of fraction. Ortho hydroxy-benzoic (like in salicylic acid),  $\beta$ -naphthol, aromatic nitrogen rings, tryptophan and amino acid peptide groups present in freshwaters DOM are the responsible of the fluorescence properties of the investigated river system. Resolved geographical distribution of these three different DOM sources over the investigated area, are discussed, showing increasing and decreasing relative contributions according to different factors like river flow, DOM natural sources origin, human activities in the area and population in put sources. Relative contributions of the different resolved DOM sources correlate well with independent experimentally estimated DOM concentrations.

## References

- Mapping Toolbox. The MathWorks Inc., U.S.
- Google Earth Google Inc., USA.
- MATLAB. The MathWorks Inc., U.S.
- Alier, M., Felipe, M., Hernández, I., Tauler, R., 2011. Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O<sub>3</sub> pollution in Barcelona. *Anal. Bioanal. Chem.* 399, 2015–2029.
- Bro, R., 1997. PARAFAC: tutorial and applications. *Chemom. Intell. Lab. Syst.* 38, 149–171.
- Bro, R., Kiers, H.A., 2003. A new efficient method for determining the number of components in PARAFAC models. *J. Chemom.* 17, 274–286.
- Brown, S.D., Tauler, R., Walczak, B., 2009. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, 2.24 Extended Multivariate Curve Resolution*. Elsevier Science Limited.
- Chen, W., Westerhoff, P., Leenheer, J.A., Booksh, K., 2003. Fluorescence excitation–emission matrix regional integration to quantify spectra for dissolved organic matter. *Environ. Sci. Technol.* 37, 5701–5710.
- Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation–emission matrix spectroscopy. *Mar. Chem.* 51, 325–346.
- Coble, P.G., Del Castillo, C.E., Avril, B., 1998. Distribution and optical properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. *Deep-Sea Res. Part II* 45, 2195–2223.
- Coble, P.G., Green, S.A., Blough, N.V., Gagosian, R.B., 1990. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. *Nature* 348, 432–435.
- Cumberland, S.A., Baker, A., 2007. The freshwater dissolved organic matter fluorescence-total organic carbon relationship. *Hydrol. Processes* 21, 2093–2099.

## Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and Multivariate Curve Resolution

128

X. Zhang et al./Chemosphere 111 (2014) 120–128

- de Juan, A., Rutan, S.C., Tauler, R., Massart, D.L., 1998. Comparison between the direct trilinear decomposition and the multivariate curve resolution-alternating least squares methods for the resolution of three-way data sets. *Chemom. Intell. Lab. Syst.* 40, 19–32.
- de Juan, A., Tauler, R., 2001. Comparison of three-way resolution methods for non-trilinear chemical data sets. *J. Chemom.* 15, 749–771.
- Determann, S., Reuter, R., Wagner, P., Willkomm, R., 1994. Fluorescent matter in the eastern Atlantic Ocean. Part 1: method of measurement and near-surface distribution. *Deep Sea Res. Part I* 41, 659–675.
- Determann, S., Reuter, R., Willkomm, R., 1996. Fluorescent matter in the eastern Atlantic Ocean. Part 2: vertical profiles and relation to water masses. *Deep Sea Res. Part I* 43, 345–360.
- Gone, D.L., Seidel, J.-L., Batiot, C., Bamory, K., Ligban, R., Biemi, J., 2009. Using fluorescence spectroscopy EEM to evaluate the efficiency of organic matter removal during coagulation-flocculation of a tropical surface water (Agbo reservoir). *J. Hazard. Mater.* 172, 693–699.
- Her, N., Amy, G., McKnight, D., Sohn, J., Yoon, Y., 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. *Water Res.* 37, 4295–4303.
- Hudson, N., Baker, A., Reynolds, D., 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review. *River Res. Appl.* 23, 631–649.
- Imberger, J., Ivey, G., 1993. Boundary mixing in stratified reservoirs. *J. Fluid Mech.* 248, 477–491.
- Jamshidian, M., Jennrich, R.I., Liu, W., 2007. A study of partial F tests for multiple linear regression models. *Comput. Stat. Data Anal.* 51, 6269–6284.
- Jaumot, J., Gargallo, R., de Juan, A., Tauler, R., 2005. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.* 76, 101–110.
- Kowalczyk, P., Stoń-Egiert, J., Cooper, W.J., Whitehead, R.F., Durako, M.J., 2005. Characterization of chromophoric dissolved organic matter (CDOM) in the Baltic Sea by excitation emission matrix fluorescence spectroscopy. *Mar. Chem.* 96, 273–292.
- Leenheer, J.A., Croué, J.-P., 2003. Peer reviewed: characterizing aquatic dissolved organic matter. *Environ. Sci. Technol.* 37, 18A–26A.
- Lochmueller, C.H., Saavedra, S.S., 1986. Conformational changes in a soil fulvic acid measured by time-dependent fluorescence depolarization. *Anal. Chem.* 58, 1978–1981.
- McDonald, J.H., 2009. *Handbook of Biological Statistics*. Sparky House Publishing Baltimore, MD.
- Mopper, K., Schultz, C.A., 1993. Fluorescence as a possible tool for studying the nature and water column distribution of DOC components. *Mar. Chem.* 41, 229–238.
- Moran, M.A., Sheldon, W.M., Zepp, R.G., 2000. Carbon loss and optical property changes during long-term photochemical and biological degradation of estuarine dissolved organic matter. *Limnol. Oceanogr.* 45, 1254–1264.
- Senesi, N., 1990. Molecular and quantitative aspects of the chemistry of fulvic acid and its interactions with metal ions and organic chemicals: Part II. The fluorescence spectroscopy approach. *Anal. Chim. Acta* 232, 77–106.
- Sharma, A., Schulman, S.G., 1999. *Introduction to Fluorescence Spectroscopy*. Wiley, New York.
- Stedmon, C., Markager, S., 2001. The optics of chromophoric dissolved organic matter (CDOM) in the Greenland Sea: an algorithm for differentiation between marine and terrestrially derived organic matter. *Limnol. Oceanogr.* 2087–2093.
- Stedmon, C.A., Markager, S., 2005a. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.* 50, 686–697.
- Stedmon, C.A., Markager, S., 2005b. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.* 686–697.
- Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* 82, 239–254.
- Tauler, R., Marqués, I., Casassas, E., 1998. Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths. *J. Chemom.* 12, 55–75.
- Tauler, R., Smilde, A., Kowalski, B., 1995. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* 9, 31–58.
- Weissenberger, S., Lucotte, M., Houel, S., Soumis, N., Duchemin, É., Canuel, R., 2010. Modelling the carbon dynamics of the La Grande hydroelectric complex in northern Quebec. *Ecol. Model.* 221, 610–620.
- Yamashita, Y., Jaffé, R., Maie, N., Tanoue, E., 2008. Assessing the dynamics of dissolved organic matter (DOM) in coastal environments by excitation emission matrix fluorescence and parallel factor analysis (EEM-PARAFAC). *Limnol. Oceanogr.* 1900–1908.
- Yamashita, Y., Tanoue, E., 2003. Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. *Mar. Chem.* 82, 255–271.
- Yan, Y., Li, H., Myrick, M., 2000. Fluorescence fingerprint of waters: excitation-emission matrix spectroscopy as a tracking tool. *Appl. Spectrosc.* 54, 1539–1542.
- Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. *Chemosphere* 38, 45–50.

## Discussion

DOM sources have different fluorophore groups which can be identified and characterized. In this Thesis, Parallel Factor Analysis (PARAFAC) and Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) have been satisfactorily applied to the EEM data obtained from the Ter River-Reservoir system. The results obtained in this work can be useful for establishing analytical methodologies protocols for DOM monitoring in surface water systems.

According to PARAFAC and MCR-ALS results, three major DOM fluorescence sources (UVC, UVA, and UVT) were resolved. The first fluorescence source consisted of UVC humic derived group type of DOM fraction. Humic substances (HS) are natural macromolecular, heterogeneous substances with a high degree of polydispersity. They occur ubiquitously in soil, aquatic environment and certain sediments. They are known to contain similar functional groups i.e. C=C, COOH, OH, OCH<sub>3</sub>, C=O, NH or NH<sub>2</sub>, arrangements such as redox quinone-semiquinone-hydroquinone, charge-transfer planar complexes etc.<sup>315</sup>.

The second fluorescence source consisted of UVA humic/fulvic type of DOM fraction. Its presence is associated with organic matter generated within the stream by decomposition of organic matter. It can be considered to fulvic acid fluorophore groups present in all environments<sup>316</sup>.

And the third fluorescence source consisted of mixed contributions of UVT protein-like type of fraction, with tryptophan-groups. The agreement between the results (wavelength of UVC, UVA, and UVT) obtained in this work with those previously obtained in other studies (introduced in the first section of this chapter) confirmed the identification of different DOM sources.

Moreover their distribution over the monitoring area under study was also estimated. The concentration distributions have been illustrated on the map using MATLAB mapping toolbox (MATLAB R2013a. The MathWorks, Inc.). When the MCR-ALS resolved contributions are examined, it is observed that UVC and UVT contributions describe the increase of DOM concentrations at Manlleu (large town), Roda de Ter (WWTP) and Riera de Sant Marti (tributary river). After these three locations and along the river flow, DOM concentrations revealed by UVA showed a gradual decreasing trend. The main difference of UVA to the others is that after the station 23 (where river flows slower from this station than upstream), it decreased sharply and should be related mostly to organic matter caused by human (population) impact.

Hydroxy-benzoic acids (like in salicylic acid),  $\beta$ -naphthol, aromatic nitrogen rings, tryptophan, amino acids and proteins can be present in DOM of the fresh water systems and are responsible of most of the fluorescence properties investigated in this work. Main

trends of the organic matter processes taking place along the fresh water system were described in detail in the results section of paper<sup>307</sup>.

The results of this work show that EEM fluorescence spectra provide a powerful and convenient way to investigate DOM source contributions in watersheds. Major DOM source contributions in watershed can be obtained by means of EEM fluorescence spectra combined with chemometric methods, such as MCR-ALS with the trilinearity constraint or PARAFAC. Increasing and decreasing relative contributions of DOM in river can be related with factors like river flow, DOM natural sources origin, human activities in the area and population input sources along the river. MCR-ALS with the trilinear constraint is an effective way to characterize and resolve DOM sources in natural fresh water systems from EEM data.

Distribution of Dissolved Organic Matter in freshwaters using Excitation Emission fluorescence and  
Multivariate Curve Resolution

---

# **Chapter 7**

## **Conclusion**



### **Conclusions of the application of chemometric methods**

1) The results obtained in this Thesis show that MCR-ALS method can be successfully used for hyperspectral image resolution purposes. The spectra signatures of the pure constituents present in hyperspectral images and their concentration distribution at a pixel level can be estimated. Constituents identification can be performed using the resolved pure spectra signatures and comparing them to reference spectra from spectral libraries or from experimental spectra of reference samples.

2) Application of image data pretreatment methods such as Multiplicative Scatter Correction (MSC), Asymmetric Least Squares (AsLS) and Savitzky-Golay (SG) smoothing reduce significantly the presence of strong fluorescence background in Raman hyperspectral images. In contrast, infrared hyperspectral imaging is not affected by fluorescence. Kramers-Kronig transform enables to calculate absorption spectra in case only reflectance spectra can be measured for infrared spectra.

3) The extent of rotation ambiguity associated to MCR-ALS and other resolution methods (like VCA (Vertex Component Analysis) or MVSA (Minimum Volume Simplex Analysis)) can be rather high when they are applied for hyperspectral image resolution with high noise. The correct resolution of hyperspectral images can only be guaranteed if additional constraints are applied, such as those providing information about the local rank properties of the image, i.e. about the presence or absence of the different constituents (components) in the image pixels. Only in this way it is possible to increase the reliability of the solutions provided by MCR methods and decrease the uncertainties associated to them. Appropriate use of local rank and selectivity constraints can improve significantly the quality of the pure spectra (signatures) and of the constituent distribution maps resolved by MCR-ALS analysis of hyperspectral images in remote sensing studies.

4) Use of correlation coefficients between selected spectra and image pixel spectra is shown to provide an alternative way for the application of the selectivity constraint in hyperspectral images for the first time. This alternative method resulted to be satisfactory when pure pixels exist.

5) MCR-BANDS method can be used to get estimations of the extension of rotation ambiguities in MCR resolved results. The Area of Feasible Solutions (AFS) represents

feasible solutions geometrically. The range of rotation ambiguity calculated by MCR-BANDS and AFS are in agreement.

6) MCR-ALS with the trilinearity constraint is an effective way to characterize and resolve Excitation-Emission Matrix fluorescence spectra (EEM). Georeferenced mapping using MATLAB is an adequate method to visualize graphically and geographical distribution maps of the constituents resolved by MCR-ALS analysis of environmental monitoring data sets using fluorescence EEM data.

### **Conclusions from environmental and food samples analysis using hyperspectral imaging and Multivariate Curve Resolution methods**

1) In the analysis of remote sensing NIR hyperspectral images obtained from the region of Cuprite in Nevada, in the San Diego airport and in the Moffet field in California, MCR-ALS results showed clearly the presence of multiple objects such as airplanes, green plant areas, lakes, hard ground field and other related objects in the image. Decisions can be taken from the detailed information (distribution maps) contained in these images after examining them by MCR-ALS.

2) Micro-hyperspectral imaging coupled to MCR-ALS is confirmed to be an appropriate tool for the direct analysis of the constituents of chocolate samples, and by extension, it is proposed for the analysis of other mixture constituents in commercial food samples. Results achieved in this work are representative of the possibilities offered by hyperspectral imaging analysis of food samples, and confirmed the potential use of the MCR-ALS as a complementary tool method for this type of analysis. Application of this combination of methods (hyperspectral imaging and MCR-ALS) can be proposed for testing samples from other similar fields including the analysis of material surfaces, agriculture products as well as biological tissue analysis in medicine.

3) Environmental monitoring studies of natural fresh water systems by Excitation-Emission Matrix (EEM) fluorescence spectra provide a powerful and convenient way to investigate Dissolved Organic Matter (DOM) source contributions in watersheds. Major DOM fluorescence sources were resolved, and assigned to UVC humic type of fraction, UVA humic/fulvic type of fraction, UVT, protein-like, and tryptophan-like type of fraction. Resolved geographical distribution of these three different DOM sources over the investigated area showed increasing and decreasing relative contributions of these

## Conclusions

---

sources according to different factors like river flow, DOM natural sources origin, human activities in the area and population in input sources.



## **Resumen de la Tesis**



## Resumen

El análisis de imágenes hiperespectrales está relativamente avanzado dentro del análisis de alimentos y de medio ambiente. Las imágenes hiperespectrales permiten obtener un conocimiento preciso y fiable sobre la composición química y la distribución de los componentes químicos en la superficie de una muestra a investigar. Los resultados del análisis de imágenes hiperespectrales pueden utilizarse para adquirir una comprensión fundamental de los sistemas químicos complejos en investigación y desarrollo, con repercusiones comerciales, como métodos de validación de la adulteración en el análisis de alimentos y derivados, y en el análisis de los procesos industriales y de su control. El análisis de los datos de imágenes hiperespectrales son un gran reto debido a su tamaño y complejidad. En esta Tesis se propone el desarrollo y utilización de métodos quimiométricos que permitan revelar la información contenida en las imágenes espectrales analizadas tanto como sea posible.

Esta Tesis trata de la resolución de datos de imágenes hiperespectrales utilizando métodos quimiométricos, en particular mediante el uso de métodos de pretratamiento de datos y utilizando métodos de resolución multivariante de curvas (MCR). La principal contribución de la presente Tesis es el estudio y la aplicación del método MCR-ALS (resolución multivariante de curvas mediante mínimos cuadrados alternados) para la resolución de imágenes hiperespectrales, adquiridas mediante técnicas de teledetección (en el aire o en el espacio, transmitidas a la Tierra mediante instrumentos de observación) y mediante técnicas de micro-espectroscopia. Específicamente, en el trabajo de esta Tesis, se explora la combinación de los métodos quimiométricos y de los métodos de análisis de imágenes hiperespectrales, para la resolución de los espectros (firmas) y de los mapas de distribución de los componentes químicos de la muestra. El objetivo final de este estudio es mejorar el análisis y la interpretación de los datos de imágenes hiperespectrales mediante el aprovechamiento de diferentes herramientas quimiométricas poderosas. La detección del rango local y las propiedades de selectividad que describen la información espacial de los componentes presentes en las imágenes espectroscópicas se puede utilizar como restricción para aumentar significativamente el rendimiento del método MCR, disminuyendo las incertidumbres asociadas a las ambigüedades rotacionales. Se han comparado diferentes métodos de resolución, tales como MCR-ALS, ICA (Análisis de Componentes Independientes), MVSA (Mínimo Volumen Simplex Análisis), PCA (Análisis de Componentes Principales), y MCR-FMIN. Los métodos MCR-BANDS y FAC-PACK se han utilizado para la evaluación de la extensión de las ambigüedades rotacionales existentes en los resultados después de la aplicación de estos métodos de resolución multivariante.

En esta Tesis se han analizado diversos conjuntos de datos compuestos por varias imágenes hiperespectrales proporcionadas por instrumentos estándar tales como el espectrómetro de imágenes hiperespectrales en el visible y en el infrarrojo AVIRIS de la NASA, y diversos espectrómetros de imágenes hiperespectrales Raman y infrarrojo de laboratorio. La eficacia del procedimiento MCR-ALS se ilustra proporcionando comparaciones exhaustivas con otros métodos de resolución de mezclas espectrales a partir de conjuntos de datos hiperespectrales simulados y reales.

## **Objetivos generales de la Tesis**

El objetivo principal de esta Tesis ha sido el desarrollo y la aplicación de métodos quimiométricos para el análisis de imágenes hiperespectrales obtenidos a partir de técnicas de teledetección o de micro-espectroscopia de muestras de medio ambiente y de alimentos. Se han estudiado en detalle diferentes conjuntos de datos simulados y experimentales, obtenidos bien a partir de repositorios de datos de teledetección públicos, o a partir de medidas experimentales. Se han analizado y caracterizado diferentes conjuntos de datos simulados y experimentales obtenidos a partir de repositorios de datos de teledetección públicos y a partir de mediciones experimentales.

Especialmente importante ha sido la extensión y aplicación del método de MCR-ALS para la resolución de datos de imágenes hiperespectrales obtenidas mediante teledetección, con el objetivo de obtener los espectros y mapas de distribución de los constituyentes de las imágenes analizadas.

## **Objetivos en el análisis de imágenes hiperespectrales**

- Aplicar el método MCR-ALS al análisis de imágenes hiperespectrales a partir de conjuntos de datos simulados y de conjuntos de datos experimentales, que incluyen imágenes obtenidas a partir de procedimientos de teledetección del medio ambiente, y de imágenes obtenidas a partir de microespectroscopia de alimentos de muestras comerciales de chocolate. También se realiza la aplicación del método MCR-ALS al análisis fluorimétrico de muestras de aguas superficiales del río Ter y de sus embalses en su curso medio (embalses de Sau y Susqueda). En todos estos casos, se propone la aplicación del método MCR-ALS para la resolución de los espectros de los componentes de las mezclas analizadas, para su caracterización y estimación de sus correspondientes contribuciones y mapas de distribución.
- Aplicar el método MCR-ALS a las imágenes hiperespectrales obtenidas por teledetección, para resolver los diferentes objetos presentes en las imágenes estudiadas, tales como lagos, suelos, vegetación, edificios, etc., y la determinación de su ubicación en la imagen estudiada.

- Aplicar el método MCR-ALS al análisis hiperespectral de los componentes del chocolate para su resolución a nivel micro según su tamaño de partícula, aspecto importante relacionado con el control de la calidad del producto en el caso de estudio de alimentos.
- Aplicar el método MCR-ALS al análisis de los datos de fluorescencia para el seguimiento de las fuentes de materia orgánica (DOM) en el río Ter para la resolución de sus diferentes contribuciones, sus niveles de concentración, su distribución geográfica y la relación que existen entre ellas y la actividad humana a lo largo del cauce del río y de sus embalses.

### **Objetivos de los análisis quimiométrico**

- Desarrollar y aplicar el método MCR-ALS para el análisis de datos de imágenes hiperespectrales, para obtener los espectros puros y la distribución de los constituyentes de las imágenes.
- Adaptar el método de MCR-ALS con las restricciones de selectividad/rango locales para la mejora de los resultados de resolución. Utilización del procedimiento FSIW-EFA y del método del coeficiente de correlación para la aplicación de la restricción de selectividad en imágenes espectroscópicas.
- Aplicar los métodos de pretratamiento espectral para reducir la influencia de la dispersión (scattering) de luz en NIR, de la fluorescencia de fondo en la espectroscopia Raman cuando se irradia la muestra, y de la presencia de contribuciones de ruido de fondo y de cambios de su contribución (línea base).
- Aplicar el método MCR-ALS al análisis simultáneo de conjuntos múltiples de datos, dispuestos en matrices aumentadas o en estructuras multidireccionales (multiway), y aplicación de la restricción asociada a modelos trilineales.
- Discutir y comparar las diferentes formas de calcular la extensión de las ambigüedades rotacionales de los métodos MCR, como las que se proponen en los procedimientos MCR-BANDS y FAC-PACK, que permiten la evaluación de la calidad de los resultados obtenidos por MCR-ALS.

## Estructura de la Tesis

Esta Tesis se presenta en dos grandes apartados. En la primera parte introductoria se incluye la descripción de las técnicas de análisis de imágenes hiperespectrales, su aplicación en diferentes áreas, y los métodos quimiométricos utilizados para su análisis. En la segunda parte se presentan los resultados y su discusión de los resultados, en la que se incluyen los artículos científicos publicados a lo largo de la Tesis y las referencias bibliográficas consideradas. Estas dos grandes partes de la Tesis se subdividen en siete capítulos que describen a continuación de forma concisa.

- En el primer capítulo se presentan los objetivos de la Tesis. Además, se describe su estructura y se relaciona con el trabajo científico de este informe.
- En el segundo capítulo, se revisan las técnicas de imágenes hiperespectrales más recientes, su introducción, antecedentes y su aplicación en medio ambiente y en el área del análisis de alimentos. Se introduce la parte teórica de los métodos quimiométricos aplicados al análisis de imágenes hiperespectrales.
- En el tercer capítulo, se muestra la aplicación de los métodos de MCR-ALS a imágenes obtenidas por teledetección hiperespectral. Se describen los datos generados utilizando espectros de la biblioteca de la USGS y de datos de teledetección pública (AVIRIS, Airborne Visible / Infrared Imaging Spectrometer) de la NASA. Se muestra también el efecto del uso de las restricciones de rango local y de selectividad basadas en la información espacial de los componentes presentes en las imágenes espectroscópicas, para así aumentar el rendimiento de los métodos MCR y disminuir su ambigüedad.
- En el capítulo cuarto, se muestra la aplicación de la espectroscopia Raman y infrarroja de imágenes hiperespectrales combinadas con los métodos de pretratamiento y de resolución multivariante de curvas (MCR) con la restricción de selectividad para el análisis de los constituyentes de las muestras de chocolate comerciales estudiadas.
- En el quinto capítulo, se comparan diversos métodos quimiométricos de resolución que utilizan modelos bilineales de descripción de los datos. Se

discuten y comparan diversas formas de calcular la extensión de las ambigüedades de rotación.

- En el Capítulo sexto, se aplica el procedimiento MCR-ALS con la restricción de trilinealidad y se propone su utilización para datos de fluorescencia de excitación-emisión para el análisis de la materia orgánica disuelta (DOM) en sistemas naturales de agua dulce, y los resultados obtenidos se comparan con los obtenidos con PARAFAC.
- En el capítulo séptimo se presentan las conclusiones de esta Tesis.

## Introducción

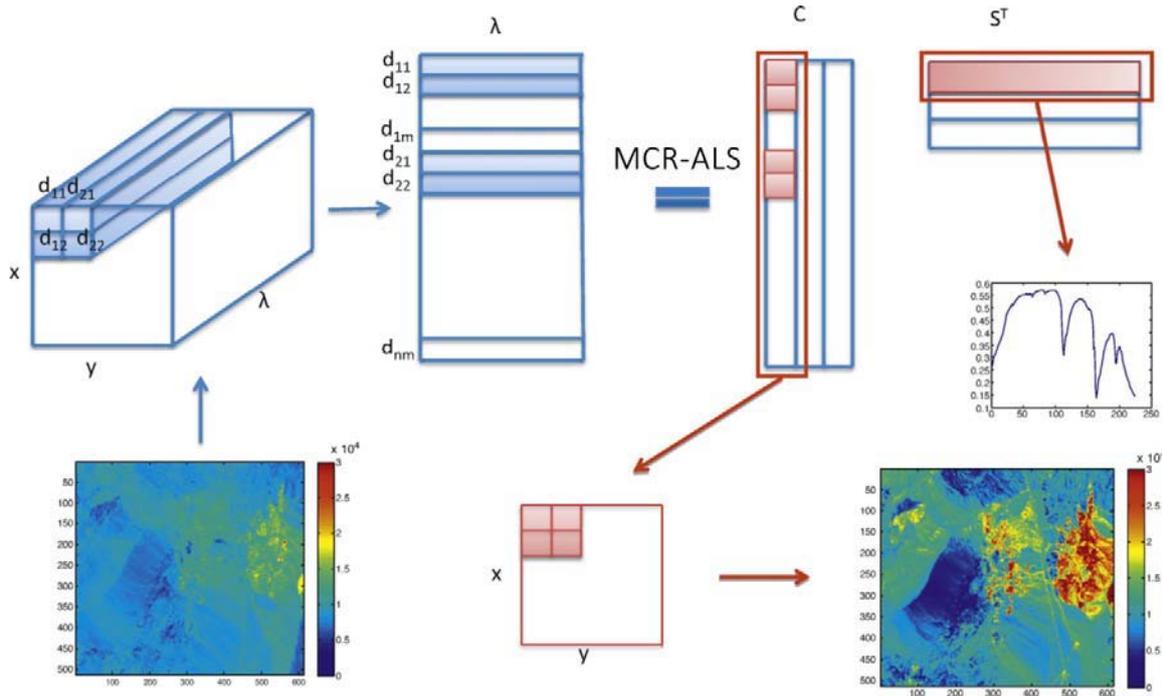
La utilización de la espectroscopia de imágenes representa un gran avance para la tecnología de teledetección<sup>317</sup>, que incluyen la utilización de las espectroscopias Raman, infrarroja y de fluorescencia<sup>318</sup>. Representa una metodología útil que puede aplicarse con fines analíticos en la agricultura<sup>10</sup>, la biología<sup>11</sup>, los estudios ambientales<sup>12</sup>, las ciencias de la tierra y en otros campos<sup>13</sup>. Se ha convertido en una disciplina muy importante en los últimos años<sup>89</sup>.

El análisis de imágenes hiperespectrales proporciona grandes conjuntos de datos en general, integrados en 100 a 200 bandas espectrales con anchuras de banda relativamente estrechas, de aproximadamente entre 5 y 10 nm. Las imágenes hiperespectrales se pueden representar en forma de cubos de datos. La información espacial se recoge en el plano X-Y, y la información espectral se representa en la dirección Z. El análisis de estos conjuntos de datos multidimensionales requiere ordenadores rápidos y grandes capacidades de almacenamiento, superiores a cientos de megabytes.

Para la investigación, desarrollo, utilización comercial, validación, y comprobación de adulteración, son necesarios métodos de referencia adecuados. Existen métodos de referencia para la seguridad alimentaria, pero el análisis y el control de calidad tienen a menudo limitaciones, debido a su adecuación a las diferentes etapas de la cadena alimentaria. Las diversas aplicaciones desarrolladas han demostrado claramente los beneficios de la utilización de la técnica de imágenes hiperespectrales para la caracterización de muestras y para la investigación de la distribución de las especies químicas presentes en las imágenes analizadas. El análisis de imágenes hiperespectrales ha demostrado ser una herramienta muy valiosa para una amplia gama de aplicaciones en el análisis de alimentos y en estudios de medio ambiente. La mayor capacidad de las tecnologías hiperespectrales en el seguimiento ambiental utilizando técnicas de teledetección permite a los gestores ambientales tomar decisiones correctas con el detalle necesario y en un marco de tiempo eficiente. Proporciona soluciones analíticas atractivas para satisfacer las demandas de los consumidores en relación a la calidad y inocuidad de los alimentos.

Debido al enorme tamaño y complejidad de los datos espectrales obtenidos en el análisis de imágenes hiperespectrales, los métodos quimiométricos son una buena opción para extraer la información acerca de la naturaleza, cantidad y ubicación de los componentes

en las muestras a partir del análisis de las imágenes medidas. El Análisis de Componentes Principales (PCA) <sup>319</sup> permite el procesamiento de conjuntos de datos de imágenes espectroscópicas y la reducción de sus dimensiones sin una pérdida significativa de la información. Sin embargo, debido al criterio de máxima varianza y de las restricciones ortogonales aplicadas, PCA no proporcionan directamente los espectros puros (verdaderos) ni los perfiles de concentración relativos de los constituyentes de imagen.



**Figure 8-1.** Descomposición de la matriz bilineal de una imagen hiperespectral utilizando el método de MCR-ALS. Resolución de los espectros (firmas) y de las concentraciones de imagen 2D de los componentes puros.

El método de resolución multivariante de curvas mediante mínimos cuadrados alternados (MCR-ALS) ha sido propuesto y utilizado ampliamente para resolver las respuestas instrumentales y concentraciones de los componentes presentes en mezclas desconocidas<sup>320</sup>. Se ha aplicado para analizar sistemas químicos multicomponentes tales como reacciones químicas<sup>321</sup>, procesos industriales<sup>322</sup>, problemas de coelución cromatográfica<sup>323</sup>, mezclas espectroscópicas<sup>324</sup>, datos de seguimiento ambiental<sup>325</sup>, y se puede aplicar a muchos otros tipos de casos y datos de análisis de mezclas. MCR-ALS ha sido propuesto también como método útil para la resolución y segmentación de imágenes hiperespectrales biomédicas y para otros tipos de imágenes hiperespectrales<sup>326-330</sup>. En MCR-ALS, se supone que las medidas (señales) analíticas siguen un modelo bilineal aditivo generalizado (tal como el de la extensión de la Ley de Lambert Beer en la

espectroscopia de absorción<sup>331</sup> para análisis multicomponente a múltiples longitudes de onda). La contribución de cada componente a la señal medida depende de su concentración y de su propia sensibilidad espectral (espectro puro). MCR-ALS también se puede aplicar para obtener información cuantitativa y proporciona soluciones con significado físico y químico (Figura 8-1). Esto se consigue porque en MCR-ALS, las restricciones que se aplican se cumplen de forma natural en las soluciones que se buscan, como son por ejemplo la no negatividad<sup>229, 332</sup>, la unimodalidad<sup>333, 334</sup>, el rango local<sup>335, 336</sup>, y la trilinealidad<sup>244, 333</sup>.

Sin embargo, aunque las soluciones MCR tienen un significado más físico y una interpretación más fácil que los obtenidos por PCA, no son soluciones únicas en general, y tienen un cierto grado de ambigüedad. La ambigüedad más crítica y difícil de evitar es la ambigüedad de rotación o rotacional. En este caso, un conjunto de soluciones diferentes (y combinaciones lineales de las mismas), se ajusta a los datos experimentales igualmente bien. Por lo tanto estas soluciones serán equivalentes desde un punto de vista matemático aunque serán completamente diferentes desde un punto de vista físico. Una manera para reducir este tipo de ambigüedad es por medio de la aplicación de más restricciones a las soluciones, pero esto debe realizarse manteniendo la plausibilidad física de las soluciones.

Además de las restricciones naturales, las estrategias más poderosas para evitar la presencia de las ambigüedades rotacionales en los métodos de MCR son el uso de la información del rango local y de la selectividad, la extensión MCR para el análisis simultáneo de múltiples conjuntos de datos y de datos multidireccionales, el uso de modelos basados en principios físico-químicos (determinista). Usando apropiadamente estas estrategias, se consigue la obtención de soluciones únicas como se ha demostrado en diversos trabajos anteriores.

Entre todas estas posibilidades, el uso de la restricción de tri-linealidad en el análisis simultáneo de múltiples conjuntos de datos es la mejor opción para el análisis de conjuntos de datos que cumplen con este tipo de modelo. Se estudia también a continuación la obtención de las mismas soluciones a partir del método PARAFAC.

Cuando se reconoce que las soluciones MCR pueden tener un cierto grado de ambigüedad, la pregunta es cómo se puede evaluar este grado de ambigüedad. Diferentes métodos han sido propuestos en la literatura para su evaluación, incluyendo el desarrollo de métodos para el cálculo de los límites de las denominadas bandas posibles o factibles y de sus áreas de existencia (area of feasible solutions, AFS).

Con el fin de evaluar la presencia de las ambigüedades de rotación asociados a una solución MCR particular y para medir sus efectos, se han propuesto diferentes métodos en los papeles<sup>249, 257-259</sup>.

## Resúmenes

### Aplicación de MCR-ALS en imágenes hiperespectrales teledetección

La teledetección es la adquisición de información acerca de un objeto o fenómeno, sin hacer contacto físico con el objeto. En su uso moderno, el término generalmente se refiere a la utilización de tecnologías de sensores para detectar y clasificar objetos sobre determinadas áreas o regiones de la Tierra. La detección remota mediante imágenes hiperespectrales es una combinación de ambas tecnologías, las imágenes hiperespectrales y la teledetección, que proporciona información analítica espectral y espacial sobre los objetos presentes en una zona geográfica determinada (Figura 8-2). Sobre la base de los espectros puros resueltos de los constituyentes de la imagen y de su distribución sobre la imagen, se puede obtener información adicional como la posible identificación de estos componentes a partir de las bases de espectros en bibliotecas, y su cuantificación relativa dentro de la imagen. Una de los aspectos de las imágenes hiperespectrales teledetección es que en cada píxel medido de la imagen, generalmente se encuentran presentes una mezcla de varios componentes espectrales. La resolución de la imagen hiperespectral requiere extraer los espectros puros de estos componentes y averiguar cómo se distribuyen los componentes correspondientes en la imagen.

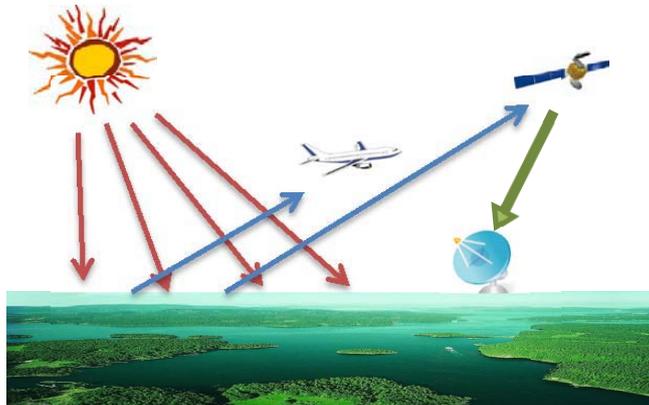


Figure 8- 2 Ilustración de la teledetección

La aplicación del método de MCR-ALS se demuestra en dos imágenes espectroscópicas de teledetección generadas por simulación y de una imagen obtenida a partir del

espectrómetro de teledetección de imágenes hiperespectrales en el infrarrojo AVIRIS de la NASA. Por aplicación de MCR-ALS, se estiman los espectros puros de los componentes presentes en la imagen y su distribución a nivel de píxel de la imagen. Los resultados obtenidos por MCR-ALS se comparan con los obtenidos por otros métodos utilizados con frecuencia en el estudio de imágenes obtenidas por espectroscopia infrarroja de teledetección como son VCA y MVSA.

En el caso del análisis de los datos experimentales, los espectros puros resueltos se compararon con espectros de referencia de la biblioteca del USGS para su identificación. En todos los casos, también se evaluaron los resultados sobre la posible presencia de ambigüedades rotacionales utilizando el método de MCR-BANDS.

Los resultados obtenidos confirmaron que el método de MCR-ALS puede ser utilizado con éxito para fines de resolución de imágenes hiperespectrales de teledetección. Sin embargo, la presencia de ambigüedad rotacional en las soluciones obtenidas por los diferentes métodos de resolución (como VCA o MVSA) puede ser en algunos casos grande y debe ser evaluados con cuidado, tratando de reducir sus efectos mediante la selección de las restricciones más apropiadas. Sólo de esta manera es posible aumentar la fiabilidad de las soluciones aportadas por estos métodos y disminuir las incertidumbres asociadas a su uso.

Los dos artículos publicados en el presente capítulo son:

Zhang, X.; Tauler, R., Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. *Analytica chimica acta*. 2013, 762, 25-38.

Zhang, X.; Juan, A.; Tauler, R., Local rank-based spatial information for improvement of remote sensing hyperspectral imaging resolution. Submitted to *Talanta*.

## Aplicación de imágenes hiperespectrales combinado con quimiometría en el análisis de alimentos

El chocolate es un producto común en la industria alimentaria, cuyo análisis es lento y tedioso. Durante el fabricación de chocolate, refinación y coaching, propiedades como el tamaño de las partículas, la consistencia y la viscosidad de la suspensión, dan la textura específica y las cualidades sensoriales del producto final. El control de la distribución de las partículas de grasa (y su tamaño de partícula) puede afectar a las propiedades reológicas y a la textura del chocolate, y también puede tener un impacto en el contenido final de grasa en el chocolate, cuya absorción en el cuerpo de los consumidores puede tener efectos sobre su salud.

La detección de las partículas constituyentes del chocolate y la estimación de su tamaño puede obtenerse a partir de las imágenes hiperespectrales, y utilizarse para el control de la calidad chocolate.



**Figura 8-3** Instrumentos para recolección de imágenes hiperespectrales utilizados en el análisis experimental de muestras de chocolate (HORIBA JobinYvon Raman microscope (HORIBA, Ltd.), Nicolet iN10 MX Infrared Imaging Microscope)

En este capítulo se muestra la aplicación de las hiperespectroscopias Raman e infrarroja (Figura 8-3) combinadas con el método de resolución multivariante de curvas (MCR) en el análisis de los constituyentes de muestras de chocolate comerciales. La combinación de diferentes métodos de pretratamiento datos espectrales ha permitido la disminución de la contribución fluorescente elevada de la señal Raman en las muestras de chocolate

investigados. A partir del uso de las restricciones de igualdad (tales como de rango local y de selectividad) durante el análisis de MCR, se mejoraron apreciablemente las estimaciones de los espectros puros de los constituyentes de la muestra de chocolate, así como sus contribuciones relativas y su distribución espacial en las muestras analizadas. Además, componentes desconocidos de la muestra podían ser también resueltos. La distribución de los constituyentes del chocolate blanco resueltos a partir de imágenes hiperespectrales Raman indican que a escala macro, la sacarosa, la lactosa, la grasa y son constituyentes que fueron mezclados en forma de partículas.

Las imágenes hiperespectrales infrarrojas no sufren de fluorescencia y se podían utilizar para el análisis de muestras de chocolate blanco y de chocolate con leche. Como conclusión de este estudio, las imágenes micro-hiperespectrales analizadas con el método de MCR se confirman como una herramienta adecuada para el análisis directo de los constituyentes de muestras de chocolate y, por extensión, se propone el análisis de otros componentes de mezclas de alimentos en productos comerciales.

El artículo publicado en el presente capítulo es:

Zhang, X.; Juan, A.; Tauler, R., Multivariate Curve Resolution applied to hyperspectral imaging analysis of chocolate samples. *Applied Spectroscopy*. 2015, 69(8).

## **Medir y comparar el rendimiento de resolución y la extensión de las ambigüedades de rotación en los métodos de modelado bilineales**

Los métodos basados en modelos bilineales se utilizan frecuentemente en el análisis de conjuntos de datos de espectroscopia y de cromatografía. Cuando se aplican estos métodos de modelización bilineales, las ambigüedades rotacionales se encuentran presentes y deben ser consideradas. En este trabajo, se examinan los resultados obtenidos por la aplicación de diferentes métodos, tales como el análisis de componentes independientes (Independent Component Analysis, ICA), el análisis de componentes principales (PCA), y el Minimum Volum Simplex Analysis (MVSA), y se comparan con los obtenidos por el método de resolución multivariante de curvas (MCR). Con el fin de hacer esta comparación, se utilizan diversos parámetros tales como el que mide la información mutua (Mutual Information, MI) el índice de Amari (AI) y el de falta de ajuste (lack of fit, lof) para la evaluación de los resultados obtenidos por los diferentes métodos. Las áreas o regiones de soluciones posibles (AFS) y sus límites son evaluados por diferentes métodos que investigan el problema ambigüedad rotacional a partir de diferentes conceptos. Los resultados obtenidos por el método de MCR-BANDS en el cálculo de la extensión de las ambigüedades rotacionales se analizan y se comparan con los recientemente propuestos para la estimación de toda la gama de soluciones posibles y para la definición del área de soluciones posibles (AFS).

Los métodos basados en modelos de descomposición bilineal, como por ejemplo, MCR-ALS, MCR-FMIN, ICA, PCA, o MVSA (todos ellos utilizados en este trabajo) son métodos adecuados para la resolución multivariante de curvas. Pero no todos ellos realizan la descomposición de la misma manera. Cuando estos métodos se aplican a datos en dos direcciones (two-way) y no existen condiciones de resolución única (por ejemplo de rango local o de selectividad), no hay forma en general de saber si estos métodos proporcionan la solución verdadera. Diferentes métodos de resolución darán una de las soluciones posibles equivalentes para el problema en estudio.

Los resultados obtenidos a partir de todos estos métodos muestran que es posible obtener un rango de soluciones posibles que pueden describir y ajustar los datos igualmente bien, sin dejar de cumplir el modelo bilineal y las restricciones del sistema estudiado. No es posible por lo tanto decir que un método es mejor que otro a partir de los resultados obtenidos por los diferentes métodos. Todos ellos pueden ser considerados como una herramienta alternativa para la resolución de las señales espectroscópicas

(espectros puros) en ciertos casos. Sin embargo, la aplicación de restricciones físicas como son la no-negatividad en los perfiles espectrales y de concentración es más flexible y fiable en métodos de MCR como el MCR-ALS.

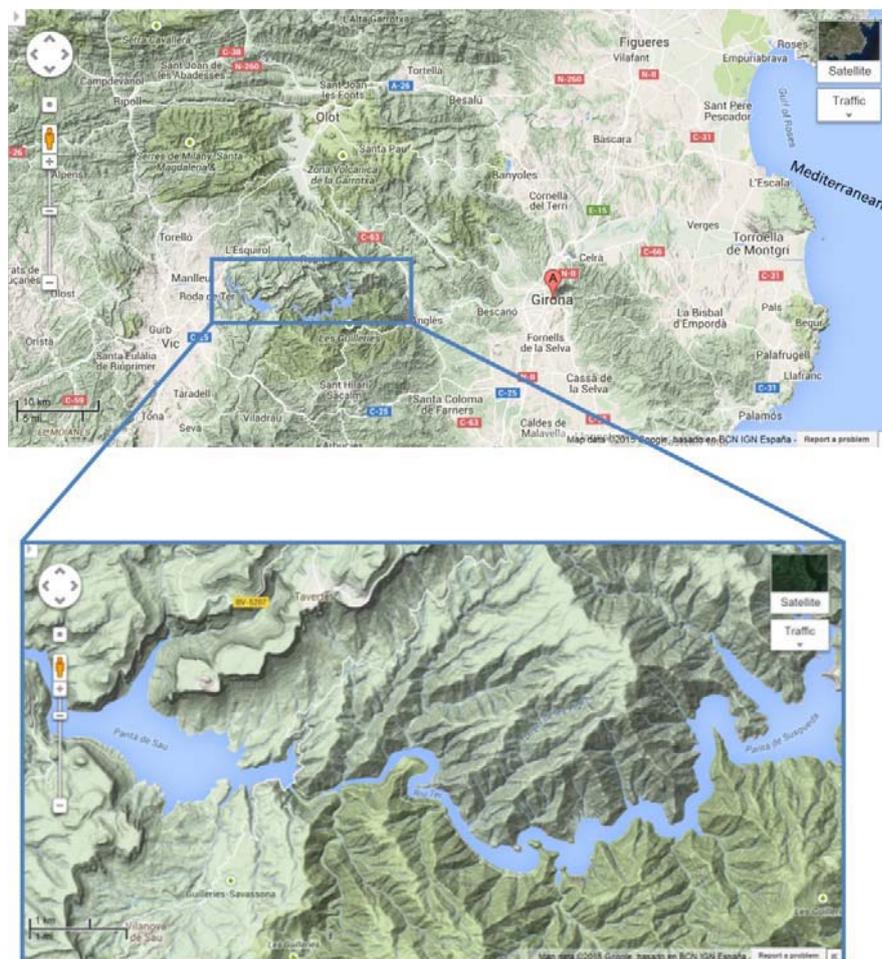
El método MCR-BANDS ofrece una primera aproximación al cálculo de la región o área de todas las soluciones MCR posibles, AFS. La representación geométrica de todas las soluciones posibles puede realizarse a partir de métodos como FAC-PACK para un número limitado de componentes. Es interesante remarcar que la extensión de las ambigüedades rotacionales calculada por el método MCR-BANDS y las regiones AFS calculados por un método como FAC-PACK son concordantes.

El artículo en el presente capítulo es:

X. Zhang, R. Tauler. Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods. Submitted to Journal of Chemometrics.

## Distribución de materia orgánica disuelta en las aguas dulces utilizando fluorescencia de excitación emisión y Multivariante Curve Resolución

Se propone el método la resolución multivariante de curvas mediante mínimos cuadrados alternados (MCR-ALS) con la restricción de trilinealidad para el análisis de los datos de excitación-emisión de fluorescencia causados por la materia orgánica disuelta (DOM) en los sistemas naturales de agua dulce (Figura 8-4 i Figura 8-5), y los resultados obtenidos se comparan con los obtenidos con PARAFAC.



**Figure 8- 4** La parte superior es el mapa de la cuenca del río Ter y la parte inferior corresponde a los principales embalses de Susqueda y Sau (La figura se obtuvo de mapa de Google).

Los efectos del diferente número de componentes y las limitaciones sobre la estabilidad de los modelos propuestos se comparan. Se demuestra que MCR-ALS es una manera eficaz para caracterizar y resolver las fuentes de DOM en los sistemas naturales de agua dulce a partir de datos de EEM, con buena correlación con los valores de concentración de DOM medidos experimentalmente.



**Figure 8- 5** La ubicación de la cuenca del río Ter en Catalunya, España.

La utilidad de mapas (mapping) georeferenciados de MATLAB se utiliza para ilustrar la distribución geográfica de las contribuciones DOM resueltos. Los espectros resueltos por MCR-ALS se utilizan para reconocer los grupos químicos correspondientes a las posibles fuentes de DOM. Las relaciones entre las actividades humanas y la situación ambiental del sistema fluvial se discuten a partir de estas posibles fuentes de DOM.

El artículo publicado en el presente capítulo es:

ZHANG, Xin., Marcé, R., Armengol, J., & Tauler, R. Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution. *Chemosphere*, 2014, 111, 120-128

## Conclusiones

### Conclusiones de la aplicación de métodos quimiométricos

1) Los resultados obtenidos en esta Tesis muestran que el método de MCR-ALS puede ser utilizado con éxito para fines de resolución de imágenes hiperespectrales. Los espectros puros de los constituyentes presentes en las imágenes hiperespectrales y la distribución de sus concentraciones a nivel de píxel pueden estimarse.

2) La aplicación de los métodos de pretratamiento de datos de imágenes como por ejemplo el método Multiplicative Scatter Correction (MSC), el método Asymmetric Least Squares (AsLS) y el método de suavización Savitzky-Golay (SG), reduce significativamente la presencia de algunos inconvenientes como son la fuerte fluorescencia de fondo en las imágenes hiperespectrales Raman. En contraste, las imágenes hiperespectrales de infrarrojos no se ven afectadas por la fluorescencia. El método de transformación de Kramers-Kronig permite calcular los espectros de absorción en caso de que solamente se puedan medir los espectros de reflectancia infrarroja.

3) El grado de ambigüedad rotacional asociada al procedimiento MCR-ALS y a otros métodos de resolución (como VCA (Vertex Component Analysis) o MVSA (Minimal Volume Simplex Analysis) puede ser bastante alta cuando se aplican a la resolución de imágenes hiperespectrales que tienen un elevado grado de ruido experimental. La resolución correcta de las imágenes hiperespectrales sólo puede garantizarse si se aplican restricciones adicionales, tales como las que proporcionan información sobre el rango local de las imágenes, es decir sobre la presencia o ausencia de los diferentes constituyentes (componentes) en los píxeles de la imagen. Sólo de esta manera es posible aumentar la fiabilidad de las soluciones proporcionadas por los métodos MCR y disminuir las incertidumbres asociadas a ellos. El uso apropiado del rango local y de la información de selectividad puede mejorar significativamente la calidad de los espectros puros (firmas) y de los mapas de distribución de los constituyentes resueltos por el análisis MCR-ALS de imágenes hiperespectrales en los estudios de teledetección.

4) Se muestra por primera vez el uso de los coeficientes de correlación entre espectros seleccionados y los espectros de píxeles de las imágenes para proporcionar una forma alternativa para la aplicación de la restricción de selectividad en imágenes hiperespectrales. Este método alternativo resultó ser satisfactorio cuando existen píxeles puros.

5) El método Area of Feasible Solutions (AFS) representa geoméricamente el conjunto de soluciones posibles. La extensión de la ambigüedad rotacional calculada por MCR-BANDS y por AFS están de acuerdo.

6) MCR-ALS con la restricción de trilinealidad es una manera eficaz para caracterizar y resolver los componentes fluorescentes de mezclas a partir de datos de matriz de Excitación-Emisión (EEM). El procedimiento de Mapping georeferenciado utilizando MATLAB es un método de visualización muy útil para describir de forma concisa la distribución geográfica de las fuentes de materia orgánica disuelta (DOM) a partir de los resultados MCR-ALS aplicados a datos EEM.

### **Conclusiones de análisis de muestras de alimentos y ambiental utilizando imágenes hiperespectrales y los métodos de Multivariante Curve Resolución**

1) En el análisis de imágenes hiperespectrales obtenidas a partir de procedimientos de teledetección procedentes de la región de Cuprita en Nevada, del aeropuerto de San Diego y del campo Moffet en California, los resultados MCR-ALS muestran claramente la presencia de varios objetos como aviones, zonas de plantas verdes, lagos, campos de tierra dura y otros objetos relacionados con la imagen. El examen de los resultados obtenidos a partir del análisis MCR-ALS de datos de teledetección de imágenes hiperespectrales sobre mapas de distribución permite la toma de decisiones.

2) El acoplamiento del análisis micro-hiperespectral de imágenes y del análisis MCR-ALS se confirma ser un instrumento adecuado para el análisis directo de los constituyentes de muestras de chocolate, y, por extensión, se propone también para el análisis de otros componentes de mezclas en muestras de alimentos comerciales. Los resultados obtenidos

en este trabajo son representativos de las posibilidades que ofrece el análisis de imágenes hiperespectrales de muestras de alimentos, y ha confirmado el uso potencial del procedimiento MCR-ALS como herramienta complementaria para este tipo de análisis. La aplicación de esta combinación de métodos (de análisis de imágenes hiperespectrales y de análisis MCR-ALS) se puede proponer para el análisis de muestras en otros campos similares, incluyendo el análisis de superficies de materiales, de productos de la agricultura, así como el análisis de tejidos biológicos en medicina.

3) Los espectros de fluorescencia (EEM) proporciona una forma poderosa y conveniente para la investigación de las fuentes de materia orgánica disuelta (DOM) en las cuencas hidrográficas. Las fuentes de materia orgánica disuelta (DOM) fluorescente resueltas y asignadas son la fracción UVC húmica, la fracción UVA húmica/fúlvica, la fracción UVT, y la fracción de proteínas y triptófano. La distribución geográfica resuelta de estas tres fuentes DOM diferentes a lo largo del área investigada en este trabajo (cuenca del río Ter), muestra aumentos y disminuciones de las contribuciones relativas de DOM de fuentes naturales de acuerdo con diferentes factores tales como el flujo del río, las actividades humanas en la zona y su población.



## References

1. G. ElMasry, D.-W. Sun. "Principles of hyperspectral imaging technology". *Hyperspectral imaging for food quality analysis and control* 2010. 3-43.
2. A. Gowen, C. O'Donnell, P. Cullen, G. Downey, J. Frias. "Hyperspectral imaging—an emerging process analytical tool for food quality and safety control". *Trends Food Sci. Technol.* 2007. 18(12): 590-598.
3. D.-W. Sun. "Hyperspectral imaging for food quality analysis and control". Elsevier: 2010.
4. A. O'SULLIVAN, B. O'CONNOR, A. Kelly, M.J. McGRATH. "The use of chemical and infrared methods for analysis of milk and dairy products". *Int. J. Dairy Technol.* 1999. 52(4): 139-148.
5. N.B. Colthup, L.H. Daly, S.E. Wiberley. "Introduction to infrared and Raman spectroscopy". Elsevier: 1990.
6. A.F.H. Goetz. "Three decades of hyperspectral remote sensing of the Earth: A personal view". *Remote Sens. Environ.* 2009. 113, Supplement 1(0): S5-S16.
7. H.J. Kramer. "Observation of the Earth and its Environment: Survey of Missions and Sensors". Springer Science & Business Media: 2002.
8. A.F. Goetz, B.N. Rock, L.C. Rowan. "Remote sensing for exploration; an overview". *Economic Geology* 1983. 78(4): 573-590.
9. R.O. Green, M.L. Eastwood, C.M. Sarture, T.G. Chrien, M. Aronsson, B.J. Chippendale, J.A. Faust, B.E. Pavri, C.J. Chovit, M. Solis. "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)". *Remote Sens. Environ.* 1998. 65(3): 227-248.
10. A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias. "Hyperspectral imaging – an emerging process analytical tool for food quality and safety control". *Trends in Food Science & Technology* 2007. 18(12): 590-598.
11. R.A. Schultz, T. Nielsen, J.R. Zavaleta, R. Ruch, R. Wyatt, H.R. Garner. "Hyperspectral imaging: A novel approach for microscopic analysis". *Cytometry* 2001. 43(4): 239-247.
12. A.F.H. Goetz, B. Curtiss. "Hyperspectral imaging of the earth: remote analytical chemistry in an uncontrolled environment". *Field Analytical Chemistry & Technology* 1996. 1(2): 67-76.
13. F.D. van der Meer, H.M.A. van der Werff, F.J.A. van Ruitenbeek, C.A. Hecker, W.H. Bakker, M.F. Noomen, M. van der Meijde, E.J.M. Carranza, J.B.d. Smeth, T. Woldai. "Multi- and hyperspectral geologic remote sensing: A review". *International Journal of Applied Earth Observation and Geoinformation* 2012. 14(1): 112-128.
14. A. Buades, B. Coll, J.-M. Morel. "A review of image denoising algorithms, with a new one". *Multiscale Modeling & Simulation* 2005. 4(2): 490-530.
15. G. Lu, B. Fei. "Medical hyperspectral imaging: a review". *Journal of Biomedical Optics* 2014. 19(1): 010901-010901.
16. Y. Garini, I.T. Young, G. McNamara. "Spectral imaging: Principles and applications". *Cytometry Part A* 2006. 69A(8): 735-747.
17. M.E. Hankus, H. Li, G.J. Gibson, B.M. Cullum. "Surface-enhanced Raman scattering-based nanoprobe for high-resolution, non-scanning chemical imaging". *Anal. Chem.* 2006. 78(21): 7535-7546.

## References

---

18. M.R. Descour, C.E. Volin, E.L. Dereniak, K.J. Thome, A. Schumacher, D.W. Wilson, P.D. Maker. "Demonstration of a high-speed non-scanning imaging spectrometer". *Opt. Lett.* 1997. 22(16): 1271-1273.
19. M.D. Schaeberle, H.R. Morris, J.F.T. Li, P.J. Treado. "Peer Reviewed: Raman Chemical Imaging Spectroscopy". *Analytical Chemistry* 1999. 71(5): 175A-181A.
20. E. Lewis. "Education for sustainability at a primary school: From silos to systems thinking". *Environmental Education Research* 2014. 20(3): 432-433.
21. T. Awad, H. Moharram, O. Shaltout, D. Asker, M. Youssef. "Applications of ultrasound in analysis, processing and quality control of food: A review". *Food Res. Int.* 2012. 48(2): 410-427.
22. P. Mandal, A. Biswas, K. Choi, U. Pal. "Methods for rapid detection of foodborne pathogens: an overview". *American Journal Of Food Technology* 2011. 6(2): 87-102.
23. T. Kourti. "Application of latent variable methods to process control and multivariate statistical process control in industry". *International Journal of Adaptive Control and Signal Processing* 2005. 19(4): 213-246.
24. J.A. Fernández Pierna, P. Vermeulen, O. Amand, A. Tossens, P. Dardenne, V. Baeten. "NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed". *Chemometrics Intellig. Lab. Syst.* 2012. 117(0): 233-239.
25. H. Kataoka, H.L. Lord, J. Pawliszyn. "Applications of solid-phase microextraction in food analysis". *J. Chromatogr. A* 2000. 880(1): 35-62.
26. J. Rosén, K.-E. Hellenäs. "Analysis of acrylamide in cooked foods by liquid chromatography tandem mass spectrometry". *Analyst* 2002. 127(7): 880-882.
27. J.H. Clark. "Green chemistry: challenges and opportunities". *Green Chemistry* 1999. 1(1): 1-8.
28. G.M. Strasburg, R.D. Ludescher. "Theory and applications of fluorescence spectroscopy in food research". *Trends Food Sci. Technol.* 1995. 6(3): 69-75.
29. P. Williams, K. Norris. "Near-infrared technology in the agricultural and food industries". *American Association of Cereal Chemists, Inc.*: 1987.
30. Y.-W. Wu, S.-Q. Sun, Q. Zhou, H.-W. Leung. "Fourier transform mid-infrared (MIR) and near-infrared (NIR) spectroscopy for rapid quality assessment of Chinese medicine preparation Honghua Oil". *J. Pharm. Biomed. Anal.* 2008. 46(3): 498-504.
31. R. Kizil, J. Irudayaraj, D. Sun. "Spectroscopic technique: Fourier transform Raman (FT Raman) spectroscopy". *Modern techniques for food authentication* 2008. 185-203.
32. M. Blanco, I. Villarroya. "NIR spectroscopy: a rapid-response analytical tool". *TrAC, Trends Anal. Chem.* 2002. 21(4): 240-250.
33. B.G. Osborne. "Near - infrared spectroscopy in food analysis". *Encyclopedia of analytical Chemistry* 2000.
34. J.F. Pierna, P. Vermeulen, O. Amand, A. Tossens, P. Dardenne, V. Baeten. "NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed". *Chemometrics Intellig. Lab. Syst.* 2012. 117: 233-239.
35. M. Delhaye, P. Dhamelinourt. "Raman microprobe and microscope with laser excitation". *Journal of Raman Spectroscopy* 1975. 3(1): 33-43.
36. I.W. Levin, R. Bhargava. "FOURIER TRANSFORM INFRARED VIBRATIONAL SPECTROSCOPIC IMAGING: Integrating Microscopy and Molecular Recognition\*". *Annu. Rev. Phys. Chem.* 2005. 56: 429-474.
37. L. Opilik, T. Schmid, R. Zenobi. "Modern Raman imaging: vibrational spectroscopy on the micrometer and nanometer scales". *Analytical Chemistry* 2013. 6.

## References

---

38. Y. Jung, J. Tam, H.R. Jalian, R.R. Anderson, C.L. Evans. "Longitudinal, 3D In Vivo Imaging of Sebaceous Glands by Coherent Anti-Stokes Raman Scattering Microscopy: Normal Function and Response to Cryotherapy". *Journal of Investigative Dermatology* 2014.
39. N.L. Garrett, L. Godfrey, A. Lalatsa, D.R. Serrano, I.F. Uchegbu, A. Schatzlein, J. Moger. In "Detecting polymeric nanoparticles with coherent anti-stokes Raman scattering microscopy in tissues exhibiting fixative-induced autofluorescence", 2015; 2015; Pp 932922-932922-8.
40. C.W. Freudiger, W. Min, B.G. Saar, S. Lu, G.R. Holtom, C. He, J.C. Tsai, J.X. Kang, X.S. Xie. "Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy". *Science* 2008. 322(5909): 1857-1861.
41. M.N. Slipchenko, T.T. Le, H. Chen, J.-X. Cheng. "High-speed vibrational imaging and spectral analysis of lipid bodies by compound Raman microscopy". *The Journal of Physical Chemistry B* 2009. 113(21): 7681-7686.
42. D. Fu, G. Holtom, C. Freudiger, X. Zhang, X.S. Xie. "Hyperspectral imaging with stimulated Raman scattering by chirped femtosecond lasers". *The Journal of Physical Chemistry B* 2013. 117(16): 4634-4640.
43. H. Yuan, A.M. Fales, C.G. Khoury, J. Liu, T. Vo - Dinh. "Spectral characterization and intracellular detection of Surface - Enhanced Raman Scattering (SERS) - encoded plasmonic gold nanostars". *Journal of Raman Spectroscopy* 2013. 44(2): 234-239.
44. S. Lee, H. Chon, J. Lee, J. Ko, B.H. Chung, D.W. Lim, J. Choo. "Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging". *Biosens. Bioelectron.* 2014. 51: 238-243.
45. M. Vendrell, K.K. Maiti, K. Dhaliwal, Y.-T. Chang. "Surface-enhanced Raman scattering in cancer detection and imaging". *Trends Biotechnol.* 2013. 31(4): 249-257.
46. H. Kneipp, N. Møbjerg, A. Jørgensen, H.G. Bohr, C. Hélix - Nielsen, J. Kneipp, K. Kneipp. "Surface enhanced Raman scattering on Tardigrada - towards monitoring and imaging molecular structures in live cryptobiotic organisms". *journal of biophotonics* 2013. 6(10): 759-764.
47. C.H. Moran, M. Rycenga, X. Xia, C.M. Cobley, Y. Xia. "Using well-defined Ag nanocubes as substrates to quantify the spatial resolution and penetration depth of surface-enhanced Raman scattering imaging". *Nanotechnology* 2014. 25(1): 014007.
48. E.L. Izake, B. Cletus, W. Olds, S. Sundarajoo, P.M. Fredericks, E. Jaatinen. "Deep Raman spectroscopy for the non-invasive standoff detection of concealed chemical threat agents". *Talanta* 2012. 94(0): 342-347.
49. B. Zachhuber, H. Östmark, T. Carlsson. In "Spatially offset hyperspectral stand-off Raman imaging for explosive detection inside containers", *SPIE Defense+ Security*, 2014; International Society for Optics and Photonics: 2014; Pp 90730J-90730J-5.
50. S. Wallin, A. Pettersson, H. Önnnerud, H. Östmark, M. Nordberg, E. Ceco, A. Ehlerding, I. Johansson, P. Käck. In "Possibilities for standoff Raman detection applications for explosives", 2012; 2012; Pp 83580P-83580P-9.
51. W.J. Olds, E. Jaatinen, P. Fredericks, B. Cletus, H. Panayiotou, E.L. Izake. "Spatially offset Raman spectroscopy (SORS) for the analysis and detection of packaged pharmaceuticals and concealed drugs". *Forensic Sci. Int.* 2011. 212(1-3): 69-77.
52. X. Zhang, A.d. Juan, R. Tauler. "Multivariate Curve Resolution applied to hyperspectral imaging analysis of chocolate samples". *Appl. Spectrosc.* 2015.
53. N. Jamin, P. Dumas, J. Moncuit, W.-H. Fridman, J.-L. Teillaud, G.L. Carr, G.P. Williams. "Highly resolved chemical imaging of living cells by using synchrotron infrared microspectrometry". *Proceedings of the National Academy of Sciences* 1998. 95(9): 4837-4840.

## References

---

54. L. Kidder, I. Levin, E.N. Lewis, V. Kleiman, E. Heilweil. "Mercury cadmium telluride focal-plane array detection for mid-infrared Fourier-transform spectroscopic imaging". *Opt. Lett.* 1997. 22(10): 742-744.
55. L.M. Kehlet, P. Tidemand-Lichtenberg, J.S. Dam, C. Pedersen. "Infrared upconversion hyperspectral imaging". *Opt. Lett.* 2015. 40(6): 938-941.
56. K. Yeh, S. Kenkel, J.-N. Liu, R. Bhargava. "Fast Infrared Chemical Imaging with a Quantum Cascade Laser". *Anal. Chem.* 2015. 87(1): 485-493.
57. M. Manley. "Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials". *Chem. Soc. Rev.* 2014. 43(24): 8200-8214.
58. P.R. Griffiths. "Infrared and Raman instrumentation for mapping and imaging". *Infrared and Raman spectroscopic imaging 2009.* 3-64.
59. G. Tøgersen, T. Isaksson, B. Nilsen, E. Bakker, K. Hildrum. "On-line NIR analysis of fat, water and protein in industrial scale ground meat batches". *Meat Science* 1999. 51(1): 97-102.
60. P. Williams. "Secondary ion mass spectrometry". *Annu. Rev. Mater. Sci.* 1985. 15(1): 517-548.
61. J. Baker. "Secondary Ion Mass Spectrometry". In *Practical Materials Characterization*, Sardela, M., Ed. Springer New York: 2014; Pp 133-187.
62. J.S. Fletcher, J.C. Vickerman, N. Winograd. "Label free biochemical 2D and 3D imaging using secondary ion mass spectrometry". *Curr. Opin. Chem. Biol.* 2011. 15(5): 733-740.
63. M. Mohri, K. Watanabe, T. Yamashina. "Sputtering process of a silicon carbide surface with energetic ions by means of an AES-SIMS-FDS combined system". *J. Nucl. Mater.* 1978. 75(1): 7-13.
64. M. Holzweber, A. Lippitz, K. Krueger, J. Jankowski, W.E. Unger. "Surface characterization of dialyzer polymer membranes by imaging ToF-SIMS and quantitative XPS line scans". *Biointerphases* 2015. 10(1): 019011.
65. L. Carlred, A. Gunnarsson, S. Solé-Domènech, B. Johansson, V. Vukojević, L. Terenius, A. Codita, B. Winblad, M. Schalling, F. Höök, P. Sjövall. "Simultaneous Imaging of Amyloid- $\beta$  and Lipids in Brain Tissue Using Antibody-Coupled Liposomes and Time-of-Flight Secondary Ion Mass Spectrometry". *J. Am. Chem. Soc.* 2014. 136(28): 9973-9981.
66. C. Bich, D. Touboul, A. Brunelle. "Cluster TOF-SIMS imaging as a tool for micrometric histology of lipids in tissue". *Mass Spectrom. Rev.* 2014. 33(6): 442-451.
67. N. Desbenoit, I. Schmitz-Afonso, C. Baudouin, O. Laprèvote, D. Touboul, F. Brignole-Baudouin, A. Brunelle. "Localisation and quantification of benzalkonium chloride in eye tissue by TOF-SIMS imaging and liquid chromatography mass spectrometry". *Anal. Bioanal. Chem.* 2013. 405(12): 4039-4049.
68. L. Piruzyan. "Feasibility of Novel Atomic-Molecular Resuscitation Tools". *Bulletin of experimental biology and medicine* 2012. 152(5): 646-652.
69. B. Bleaney. "Physics at the University of Oxford". *European Journal of Physics* 1988. 9(4): 283.
70. M. Sathiya, J.B. Leriche, E. Salager, D. Gourier, J.M. Tarascon, H. Vezin. "Electron paramagnetic resonance imaging for real-time monitoring of Li-ion batteries". *Nat Commun* 2015. 6.
71. M. Abou Fadel, X. Zhang, A. de Juan, R. Tauler, H. Vezin, L. Duponchel. "Extraction of Pure Spectral Signatures and Corresponding Chemical Maps from EPR Imaging Data Sets: Identifying Defects on a CaF<sub>2</sub> Surface Due to a Laser Beam Exposure". *Anal. Chem.* 2015. 87(7): 3929-3935.
72. H.K. Noh, R. Lu. "Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality". *Postharvest Biol. Technol.* 2007. 43(2): 193-201.

## References

---

73. V. Studer, J. Bobin, M. Chahid, H.S. Mousavi, E. Candes, M. Dahan. "Compressive fluorescence microscopy for biological and hyperspectral imaging". *Proceedings of the National Academy of Sciences* 2012. 109(26): E1679-E1687.
74. G. Zavattini, S. Vecchi, G. Mitchell, U. Weisser, R.M. Leahy, B.J. Pichler, D.J. Smith, S.R. Cherry. "A hyperspectral fluorescence system for 3D in vivo optical imaging". *Physics in medicine and biology* 2006. 51(8): 2029.
75. W.F.J. Vermaas, J.A. Timlin, H.D.T. Jones, M.B. Sinclair, L.T. Nieman, S.W. Hamad, D.K. Melgaard, D.M. Haaland. "In vivo hyperspectral confocal fluorescence imaging to determine pigment localization and distribution in cyanobacterial cells". *Proceedings of the National Academy of Sciences* 2008. 105(10): 4050-4055.
76. S.G. Kong, Y.-R. Chen, I. Kim, M.S. Kim. "Analysis of hyperspectral fluorescence images for poultry skin tumor inspection". *Appl. Opt.* 2004. 43(4): 824-833.
77. E.M. Sevick-Muraca. "Translation of Near-Infrared Fluorescence Imaging Technologies: Emerging Clinical Applications". *Annu. Rev. Med.* 2012. 63(1): 217-231.
78. W.R. Johnson, D.W. Wilson, W. Fink, M. Humayun, G. Bearman. "Snapshot hyperspectral imaging in ophthalmology". *Journal of biomedical optics* 2007. 12(1): 014036-014036-7.
79. A. Bodkin, A. Sheinis, A. Norton, J. Daly, S. Beaven, J. Weinheimer. In "Snapshot hyperspectral imaging: the hyperpixel array camera", 2009; 2009; Pp 73340H-73340H-11.
80. R.T. Kester, N. Bedard, L. Gao, T.S. Tkaczyk. "Real-time snapshot hyperspectral imaging endoscope". *Journal of Biomedical Optics* 2011. 16(5): 056005-056005-12.
81. Z. Gao, L. Zeng, Q. Wu. In "AOTF-based near-infrared imaging spectrometer for rapid identification of camouflaged target", 2014; 2014; Pp 92980R-92980R-8.
82. Q. Li, D. Xu, X. He, Y. Wang, Z. Chen, H. Liu, Q. Xu, F. Guo. "AOTF based molecular hyperspectral imaging system and its applications on nerve morphometry". *Appl. Opt.* 2013. 52(17): 3891-3901.
83. G. Konda Naganathan, K. Cluff, A. Samal, C. Calkins, D.D. Jones, C. Lorenzen, J. Subbiah. "A prototype on-line AOTF hyperspectral image acquisition system for tenderness assessment of beef carcasses". 2015.
84. J. Gómez-Sanchis, D. Lorente, E. Soria-Olivas, N. Aleixos, S. Cubero, J. Blasco. "Development of a hyperspectral computer vision system based on two liquid crystal tuneable filters for fruit inspection. Application to detect citrus fruits decay". *Food and bioprocess technology* 2014. 7(4): 1047-1056.
85. F. Scholkmann, S. Kleiser, A.J. Metz, R. Zimmermann, J.M. Pavia, U. Wolf, M. Wolf. "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology". *NeuroImage* 2014. 85: 6-27.
86. D.J. Kissick, D. Wanapun, G.J. Simpson. "Second-Order Nonlinear Optical Imaging of Chiral Crystals". *Annu. Rev. Anal. Chem.* 2011. 4(1): 419-437.
87. F. Blanchard, A. Doi, T. Tanaka, K. Tanaka. "Real-time, subwavelength terahertz imaging". *Annual Review of Materials Research* 2013. 43: 237-259.
88. M. Abou Fadel, X. Zhang, A. de Juan, R. Tauler, H. Vezin, L. Duponchel. "Extraction of Pure Spectral Signatures and Corresponding Chemical Maps from EPR Imaging Data Sets: Identifying Defects on a CaF<sub>2</sub> Surface Due to a Laser Beam Exposure". *Anal. Chem.* 2015.
89. T. Blaschke. "Object based image analysis for remote sensing". *ISPRS journal of photogrammetry and remote sensing* 2010. 65(1): 2-16.
90. R.A. Schowengerdt. "Remote sensing: models and methods for image processing". Academic press: 2006.
91. N.B. Grimm, J.G. Grove, S.T. Pickett, C.L. Redman. "Integrated Approaches to Long-Term Studies of Urban Ecological Systems Urban ecological systems present multiple challenges to

## References

---

- ecologists—pervasive human impact and extreme heterogeneity of cities, and the need to integrate social and ecological approaches, concepts, and theory". *Bioscience* 2000. 50(7): 571-584.
92. M. Govender, K. Chetty, H. Bulcock. "A review of hyperspectral remote sensing and its application in vegetation and water resource studies". *Water Sa* 2007. 33(2).
93. W.B. Cohen, S.N. Goward. "Landsat's role in ecological applications of remote sensing". *Bioscience* 2004. 54(6): 535-545.
94. X. Zhang, A.d. Juan, R. Tauler. "Local rank-based spatial information for improvement of remote sensing hyperspectral imaging resolution".
95. X. Zhang, R. Tauler. "Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging". *Anal. Chim. Acta* 2013. 762(0): 25-38.
96. M. Moeller. "Remote sensing for the monitoring of urban growth patterns". *world* 1950. 1975: 2000.
97. J. Im, J.R. Jensen, R.R. Jensen, J. Gladden, J. Waugh, M. Serrato. "Vegetation cover analysis of hazardous waste sites in Utah and Arizona using hyperspectral remote sensing". *Remote Sensing* 2012. 4(2): 327-353.
98. P.S. Thenkabail, R.B. Smith, E. De Pauw. "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics". *Remote Sens. Environ.* 2000. 71(2): 158-182.
99. D. Haboudane, J.R. Miller, N. Tremblay, P.J. Zarco-Tejada, L. Dextraze. "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture". *Remote Sens. Environ.* 2002. 81(2): 416-426.
100. T. Selige, J. Böhner, U. Schmidhalter. "High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures". *Geoderma* 2006. 136(1): 235-244.
101. F.A. Kruse. "Use of Airborne Imaging Spectrometer data to map minerals associated with hydrothermally altered rocks in the northern Grapevine Mountains, Nevada, and California". *Remote Sens. Environ.* 1988. 24(1): 31-51.
102. J. Ellis, H. Davis, J. Zamudio. "Exploring for onshore oil seeps with hyperspectral imaging". *Oil Gas J.* 2001. 99(37): 49-58.
103. K. Smith, M. Steven, J. Colls. "Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks". *Remote Sens. Environ.* 2004. 92(2): 207-217.
104. A. Barducci, D. Guzzi, P. Marcoinni, I. Pippi. "Aerospace wetland monitoring by hyperspectral imaging sensors: A case study in the coastal zone of San Rossore Natural Park". *J. Environ. Manage.* 2009. 90(7): 2278-2286.
105. G.P. Asner, J. Boardman, C.B. Field, D.E. Knapp, T. Kennedy-Bowdoin, M.O. Jones, R.E. Martin. "Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems". *Journal of Applied Remote Sensing* 2007. 1(1): 013536-013536-21.
106. B.-C. Gao, M.J. Montes, Z. Ahmad, C.O. Davis. "Atmospheric correction algorithm for hyperspectral remote sensing of ocean color from space". *Appl. Opt.* 2000. 39(6): 887-896.
107. D. Manolakis, G. Shaw. "Detection algorithms for hyperspectral imaging applications". *Signal Processing Magazine, IEEE* 2002. 19(1): 29-43.
108. W.L. Smith, D.K. Zhou, F.W. Harrison, H.E. Revercomb, A.M. Larar, H.-L. Huang, B. Huang. In "Hyperspectral remote sensing of atmospheric profiles from satellites and aircraft", *Second International Asia-Pacific Symposium on Remote Sensing of the Atmosphere, Environment, and Space, 2001; International Society for Optics and Photonics: 2001; Pp 94-102.*

## References

---

109. L.J. Rickard, R.W. Basedow, E.F. Zalewski, P.R. Silverglate, M. Landers. In "HYDICE: An airborne system for hyperspectral imaging", Optical Engineering and Photonics in Aerospace Sensing, 1993; International Society for Optics and Photonics: 1993; Pp 173-179.
110. P. Colarusso, L.H. Kidder, I.W. Levin, J.C. Fraser, J.F. Arens, E.N. Lewis. "Infrared spectroscopic imaging: from planetary to cellular systems". *Appl. Spectrosc.* 1998. 52(3): 106A.
111. S.L. Ustin, D.A. Roberts, J.A. Gamon, G.P. Asner, R.O. Green. "Using Imaging Spectroscopy to Study Ecosystem Processes and Properties". *Bioscience* 2004. 54(6): 523-534.
112. G. Vane, A.F. Goetz. "Terrestrial imaging spectroscopy". *Remote Sens. Environ.* 1988. 24(1): 1-29.
113. L.C. Rowan, A.F. Goetz, R.P. Ashley. "Discrimination of hydrothermally altered and unaltered rocks in visible and near infrared multispectral images". *Geophysics* 1977. 42(3): 522-535.
114. J.W. Boardman. In "Automating spectral unmixing of AVIRIS data using convex geometry concepts", *Summaries 4th Annu. JPL Airborne Geoscience Workshop, 1993; 1993; Pp 11-14.*
115. M. Moroni, E. Lupo, E. Marra, A. Cenedese. "Hyperspectral Image Analysis in Environmental Monitoring: Setup of a New Tunable Filter Platform". *Procedia Environmental Sciences* 2013. 19(0): 885-894.
116. G.A. Blackburn. "Hyperspectral remote sensing of plant pigments". *J. Exp. Bot.* 2007. 58(4): 855-867.
117. G.A. Blackburn, J.G. Ferwerda. "Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis". *Remote Sens. Environ.* 2008. 112(4): 1614-1632.
118. R.F. Kokaly, R.N. Clark. "Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression". *Remote Sens. Environ.* 1999. 67(3): 267-287.
119. K.S. He, D. Rocchini, M. Neteler, H. Nagendra. "Benefits of hyperspectral remote sensing for tracking plant invasions". *Divers. Distrib.* 2011. 17(3): 381-392.
120. M. Herold, M.E. Gardner, D.A. Roberts. "Spectral resolution requirements for mapping urban areas". *Geoscience and Remote Sensing, IEEE Transactions on* 2003. 41(9): 1907-1919.
121. G.A. Blackburn. "Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves". *Int. J. Remote Sens.* 1998. 19(4): 657-675.
122. D.A. Sims, J.A. Gamon. "Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features". *Remote Sens. Environ.* 2003. 84(4): 526-537.
123. L.G. Olmanson, P.L. Brezonik, M.E. Bauer. "Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The Mississippi River and its tributaries in Minnesota". *Remote Sens. Environ.* 2013. 130(0): 254-265.
124. F. Salem, M. Kafatos, T. El - Ghazawi, R. Gomez, R. Yang. "Hyperspectral image assessment of oil - contaminated wetland". *Int. J. Remote Sens.* 2005. 26(4): 811-821.
125. J.B. German. "Looking into the future of foods and health". *Innovation: Management, Policy & Practice* 2008. 10(1): 109-120.
126. J.E. Hobbs, L.M. Young. "Closer vertical co-ordination in agri-food supply chains: a conceptual framework and some preliminary evidence". *Supply Chain Management: An International Journal* 2000. 5(3): 131-143.
127. R. Karoui, J. De Baerdemaeker. "A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products". *Food Chem.* 2007. 102(3): 621-640.
128. C.-J. Du, D.-W. Sun. "Recent developments in the applications of image processing techniques for food quality evaluation". *Trends Food Sci. Technol.* 2004. 15(5): 230-249.

## References

---

129. V. Narendra, K. Hareesha. "Prospects of computer vision automated grading and sorting systems in agricultural and food products for quality evaluation". *International Journal of Computer Applications* 2010. 1(4): 1-9.
130. P. Butz, C. Hofmann, B. Tauscher. "Recent developments in noninvasive techniques for fresh fruit and vegetable internal quality analysis". *J. Food Sci.* 2005. 70(9): R131-R141.
131. E. Audesse, L. Huston, L. Palmer, A. Westlund, Tungsten-halogen lamp. In Google Patents: 1974.
132. J.B. Buschmann, K.R. Davies, J.E. Oetken, Tungsten halogen lamp. In Google Patents: 2005.
133. G. Polder, G.W. van der Heijden. In "Calibration and characterization of spectral imaging systems", *Multispectral Image Processing and Pattern Recognition, 2001; International Society for Optics and Photonics: 2001; Pp 10-17.*
134. M. Bigas, E. Cabruja, J. Forest, J. Salvi. "Review of CMOS image sensors". *Microelectron. J.* 2006. 37(5): 433-451.
135. W.W. Morey, G. Meltz, W.H. Glenn. In "Fiber Optic Bragg Grating Sensors", 1990; 1990; Pp 98-107.
136. T.K. Gangopadhyay. "Prospects for fibre Bragg gratings and Fabry-Perot interferometers in fibre-optic vibration sensing". *Sensors and Actuators A: Physical* 2004. 113(1): 20-38.
137. M. Kamruzzaman, G. ElMasry, D.-W. Sun, P. Allen. "Application of NIR hyperspectral imaging for discrimination of lamb muscles". *J. Food Eng.* 2011. 104(3): 332-340.
138. P.M. Mehl, Y.-R. Chen, M.S. Kim, D.E. Chan. "Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations". *J. Food Eng.* 2004. 61(1): 67-81.
139. G. ElMasry, N. Wang, A. ElSayed, M. Ngadi. "Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry". *J. Food Eng.* 2007. 81(1): 98-107.
140. D.P. Ariana, R. Lu, D.E. Guyer. "Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers". *Comput. Electron. Agric.* 2006. 53(1): 60-70.
141. N. Wang, G. ElMasry, S. Sevakarampalayam, J. Qiao. "Spectral imaging techniques for food quality evaluation". *Stewart Postharvest Review* 2007. 3(1): 1-8.
142. B.S. Sorg, O. Donovan, Y. Cao, M.W. Dewhirst, B.J. Moeller. "Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development". *Journal of biomedical optics* 2005. 10(4): 044004-044004-11.
143. K. Chao, C. Yang, Y. Chen, M. Kim, D. Chan. "Hyperspectral-multispectral line-scan imaging system for automated poultry carcass inspection applications for food safety". *Poultry science* 2007. 86(11): 2450-2460.
144. J. Qiao, M.O. Ngadi, N. Wang, C. Gariépy, S.O. Prasher. "Pork quality and marbling level assessment using a hyperspectral imaging system". *J. Food Eng.* 2007. 83(1): 10-16.
145. G.K. Naganathan, L.M. Grimes, J. Subbiah, C.R. Calkins, A. Samal, G.E. Meyer. "Visible/near-infrared hyperspectral imaging for beef tenderness prediction". *Comput. Electron. Agric.* 2008. 64(2): 225-233.
146. M. Kamruzzaman, G. ElMasry, D.-W. Sun, P. Allen. "Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression". *Innovative Food Science & Emerging Technologies* 2012. 16: 218-226.
147. X. Fu, Y. Ying. "Food safety evaluation based on near infrared spectroscopy and imaging: a review". *Crit. Rev. Food Sci. Nutr.* 2014. (just-accepted): 00-00.
148. A.A. Gowen, J. Burger, D. O'Callaghan, C. O'Donnell. In "Potential applications of hyperspectral imaging for quality control in dairy foods", 1st international workshop on computer image analysis in agriculture, Potsdam, Germany, 2009; 2009.

## References

---

149. P. Geladi, K. Esbensen. "The start and early history of chemometrics: Selected interviews. Part 1". *J. Chemom.* 1990. 4(5): 337-354.
150. B.R. Kowalski. "Chemometrics: views and propositions". *Journal of Chemical Information and Computer Sciences* 1975. 15(4): 201-203.
151. J.W. Einax. "Chemometrics in analytical chemistry". *Anal. Bioanal. Chem.* 2008. 390(5): 1225-1226.
152. P.K. Hopke. "The evolution of chemometrics". *Anal. Chim. Acta* 2003. 500(1): 365-377.
153. B.K. Lavine, J. Workman. "Chemometrics". *Anal. Chem.* 2002. 74(12): 2763-2770.
154. H.L. Wu, M. Shibukawa, K. Oguma. "An alternating trilinear decomposition algorithm with application to calibration of HPLC–DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons". *J. Chemom.* 1998. 12(1): 1-26.
155. B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski. "Second - order chemometric standardization for high - speed hyphenated gas chromatography: Analysis of GC/MS and comprehensive GC× GC data". *J. Microcolumn Sep.* 1999. 11(2): 97-107.
156. M. Praisler, J. Van Bocxlaer, A. De Leenheer, D. Massart. "Chemometric detection of thermally degraded samples in the analysis of drugs of abuse with gas chromatography–Fourier-transform infrared spectroscopy". *J. Chromatogr. A* 2002. 962(1): 161-173.
157. A. Nordström, G. O'Maille, C. Qin, G. Siuzdak. "Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum". *Anal. Chem.* 2006. 78(10): 3289-3295.
158. S. Sentellas, J. Saurina. "Chemometrics in capillary electrophoresis. Part A: Methods for optimization". *J. Sep. Sci.* 2003. 26(9 - 10): 875-885.
159. F.-T. Chau, Y.-Z. Liang, J. Gao, X.-G. Shao. "Chemometrics: from basics to wavelet transform". John Wiley & Sons: 2004; Vol. 234.
160. M. Kim, Y. Chen, P. Mehl. "Hyperspectral reflectance and fluorescence imaging system for food quality and safety". *Transactions-American Society of Agricultural Engineers* 2001. 44(3): 721-730.
161. D.T. Dicker, J. Lerner, P. Van Belle, t. Guerry, DuPont, M. Herlyn, D.E. Elder, W.S. El-Deiry. "Differentiation of normal skin and melanoma using high resolution hyperspectral imaging". *Cancer biology & therapy* 2006. 5(8): 1033-1038.
162. Q. Wang. "Raman spectroscopic characterization and analysis of agricultural and biological systems". 2013.
163. C.J. Behrend, C.P. Tarnowski, M.D. Morris. "Identification of outliers in hyperspectral Raman image data by nearest neighbor comparison". *Appl. Spectrosc.* 2002. 56(11): 1458-1461.
164. L. Zhang, M.J. Henson. "A practical algorithm to remove cosmic spikes in Raman imaging data for pharmaceutical applications". *Appl. Spectrosc.* 2007. 61(9): 1015-1020.
165. F. Ehrentreich, L. Sümmchen. "Spike removal and denoising of Raman spectra by wavelet transform methods". *Anal. Chem.* 2001. 73(17): 4364-4373.
166. M. Vidal, J.M. Amigo. "Pre-processing of hyperspectral images. Essential steps before image analysis". *Chemometrics Intellig. Lab. Syst.* 2012. 117: 138-148.
167. P.H. Eilers. "Parametric time warping". *Anal. Chem.* 2004. 76(2): 404-411.
168. P.H. Eilers. "A perfect smoother". *Anal. Chem.* 2003. 75(14): 3631-3636.
169. P. Geladi, D. MacDougall, H. Martens. "Linearization and scatter-correction for near-infrared reflectance spectra of meat". *Appl. Spectrosc.* 1985. 39(3): 491-500.
170. H. Martens, J.P. Nielsen, S.B. Engelsen. "Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures". *Anal. Chem.* 2003. 75(3): 394-404.

## References

---

171. J. Burger, P. Geladi. "Spectral pre-treatments of hyperspectral near infrared images: analysis of diffuse reflectance scattering". *J. Near Infrared Spectrosc.* 2007. 15(1): 29.
172. A.A. Gowen, M. Taghizadeh, C.P. O'Donnell. "Identification of mushrooms subjected to freeze damage using hyperspectral imaging". *J. Food Eng.* 2009. 93(1): 7-12.
173. Q. Guo, W. Wu, D. Massart. "The robust normal variate transform for pattern recognition with near-infrared data". *Anal. Chim. Acta* 1999. 382(1): 87-103.
174. A. Savitzky, M.J.E. Golay. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures". *Anal. Chem.* 1964. 36(8): 1627-1639.
175. P.A. Gorry. "General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method". *Anal. Chem.* 1990. 62(6): 570-573.
176. M.U. Bromba, H. Ziegler. "Application hints for Savitzky-Golay digital smoothing filters". *Anal. Chem.* 1981. 53(11): 1583-1586.
177. T. Isaksson, T. Næs. "The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy". *Appl. Spectrosc.* 1988. 42(7): 1273-1284.
178. S. Romero-Torres, J.D. Pérez-Ramos, K.R. Morris, E.R. Grant. "Raman spectroscopic measurement of tablet-to-tablet coating variability". *J. Pharm. Biomed. Anal.* 2005. 38(2): 270-274.
179. W.K. Newey, J.L. Powell. "Asymmetric least squares estimation and testing". *Econometrica: Journal of the Econometric Society* 1987. 819-847.
180. J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, J. Tan. "Asymmetric least squares for multiple spectra baseline correction". *Anal. Chim. Acta* 2010. 683(1): 63-68.
181. F. Vogt, M. Tacke. "Fast principal component analysis of large data sets". *Chemometrics and intelligent laboratory systems* 2001. 59(1): 1-18.
182. J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, J. Han. "Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing". *ISPRS Journal of Photogrammetry and Remote Sensing* 2014. 93(0): 112-122.
183. C.F. Caiafa, E. Salerno, A.N. Proto, L. Fiumi. "Blind spectral unmixing by local maximization of non-Gaussianity". *Signal Processing* 2008. 88(1): 50-68.
184. A. Hyvärinen, J. Karhunen, E. Oja. "Independent component analysis". John Wiley & Sons: 2004; Vol. 46.
185. J.L. Rodgers, W.A. Nicewander, L. Toothaker. "Linearly independent, orthogonal, and uncorrelated variables". *The American Statistician* 1984. 38(2): 133-134.
186. W. Windig, M.R. Keenan. "Homeopathic ICA: A simple approach to expand the use of independent component analysis (ICA)". *Chemometrics Intellig. Lab. Syst.* 2015. 142(0): 54-63.
187. J.M. Nascimento, J.M. Bioucas Dias. "Vertex component analysis: A fast algorithm to unmix hyperspectral data". *Geoscience and Remote Sensing, IEEE Transactions on* 2005. 43(4): 898-910.
188. J. Li, J.M. Bioucas-Dias. In "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data", *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, 2008; IEEE: 2008; Pp III250-III253.*
189. J. Li, J.M. Bioucas-Dias. In "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data", *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, 2008; IEEE: 2008; Pp III-250-III-253.*
190. F. Chaudhry, C.-C. Wu, W. Liu, C.-I. Chang, A. Plaza. "Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery". *Recent Advances in Hyperspectral Signal and Image Processing* 2006. 37(661): 2.
191. J.W. Boardman. "Automating spectral unmixing of AVIRIS data using convex geometry concepts". 1993.

## References

---

192. R. Neville, K. Staenz, T. Szeredi, J. Lefebvre, P. Hauff. In "Automatic endmember extraction from hyperspectral data for mineral exploration", Proc. 21st Can. Symp. Remote Sens, 1999; 1999; Pp 21-24.
193. C.-I. Chang, C.-C. Wu, W.-m. Liu, Y.-C. Ouyang. "A new growing method for simplex-based endmember extraction algorithm". *Geoscience and Remote Sensing, IEEE Transactions on* 2006. 44(10): 2804-2819.
194. T.-H. Chan, C.-Y. Chi, Y.-M. Huang, W.-K. Ma. "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing". *Signal Processing, IEEE Transactions on* 2009. 57(11): 4418-4432.
195. J.M. Bioucas-Dias. In "A variable splitting augmented Lagrangian approach to linear spectral unmixing", *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on, 2009; IEEE: 2009; Pp 1-4.*
196. Y. Altmann, N. Dobigeon, J.Y. Tournet. "Unsupervised Post-Nonlinear Unmixing of Hyperspectral Images Using a Hamiltonian Monte Carlo Algorithm". *Image Processing, IEEE Transactions on* 2014. 23(6): 2663-2675.
197. Y. Altmann, N. Dobigeon, J.Y. Tournet, S. McLaughlin. In "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares", *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, 24-29 July 2011, 2011; 2011; Pp 1151-1154.*
198. Y. Altmann. *Nonlinear unmixing of hyperspectral images. Institut National Polytechnique de Toulouse - INPT, 2013.*
199. C. Jie, C. Richard, P. Honeine. "Nonlinear Unmixing of Hyperspectral Data Based on a Linear-Mixture/Nonlinear-Fluctuation Model". *Signal Processing, IEEE Transactions on* 2013. 61(2): 480-492.
200. R. Tauler, A. Smilde, B. Kowalski. "Selectivity, local rank, three - way data analysis and ambiguity in multivariate curve resolution". *J. Chemom.* 1995. 9(1): 31-58.
201. N.E. Llamas, M. Garrido, M.S. Di Nezio, B.S.F. Band. "Second order advantage in the determination of amaranth, sunset yellow FCF and tartrazine by UV-vis and multivariate curve resolution-alternating least squares". *Anal. Chim. Acta* 2009. 655(1): 38-42.
202. M. López-Pastor, M.J. Ayora-Cañada, M. Valcárcel, B. Lendl. "Association of methanol and water in ionic liquids elucidated by infrared spectroscopy using two-dimensional correlation and multivariate curve resolution". *The Journal of Physical Chemistry B* 2006. 110(22): 10896-10902.
203. R. Tauler, D. Barceló. "Multivariate curve resolution applied to liquid chromatography—diode array detection". *TrAC, Trends Anal. Chem.* 1993. 12(8): 319-327.
204. N.B. Gallagher, J.M. Shaver, E.B. Martin, J. Morris, B.M. Wise, W. Windig. "Curve resolution for multivariate images with applications to TOF-SIMS and Raman". *Chemometrics Intellig. Lab. Syst.* 2004. 73(1): 105-117.
205. W. Windig, B. Antalek, M.J. Robbins, N. Zumbulyadis, C.E. Heckler. "Applications of the direct exponential curve resolution algorithm (DECRA) to solid state nuclear magnetic resonance and mid - infrared spectra". *J. Chemom.* 2000. 14(3): 213-227.
206. D.M. Haaland, J.A. Timlin, M.B. Sinclair, M.H. Van Benthem, M.J. Martinez, A.D. Aragon, M. Werner-Washburne. In "Multivariate curve resolution for hyperspectral image analysis: applications to microarray technology", *Biomedical Optics 2003, 2003; International Society for Optics and Photonics: 2003; Pp 55-66.*
207. M. Esteban, C. Arino, J. Diaz-Cruz, M. Diaz-Cruz, R. Tauler. "Multivariate curve resolution with alternating least squares optimisation: a soft-modelling approach to metal complexation studies by voltammetric techniques". *TrAC, Trends Anal. Chem.* 2000. 19(1): 49-61.

## References

---

208. P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne. "Multivariate curve resolution of time course microarray data". *BMC Bioinformatics* 2006. 7(1): 343.
209. M. Farrés, B. Piña, R. Tauler. "Chemometric evaluation of *Saccharomyces cerevisiae* metabolic profiles using LC-MS". *Metabolomics* 2015. 11(1): 210-224.
210. J. Jaumot, A. de Juan, R. Tauler. "MCR-ALS GUI 2.0: New features and applications". *Chemometrics Intellig. Lab. Syst.* 2015. 140: 1-12.
211. A. Malik, R. Tauler. "Performance and validation of MCR-ALS with quadrilinear constraint in the analysis of noisy datasets". *Chemometrics Intellig. Lab. Syst.* 2014. 135(0): 223-234.
212. A.L. Pomerantsev, Y.V. Zontov, O.Y. Rodionova. "Nonlinear multivariate curve resolution alternating least squares (NL-MCR-ALS)". *J. Chemom.* 2014. 28(10): 740-748.
213. A. Atifi, K. Czarnecki, H. Mountacer, M.D. Ryan. "In Situ Study of the Photodegradation of Carbofuran Deposited on TiO<sub>2</sub> Film under UV Light, Using ATR-FTIR Coupled to HS-MCR-ALS". *Environ. Sci. Technol.* 2013. 47(15): 8650-8657.
214. A. Malik, R. Tauler. "Exploring the interaction between O<sub>3</sub> and NO<sub>x</sub> pollution patterns in the atmosphere of Barcelona, Spain using the MCR-ALS method". *Sci. Total Environ.* 2015. 517(0): 151-161.
215. C. Fernández, M. Pilar Callao, M. Soledad Larrechi. "UV-visible-DAD and 1H-NMR spectroscopy data fusion for studying the photodegradation process of azo-dyes using MCR-ALS". *Talanta* 2013. 117(0): 75-80.
216. M.J. Culzoni, A. Mancha de Llanos, M.M. De Zan, A. Espinosa-Mansilla, F. Cañada-Cañada, A. Muñoz de la Peña, H.C. Goicoechea. "Enhanced MCR-ALS modeling of HPLC with fast scan fluorimetric detection second-order data for quantitation of metabolic disorder marker pteridines in urine". *Talanta* 2011. 85(5): 2368-2374.
217. K. Kumar, A.K. Mishra. "Application of 'multivariate curve resolution alternating least square (MCR-ALS)' analysis to extract pure component synchronous fluorescence spectra at various wavelength offsets from total synchronous fluorescence spectroscopy (TSFS) data set of dilute aqueous solutions of fluorophores". *Chemometrics Intellig. Lab. Syst.* 2012. 116(0): 78-86.
218. G. Ahmadi, R. Tauler, H. Abdollahi. "Multivariate calibration of first-order data with the correlation constrained MCR-ALS method". *Chemometrics Intellig. Lab. Syst.* 2015. 142(0): 143-150.
219. M. Abou Fadel, A. de Juan, N. Touati, H. Vezin, L. Duponchel. "New chemometric approach MCR-ALS to unmix EPR spectroscopic data from complex mixtures". *J. Magn. Reson.* 2014. 248(0): 27-35.
220. B. Hemmateenejad, A. Safavi, F. Honarasa. "Electrochemical study of weak inclusion complex interactions by simultaneous MCR-ALS analyses of potential step-chronoamperometric data matrices". *Anal. Methods* 2012. 4(6): 1776-1782.
221. A. Folch-Fortuny, M. Tortajada, J.M. Prats-Montalbán, F. Llaneras, J. Picó, A. Ferrer. "MCR-ALS on metabolic networks: Obtaining more meaningful pathways". *Chemometrics Intellig. Lab. Syst.* 2015. 142(0): 293-303.
222. S. Grassi, C. Alamprese, V. Bono, E. Casiraghi, J. Amigo. "Modelling Milk Lactic Acid Fermentation Using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS)". *Food and Bioprocess Technology* 2014. 7(6): 1819-1829.
223. M.A. Hegazy, N.S. Abdelwahab, A.S. Fayed. "A novel spectral resolution and simultaneous determination of multicomponent mixture of Vitamins B1, B6, B12, Benfotiamine and Diclofenac in tablets and capsules by derivative and MCR-ALS". *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 2015. 140(0): 524-533.

## References

---

224. S. Piqueras, L. Duponchel, R. Tauler, A. De Juan. "Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares". *Anal. Chim. Acta* 2011. 705(1): 182-192.
225. W. Windig, J. Guilment. "Interactive self-modeling mixture analysis". *Anal. Chem.* 1991. 63(14): 1425-1432.
226. M.H. Van Benthem, M.R. Keenan. "Fast algorithm for the solution of large - scale non - negativity - constrained least squares problems". *J. Chemom.* 2004. 18(10): 441-450.
227. W.H. Lawton, E.A. Sylvestre. "Self modeling curve resolution". *Technometrics* 1971. 13(3): 617-633.
228. C.L. Lawson, R.J. Hanson. "Solving least squares problems". SIAM: 1974; Vol. 161.
229. R. Bro, S. De Jong. "A fast non-negativity-constrained least squares algorithm". *J. Chemom.* 1997. 11(5): 393-401.
230. A. De Juan, Y. Vander Heyden, R. Tauler, D. Massart. "Assessment of new constraints applied to the alternating least squares method". *Anal. Chim. Acta* 1997. 346(3): 307-318.
231. P. Geladi, S. Wold. "Local principal component models, rank maps and contextuality for curve resolution and multi-way calibration inference". *Chemometrics Intellig. Lab. Syst.* 1987. 2(4): 273-281.
232. F.R. Gantmacher. "The theory of matrices". Taylor & Francis: 1964.
233. E. Sanchez, B.R. Kowalski. "Generalized rank annihilation factor analysis". *Anal. Chem.* 1986. 58(2): 496-499.
234. A. de Juan, M. Maeder, T. Hancewicz, R. Tauler. "Local rank analysis for exploratory spectroscopic image analysis. Fixed size image window-evolving factor analysis". *Chemometrics Intellig. Lab. Syst.* 2005. 77(1): 64-74.
235. P. Saikkonen, H. Lütkepohl. "Local power of likelihood ratio tests for the cointegrating rank of a VAR process". *Econometric Theory* 1999. 15(01): 50-78.
236. H. Keller, D. Massart. "Evolving factor analysis". *Chemometrics Intellig. Lab. Syst.* 1991. 12(3): 209-224.
237. A. De Juan, R. Tauler. "Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution". *Anal. Chim. Acta* 2003. 500(1): 195-210.
238. R. Tauler, B. Walczak, S.D. Brown. "Comprehensive chemometrics: chemical and biochemical data analysis". Elsevier: 2009.
239. "MATLAB. The Mathworks, Inc."
240. K.S. Booksh, B.R. Kowalski. "Theory of analytical chemistry". *Anal. Chem.* 1994. 66(15): 782A-791A.
241. R.A. Harshman, M.E. Lundy. "PARAFAC: Parallel factor analysis". *Comput. Stat. Data Anal.* 1994. 18(1): 39-72.
242. R. Tauler, I. Marqués, E. Casassas. "Multivariate curve resolution applied to three - way trilinear data: Study of a spectrofluorimetric acid - base titration of salicylic acid at three excitation wavelengths". *Journal of Chemometrics* 1998. 12(1): 55-75.
243. R. Bro. "PARAFAC. Tutorial and applications". *Chemometrics Intellig. Lab. Syst.* 1997. 38(2): 149-171.
244. R. Tauler, I. Marqués, E. Casassas. "Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths". *Journal of Chemometrics* 1998. 12(1): 55-75.
245. E. Spjøtvoll, H. Martens, R. Volden. "Restricted least squares estimation of the spectra and concentration of two unknown constituents available in mixtures". *Technometrics* 1982. 24(3): 173-180.

## References

---

246. J. Jaumot, R. Tauler. "MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution". *Chemometrics and Intelligent Laboratory Systems* 2010. 103(2): 96-107.
247. H. Abdollahi, R. Tauler. "Uniqueness and rotation ambiguities in multivariate curve resolution methods". *Chemometrics Intellig. Lab. Syst.* 2011. 108(2): 100-111.
248. R. Tauler. "Comments on a recently published paper 'some surprising properties of multivariate curve resolution - alternating least squares (MCR - ALS) algorithms' ". *J. Chemom.* 2010. 24(2): 87-90.
249. H. Abdollahi, M. Maeder, R. Tauler. "Calculation and Meaning of Feasible Band Boundaries in Multivariate Curve Resolution of a Two-Component System". *Anal. Chem.* 2009. 81(6): 2115-2122.
250. A. Golshan, H. Abdollahi, M. Maeder. "The reduction of rotational ambiguity in soft-modeling by introducing hard models". *Anal. Chim. Acta* 2012. 709: 32-40.
251. O.S. Borgen, B.R. Kowalski. "An extension of the multivariate component-resolution method to three components". *Analytica Chimica Acta* 1985. 174(0): 1-26.
252. M. Sawall, C. Kubis, D. Selent, A. Börner, K. Neymeyr. "A fast polygon inflation algorithm to compute the area of feasible solutions for three - component systems. I: concepts and applications". *J. Chemom.* 2013. 27(5): 106-116.
253. O.S. Borgen, B.R. Kowalski. "An extension of the multivariate component-resolution method to three components". *Anal. Chim. Acta* 1985. 174: 1-26.
254. R.C. Henry, B.M. Kim. "Extension of self-modeling curve resolution to mixtures of more than three components: Part 1. Finding the basic feasible region". *Chemometrics Intellig. Lab. Syst.* 1990. 8(2): 205-216.
255. M.N. Leger, P.D. Wentzell. "Dynamic Monte Carlo self-modeling curve resolution method for multicomponent mixtures". *Chemometrics Intellig. Lab. Syst.* 2002. 62(2): 171-188.
256. R. Rajkó, K. István. "Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry". *J. Chemom.* 2005. 19(8): 448-463.
257. A. Golshan, M. Maeder, H. Abdollahi. "Determination and visualization of rotational ambiguity in four-component systems". *Anal. Chim. Acta* 2013. 796: 20-26.
258. M. Sawall, K. Neymeyr. "A fast polygon inflation algorithm to compute the area of feasible solutions for three - component systems. II: Theoretical foundation, inverse polygon inflation, and FAC - PACK implementation". *J. Chemom.* 2014. 28(8): 633-644.
259. R. Tauler. "Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution". *Journal of Chemometrics* 2001. 15(8): 627-646.
260. A. de Juan, E. Casassas, R. Tauler. "Soft Modeling of Analytical Data". In *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Ltd: 2006.
261. R.N. Clark, G.A. Swayze, R. Wise, K.E. Livo, T.M. Hoefen, R.F. Kokaly, S.J. Sutley. "USGS digital spectral library splib06a". US Geological Survey Reston, VA: 2007.
262. [http://aviris.jpl.nasa.gov/data/free\\_data.html](http://aviris.jpl.nasa.gov/data/free_data.html), In NASA, Jet Propulsion Laboratory: 1997.
263. A.F. Goetz, V. Srivastava. In "Mineralogical mapping in the Cuprite mining district, Nevada", *Proc. Airborne Imaging Spectrometer Data Analysis Workshop*, 1985; 1985; Pp 22-29.
264. J.M.P. Nascimento, J.M.B. Dias. "Vertex component analysis: A fast algorithm to unmix hyperspectral data". *Geoscience and Remote Sensing, IEEE Transactions on* 2005. 43(4): 898-910.
265. A. Plaza, P. Martínez, R. Pérez, J. Plaza. "Spatial/spectral endmember extraction by multidimensional morphological operations". *Geoscience and Remote Sensing, IEEE Transactions on* 2002. 40(9): 2025-2041.

## References

---

266. X. Zhang, R. Tauler. "Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging". *Analytica chimica acta* 2013. 762: 25-38.
267. X. Zhang, R. Tauler. "Measuring and comparing the resolution performance and the extend of rotation ambiguities in bilinear modelling methods". 2015.
268. A. Plaza, P. Martinez, R. Pérez, J. Plaza. "Spatial/spectral endmember extraction by multidimensional morphological operations". *Geoscience and Remote Sensing, IEEE Transactions on* 2002. 40(9): 2025-2041.
269. J.M. Bioucas-Dias. In "A variable splitting augmented Lagrangian approach to linear spectral unmixing", *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop on, 26-28 Aug. 2009, 2009; 2009; Pp 1-4.*
270. Z. Xiaobo, Z. Jiewen, M. Holmes, M. Hanpin, S. Jiyong, Y. Xiaopin, L. Yanxiao. "Independent component analysis in information extraction from visible/near-infrared hyperspectral imaging data of cucumber leaves". *Chemometrics Intellig. Lab. Syst.* 2010. 104(2): 265-270.
271. M.Z. Baghbidj, K. Jamshidi, A.R.N. Nilchi, S. Homayouni. "Improvement of Anomaly Detection Algorithms in Hyperspectral Images using Discrete Wavelet Transform". *arXiv preprint arXiv:1201.2025* 2012.
272. D. Manolakis, D. Marden, G.A. Shaw. "Hyperspectral image processing for automatic target detection applications". *Lincoln Laboratory Journal* 2003. 14(1): 79-116.
273. K. Tiwari, M. Arora, D. Singh. "An assessment of independent component analysis for detection of military targets from hyperspectral images". *International Journal of Applied Earth Observation and Geoinformation* 2011. 13(5): 730-740.
274. K. Tiwari, M. Arora, D. Singh. In "Subpixel target detection and enhancement in hyperspectral images", *SPIE Defense, Security, and Sensing, 2011; International Society for Optics and Photonics: 2011; Pp 80481T-80481T-11.*
275. J.B. Campbell. "Introduction to remote sensing". CRC Press: 2002.
276. Y.-Z. Feng, D.-W. Sun. "Application of hyperspectral imaging in food safety inspection and control: a review". *Crit. Rev. Food Sci. Nutr.* 2012. 52(11): 1039-1058.
277. E.O. Afoakwa, A. Paterson, M. Fowler. "Factors influencing rheological and textural qualities in chocolate—a review". *Trends in Food Science & Technology* 2007. 18(6): 290-298.
278. C. Ieggli, D. Bohrer, P. Do Nascimento, L. De Carvalho. "Determination of sodium, potassium, calcium, magnesium, zinc and iron in emulsified chocolate samples by flame atomic absorption spectrometry". *Food Chem.* 2011. 124(3): 1189-1193.
279. S.T. Beckett. "Is the taste of British milk chocolate different?". *Int. J. Dairy Technol.* 2003. 56(3): 139-142.
280. J.J. O'rourke, *Chocolate product and process*. In Google Patents: 1959.
281. E.O. Afoakwa, A. Paterson, M. Fowler, A. Ryan. "Matrix effects on flavour volatiles release in dark chocolates varying in particle size distribution and fat content using GC–mass spectrometry and GC–olfactometry". *Food Chem.* 2009. 113(1): 208-215.
282. A. Krishna, M. Morrin. "Does touch affect taste? The perceptual transfer of product container haptic cues". *Journal of Consumer Research* 2008. 34(6): 807-818.
283. I. Larmour, K. Faulds, D. Graham. "Rapid Raman mapping for chocolate analysis". *Anal. Methods* 2010. 2(9): 1230-1232.
284. J.M. Amigo, I. Martí, A. Gowen. "Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality". In *Data Handling in Science and Technology. Vol 28, 2013; Pp 343-370.*
285. G. Elmasry, M. Kamruzzaman, D.-W. Sun, P. Allen. "Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review". *Crit. Rev. Food Sci. Nutr.* 2012. 52(11): 999-1023.

## References

---

286. E.O. Afoakwa, A. Paterson, M. Fowler, J. Vieira. "Particle size distribution and compositional effects on textural properties and appearance of dark chocolates". *J. Food Eng.* 2008. 87(2): 181-190.
287. J.M. Amigo. "Practical issues of hyperspectral imaging analysis of solid dosage forms". *Anal. Bioanal. Chem.* 2010. 398(1): 93-109.
288. A. Gowen, C. O'Donnell, P.J. Cullen, S. Bell. "Recent applications of chemical imaging to pharmaceutical process monitoring and quality control". *European Journal of Pharmaceutics and Biopharmaceutics* 2008. 69(1): 10-22.
289. S. Moussaoui, H. Hauksdottir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, J.A. Benediktsson. "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation". *Neurocomputing* 2008. 71(10): 2194-2208.
290. S. Beyramysoltan, R. Rajkó, H. Abdollahi. "Investigation of the equality constraint effect on the reduction of the rotational ambiguity in three-component system using a novel grid search method". *Anal. Chim. Acta* 2013. 791: 25-35.
291. C.J. Vörösmarty, P. McIntyre, M.O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S.E. Bunn, C.A. Sullivan, C.R. Liermann. "Global threats to human water security and river biodiversity". *Nature* 2010. 467(7315): 555-561.
292. P. Debels, R. Figueroa, R. Urrutia, R. Barra, X. Niell. "Evaluation of water quality in the Chillán River (Central Chile) using physicochemical parameters and a modified water quality index". *Environ. Monit. Assess.* 2005. 110(1-3): 301-322.
293. C. Evans, D. Monteith, D. Cooper. "Long-term increases in surface water dissolved organic carbon: observations, possible causes and environmental impacts". *Environ. Pollut.* 2005. 137(1): 55-71.
294. C.A. Stedmon, S. Markager, R. Bro. "Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy". *Mar. Chem.* 2003. 82(3): 239-254.
295. D.P. Morris, H. Zagarese, C.E. Williamson, E.G. Balseiro, B.R. Hargreaves, B. Modenutti, R. Moeller, C. Queimalinos. "The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon". *Limnology and Oceanography* 1995. 1381-1391.
296. C. Stedmon, S. Markager. "The optics of chromophoric dissolved organic matter (CDOM) in the Greenland Sea: An algorithm for differentiation between marine and terrestrially derived organic matter". *Limnology and Oceanography* 2001. 2087-2093.
297. C.H. Lochmueller, S.S. Saavedra. "Conformational changes in a soil fulvic acid measured by time-dependent fluorescence depolarization". *Analytical Chemistry* 1986. 58(9): 1978-1981.
298. P.G. Coble. "Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy". *Marine Chemistry* 1996. 51(4): 325-346.
299. S. Opsahl, R. Benner, R.M.W. Amon. "Major flux of terrigenous dissolved organic matter through the Arctic Ocean". *Limnology and Oceanography* 1999. 2017-2023.
300. P.R. Kannel, S. Lee, S.R. Kanel, S.P. Khan. "Chemometric application in classification and assessment of monitoring locations of an urban river system". *Anal. Chim. Acta* 2007. 582(2): 390-399.
301. X.S. Xie, J.K. Trautman. "Optical studies of single molecules at room temperature". *Annu. Rev. Phys. Chem.* 1998. 49(1): 441-480.
302. J.C.E. da Silva, M.J. Tavares, R. Tauler. "Multivariate curve resolution of multidimensional excitation-emission quenching matrices of a Laurentian soil fulvic acid". *Chemosphere* 2006. 64(11): 1939-1948.

## References

---

303. Y. Yamashita, R. Jaffé, N. Maie, E. Tanoue. "Assessing the dynamics of dissolved organic matter (DOM) in coastal environments by excitation emission matrix fluorescence and parallel factor analysis (EEM - PARAFAC)". *Limnol. Oceanogr.* 2008. 53(5): 1900-1908.
304. A. Astel, S. Tsakovski, P. Barbieri, V. Simeonov. "Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets". *Water Res.* 2007. 41(19): 4566-4578.
305. T. Kowalkowski, R. Zbytniewski, J. Szpejna, B. Buszewski. "Application of chemometrics in river water classification". *Water Res.* 2006. 40(4): 744-752.
306. S. Mas, A. de Juan, R. Tauler, A.C. Olivieri, G.M. Escandar. "Application of chemometric methods to environmental analysis of organic pollutants: a review". *Talanta* 2010. 80(3): 1052-1067.
307. X. Zhang, R. Marcé, J. Armengol, R. Tauler. "Distribution of dissolved organic matter in freshwaters using excitation emission fluorescence and Multivariate Curve Resolution". *Chemosphere* 2014. 111: 120-128.
308. A. de Juan, S.C. Rutan, R. Tauler, D.L. Massart. "Comparison between the direct trilinear decomposition and the multivariate curve resolution-alternating least squares methods for the resolution of three-way data sets". *Chemometrics and Intelligent Laboratory Systems* 1998. 40(1): 19-32.
309. A. de Juan, R. Tauler. "Comparison of three - way resolution methods for non - trilinear chemical data sets". *Journal of Chemometrics* 2001. 15(10): 749-771.
310. R. Cespedes, S. Lacorte, A. Ginebreda, D. Barceló. "Occurrence and fate of alkylphenols and alkylphenol ethoxylates in sewage treatment plants and impact on receiving waters along the Ter River (Catalonia, NE Spain)". *Environ. Pollut.* 2008. 153(2): 384-392.
311. "[http://ccaa.elpais.com/ccaa/2013/03/25/catalunya/1364245730\\_525669.html](http://ccaa.elpais.com/ccaa/2013/03/25/catalunya/1364245730_525669.html)".
312. J. Armengol, R. Merce The Characteristics of Spanish Reservoirs; DTIC Document: 2003.
313. R. Céspedes, S. Lacorte, A. Ginebreda, D. Barceló. "Chemical monitoring and occurrence of alkylphenols, alkylphenol ethoxylates, alcohol ethoxylates, phthalates and benzothiazoles in sewage treatment plants and receiving waters along the Ter River basin (Catalonia, NE Spain)". *Anal. Bioanal. Chem.* 2006. 385(6): 992-1000.
314. R. Céspedes, S. Lacorte, A. Ginebreda, D. Barceló. "Occurrence and fate of alkylphenols and alkylphenol ethoxylates in sewage treatment plants and impact on receiving waters along the Ter River (Catalonia, NE Spain)". *Environ. Pollut.* 2008. 153(2): 384-392.
315. D. Sławińska, K. Polewski, P. Rolewski, P. Pluciński, J. Sławiński. "Spectroscopic studies on UVC-induced photodegradation of humic acids". *Development* 2002. 5(2).
316. C.A. Stedmon, S. Markager. "Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis". *Limnology and Oceanography* 2005. 50(2): 686-697.
317. J. Rogan, D. Chen. "Remote sensing technology for mapping and monitoring land-cover and land-use change". *Progress in Planning* 2004. 61(4): 301-325.
318. C. Fischer, I. Kakoulli. "Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications". *Reviews in conservation* 2006. 7: 3-16.
319. S. Wold, K. Esbensen, P. Geladi. "Principal component analysis". *Chemometrics Intellig. Lab. Syst.* 1987. 2(1): 37-52.
320. T. Azzouz, R. Tauler. "Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples". *Talanta* 2008. 74(5): 1201-1210.

## References

---

321. M. Garrido, F. Rius, M. Larrechi. "Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes". *Analytical and Bioanalytical Chemistry* 2008. 390(8): 2059-2066.
322. R. Tauler, B. Kowalski, S. Fleming. "Multivariate curve resolution applied to spectral data from multiple runs of an industrial process". *Anal. Chem.* 1993. 65(15): 2040-2047.
323. E. Peré-Trepat, S. Lacorte, R. Tauler. "Solving liquid chromatography mass spectrometry coelution problems in the analysis of environmental samples by multivariate curve resolution". *Journal of Chromatography A* 2005. 1096(1–2): 111-122.
324. J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa. "Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions". *Journal of Chemometrics* 1998. 12(3): 183-203.
325. R. Tauler, S. Lacorte, M. Guillaumon, R. Cespedes, P. Viana, D. Barceló. "Chemometric modeling of main contamination sources in surface waters of Portugal". *Environmental Toxicology and Chemistry* 2004. 23(3): 565-575.
326. S. Piqueras, L. Duponchel, R. Tauler, A. De Juan. "Resolution and segmentation of hyperspectral biomedical images by Multivariate Curve Resolution-Alternating Least Squares". *Analytica chimica acta* 2011. 705(1-2): 182-192.
327. A. de Juan, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler. "Chemometric Tools for Image Analysis". In *Infrared and Raman Spectroscopic Imaging*, Wiley-VCH Verlag GmbH & Co. KGaA: 2009; Pp 65-109.
328. A. de Juan, M. Maeder, T. Hancewicz, R. Tauler. "Use of local rank-based spatial information for resolution of spectroscopic images". *J. Chemom.* 2008. 22(5): 291-298.
329. J.M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco. "Study of pharmaceutical samples by NIR chemical-image and multivariate analysis". *TrAC Trends in Analytical Chemistry* 2008. 27(8): 696-713.
330. S. Piqueras, J. Burger, R. Tauler, A. de Juan. "Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution". *Chemometrics and Intelligent Laboratory Systems* 2011. DOI: 10.1016/j.chemolab.2011.12.004(DOI: 10.1016/j.chemolab.2011.12.004).
331. J. Jaumot, R. Gargallo, A. de Juan, R. Tauler. "A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB". *Chemometrics Intellig. Lab. Syst.* 2005. 76(1): 101-110.
332. D.J. Leggett. "Numerical analysis of multicomponent spectra". *Analytical Chemistry* 1977. 49(2): 276-281.
333. R. Tauler, A. Smilde, B. Kowalski. "Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution". *Journal of Chemometrics* 1995. 9(1): 31-58.
334. A. de Juan, Y. Vander Heyden, R. Tauler, D.L. Massart. "Assessment of new constraints applied to the alternating least squares method". *Analytica chimica acta* 1997. 346(3): 307-318.
335. A.d. Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder. "Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis". *TrAC, Trends Anal. Chem.* 2004. 23(1): 70-79.
336. A. de Juan, S. Navea, J. Diewok, R. Tauler. "Local rank exploratory analysis of evolving rank-deficient systems". *Chemometrics and Intelligent Laboratory Systems* 2004. 70(1): 11-21.