Multivariate Curve Resolution of Hyperspectral Images

Neal B. Gallagher[1], Jeremy M. Shaver[1], Barry M. Wise[1] and Elaine B. Martin[2]
[1]Eigenvector Research, P.O. Box 561, Manson, WA, USA [2]University of Newcastle, Merz Court, Newcastle upon Tyne, United Kingdom

EIGENVECTOR RESEARCH INCORPORATED

Introduction

Multivariate curve resolution (MCR) is a powerful technique for extracting estimates of pure component concentrations **C** and spectra **S** from multivariate images (MI). I.e. MCR extracts *chemical information* from a data set. This is in contrast to principal components analysis (PCA) that extracts orthogonal factors that simply capture variance in a data set.

Advantages of the MCR decomposition are 1) chemical specific visualization, 2) statistical monitoring based on chemistry, 3) estimates of **S** that are chemical matrix dependent i.e. are more relevant for the problem than estimates from pure analytes, and 4) better interpretability compared to PCA factors. The **C** can be used to create "concentration images" can be used for process monitoring, quality control, change and anomaly detection, land mapping in remote sensing, and estimating spatial statistics.

Two challenges for MCR in MI are 1) initializing the decomposition and 2) lack of selectivity in the image. Novel methods have been developed that utilize knowledge of the chemistry and physics of the sensing systems.

Analysis was performed using MATLAB and the PLS_Toolbox[1].

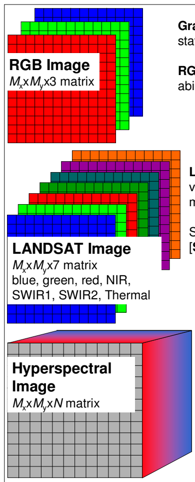
The MCR model is based on classical least squares (CLS)

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E}$$

where **X** is a M_xN matrix of measured spectra, **E** is a M_xN matrix of residuals,
M is the number of image pixels and N is the number of spectral channels,
C is a M_xK matrix of concentrations and **S** is a N_xK matrix of spectra to be determined,
and K is the number of factors in the model.

Algorithm used constrained alternating least-squares. Can also use positive matrix factorization algorithm.

Gallagher, N.B., Shaver, J.M., Martin, E.B., Morris, J., Wise, B.M. and Windig, W., "Curve resolution for images with applications to TOF-SIMS and Raman", *Chemo. and Intell. Lab. Sys.* in print (2003).



Gray-Scale Image can be used to obtain simple spatial statistics (e.g. sizes, shapes, and distances).

RGB Image. Additional channels (e.g. color) ⇒ Adds ability to discriminate different materials.

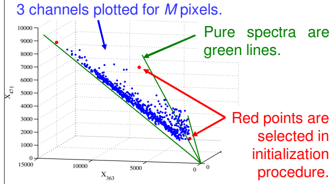
LANDSAT Image. Additional spectral channels (e.g. visible + IR). Additional ability to discriminate different materials.

Starts yielding chemical information.
[See example at top of next column.]

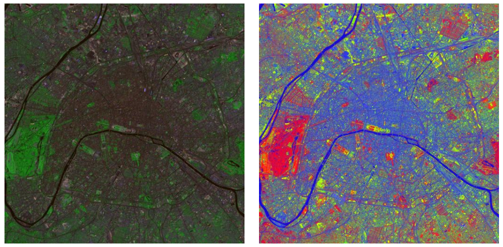
Hyperspectral refers to mulitple (mega, omni, giga) channels that yield higher selectivity ⇒ chemical discrimination.
Additional ability to discriminate different materials.
[See examples in poster.]

Combined spatial, physical and chemical information.
The objective is to extract additional information to improve understanding, process monitoring and quality control.

Initialization (using DISTSLCT)



A "purity" technique was utilized to initialize the decomposition. More recent results with a newer technique are showing even more promise. The methods are based on the physical argument that pure spectra must lie on the exterior of the data cloud and mixtures are on the interior of the data cloud.



RGB False Color Image from LANDSAT (blue, NIR, and thermal channels)

LANDSAT images of Paris

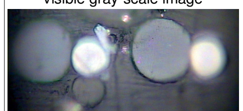
RGB False Color Image using initial MCR solution of all 7 channel from LANDSAT.

Of the 7 LANDSAT channels which 3 should be used for visualization?
Visualization is enhanced by a) using a weighted average of ALL 7 channels
(i.e. making concentration images from DISTSLCT estimate), and b) using contrasting.

Raman Image of Embedded Polymer Fibers

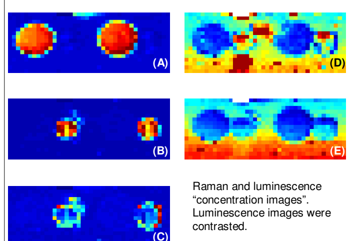
(DISTSLCT initial guess, Non-negative + baseline functional constraints)

visible gray-scale image



Raman image of a sample consisting of three types of fibers (PET, Twaron, and cellulose) embedded in a Raman-inactive medium. 16 by 40 pixel. Mostly Raman signal with some low-level luminescence from the fibers and imbedding medium. Functional constraints used to account for luminescence offset. Where are the different materials?

Functional constraints (e.g. utilizing kinetics- or physics-based constraints) helps when measurements have a lack of selectivity (e.g. due to luminescence in Raman).



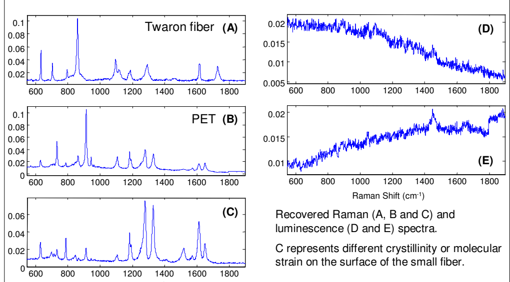
Raman components for large PET fibers (A) and smaller Twaron fibers.

Luminescence components (D and E).

Cellulotic fiber is observed only in the luminescence image (D).

Raman component (C) corresponded to different crystallinity or molecular strain on the outside of the smaller Twaron fiber.

Raman and luminescence "concentration images". Luminescence images were contrasted.



Recovered Raman (A, B and C) and luminescence (D and E) spectra.

C represents different crystillinity or molecular strain on the surface of the small fiber.

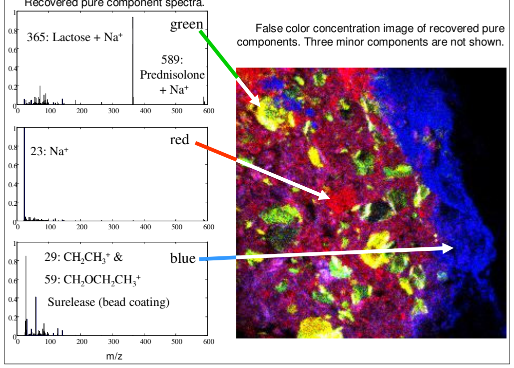
TOF-SIMS Image of Drug Bead

(DISTSLCT initial guess, Non-negative constraints only)

Time-of flight secondary ion mass spectrometry (TOF-SIMS) of a drug bead. The drug bead contains a core of prednisolone sodium metasulfobenzoate in microcrystalline cellulose and lactose coated with 1 part amylose and 4 parts surelease (ethylcellulose, ammonium hydroxide, coconut oil and oleic acid). Measurements were made using a TRIFT II instrument (Physical Electronics, Eden Prarie, MN). The image of interest is 250 x 250 μm² region of a cross-sectioned bead. Where is the drug, coating and filler? What is the distribution?

A.M. Belu, M.C. Davies, J.M. Newton, and N. Patel, "TOF-SIMS Characterization and Imaging of Controlled-Release Drug Delivery Systems", *Anal. Chem.*, **72**(22), 5625-5638 (2000).

Recovered pure component spectra.

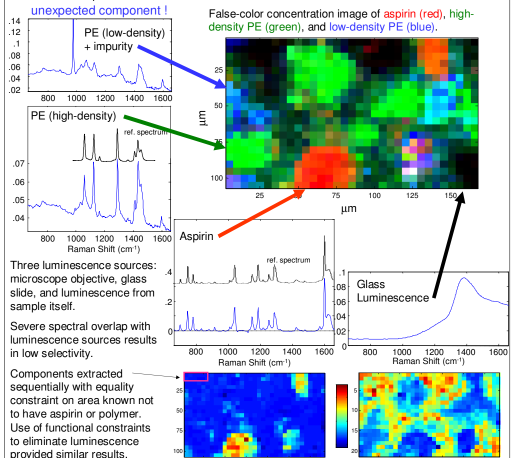


False color concentration image of recovered pure components. Three minor components are not shown.

Raman Image of Aspirin / Polymer Mixture

(DISTSLCT initial guess, Sequential ALS, >0 + equality constraints)

Mixture of aspirin and high-density polyethylene (PE) particles on a glass slide. Data collected on a fiber-coupled Raman microscope (Hololab 5000, Kaiser Optical Systems, Inc.) over a 105 x 165 μm² region at 5 μm steps using a 10x Olympus objective. Laser wavelength was 785 nm and spectra were measured from 0 to 1926 cm⁻¹ at 1 cm⁻¹ intervals. Data were truncated to 600 to 1660 cm⁻¹ and re-sampled to 2 cm⁻¹. Where and what are the materials.



False-color concentration image of aspirin (red), high-density PE (green), and low-density PE (blue).

Three luminescence sources: microscope objective, glass slide, and luminescence from sample itself.

Severe spectral overlap with luminescence sources results in low selectivity.

Components extracted sequentially with equality constraint on area known not to have aspirin or polymer. Use of functional constraints to eliminate luminescence provided similar results.

Mid-IR Image of Corn Kernel Cross-Section

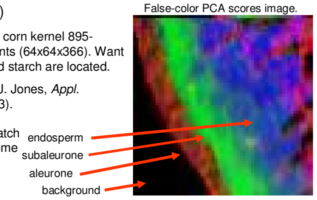
(DISTSLCT initial guess)

Mid-IR of a cross-section of a corn kernel 895-3705.5 cm⁻¹ 7.7 cm⁻¹ increments (64x64x366). Want to know where the protein and starch are located.

B.O. Budevska, S.T. Sum, T.J. Jones, *Appl. Spectrosc.*, **57**, 124-131 (2003).

Recovered spectra closely match library spectra. Unknowns (some not shown) were also found.

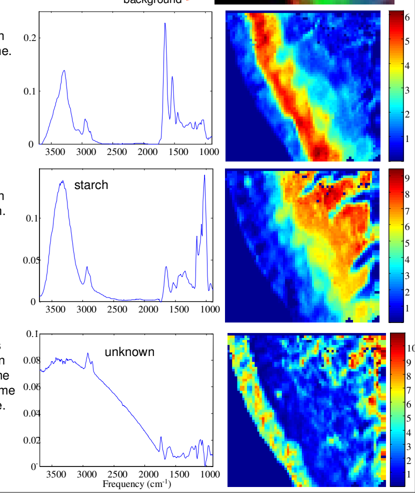
False-color PCA scores image.



The protein is concentrated in the subaleruone.

The starch is concentrated in the endosperm.

An unknown is concentrated in the center of the kernel, and some in the alerione.



TOF-SIMS of Cloth

(DISTSLCT only)

Time-of flight secondary ion mass spectrometry (TOF-SIMS) of a cloth sample.

Highly contrasted (low signal-to-noise).

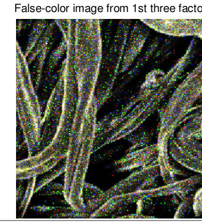
Mass: 19 (red), 213 (green), 83 (blue).

White indicates presence of all three masses.

Sampling likely affected by sample geometry.

What is in the sample and where?

False-color image from 1st three factors.



Conclusions and Future Work

Multivariate curve resolution (MCR) can be used to extract ***chemical information*** from multivariate images (e.g. using NIR, MID-IR, Visible+, Raman, mass spec, NMR, ...). This yields 1) chemical specific visualization, 2) statistical monitoring based on chemistry, 3) estimates of **S** that are chemical matrix dependent i.e. are more relevant for the problem than estimates from pure analytes, and 4) better interpretability compared to PCA factors.

Constraints can be used to incorporate known physics and chemistry into the decomposition. Results can be coupled with traditional image analysis tools (e.g. edge detection, size and size distribution analysis, texture analysis, ...). Novel methods can couple MCR and traditional image analysis tools.

Novel initialization procedure 1) provides a good first estimate that is obtained quickly and can be interpreted on it's own merits, 2) yields faster convergence of the decomposition.