Resolution of hyperspectral images. Pre-, in- and post-processing.

Anna de Juan\textsuperscript{1}, Marcel Maeder\textsuperscript{2}, Tom Hancewicz\textsuperscript{3}, Ludovic Duponchel\textsuperscript{4} and Romà Tauler\textsuperscript{5}.

1. Chemometrics group. Dept. Analytical Chemistry. Universitat de Barcelona. anna.dejuan@ub.edu
2. Dept. of Chemistry. The University of Newcastle. Newcastle (Australia).
3. Unilever Research & Development, NJ 0702, U.S.
4. LASIR, Université des Sciences et Technologies de Lille (France).
5. IIQAB-CSIC, Barcelona.
Outline

1. The data sets.
2. Single and multiset image resolution.
3. Resolution pre-processing
   a. Signal processing.
   b. Image compression
4. Resolution in-processing
   a. Introduction of constraints.
   b. Multiset image analysis
5. Resolution post-processing
   a. Identification of image constituents
   b. Quantitative information.
Data sets

**Single images**

**Emulsion**

- **Size** $(x \times y \times \nu)$
  - $60 \times 60 \times 256$
- **Wavenumber range**
  - $(950-1800 \text{ cm}^{-1})$

**Kidney calculus**

- **Size** $(x \times y \times \nu)$
  - $60 \times 70 \times 256$
- **Wavenumber range**
  - $(400-3500 \text{ cm}^{-1})$

Available spectra library

Data sets

Multiset images

Pill data set

<table>
<thead>
<tr>
<th>Excipient</th>
<th>20% active</th>
<th>Size $$(x \times y \times v)$$</th>
<th>Wavenumber range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>45 $\times$ 45 $\times$ 576</td>
<td>(609-1173 cm)</td>
</tr>
<tr>
<td>40% active</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60% active</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80% active</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Independent 2D images

RAMAN intensity

Wavenumbers

Multilayer emulsion image

<table>
<thead>
<tr>
<th>Size $$(x \times y \times z \times v)$$</th>
<th>Wavenumber range</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 $\times$ 23 $\times$ 10 $\times$ 229</td>
<td>(950-1800 cm$^{-1}$)</td>
</tr>
</tbody>
</table>

Layer mean spectra

Depth

3D image

Single image resolution

Recovery of pure spectra ($S^T$) and distribution maps ($C$) of pure constituents from raw image data ($D$)
Multiset image resolution

Distribution maps are obtained for each image/layer

Constituent identity is preserved among layers/images (single $S^T$ matrix)
Resolution pre-processing

Signal preprocessing

✓ De-noising and/or weighting.
✓ Background correction.

Image compression

✓ General approaches
  • Binning, wavelet-based,…
✓ Resolution-oriented approaches.
  • Based on exploratory tools
Resolution pre-processing

- Signal pre-processing
  - De-noising and/or weighting.
  - Background correction.

![Graphs showing raw data, de-noising, and background correction](image)

Take into account noise presence and structure
Flexible background correction methods
Pre-processed raw data improve final results

Resolution pre-processing

○ Image compression for resolution

Spatial compression

\[ D_{\text{comp}} \]

\[ D \]

Resolution of the compressed data

\[ D_{\text{comp}} \]

\[ C_{\text{comp}} \]

\[ S^T \]

Spectral compression

\[ D_{\text{comp}} \]

\[ C \]

\[ S^T_{\text{comp}} \]

\[ (LS D, S) \]

\[ (LS D, C) \]

LS recovery of the compressed direction
Image compression for resolution

- General approaches
  - ✓ Decrease image size.
    - • Binning, wavelet selection,…

- Resolution-oriented approaches.
  - ✓ Most unmixed pixels or spectral channels.
    - • SIMPLISMA (purest pixels/spectral channels and neighbors are selected)
    - • Local rank analysis (pixels with lowest rank, least compound overlap are selected)
Image compression for resolution

- Resolution-oriented approaches.
  - SIMPLISMA. Selects purest pixels or spectral channels in the image.

Compressed image is formed by purest pixels \((a,b,c,d)\) and surrounding neighbors in 2D.
Image compression for resolution

‘Resolution-oriented approaches. SIMPLISMA

Resolution on compressed data (pixel subset)
\( D_{\text{comp}} (36 \times 253) \)

Resolution on full image
\( D (3600 \times 253) \)
Image compression for resolution

Resolution-oriented approaches.

✓ Local rank analysis (Fixed size Image Window- Evolving Factor Analysis, FSIW-EFA). Provides local rank maps.

Kidney image

Local rank map

Yellow (1) Orange (2) Red (3)

Compressed image is formed by pixels with lowest rank (1 and 2)
Image compression for resolution

- Resolution-oriented approaches. FSIW-EFA.

Results on rank = 1,2 pixels

Raw data results

- LS
  
  \( (D, S^T) \)

- \( S^T \) (not shown)

- \( C_{comp} \) (not shown)

- \( U \)

- \( B \)

UNIVERSITAT DE BARCELONA
Resolution in-processing

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

\[ D = CS^T + E \]

1) Determination of the number of compounds (D).
2) Construction of an initial estimate (C or ST).

Purest spectra (SIMPLISMA,…).

3) Given D and ST, constrained calculation of C.
4) Given D and C, constrained calculation of ST.
5) Check whether CST is similar enough to D_{PCA}. If not, go to 3.
Resolution ‘in-processing’

- **Introduction of constraints.**
  - ✓ Non-negativity
  - ✓ Equality constraints (fixed spectra)
  - ✓ Local rank/selectivity constraints

- **Multiset arrangements**
  - ✓ Spectral consistency
  - ✓ Quantitative information.
Resolution ‘in-processing’

- Introduction of local rank/selectivity constraints.

Emulsion data

Local rank map

Reference spectral information

Equality constraint matrix

- A absent
- B absent
- C absent
- D absent

Allows coding absence of constituents in pixels

Total number of pixels: 3600
Constrained pixels: 1177

Better definition of compounds
Decrease of rotational ambiguity

Resolution ‘in-processing’

- Multiset arrangements

\[
\begin{align*}
\text{Image 1} & \quad \text{Image 2} & \quad \text{Image 3} & \quad \ldots & \quad \text{Image n} \\
\end{align*}
\]

\[
\begin{align*}
D & \quad = & \quad \text{C} \\
\end{align*}
\]

- Single matrix of pure spectra (spectral consistency among images)

- Distribution maps
  - Higher robustness in data description.
  - Complementary information among images is used.
    - Easier modelling of minor compounds.
    - Decrease of ambiguity.
  - ‘Quantitative information’ is obtained.
Individual vs. multiset image analysis

Layers

Multilayer emulsion

Top

Bottom

Poor definition of compound identity

Spectral consistency
Decrease of rotational ambiguity
Resolution post-processing

- Identification of image constituents ($S^T$ matrix)

- Quantitative information
  - Within a distribution map (pixel-to-pixel relative quantitative information) ($C$ matrix)
  - Within image (signal contribution) ($C$ and $S^T$ matrices)
    - Relates components among them.
  - Between-images (relative quantitative information) ($C$ matrix)
    - Relates the concentration of a constituent in different images.
Resolution post-processing

Identification of image constituents

Projection on the space of resolved spectra ($S^T$ matrix)

- Library spectrum (target)
- Error in Projection (EP)
- Projected target
- Defined by the resolved spectra
- Targets with the smallest EP
- Identified compounds

- Small data set ($S$) defines the image space.
- ‘Cleaner’ than the original image space
- The image space is defined by real spectra
Resolution post-processing

Identification of image constituents

Projection on the space of resolved spectra (Sᵀ matrix)

Plot of error in projection (EP)

Kidney image

- Easy matching of accepted candidates with resolved spectra.
- Identified compounds can be used as equality constraints to improve resolution of unknowns.

Potential image constituents

Targets

EP

Resolved spectra

Library spectra

Ca oxalate (whewellite)

Ca oxalate (wedelllite)

Dahllite

Universitat de Barcelona
Resolution post-processing

- Identification of image constituents ($S^T$ matrix)
- Quantitative information
  - **Within a distribution map** ($C$ matrix)
    - Pixel-to-pixel relative quantitative information (heterogeneity)
  - **Within image** (signal contribution) ($C$ and $S^T$ matrices)
    - Relates components among them.
  - **Between-images** (relative quantitative information) ($C$ matrix)
    - Relates the bulk concentration of a constituent in different images.
Resolution post-processing

‘Quantitative’ information

✓ Within image (signal contribution)

\[ \sum c_i s_i^T = \frac{\sum c_i s_i^T}{\sum CS^T} \]

Indicates approximately the prevalence of each constituent in the image.

✓ Cannot be straightforwardly interpreted as quantitative information (different absorptivities among constituents).
Resolution post-processing

- ‘Quantitative’ information
- Between-images (relative quantitative information)

\[
\text{Image 1} = \text{Image 2} + \ldots + \text{Image n}
\]

\[
\sum_{i=1}^{n} c_{ik} = \sum_{i=1}^{n} c_{i1}
\]

Absolute quantitation

Bulk reference concentrations

Relative quantitation

Representative of the scanned surface

Applicable to series of independent images, to a process monitored by imaging,...
Resolution post-processing

- Quantitative information
- ✓ Between-images (relative quantitative information)

**Pill multiset example**

<table>
<thead>
<tr>
<th>% active</th>
<th>% excipient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Calculated</td>
</tr>
<tr>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>23.23</td>
</tr>
<tr>
<td>40</td>
<td>39.15</td>
</tr>
<tr>
<td>60</td>
<td>52.00</td>
</tr>
<tr>
<td>80</td>
<td>85.62</td>
</tr>
</tbody>
</table>

![Graph showing absolute quantitation](image)

- $r(\text{active}) = 0.974$
- $r(\text{excipient}) = 0.985$
Conclusions

- Resolution-oriented image compression provides smaller images with efficient and more easily resolvable information.

- Image constraints go beyond non-negativity and allow including selective information.

- Multiset image resolution helps to the definition of minor constituents in certain layers and provides less ambiguous solutions. More robust characterisation of compounds is achieved due to the single $S^T$ matrix used.

- Resolution post-processing takes advantage of the information on $S^T$ and $C$ for identification and quantitative information about the image constituents.