Joint Research Centre (JRC)

*Multi-way models in chemometrics and (mostly CP) algorithms*

Giorgio Tomasi
Giorgio.tomasi@ec.europa.eu

IHCP - Institute for Health and Consumer Protection
*Ispra - Italy*

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/
Multilinear modelling

Data tensors produced by from several analytical techniques

- Spectroscopy (fluorescence, 2D/3D NMR)
- Hyphenated methods (chromatography)
- Process monitoring (industry, environment, chemistry)

Standard methods like PCA or PLS can be used but

- The number of model parameters is too large
- Additional relations given by $N$-way structure are lost

Extend these methods to $N$- way case

- PCA $\rightarrow$ Tucker models ($N$-way PCA)
- Factor Analysis $\rightarrow$ CP / PARAFAC2
- $\rightarrow$ Restricted Tucker / CP
- PCR $\rightarrow$ Regression on scores
- $PLSn$ (for $n \leq 2$) $\rightarrow$ $N$ PLS$n$ (arbitrary $n$ and $N$)
CP is an $N$-linear model for an $N$-way array

For an $I_1 \times I_2 \times \ldots \times I_N$ array $\mathbf{X}$, it is expressed as

$$\mathbf{X}_{i_1i_2\ldots i_N} = \sum_{f=1}^{F} \prod_{n=1}^{N} \mathbf{a}_{i_n}^{(n)} + \mathbf{r}_{i_1i_2\ldots i_N}$$

$\mathbf{X}_{i_1i_2\ldots i_N}$ denotes the array elements

$\mathbf{a}_{i_n}^{(n)}$ are the model parameters

$F$ is the number of fitted components

$\mathbf{r}_{i_1i_2\ldots i_N}$ denotes the residuals

$\mathbf{A}_n$ loading matrix

$$\mathbf{A}_n \equiv \{ \mathbf{a}_{i_n}^{(n)} | i_n = 1 \ldots I_n, f = 1 \ldots F \}$$
Solution is unique under mild conditions

- Non-trivial transformations of the loading matrices lead to loss of fit
- Scaling, sign, permutation indeterminacy are easily fixed

The model corresponds to the physical model underlying certain types of chemical/physical data

- Fluorescence
- Chromatographic data
- Kinetic data
- ...

Model’s parameters are readily interpretable
Absorbed light at $i$-th wavelength

$$x_i \sim x_i \varepsilon_i$$

Part of the light is reemitted at $j$-th wavelength

$$x_{ij} \sim x_i \varepsilon_i \lambda_j$$

For $K$ solutions and $F$ fluorophores it becomes

$$x_{ijk} \sim \sum_{f=1}^{F} x_{kj} \varepsilon_{if} \lambda_{jf}$$

Including time decay $\tau$ and quantum yield $\phi$

$$x_{ijkb} \sim \sum_{f=1}^{F} x_{kj} \varepsilon_{if} \lambda_{jf} \phi_{ij} \tau_{mf}$$
PAH Metabolites in Fish Bile

Fish bile data set

Polycyclic Aromatic Hydrocarbons are metabolised (oxydised)

The metabolites are discharged in the fish bile

Array $42 \times 126 \times 43$

Four factors explain 99% of TSS

Two factors are related to biological factors

Two factors are related to the exposure to PAHs (metabolites)
Fish bile data set

Polycyclic Aromatic Hydrocarbons are metabolised (oxydised)

The metabolites are discharged in the fish bile

Array 42 × 126 × 43

Four factors explain 99% of TSS

Two factors are related to biological factors

Two factors are related to the exposure to PAHs (metabolites)
Fish bile data set

Polycyclic Aromatic Hydrocarbons are metabolised (oxydised)

The metabolites are discharged in the fish bile

Array $42 \times 126 \times 43$

Four factors explain 99% of TSS

Two factors are related to biological factors

Two factors are related to the exposure to PAHs (metabolites)
Absorbed light at $i$-th wavelength

$$x_i \cong \chi \varepsilon_i$$

Part of the light is reemitted at $j$-th wavelength

$$x_{ij} \cong \chi \varepsilon_i \lambda_j$$

For $K$ solutions and $F$ fluorophores it becomes

$$x_{ijk} \cong \sum_{f=1}^{P} \chi_{kf} \varepsilon_{if} \lambda_{jf}$$

Including time decay $\tau$ and quantum yield $\varphi$

$$x_{ijkm} \cong \sum_{f=1}^{P} \chi_{kf} \varepsilon_{if} \lambda_{jf} \varphi_{if} T_{mf}$$
A chromatograph separates compounds in a mixture based on their physical/chemical properties (polarity, size, boiling point, etc).

The order of the data depends on the detection.

Two chromatographs can be combined also increasing the order of the data.
Light absorbed at \( i \)-th wavelength

\[ x_i \approx \chi \varepsilon_i \]

Light absorbed at \( j \)-th time and \( i \)-th wavelength

\[ x_{ij} \approx \chi \varepsilon_i T_{ij} \]

For \( F \) compounds and \( K \) samples

\[ x_{ijk} \approx \sum_{f=1}^{F} \chi_{kf} \varepsilon_{ijf} T_{ijf} \]

Since retention time changes across sample modes

\[ x_{ijk} \approx \sum_{f=1}^{F} \chi_{kf} \varepsilon_{ijf} T_{ijf}^{(k)} \]
PARAFAC2 is an \((N+1)\)-linear model for an \(N\)-way array

For an \(I_1 \times I_2 \times \ldots \times I_N\) array \(\mathbf{X}\), it is expressed as

\[
x_{i_1i_2\ldots i_N} = \sum_{f=1}^{F} \left( \sum_{f'=1}^{F} p_{i_2}^{(i_2)} h_{ff'} \right) \prod_{n=2}^{N} a_{i_n}^{(n)} + r_{i_1i_2\ldots i_N}
\]

Subject to the constraint:

\[
\sum_{i_1=1}^{I_1} p_{i_1f}^{(i_2)} p_{i_1f}^{(i_2)} = \delta_{ff'}
\]

It was devised to fit a CP model to covariance matrices leaving one mode implicit

Advantage: rotationally unique if "\(I_2\" sufficiently large

Problems: can handle limited types of shift and shape changes
Light absorbed at $i$-th wavelength

$$x_i \equiv \chi \varepsilon_i$$

Light absorbed at $j$-th time and $i$-th wavelength

$$x_{ij} \equiv \chi \varepsilon_i T_{ij}$$

For $F$ compounds and $K$ samples

$$x_{ijk} \equiv \sum_{f=1}^{F} \chi_{kf} \varepsilon_{if} T_{jf}$$

Since retention time changes across sample modes

$$x_{ijk} \equiv \sum_{f=1}^{F} \chi_{kf} \varepsilon_{if} T^{(k)}_{jf}$$
The shifted CP includes shifts $\tau_{ij}$ as additional set of parameters.

For an $I_1 \times I_2 \times \ldots \times I_N$ array $\chi$, it is expressed as

$$x_{i_1 \ldots i_N} = \sum_{f=1}^{F} S_{\tau_{ij}}(a^{(i)}_{ij}) \prod_{n=2}^{N} a^{(n)}_{i_n} + \tau_{i_1 \ldots i_N}$$

$$\tilde{x}_{i_1 \ldots i_N} = \sum_{f=1}^{F} \tilde{a}_{\tilde{j},f} e^{-2\pi \nu_{\tilde{j}} I_1 \tau_{ij}} \prod_{n=2}^{N} a^{(n)}_{i_n} + \tilde{\tau}_{i_1 \ldots i_N}$$

where

$S_{\tau_{ij}}$ \hspace{2cm} Shift operator

$\tilde{x}_{i_1 \ldots i_N}, \tilde{a}_{\tilde{j},f}, \tilde{\tau}_{i_1 \ldots i_N}$ \hspace{2cm} Frequency domain equivalents to $x_{i_1 \ldots i_N}, a_{ij}, \tau_{i_1 \ldots i_N}$

$e^{-2\pi \nu_{\tilde{j}} I_1 \tau_{ij}}$ \hspace{2cm} Shift term in the frequency domain
The physical model for molecular absorbance detection is

\[ x_{ijk} \approx \sum_{f=1}^{F} \chi_{kf} \varepsilon_{if} \left( T_{jf} + \Delta T_{jf}^{(k)} \right) \]

Shifted CP assumes shape invariance of the factors between samples

Allowed shift has to be constrained

Local minima are hit in the more complex case

There are problems at the boundaries
The physical model for molecular absorbance detection is

\[ x_{ijk} \approx \sum_{f=1}^{F} \chi_{kf} e_{if} \left( T_{jf} + \Delta T_{jf}^{(k)} \right) \]

Shifted CP assumes shape invariance of the factors between samples

Allowed shift has to be constrained

Local minima are hit in the more complex case

There are problems at the boundaries
The physical model for molecular absorbance detection is

\[ x_{ijk} \approx \sum_{f=1}^{F} \chi_{kf} \varepsilon_{if} \left( T_{jf} + \Delta T_{jf}^{(k)} \right) \]

Shifted CP assumes shape invariance of the factors between samples

Allowed shift has to be constrained

Local minima are hit in the more complex case

There are problems at the boundaries
The physical model for molecular absorbance detection is

\[ x_{ijk} \approx \sum_{f=1}^{F} \chi_{kf} \varepsilon_{if} \left( T_{jf} + \Delta T^{(k)}_{jf} \right) \]

Shifted CP assumes shape invariance of the factors between samples

Allowed shift has to be constrained

Local minima are hit in the more complex case

There are problems at the boundaries
SCP and PARAFAC2: challenges

Changes in shape with SCP (shifted CANDECOMP/PARAFAC)

Constraints in PARAFAC2

Shift in two dimensions

- Hyphenated chromatographic columns (LC × LC, GC × GC)
- Gel Electrophoresis
- Currently handled with preprocessing

Both PARAFAC2 and SCP can be extended

\[ x_{ijk\ldots} = \sum_{f=1}^{F} a_{ijf}^{(k)} b_{jif}^{(k)} c_{jif} \ldots + r_{ijk\ldots} \]

\[ x_{ijk\ldots} = \sum_{f=1}^{F} S_{r_{ijf}} \left( a_{ijf} \right) S_{u_{jif}} \left( b_{jif} \right) c_{jif} \ldots + r_{ijk\ldots} \]
Constraints are essential to obtain interpretable / physically meaningful results

Typical constraints include:

*Non negativity* chromatography, spectroscopy,…

*Orthogonality*: exploratory data analysis, design of experiments (progressively removed)

*Unimodality*: chromatography

*Equality*

- Offsets (ones): design of experiments
- Selectivity (zeros): spectroscopy, chromatography, curve resolution in general
- Linear dependence: kinetics, flow injection analysis (PARALIND/CONFAC)
GEneralised Multiplicative Analysis Of Variance

In designed experiments, several factors are varied systematically to maximise information.

The variation is modeled as

\[ x_{ijkl...} = a_i b_j c_k + d_i e_j f_k + g_i h_j + r_{ijkl...} \]

Or more generally

\[ x_{i_1...i_N} = \sum_{f=1}^F \prod_{n=1}^N a_{i_n}^{(m)} + r_{i_1...i_N} \]

s.t. \( a_{i_n}^{(m)} = 1 \) for \( i_m = 1...I_m \),

\( m \in M \subset \{1,...,N\} \),

\( f \in F \subset \{1,...,F\} \)

Typical problems:

- Offset terms
- Many missing values
- Small sizes
Color changes in fresh beef during storage*

Factors:        # Levels

Storage times (a)   5
Temperature (b)     3
Exposure time to light (c) 3
O₂ content in headspace (d) 3
Muscle # (e)       6

Data set:
324 samples (reduced design)
60% missing out of 810 from full design
L, a, b colour components (only a used)

Model
Validated with LOO cross validation

\[ y_{ijklm} = a_i b_j c_k d_l + e_m + t_{ijklm} \]

*Bro R., Jakobsen M. J. Chemometrics, 2002 (16), 294 - 304
Tucker models

Tucker-N is an \((N+1)\)-linear model for an \(N\)-way array

For an \(I_1 \times I_2 \times \ldots \times I_N\) array \(\mathcal{X}\), it is expressed as

\[
x_{i_1 i_2 \ldots i_N} = \sum_{f_1=1}^{F_1} \sum_{f_N=1}^{F_N} g_{f_1 \ldots f_N} \prod_{n=1}^{N} a^{(n)}_{i_n f} + r_{i_1 i_2 \ldots i_N}
\]

- \(F_n\) is the number components in mode \(n\)
- \(G\) core array (size \(F_1 \times F_2 \times \ldots \times F_N\))

There are also Tucker-\(m\) models only \(m\) modes are reduced

\[
x_{i_1 i_2 \ldots i_N} = \sum_{f_1=1}^{F_1} \sum_{f_{N-m}=1}^{F_{N-m}} g_{f_1 \ldots i_{N-m} \ldots i_N} \prod_{n=1}^{N} a^{(n)}_{i_n f} + r_{i_1 i_2 \ldots i_N}
\]
Tucker models: a whole family

Tucker-N

\[
\mathbf{X} = \sum_{i=1}^{r} \mathbf{A}_i \mathbf{B}_i \mathbf{C}_i + \mathbf{R}
\]

Block Tucker

\[
\mathbf{X} = \sum_{i=1}^{r} \mathbf{A}_i \mathbf{B}_i \mathbf{C}_i + \mathbf{R}
\]

Restricted Tucker

\[
\mathbf{X} = \mathbf{A}_1 \mathbf{B}_1 \mathbf{C}_1 + \mathbf{R}
\]
Exploratory analysis

No need for underlying N-linear generating process for it to work (unlike CP)

Data compression to speed up fitting / constrain models

(when constrained) they correspond to the physical model of certain types of data

Tucker(N-1) models have been proposed when interactions occur between fluorophores

(L,L,1) block decomposition are suitable for MS$^2$ and 2D NMR data

Ad hoc model can be designed for "Rank deficient" data (Flow Injection Analysis data)

... and can be (at least partially) unique

Model’s parameters are readily interpretable
Biodegradation study

4 oil types (crude, LFO, HFO and bilge/ship oil)

4 bacterial treatments (Sterile, U, R and M)

5 incubation times (20, 54, 132, 224 and 364 days)

44 mass fragments measured in SIM mode.

3 m/z of interest:
  180 Methylfluoranthenes
  192 Methylphenanthrenes
  198 Methyldibenzothiophenes

144 Measurements

33 Reference oil samples (mix of crude and bilge)
Rearranging concatenated signals in a $5 \times 4 \times 4 \times 355$

Centring across treatments

Optimal dimensionality: $[2 \ 3 \ 3 \ 3]$

First five combinations explain 90% of the model variation

**Interpretation**

Combination (1,1,1,1): average degradation

Others are specific processes for treatments and oils
Tucker: exploratory analysis
In these techniques, more than one CP component is required for the same analyte

(l, l, 1) block decomposition naturally emerge

Blocks could be partially overlapping

Structure elucidation (proteins, large organic molecules)

Complex mixtures

Ion source

Q1: Mother ions selection (mass filter)

Q2: Break up of mother ion (not a mass filter)

Q3: Child ions selection (mass filter)

*http://en.wikipedia.org/wiki/Humic_acid
The structure is the same as that of Tucker, but weights and scores are found according to a different criterion.

Maximise the covariance between the scores on the array $\mathbf{X}$ and those on the predicted array $\mathbf{Y}$.

For an $I_1 \times I_2 \times \ldots \times I_N$ tensor $\mathbf{X}$ and a $J_1 \times J_2 \times \ldots \times J_N$ tensor $\mathbf{Y}$ the $N$PLS1 model is expressed as:

$$
\mathbf{x}_{i_1i_2...i_N} = \sum_{f_1=1}^{F} \cdots \sum_{f_N=1}^{F} g_{f_1...f_N} t_{i_1f_1} \prod_{n=2}^{N} w_{i_nf_n}^{(n)} + r_{i_1i_2...i_N}
$$

$$
\mathbf{y}_{j_1j_2...j_M} = \sum_{f_1=1}^{F} \cdots \sum_{f_M=1}^{F} g_{j_1...j_M} u_{i_1f_1} \prod_{m=2}^{M} q_{i_mf_m}^{(m)} + r_{j_1j_2...j_M}
$$

$w_{i_nf_n}^{(n)}$ denotes the weights in the $n$-th mode of $\mathbf{X}$.

$q_{i_mf_m}^{(m)}$ denotes the weights in the $m$-th mode of $\mathbf{Y}$.

$t$ are the scores on $\mathbf{X}$.

$u$ are the scores on $\mathbf{Y}$.

$F$ is the number of PLS components.
PARAFAC
- Alternating Least Squares (1970)
  - Rotation Enhanced ALS (2010)
- Gauss-Newton (1982)
- Preconditioned Conjugate Gradients
  - Multilinear Engine (1999)
  - Equation Oriented System (2001)
  - CPOPT/CPWOPT (2009)
- Levenberg-Marquardt
  - PMF3 (1997)
  - QR-PARAFAC (2009)
- Direct Trilinear Decomposition (1990)
- Simultaneous Matrix Diagonalisation (2006)
- Alternating Trilinear Decomposition (1998)
- Alternating Slice-wise Decomposition (2000)
- Self-Weighted Alternating TriLinear Decomposition (2000)
- Pseudo-Alternating Least Squares (2001)
- CP with Penalty Diagonalization Error (2001)

PARAFAC2
- Alternating Least Squares (direct fitting)
- Alternating Least Squares (indirect fitting, on covariances)
  - Preconditioned Conjugate Gradients (ME)
  - Other CP fitting methods

SCP / Shifted PARAFAC
- Alternating Least Squares (complete enumeration – 2003)
- Alternating Least Squares (frequency domain – 2009)

Tucker-\(m\)/Tucker-\(N\)/Block Tucker
- SVD (not least squares, 1966)
- Higher Order Orthogonal Iteration (2000)
- Slice Projection (2005)
- Multislice Projection (2007)
- Newton, Quasi-Newton on Grassman-Manifolds (2009)
- Levenberg-Marquardt (2009)

\(N\)-PLS/\(n\)
Ad hoc direct (SVD based) method for \(N=3\)
and \(n=1\), iterative for \(N \geq 4\) or \(n \geq 2\)
The basic PARAFAC-ALS can be greatly improved

For 3- and 4- way arrays no rearrangement is needed

The computational load can be reduced by a factor $I/F$

Operating column-wise for $A_2$ and $A_3$

one reduces the number of operations by a factor $F$

Partial results and recursion can be used for higher orders

The loss function can be calculated without explicitly calculating the residuals
An update to the solution is found by solving the problem

\[ \Delta p = \arg \min_{\Delta p} \| r + J \Delta p \|_2 \quad p = \text{vec}[A_1^T \cdots A_N^T] \]

where \( J \) is the Jacobian matrix of the residuals

\[ J = \left[ M_1 \left( I_{r_1} \otimes (A_3 \otimes A_2) \right) \cdots M_3 \left( I_{r_3} \otimes (A_2 \otimes A_1) \right) \right] \]

Direct methods (e.g. QR) cannot be used even though \( J \) is very sparse

We solve the normal equations (with regularisation)

\[ (J^T J + \lambda I) \Delta p = -J^T r \]

\( J \) is too large for most problems, but is not necessary to calculate it explicitly

The Cholesky factor is almost full unless \( I \gg J + K + \ldots \)
The array is projected on some truncated bases

**SVD based compressions**

*Prior knowledge (CANDELINC, PARAFAC-IV)*

The array is compressed to $F^N$ elements

Not straightforward implementation of non-negativity constraints

\[
X^{(I_1 \times I_2 I_3)} \cong U_1 \tilde{X}^{(F' \times F'' F''')} (U_3 \otimes U_2)^T \\
\tilde{X}^{(F' \times F'' F''')} \cong \tilde{A}_1 (\tilde{A}_3 \circ \tilde{A}_2)^T \\
A_n^{(i)} = U_n \tilde{A}_n \quad n = 1, 2, 3 \\
\text{Fit on } X^{(I_1 \times I_2 I_3)} \text{ from } A_1^{(0)}, A_2^{(0)}, A_3^{(0)}
\]
Tucker compression

The array is projected on some truncated bases

SVD based compressions

Tucker based compressions

Prior knowledge (CANDELINC, PARAFAC-IV)

The array is compressed to $F^N$ elements

Not straightforward implementation of non-negativity constraints
Tucker compression

The array is projected on some truncated bases

SVD based compressions

Tucker based compressions

Prior knowledge (CANDELINC, PARAFAC-IV)

The array is compressed to $F^N$ elements

Not straightforward implementation of non-negativity constraints
A QR decomposition is calculated for all the loading matrices

$J$ becomes extremely sparse

The density (for $N=3$) is

$$\beta = \frac{(F+1)(2F+1)}{6IJK} \ll \frac{3}{(I+J+K)}$$

The sparsity pattern allows use sparse QR decomposition

Only $(I \mid J \mid K \ 2F)^2$ rows contain non-zero values

The approach is similar to the INDAFAC algorithm

For large $J$’s index storage is the biggest problem
The rows with non-zeros refer to elements in $N+1$ blocks

$F \times \ldots \times F$ Common to all modes

$(I_n - F) \times F \times \ldots \times F$ Relative to mode $n$

The blocks can be calculated independently

The size of the problem is

$$\left( F \sum_{n=1}^{N} I_n \right) \times \left( F \sum_{n=1}^{N} I_n \right)$$

Complexity is reduced to that of a CP of dimension

$F^{N-1} \sum_{n} I_n$

The cost per iteration depends on the number of non-zeros in $J$

Currently $J$ is constructed explicitly
Tests

ALSe  Fast ALS with exact line search
ALS   Fast ALS with \( \alpha = \alpha t^{1/m} \) for \( m = 3, 4 \ldots \)
ALSn  Fast ALS without line search
LM    CP LM
SWA   SWATLD
cALSe  Compression and ALS with exact line search
cALS  Compression and ALS with \( \alpha = \alpha t^{1/m} \) for \( m = 3, 4 \ldots \)
cALSn  Compression and ALS without line search
cLM   Compression and PARAFAC-LM
cSWA  Compression and SWATLD
QR    QR CP with QR decomposition
QRne  QR CP with Normal equations

50 random starts for all the algorithms

<table>
<thead>
<tr>
<th>Data set</th>
<th>Size</th>
<th>( F )</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP1</td>
<td>10×8×5</td>
<td>3</td>
<td>Synthetic</td>
</tr>
<tr>
<td>PP2</td>
<td>10×8×5</td>
<td>4</td>
<td>Synthetic</td>
</tr>
<tr>
<td>Aminoacids</td>
<td>5×201×61</td>
<td>3</td>
<td>Fluorescence</td>
</tr>
<tr>
<td>Aminoacids + 1</td>
<td>22×87×13</td>
<td>4</td>
<td>Fluorescence</td>
</tr>
<tr>
<td>Aminoacids + 2</td>
<td>15×66×15</td>
<td>4</td>
<td>Fluorescence</td>
</tr>
<tr>
<td>Antibiotics</td>
<td>22×107×65</td>
<td>2</td>
<td>HPLC-DAD</td>
</tr>
</tbody>
</table>
Results

PP1

PP2
Results

Aminoacids

Antibiotics

Aminoacids + 2

Aminoacids + 1

Antibiotics
Results

Aminoacids +2
Results

Antibiotics
CP algorithms are not quite fast enough

**ALS** works well but is not suitable for difficult problems
- More affected by overfactoring
- Slow convergence
- More affected by swamps and degeneracies
- Exact line search improves convergence but does not scale well and is not always worth the additional computation cost

**LM** is too expensive for large problems
- Scales well to higher orders
- QR/Tucker compression work well for \( F \ll I_n \)
- Tucker compression makes it fast, but constraints are not trivial to implement
- QR compression does not well with weighted least squares
- Several constraints are still missing
- Unimodality constraints do not fit well in this framework

**Conjugate Gradients** are especially promising for large (very sparse and low order) tensors and \( F^S \)
- Still in their early days in CP fit
- Constraints are not quite there, yet

**Simultaneous Diagonalisation/Direct methods**
- Good for low noise case
- Refinement may be required
- No constraints are currently available
The vast majority of applications uses ALS

The algorithm is a slightly modified PARAFAC2-ALS algorithm

Improving the speed of CP will also improve the PARAFAC2

The estimation of the $I_2$ matrices $P_n$ has fixed complexity

Multilinear Engine (nonlinear CG) have been suggested as an alternative

Constraints in $N$-1 modes are easily implemented.
Relevant aspects for an algorithm

– Effect of overfactoring / underfactoring
– Missing values
– Constraints
– Time consumption (especially for resampling techniques)
– Use of resources (especially memory)
– Complex numbers

We need

a series of reliable data sets (real and simulated) for testing in an objective way

and

Methods and guidelines so that the best method can be picked depending on the problem
Acknowledgements

For the data and the help: R. Bro, I. Garcia, J.H. Christensen, M. Mørup

References (for the data)


2. Christensen J.H., Tomasi G., Strand J., Andersen O., CP Modeling of Fluorescence Excitation#Emission Spectra of Fish Bile for Rapid En Route Screening of PAC Exposure


6. Tomasi, G. Practical and computational aspects in chemometric data analysis. Ph.D. The Royal and Agricultural University, Frederiksborg, Denmark, May **2006**.
THANK YOU!

Giorgio.tomasi@ec.europa.eu