Introduction to Multi-Block Methods

October 8, 2004
ODIN course
Introduction to Multi-Block Methods

Material:
www.models.kvl.dk  →  [Algorithms]  →  [MBToolbox]

10:00-12:00h  Talk: Introduction (± 40min)
              Talk: Literature and algorithms (± 40min)

12:00-13:00h  Lunch

13:00-16:00h  Talk: Computer exercises (± 15min)
              Computer exercises
Multi-block
The idea

spectral data + weather conditions + process settings + control signals = ?
Multi-block
The idea

Answer or Answers?
Multi-block
The idea

Answer
Multi-block
The idea

Answer
Example
Wheat flourmill data *)

Seven sample point

Particle Size Distribution and Product Composition?

*) J.Pram Nielsen, D.Bertrand, E.Micklander, P.Courcoux and L.Munck
(off-line) laboratory data

Size distribution

Chemical composition

1 Dry matter
2 Ash
3 Protein
4 Starch
5 Damaged starch

Laser Diffraction on 7 samples separated in 6 size fractions a-f
Example
Wheat flourmill data

**Near InfraRed**

**SNV-NIR**

**correlation coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Laser</th>
<th>Chem</th>
<th>NIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chem</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIR</td>
<td>0.9</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>SNV-NIR</td>
<td>0.7</td>
<td>0.6</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Principal Component Analysis (PCA) Algorithm

0 - choose starting $t$
1 - $p = X'.t / (t'.t)$
2 - $p = p / ||p||$
3 - $t = X.p / (p'.p)$
4 - convergence?
   no = back to 1
5 - $X = X - t.p'$
   back to 0

(e.g. by NIPALS)

\[
X' = X \cdot (\ldots)^{-1} \cdot (\ldots)^{-1}
\]
Multi-block PCA
Super level

![Diagram of multi-block PCA with super level and associated transformations]
Multi-block PCA Algorithm

\begin{align*}
X_1^t & = X_1^t T \\
X_2^t & = X_2^t T \\
T & = \text{"super" level}
\end{align*}

\begin{align*}
p^t_1 & = p^t_1 T \\
p^t_2 & = p^t_2 T
\end{align*}
Example
Wheat flourmill data

(overall) object scores

block weights

Laser X

Chem X
Example
Off-line (laboratory) data

explained variance

Distribution
Composition
NIR
SNV-NIR
Example
Wheat flourmill data

(overall) object scores

block weights

NIR
X

SNV-NIR
X
Example
In-process data

explained variance

Distribution
Composition
NIR
SNV-NIR

object scores

explained variance (%)
Partial Least Squares (PLS) Algorithm
Multi-block PLS
Super level

(1.16)
Multi-block PLS
Algorithm

"super" level

$X_1$  $X_2$

$T$

$Y_1$  $Y_2$

$w'_1$  $w'_2$

$q'_1$  $q'_2$

$w'_t$  $w'_u$

$p'_1$  $p'_2$

$u_1$  $u_2$
Example
Wheat flourmill data

![Diagram showing object scores and block weights for NIR, SNV-NIR, Laser, and Chem data.](image)
Example
In-process $\rightarrow$ laboratory data

explained variance

Distribution
Composition
NIR
SNV-NIR

object scores (U)

(→) cross validation prediction error

explained variance (%)

size

chem.
Second example
Cheese storage

Cheese Making

- Separation → Skim milk
- Standardization → Cream
- Culture
- Pre-pressing → Whey
- Salting
- Storage → Flavor and texture development

Proteins, peptides, & aroma compounds

© Vibeke T. Povlsen
The Royal Veterinary and Agricultural University (KVL), Denmark
Dept. of Dairy and Food Science, Food Technology
Second example

Cheese ripening

- Intact casein
- Plasmin
- Rennet
- Large peptides
- Bacterial proteases and peptidases
- Small peptides
- Free amino acids
- Aroma compounds
- Bacteriel enzymes

3 (large) chemical and physical measurement series, 35 weeks of storage, plus sensory evaluation of the end product.
The building/data blocks

Following measurements are performed at week 4, 10, and 16

**Chemical measurements**
- Chemical (5 variables)
  - Fat, protein, dry matter, salt, and pH.
- Aroma (22 variables)
- Casein (8 variables)
- Peptide (16 variables)

**Physical measurements**
- Compression (3 variables)
- Stretch (2 variables)
- TPA (9 variables)
- Oscillation (36 variables).

Following measurements are performed at week 35

**Sensory Evaluation**
- 12 sensory attributes
Second example

Multi-block model

Super level

\[ \mathbf{T} \]

\[ W'_{t} \]

\[ \mathbf{Y} \]

\[ X_{1}, X_{2}, X_{3}, X_{4}, X_{5}, X_{6}, X_{7}, X_{8} \]
Second example

Selection of Y-variables (sensory attributes)
Second example

Multi-block decomposition

Different view-points:

- All there is
- Nature of the signal
- Time
- Time/Signal
- Selection

[Diagram showing multi-block decomposition with chemical and physical blocks, and selection process indicated by arrows.]
Second example

Matrix/block correlations
Second example

Y1: decomposition step 2

Y1: decomposition step 3
Optimization in $Q^2$ for the new models

Y1: Cheese smell
Y2: Cheese taste
Y3: Sharp taste
Y4: After taste.
Second example

Q² for the PLSR1 vs. PLSR2 models

All blocks

Selected blocks

7 factor PLSR2

5 factor PLSR2
Multi-block methods from literature

Examples and algorithms

Number of examples from literature:
(by no means complete, unbiased or in any particular order)

- Alternative methods (well established in other research areas!)

- Some examples (mostly from process monitoring and control)

- Some algorithm stuff (just to show the idea)
### Multi-block

(Partial) history

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>H. Wold</td>
<td>path-PLS models</td>
</tr>
<tr>
<td>1984</td>
<td>S. Wold/Martens/H. Wold</td>
<td>multiblock-PLS</td>
</tr>
<tr>
<td>1984</td>
<td>Frank/Kowalski</td>
<td>wine, chemical and sensory</td>
</tr>
<tr>
<td>1988</td>
<td>Wangen/Kowalski</td>
<td>simulations</td>
</tr>
</tbody>
</table>

Origin of PLS like we know it is rooted in Multi-block PLS!!!
Response to LISREL from social and business sciences
("one factor PLS per connection") Notice that there is a direction!

Michel Tenenhaus et.al. *PLS methodology to study relationships between hedonic judgements and product characteristics*
Path-PLS
Causality relations

Two stages: backwards (U _) + forwards (T -->) step

Two stages are almost symmetrical, with exchange around the X-block (X is 'the pivot')
Evaluation of six orange juices

**Path-PLS**
Causality relations

(2.05)

Evaluation of six orange juices

**Chemistry**
- (Imposed)
- Latent structure (path)

**Sensory**
- Hedonic consumer judgment

Different objective:
- e.g. product optimization
- or sales

\[
\begin{align*}
\text{Glucose} & \quad \text{Fructose} \\
\text{Saccharose} & \quad \text{Sweetening power} \\
\text{pH before processing} & \quad \text{pH after centrifugation} \\
\text{Titre} & \quad \text{Citric acid} \\
\text{Vitamin C} & \quad \text{Smell intensity} \\
\text{Odour typicality} & \quad \text{Pulp} \\
\text{Taste intensity} & \quad \text{Acidity} \\
\text{Bitterness} & \quad \text{Sweetness}
\end{align*}
\]

\[
\begin{align*}
\xi_1 & \quad \text{Judge 2} \quad \text{Judge 3} \\
\xi_2 & \quad \ldots \\
\xi_3 & \quad \text{Judge 96}
\end{align*}
\]
Multi-block
Same block sizes

Same information matrix/data-table for different samples
PARAFAC and Tucker
Same block sizes

Rasmus Bro *PARAFAC Tutorial and applications* - Chemometrics and Intelligent Laboratory Systems 38(1997)149-171

Hyphenated Analytical Chemistry: samples x GC x MS
Sensory Science: products x assessors x attributes
Process Monitoring: batches x process variables x time

\[ X = A \times B \times C + D \times E \]
Extensions
Several blocks and tables combined ... 

Age Smilde and Henk Kiers *Multiway covariate regression models* *Journal of Chemometrics* 13(1999)31-48
Important issues raised in these papers:
What if the different blocks have different ranks?

This is very likely when you have many
different sources of information (different blocks)!

This can only be solved by simultaneous factor estimation (e.g. ALS),
not by component-wise methods (e.g. NIPALS)
Covariate regression
Component-wise versus simultaneous estimation

Component-wise (NIPALS)

Simultaneously (ALS) – like Covariates regression

\[
\min_w \left[ \beta\|X - XWP_X^T\| + (1 - \beta)\|y - XWP_y\| \right] \quad \text{for a given } \beta \quad 0 \leq \beta \leq 1.
\]
Generalized Procrustes Analysis
Same block sizes

Generalized Procrustes Analysis
Procrustes rotation

(2.12)

- Target
- Center
- Shrink/expand
- Flip sign(s)
- Rotate

\[ (0,0) \]
Generalized Procrustes Analysis
Same block sizes

\[ \sum \]

Average

Procrustes Rotation

Average

\[ \sum \]

Until convergence
Two objectives:
- Get a good average / consensus estimate for the underlying phenomena
- Compare individual results with group average / consensus (panel training, homogeneity, etc.)
Block correlation (RV) coefficient

Intermediate


\[
RV(X, Y) = \frac{\text{tr}(X.X^T.Y.Y^T)}{\sqrt{\text{tr}(X.X^T.X.X^T).\text{tr}(Y.Y^T.Y.Y^T)}}
\]

\rightarrow RV number between 0 and 1
\rightarrow RV = 1 means c.X.H = Y or X can be rotated into Y

\[
X \ X' = \ Y \ Y'
\]

Normal correlation (e.g. predicted y versus reference x)

\[
r^2 = \left( \frac{x^T.y}{\sqrt{(x^T.x.y^T.y)}} \right)^2
\]
STATIS
Sensory data

El Mostafa Qannari et al. *A hierarchy of models for analyzing sensory data* Food Quality and Preference 40/1(1975)33-51
STATIS = Structuration des Tableaux A Trois Indices de la Statistique

Compute association matrices \( W_i = X_i X_i^T \)

Weights

\[
\begin{align*}
    e_1 & \leftarrow \begin{pmatrix} R(1,1) & R(1,2) & R(1,3) & R(1,4) \\ R(2,1) & R(2,2) & R(2,3) & R(2,4) \\ R(3,1) & R(3,2) & R(3,3) & R(3,4) \\ R(4,1) & R(4,2) & R(4,3) & R(4,4) \end{pmatrix} \\
    e_2 & \leftarrow \begin{pmatrix} R(1,1) & R(1,2) & R(1,3) & R(1,4) \\ R(2,1) & R(2,2) & R(2,3) & R(2,4) \\ R(3,1) & R(3,2) & R(3,3) & R(3,4) \\ R(4,1) & R(4,2) & R(4,3) & R(4,4) \end{pmatrix} \\
    e_3 & \leftarrow \begin{pmatrix} R(1,1) & R(1,2) & R(1,3) & R(1,4) \\ R(2,1) & R(2,2) & R(2,3) & R(2,4) \\ R(3,1) & R(3,2) & R(3,3) & R(3,4) \\ R(4,1) & R(4,2) & R(4,3) & R(4,4) \end{pmatrix} \\
    e_4 & \leftarrow \begin{pmatrix} R(1,1) & R(1,2) & R(1,3) & R(1,4) \\ R(2,1) & R(2,2) & R(2,3) & R(2,4) \\ R(3,1) & R(3,2) & R(3,3) & R(3,4) \\ R(4,1) & R(4,2) & R(4,3) & R(4,4) \end{pmatrix}
\end{align*}
\]

Eigenvector is normalized, \( \sum e_i = 1 \)

Members close to the average get high scores
Members far off (outliers) get low scores

L-PLS
X, Y and Z data

Harald Martens, et al. Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR

Three sources of information:
- Z and X both related to Y
- Z and X no common dimension
- Bi-linear models for all blocks, in accordance with PLS-concept, modeling of Y guided by X and Z
**L-PLS**

X, Y and Z data

Many relations (to many?)

Extensions possible (U-PLS, etc.)
Tennessee Eastman benchmark (famous in control community) as example in Statistical Process Monitoring by multi-block methods. 41 measurements and 12 control variables.

Multi-block PCA and PLS
Process Monitoring and Control


Blocks: conceptually meaningful (for an engineer!)

Reactors  Stripper  Separator
Multi-block PCA and PLS
Process Monitoring and Control


Fault detection: Overall process / Process divided in blocks
Multi-block PCA and PLS
Process Monitoring and Control

Theodora Kourti et.al. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS
Theodora Kourti and John MacGregor Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods
Chemometrics and Intelligent Laboratory Systems 28(1995)3-21
Multi-block PCA and PLS
Process Monitoring and Control

Theodora Kourti et al. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS

Theodora Kourti and John MacGregor Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods
Chemometrics and Intelligent Laboratory Systems 28(1995)3-21
Multi-block PCA and PLS
Process Monitoring and Control

Theodora Kourtial Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS
Theodora Kourt and John MacGregor Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods
Chemometrics and Intelligent Laboratory Systems 28(1995)3-21

Process monitoring

Find fault and identify (potential) cause (so-called contribution plot)
Multi-block PCA and PLS
Process Monitoring and Control

Carl Duchesne and John MacGregor *Multivariate analysis and optimization of process variable trajectories for batch processes* Chemometrics and Intelligent Laboratory Systems 51(2000)125-137

Intermediate quality measurements

Investigated by a path-like MBPLS algorithm
Multi-block PCA and PLS
Process Monitoring and Control

Theodora Kourtı et al. *Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS*  
Theodora Kourtı and John MacGregor *Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods*  
Chemometrics and Intelligent Laboratory Systems 28(1995)3-21
Stefan Ränner et al. *Adaptive batch monitoring using hierarchical PCA*  
Chemometrics and Intelligent Laboratory Systems 41(1998)73-81

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z$</td>
<td>Batches</td>
<td>Initial conditions (Process)</td>
</tr>
<tr>
<td>$X_0$</td>
<td>Process variables (On-line)</td>
<td></td>
</tr>
<tr>
<td>$Y$</td>
<td>Quality measures (Off-line)</td>
<td></td>
</tr>
<tr>
<td>$X_{k-1}$</td>
<td>$X_k$</td>
<td>present data</td>
</tr>
</tbody>
</table>

Comparison of old and present

$\mathbf{t}_{k-1} = \text{old state}$  
$\mathbf{t}_k = \text{present state}$

EWMA-idea
Stefan Rännere et al. *Adaptive batch monitoring using hierarchical PCA* Chemometrics and Intelligent Laboratory Systems 41(1998)73-81

**Multi-block PCA and PLS**

*Process Monitoring and Control*

---

**Full set**

**Lower local rank (more parsimonious model)**

**EWMA**

- 

\[ t_k = \text{present state} \]

- 

Comparison of old and present

- 

Weight-ratio old/new observations

Determines Average Run Length

\[
\begin{align*}
\text{ARL} & = d = \text{block weights} \\
(d &= 1.00 \text{ and } d = 0.33)
\end{align*}
\]

---

**EWMA-idea**
Multi-block PCA and PLS
Wet granulation and tablet pressing


Two-step process
Mixing                  tablet pressing                  tablet quality

Crushing strength
Disintegration time

Physical properties
of the granulates

Process variables
(of both steps)

\[ (2.28) \]
Multi-block PCA and PLS
Decentralized process monitoring


Polystyrene film manufacturing

<table>
<thead>
<tr>
<th>Block number</th>
<th>Process section</th>
<th>Variables in each block</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Drying zone</td>
<td>1–9</td>
</tr>
<tr>
<td>2</td>
<td>Extrusion zone</td>
<td>10–29</td>
</tr>
<tr>
<td>3</td>
<td>Melt pipes zone 1</td>
<td>30–40</td>
</tr>
<tr>
<td>4</td>
<td>Melt pipes zone 2</td>
<td>41–52</td>
</tr>
<tr>
<td>5</td>
<td>Die zone</td>
<td>53–61</td>
</tr>
<tr>
<td>6</td>
<td>Casting zone</td>
<td>62–77</td>
</tr>
<tr>
<td>7</td>
<td>Tenter zone</td>
<td>78–103</td>
</tr>
</tbody>
</table>

2879 time-frames 103 process variables → 7 blocks
Multi-block PCA and PLS
Decentralized process monitoring

Multi-block PCA and PLS
Decentralized process monitoring


Block-wise approach
Multi-block (consensus) PCA

The one we will use

\[ \text{PCA} \]

Choose \( t \)

Loop

\[ p = X^T.t/(t^T.t) \]

\[ p \rightarrow \|p\| = 1 \]

\[ t = X.p \]

End

\[ E = X - t.p^T \]

\[ \text{CPCA} \]

Choose \( t \)

Loop

\[ p_b = X_b^T.t/(t_b^T.t_b) \]

\[ p_b \rightarrow \|p_b\| = 1 \]

\[ t_b = X_b.p_b \]

\[ T = [t_1 \ t_2 \ \ldots \ t_b] \]

\[ w_t = T^T.t/(t_t^T.t_t) \]

\[ w_t \rightarrow \|w_t\| = 1 \]

\[ t_t = T.w_t \]

End

\[ p_b = X_b^T.t/(t_t^T.t_t) \]

\[ E_b = X_b - t_t.p_b^T \]

Relation

(\( b = \text{block}, \ t = \text{super-level} \))

\[ t_t == t \]

\[ p_b = X_b^T.t/(t_t^T.t_t) \]

\[ p_b \rightarrow \|p_b\| = 1 \]

\[ t_b = X_b.p_b \]

\[ T = [t_1 \ t_2 \ \ldots \ t_b] \]

\[ w_t = T^T.t/(t_t^T.t_t) \]

\[ w_t \rightarrow \|w_t\| = 1 \]

\[ t_t = T.w_t \]

\[ E_b = X_b - t_t.p_b^T \]
# Multi-block PLS

The one we will use

\[ (2.33) \]

## PLS

Choose \( u \)

Loop

\[
\begin{align*}
  w &= X^T.u / (u^T.u) \\
  w &\rightarrow ||w|| = 1 \\
  t &= X.w \\
  q &= Y^T.t / (t^T.t) \\
  u &= Y.q \\
\end{align*}
\]

End

\[
\begin{align*}
  p &= X^T.t / (t^T.t) \\
  E &= X - t.p^T \\
  F &= Y - t.q^T \\
\end{align*}
\]

## MBPLS

Choose \( u \)

Loop

\[
\begin{align*}
  w_b &= X_b^T.u / (u^T.u) \\
  w_b &\rightarrow ||w_b|| = 1 \\
  t_b &= X_b.w_b \\
  T &= [t_1 \ t_2 \ldots \ t_b] \\
  w_t &= T^T.u / (u^T.u) \\
  w_t &\rightarrow ||w_t|| = 1 \\
  t_t &= T.w_t \\
  q &= Y^T.t / (t^T.t) \\
  u &= Y.q \\
\end{align*}
\]

End

\[
\begin{align*}
  p_b &= X_b^T.t / (t^T.t) \\
  E_b &= X_b - t_p.p_b^T \\
  E &= X - t_t.p_b^T \\
  F &= Y - t_t.q^T \\
\end{align*}
\]

## Relation

(b = block, t = super-level)

\[
\begin{align*}
  t_t &= t \\
  u &= u \\
  w_b &= X_b^T.u / (u^T.u) \\
  t_b &= X_b.w_b \\
  T &= [t_1 \ t_2 \ldots \ t_b] \\
  w_t &= T^T.u / (u^T.u) \\
  t_t &= T.w_t \\
  q &= Y^T.t / (t^T.t) \\
  u &= Y.q \\
  p_b &= X_b^T.t / (t^T.t) \\
  E_b &= X_b - t_t.p_b^T \\
  E &= X - t_t.p_b^T \\
  F &= Y - t_t.q^T \\
\end{align*}
\]

---


Johan Westerhuis and Age Smilde *Short Communication Deflation in multiblock PLS* Journal of Chemometrics 15(2001)485-493

Multi-block PCA and PLS
The ones we will use

Conclusions:

- MBPCA and MBPLS can be computed
  from PCA and PLS on augmented data matrices!

- Use e.g. The Unscrambler or SIMCA output

- In this view multi-block methods are not a new “modeling technique”
  (no better models), but rather a new / alternative way of looking at
  large data-sets with conceptually meaningful blocks
  New plotting diagnostics at the block level

- Right scaling of the blocks turns out to be crucial!

Many names, same result; e.g. Generalized PCA and SUM-PCA
Ph. Casin A generalization of principal component analysis to K sets of variables Computational Statistics & Data Analysis 35(2001)417-428
Eduard Derks et.al. An introduction to Multi-block Component Analysis by means of a flavor language case study
Multi-block PCA and PLS
Block weighing

**Process variables**

- **Stirrer speed**: 100 – 110 rpm
- **Time**: 180 – 240 minutes (10800 – 14400 second)
- **Heat of evaporation**: $2.26 \times 10^6 - 2.41 \times 10^6$ J.kg$^{-1}$
- **Temperature**: 328 – 343 K
- **Pressure**: $15.7 \times 10^3 - 31.2 \times 10^3$ Pa

Process variables are usually not compatible. Different variances, hence different “weight” in factor models like PCA and PLS. By auto-scaling we give them equal importance:

- center = 0
- variance = 1 ($\Sigma x_i^2 / (N-1) = 1$)

But this choice is arbitrary!

E.g. it might be decided to “weigh-down” stirrer speed (likely not important), which could greatly influence the model, or Temperature and Pressure because they are infect twice the same thing.
Multi-block PCA and PLS
Block weighing

Block weighting turns out to be a difficult / important issue

We will use sum-of-squares to scale blocks

\[
\sum (x_{\text{raw}}(i,j)^2 \times \text{factor}) = \text{constant} = \sum x_{\text{raw}}(i,j)^2
\]

The sum of all squared entries in the scaled block is made equal to a chosen constant by determining the right factor

Only the relative size of the constants for different blocks counts

A good starting point is to give all blocks a constant of 1

Same holds for multi-block methods:
e.g.
- 1 Input block
- 1 Output block
- 100 Time Frames on the process variables

Depending on the objective you have to scale the different matrices

E.g. for equal weights the Input will “drown” with this many Time Frames.
NIR temperature effects
Mixture designs

Florian Wülfert, Wim Kok and Age Smilde

Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models
Analytical Chemistry 70(1998)1761-1767

Data: design mixtures Ethanol, Water and Iso-propanol

Fraction (m/m)

<table>
<thead>
<tr>
<th>E</th>
<th>W</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.66</td>
<td>0.34</td>
<td>0</td>
</tr>
<tr>
<td>0.67</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>0.67</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>0</td>
</tr>
<tr>
<td>0.50</td>
<td>0.33</td>
<td>0.17</td>
</tr>
<tr>
<td>0.50</td>
<td>0.17</td>
<td>0.33</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td>0.33</td>
<td>0.67</td>
<td>0</td>
</tr>
<tr>
<td>0.33</td>
<td>0.50</td>
<td>0.17</td>
</tr>
<tr>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>0.32</td>
<td>0.16</td>
<td>0.51</td>
</tr>
<tr>
<td>0.34</td>
<td>0</td>
<td>0.66</td>
</tr>
<tr>
<td>0.17</td>
<td>0.67</td>
<td>0.17</td>
</tr>
<tr>
<td>0.17</td>
<td>0.50</td>
<td>0.33</td>
</tr>
<tr>
<td>0.17</td>
<td>0.33</td>
<td>0.50</td>
</tr>
<tr>
<td>0.16</td>
<td>0.16</td>
<td>0.67</td>
</tr>
<tr>
<td>0</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td>0</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>0</td>
<td>0.33</td>
<td>0.67</td>
</tr>
</tbody>
</table>
Five blocks are same mixtures measured at five different temperatures
(file MB_NIR_Temp_Script.m will give you a hint how to start)
Multi-block Toolbox
Matlab functions

mbpca.m ↔ Multiblock PCA
mbpls.m ↔ Multiblock PLS

mypca.m ↔ normal PCA
mypcacv.m ↔ PCA cross validation
mypls.m ↔ normal PLS
myplscv.m ↔ PLS cross validation

autosc.m ↔ auto-scale block
blocknorm.m ↔ norm blocks
matrixblobs.m ↔ plotting routine (e.g. RV-coefficients)
matrixcorr.m ↔ series of matrix correlations
mbdata.m ↔ generates some toy data
meanc.m ↔ Mean center block
rangesc.m ↔ Range scale block
rvcoef.m ↔ RV coefficient
svdnan.m ↔ SVD for missing values

procrus2D.m ↔ Procrustes rotation (image)
procrusND.m ↔ Procrustes rotation (N dimensional)
$^1$H-NMR with shifts

Apple juice

Four commercial apple juices

Ethanol % (w/w) added

Shifted

Malic Acid

↑ Water

Shifted

(3.04)


\[ ^1 \text{H-NMR with shifts} \]

Apple juice

Giorgio Tomasi, Frans van den Berg and Claus Andersson *Correlation Optimized Warping and Dynamic Time Warping as preprocessing methods for chromatographic data* Journal of Chemometrics 18(2004)231-241
\textbf{\textsuperscript{1}H-NMR with shifts}

Apple juice

- Block schemes
- Different regions in spectrum
  - NMR shift
  - Ethanol
- Shifted versus corrected NMR
  - NMR shift
  - NMR cow
  - Ethanol
- Different brands
  (two are from the manufacturer, which ones?)
NIR Marzipan
Different instruments


32 marzipan samples
+ Moisture contents
Sugar contents
Measured on 6 instruments
Compare instruments
Block selection