Frans van den Berg
The Royal Veterinary and Agricultural University (KVL), Denmark
Dept. of Dairy and Food Science, Food Technology
Chemometrics group
fb@kvl.dk

www.models.kvl.dk

Introduction to Multi-Block Methods

October 8, 2004

ODIN course

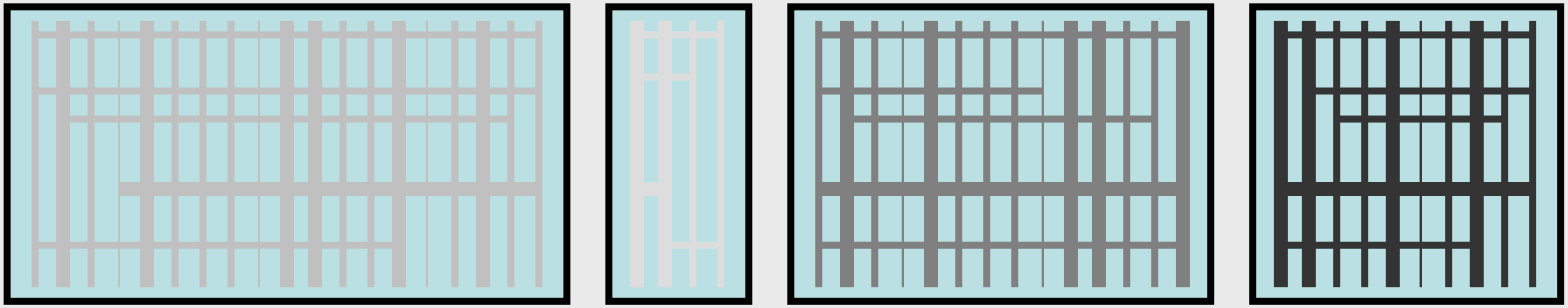
Introduction to Multi-Block Methods

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

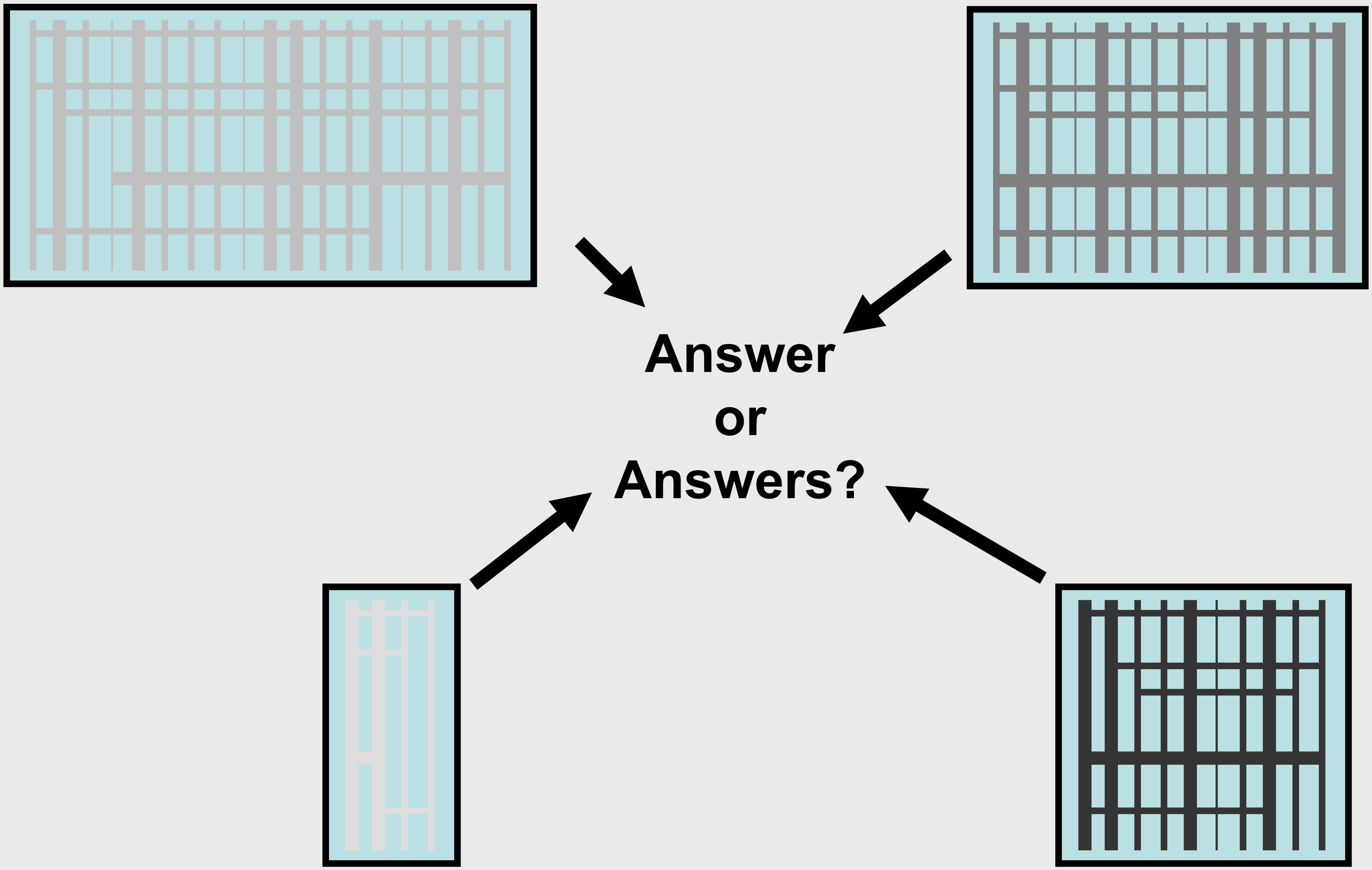
The idea



spectral data + weather conditions + process settings + control signals = ?

Multi-block

The idea

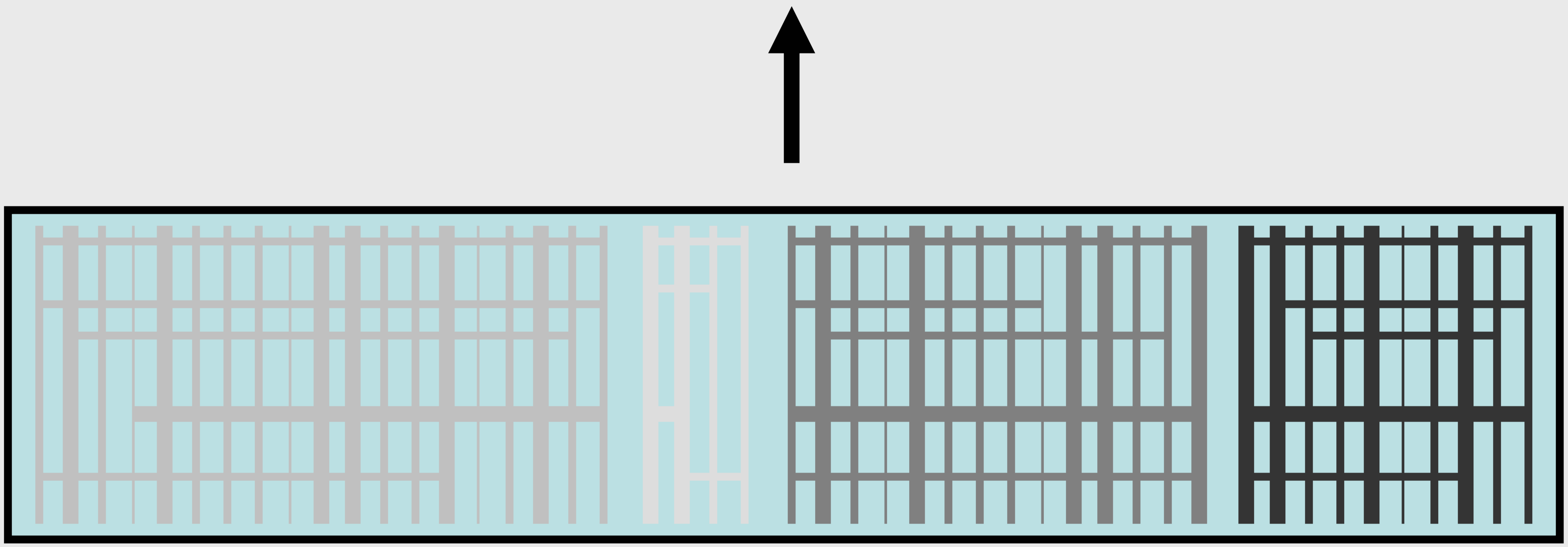


Multi-block

(1.03)

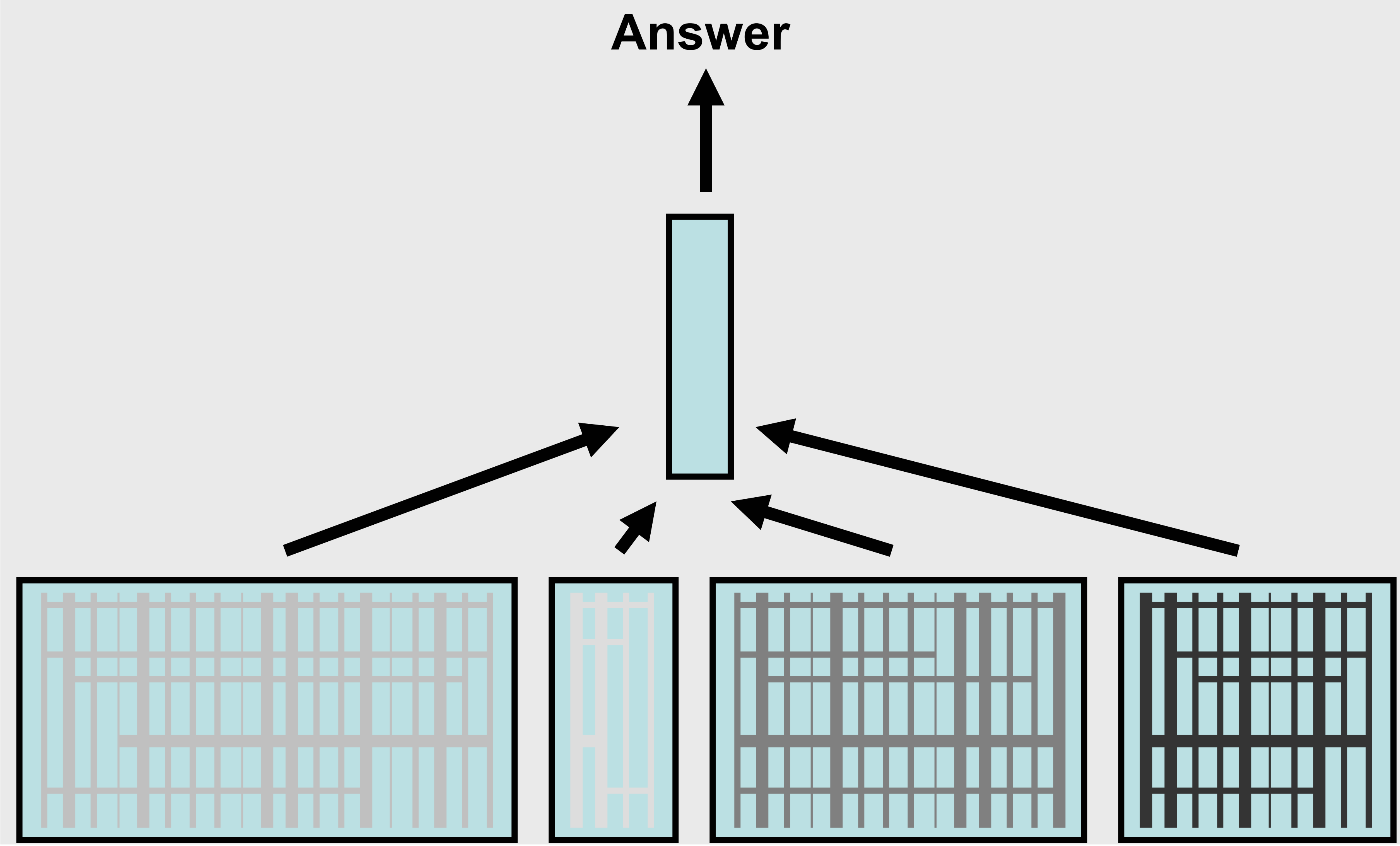
The idea

Answer



Multi-block

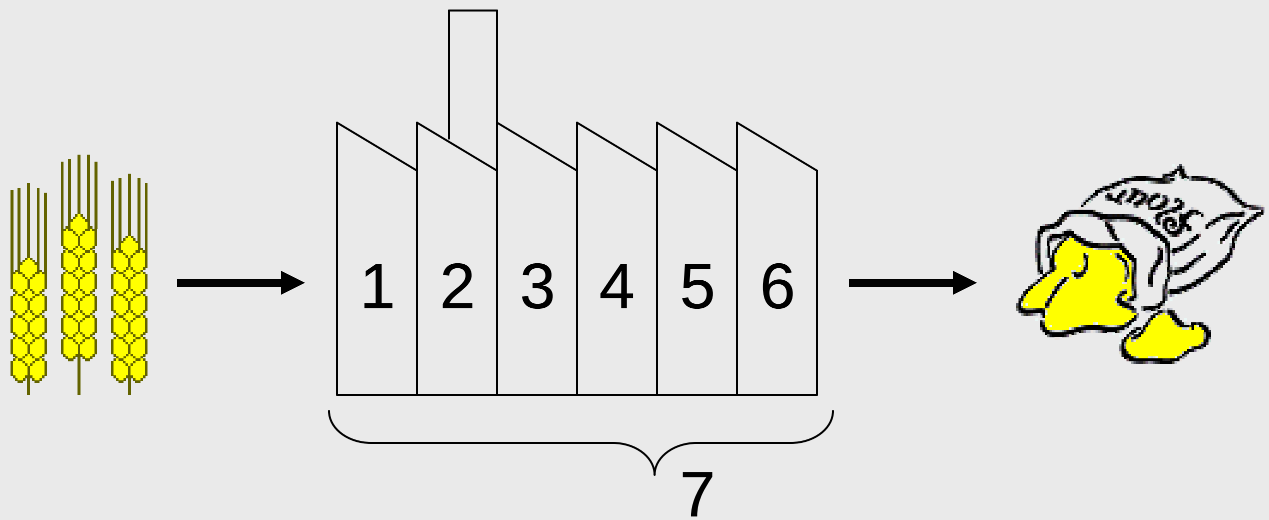
The idea



Example

Wheat flourmill data *)

Seven sample point



Particle Size Distribution and Product Composition?

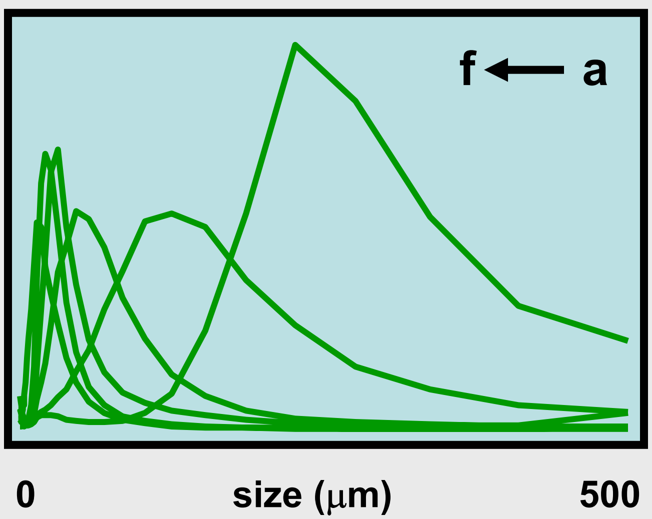
*) J.Pram Nielsen, D.Bertrand, E.Micklander, P.Courcoux and L.Munck
'Study of NIR spectra, particle size distributions and chemical parameters of wheat flours: a multi-way approach' J. Near Infrared Spectroscopy. 9(2001)275–285

Example

Wheat flourmill data

(off-line) laboratory data

Size distribution



Laser Diffraction on
7 samples separated in
6 size fractions a-f

Chemical composition

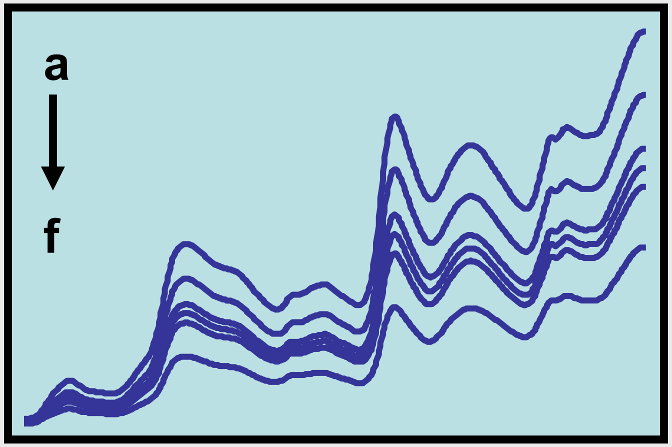
1 Dry matter
2 Ash
3 Protein
4 Starch
5 Damaged starch

Example

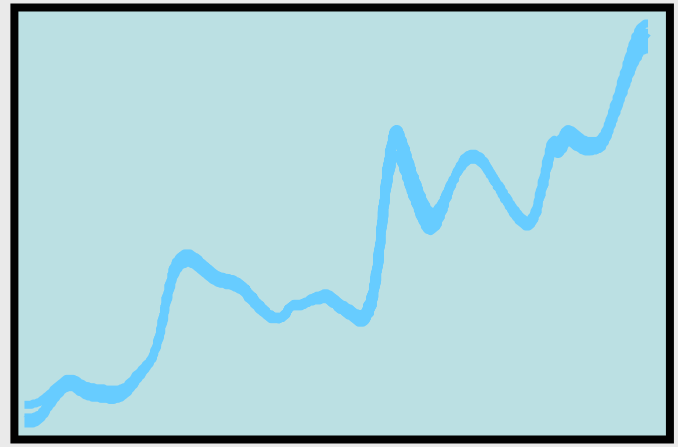
Wheat flourmill data

in-process data

Near InfraRed



SNV-NIR



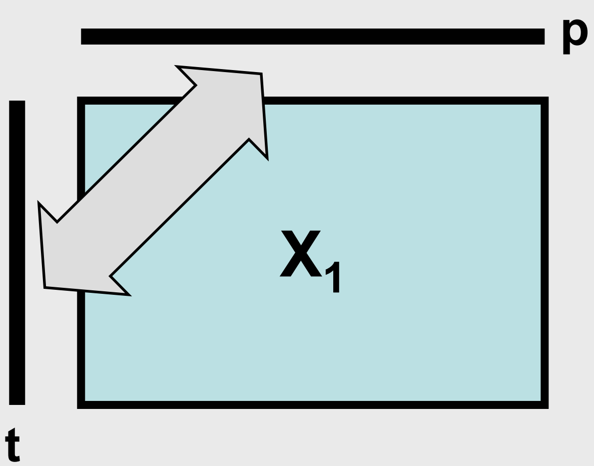
Standard Normal Variate scaling

correlation coefficients

	Laser	Chem	NIR
Chem	0.1		
NIR	0.9	0.3	
SNV-NIR	0.7	0.6	0.9

Principal Component Analysis (PCA)

Algorithm



(e.g. by NIPALS)

0 - choose starting **t**

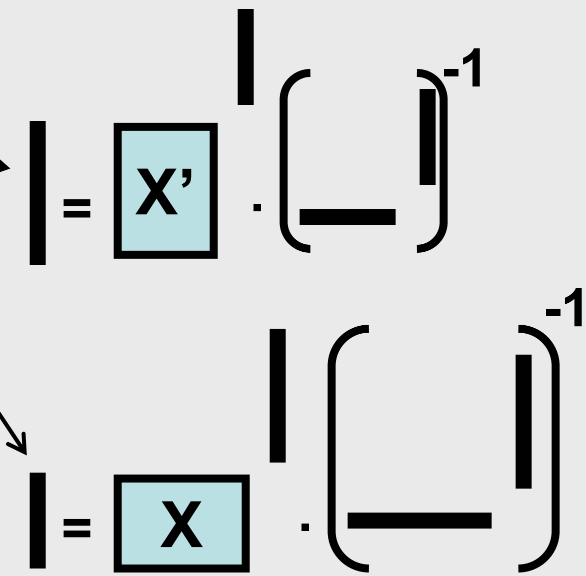
1 - **p = X'.t / (t'.t)**

2 - **p = p / ||p||**

3 - **t = X.p / (p'.p)**

4 - convergence?
  no = back to 1

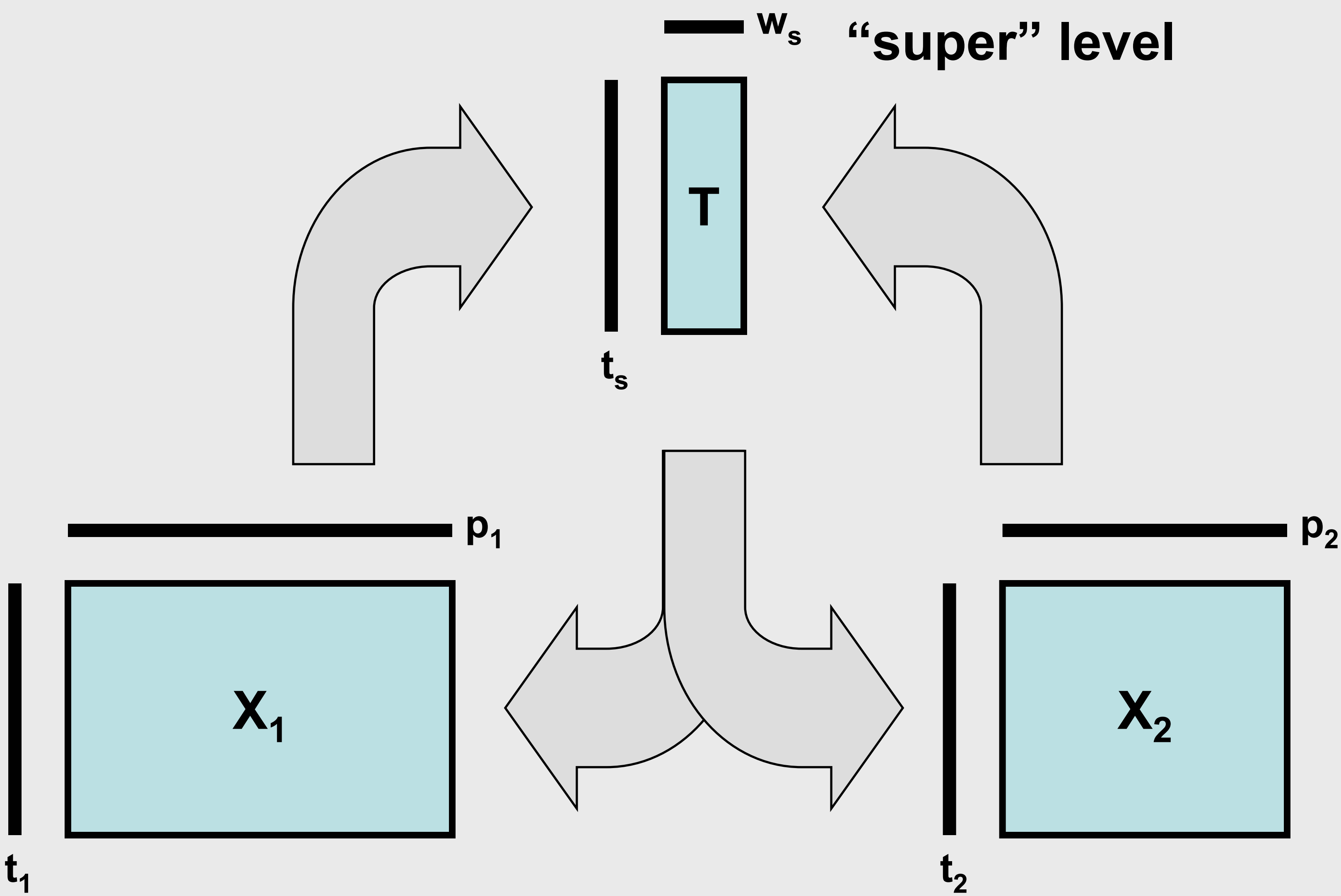
5 - **X = X - t.p'**
  back to 0



Multi-block PCA

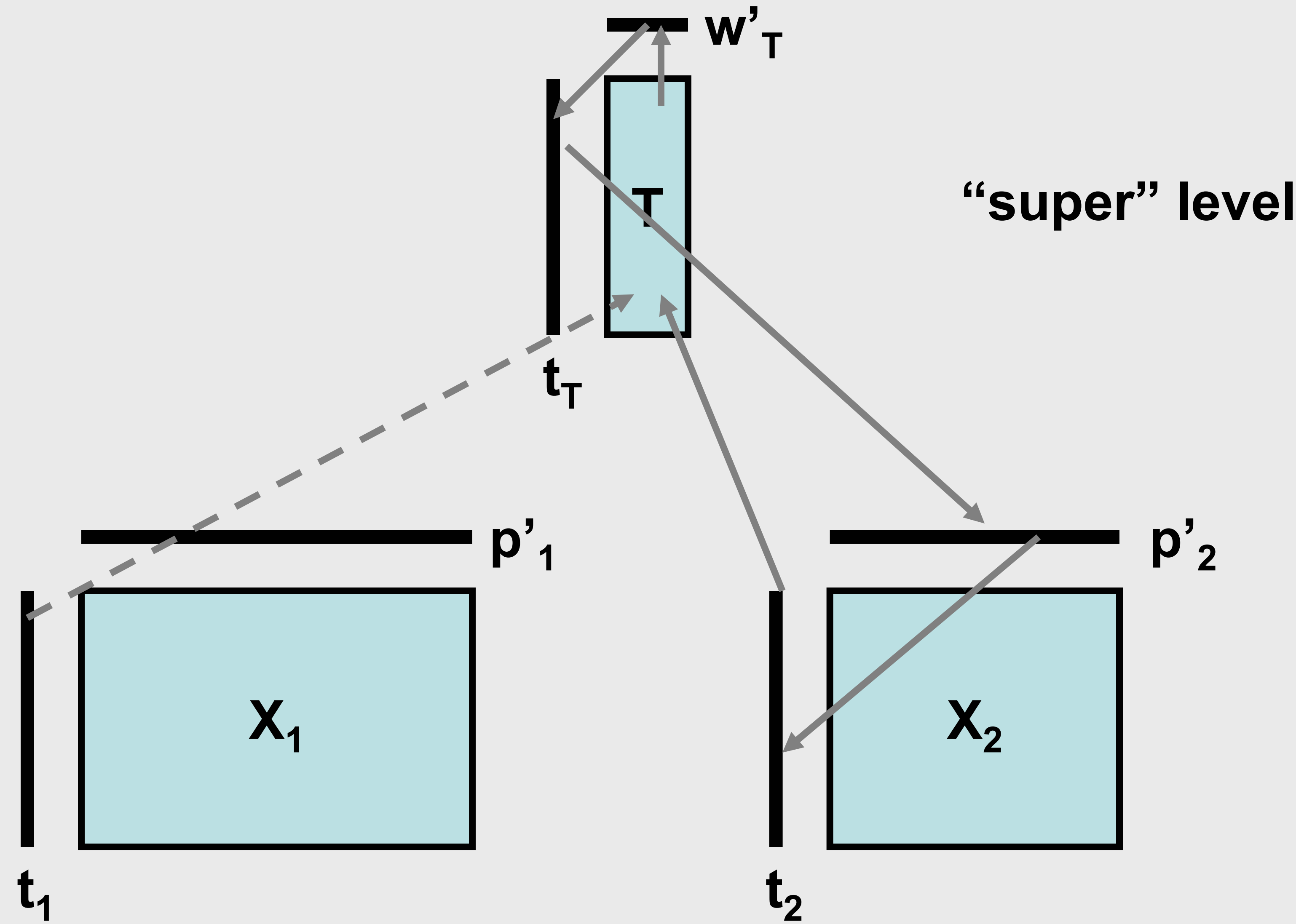
(1.09)

Super level



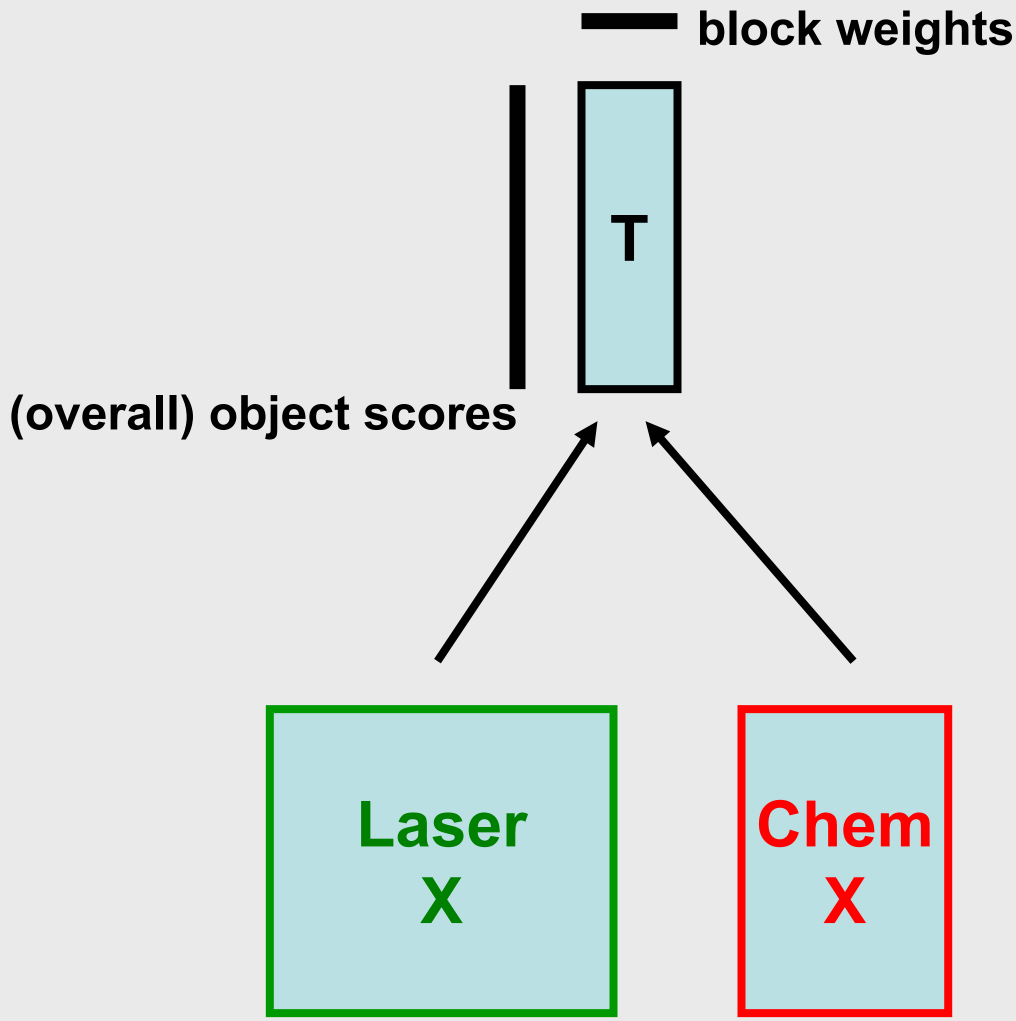
Multi-block PCA

Algorithm



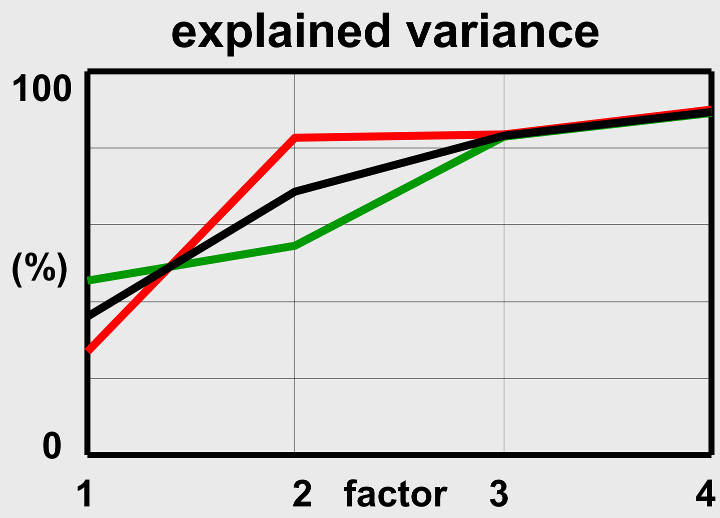
Example

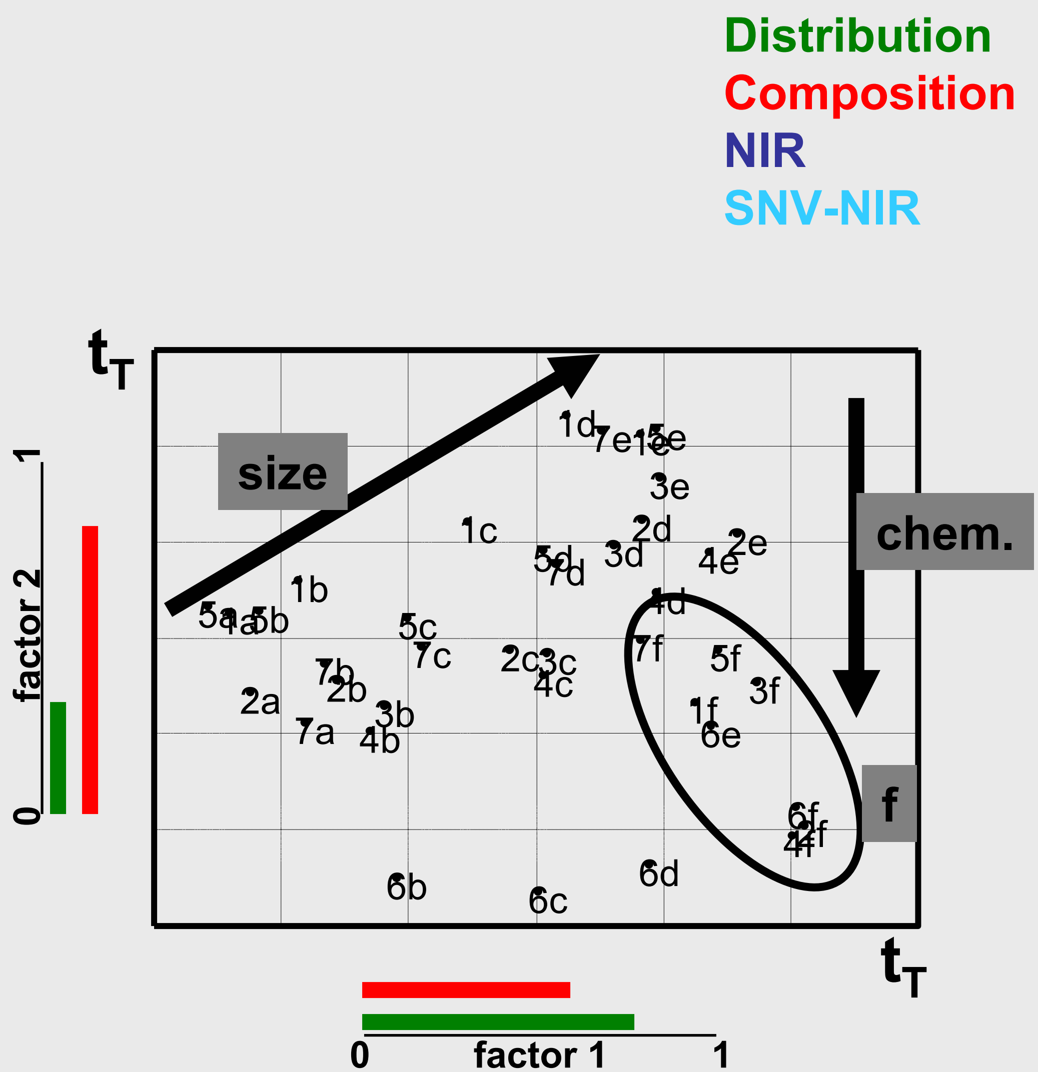
Wheat flourmill data



Example

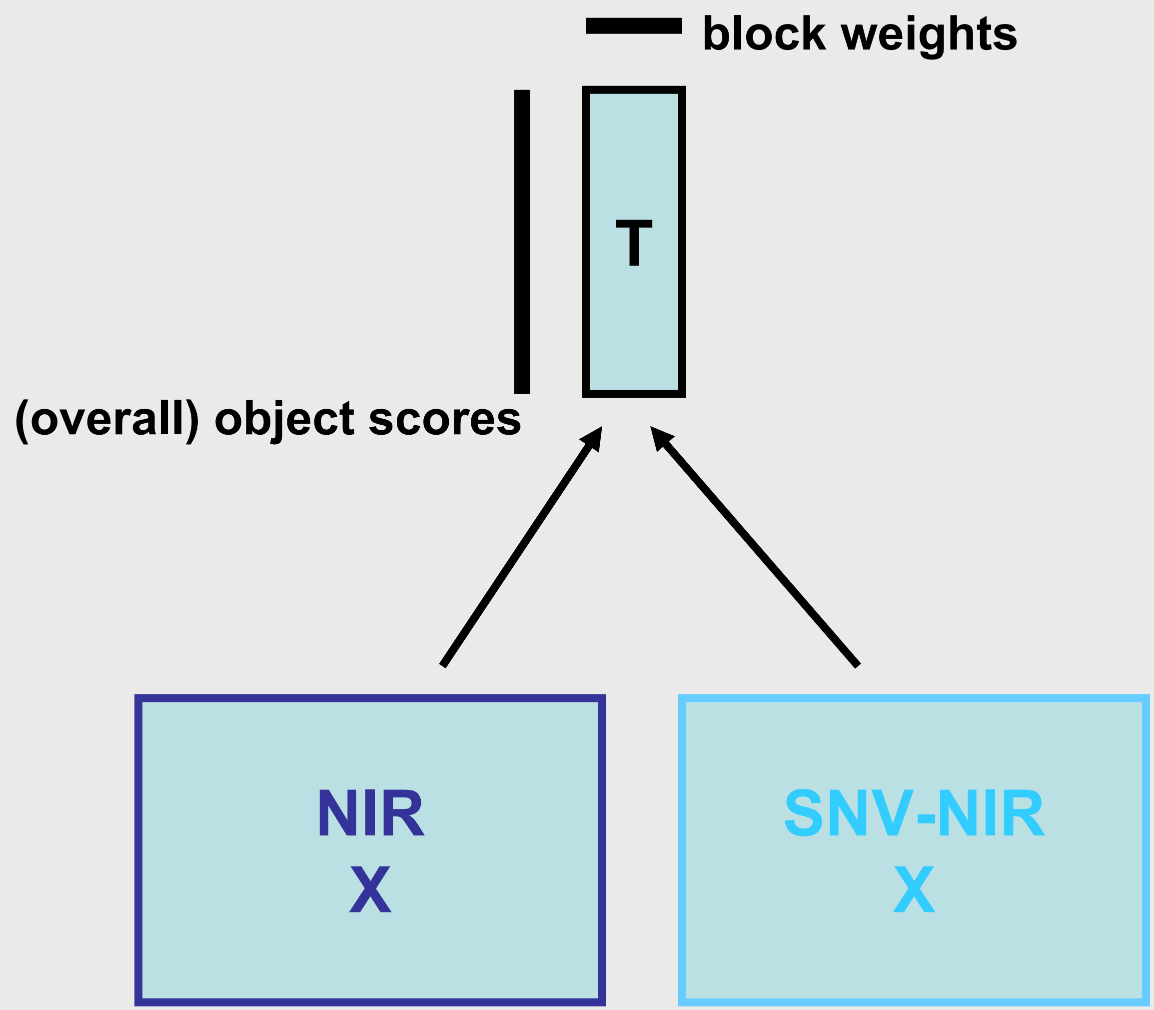
Off-line (laboratory) data





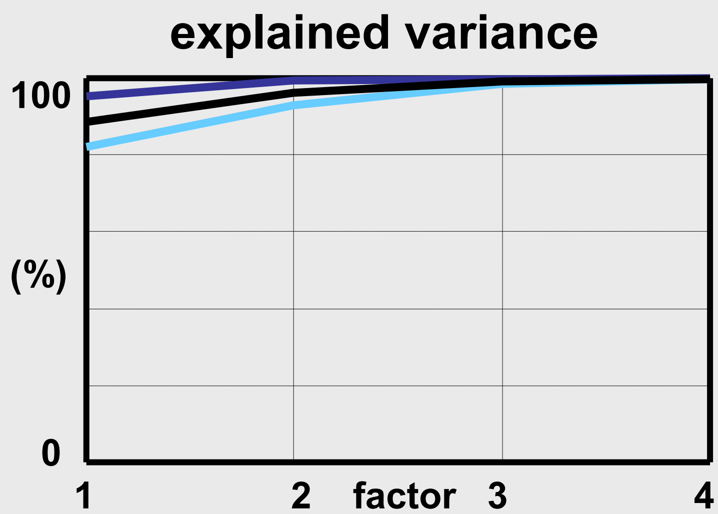
Example

Wheat flourmill data

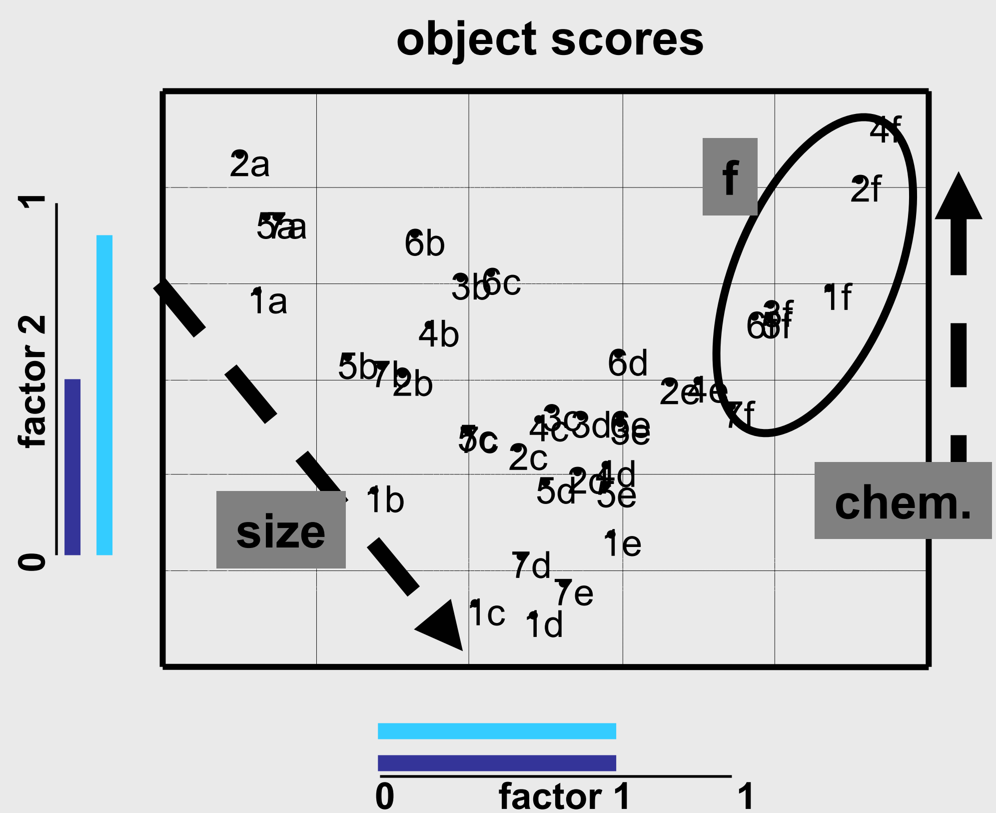


Example

In-process data

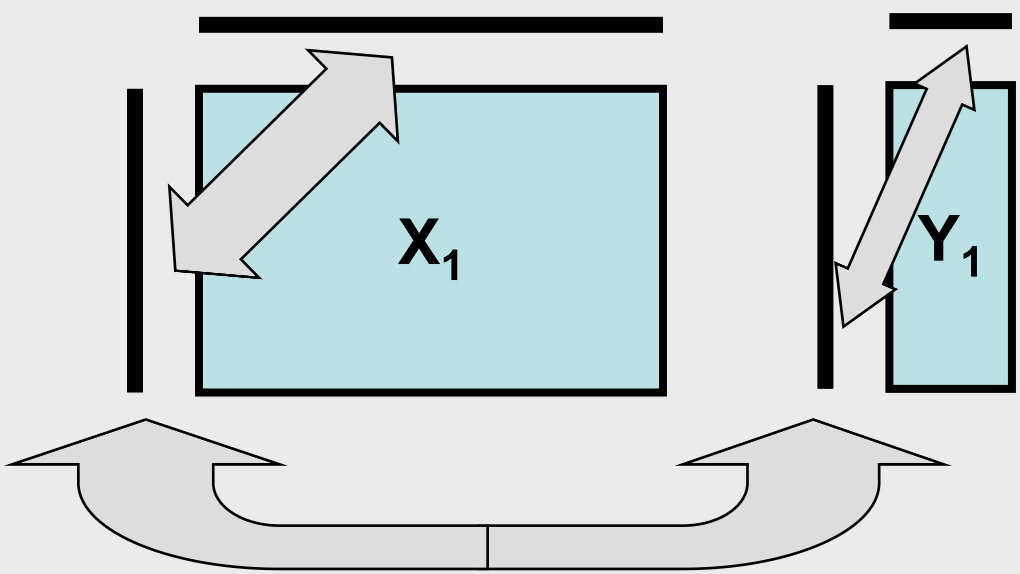


Distribution
Composition
NIR
SNV-NIR



Partial Least Squares (PLS)

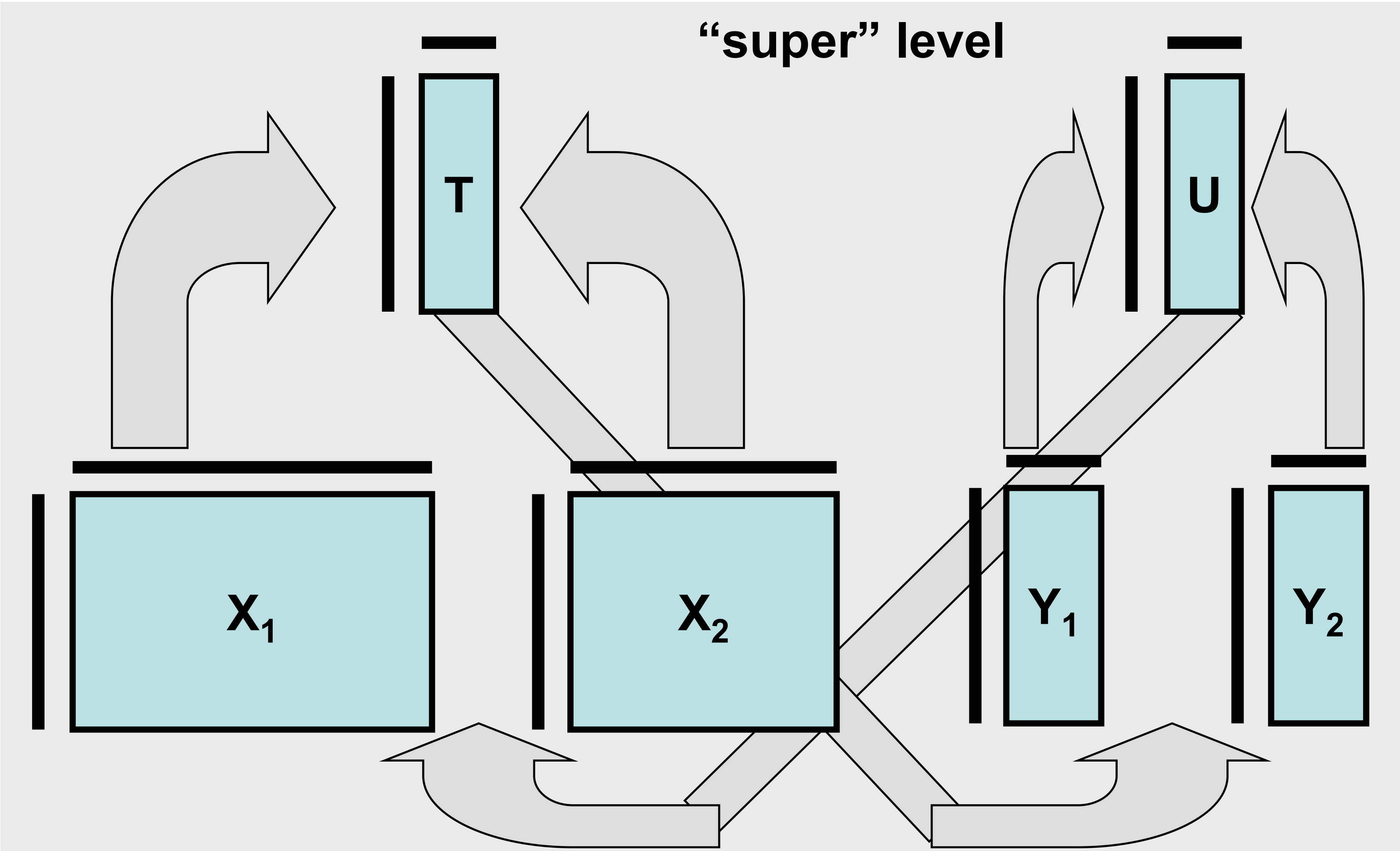
Algorithm



(1.16)

Multi-block PLS

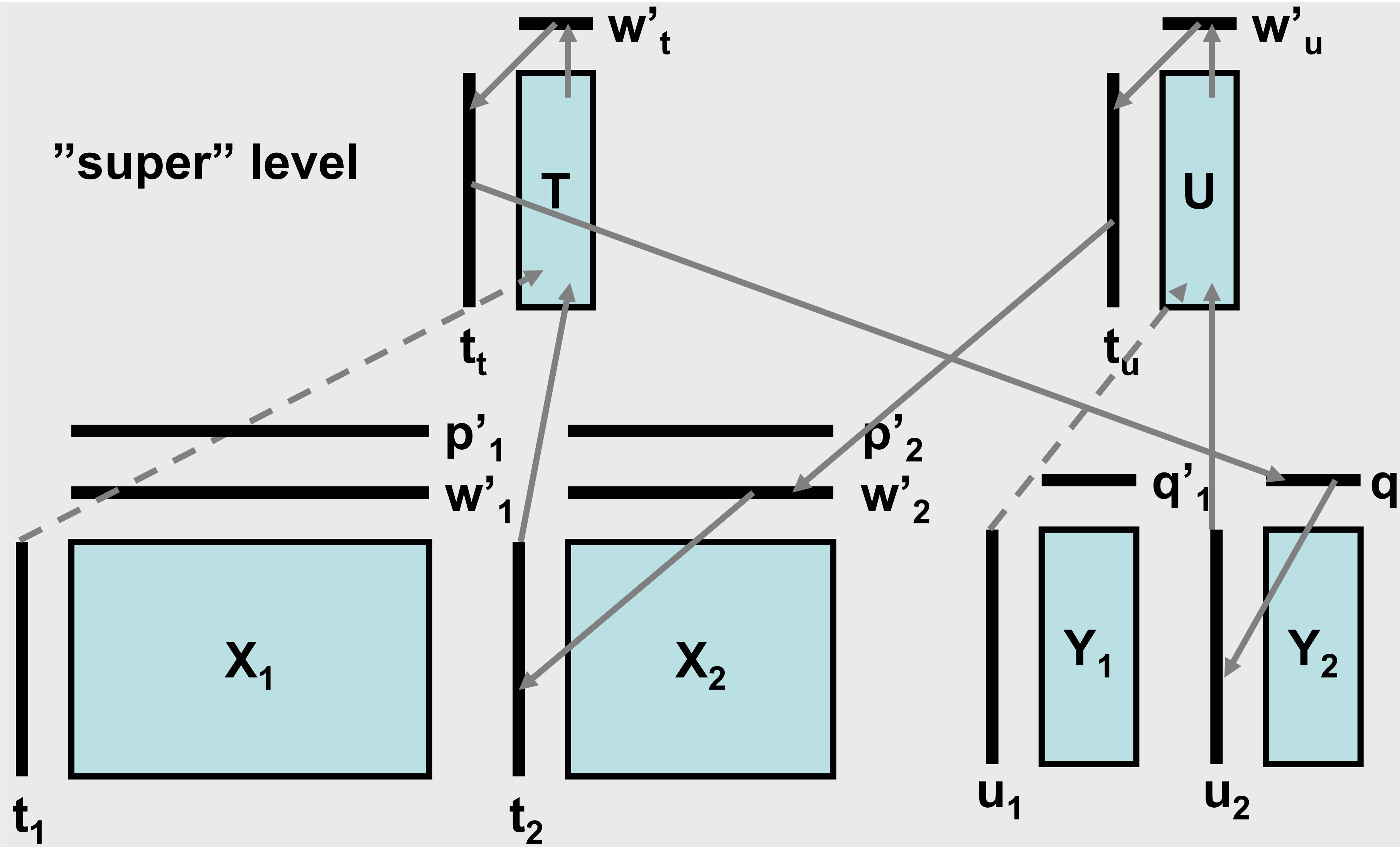
Super level



(1.17)

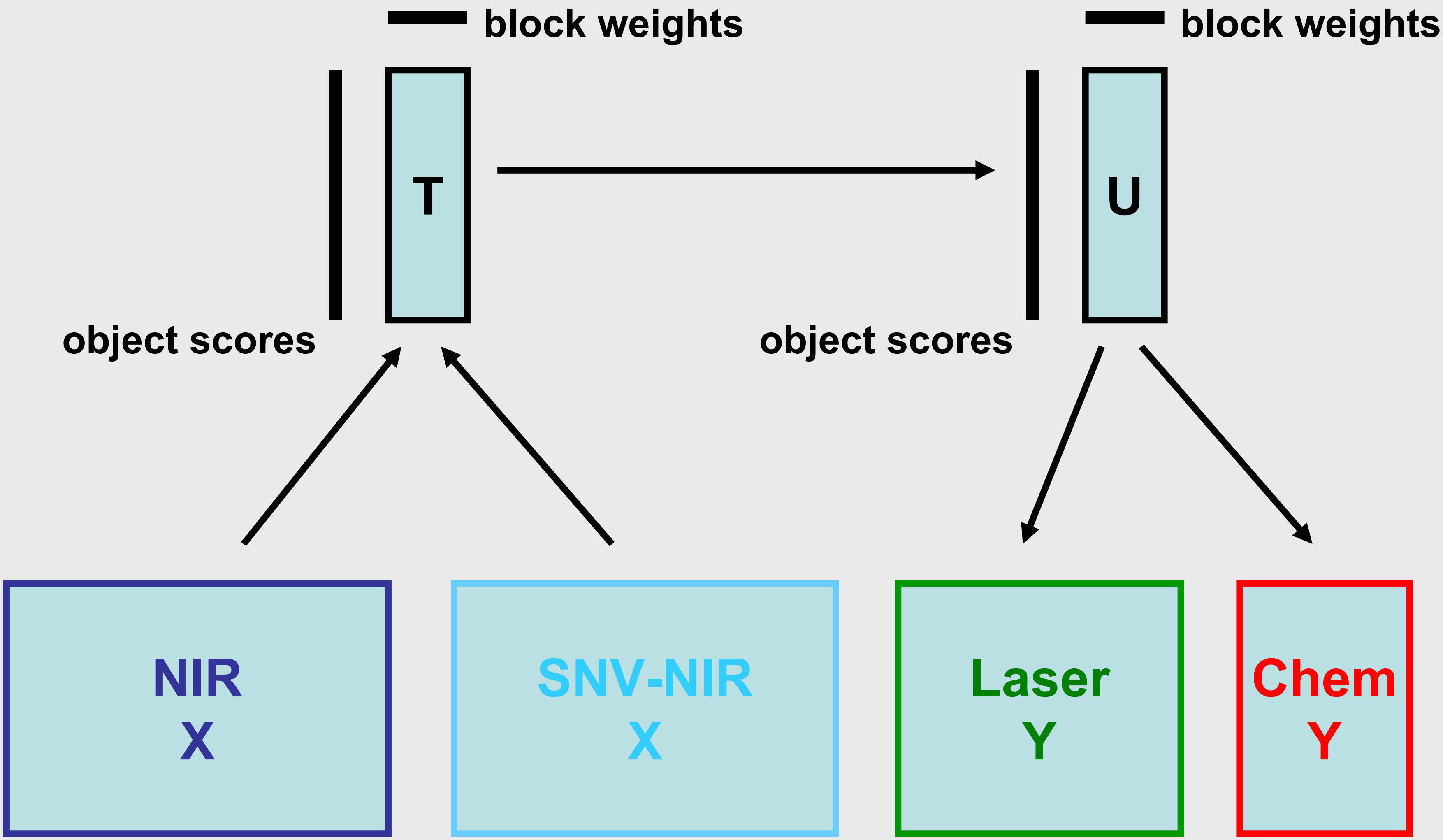
Multi-block PLS

Algorithm



Example

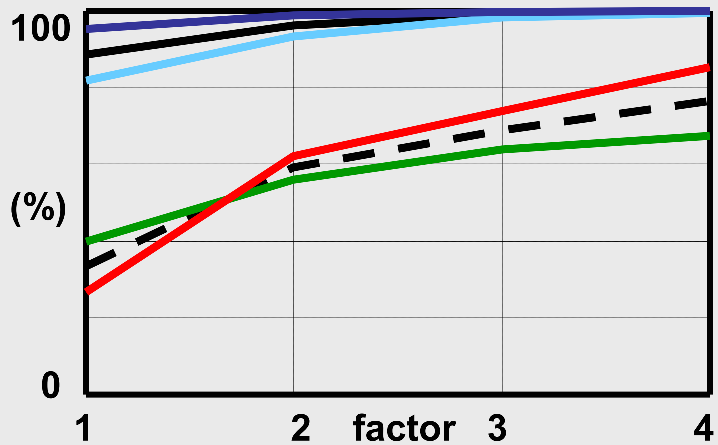
Wheat flourmill data

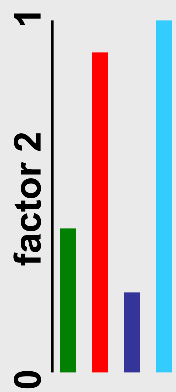


Example

In-process → laboratory data

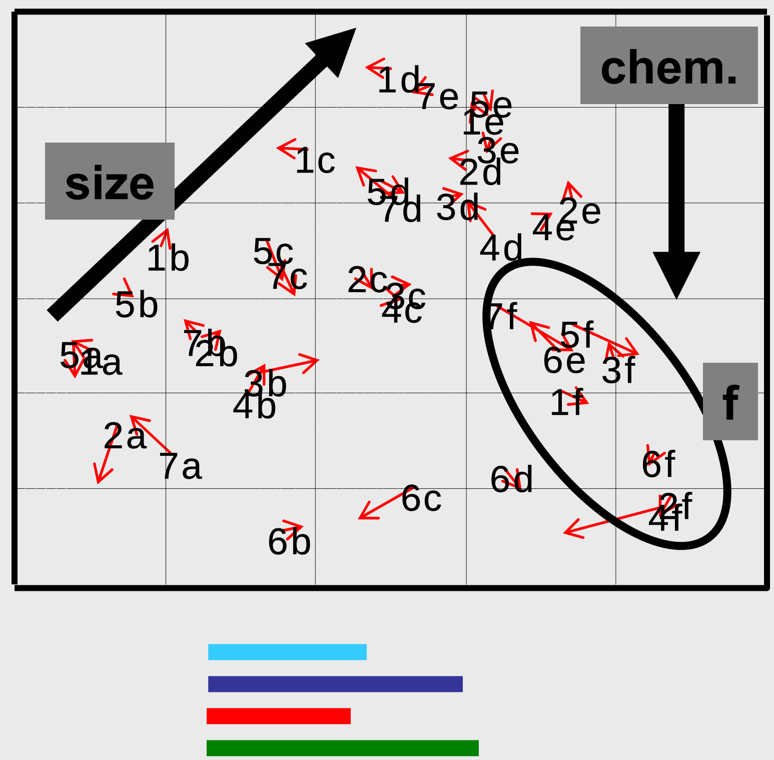
explained variance





Distribution
Composition
NIR
SNV-NIR

object scores (U)



(→) **cross validation prediction error**

Second example

Cheese storage

The Royal Veterinary and Agricultural University (KVL), Denmark
Dept. of Dairy and Food Science, Food Technology

Cheese Making

Separation → **Skim milk**, **Cream**

↓

Standardization

↓

Culture

↓

Pre-pressing → **Whey**

↓

Salting

↓

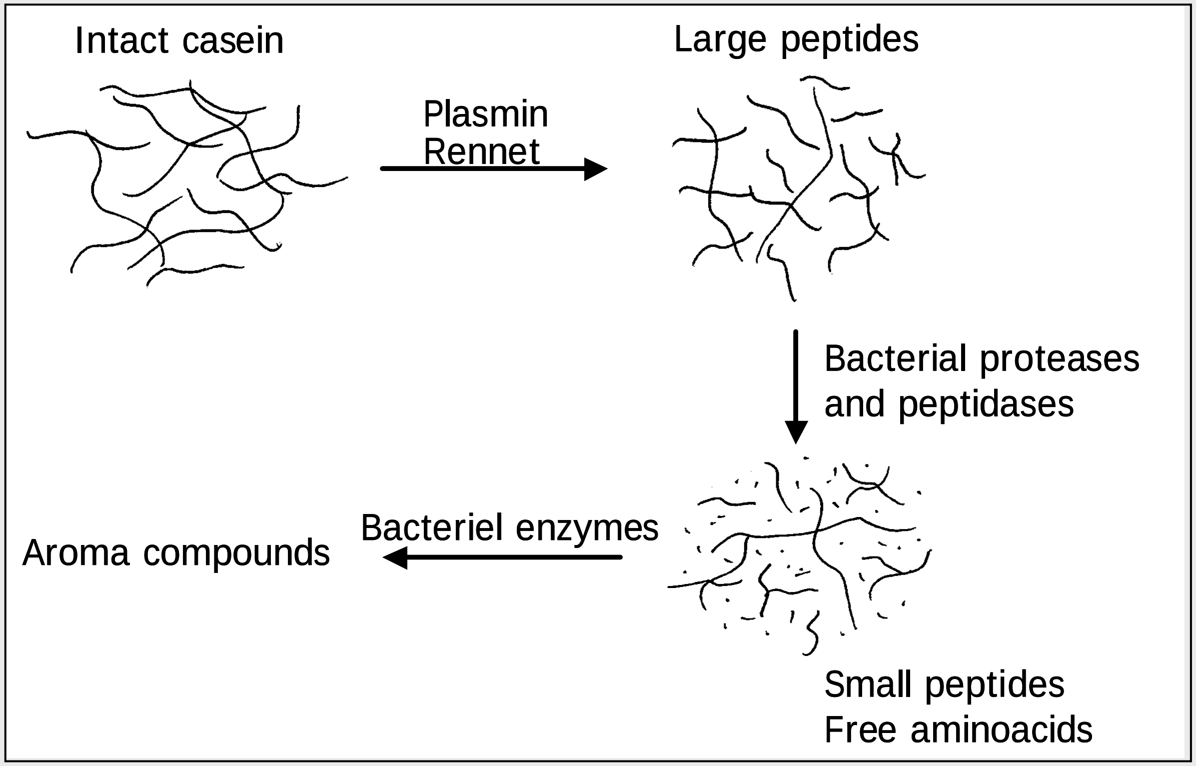
Storage → **Flavor and texture development**

Proteins, peptides, & aroma compounds

Second example

(1.21)

Cheese ripening



3 (large) chemical and physical measurement series,
35 weeks of storage,
plus sensory evaluation of the end product

Second example

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).

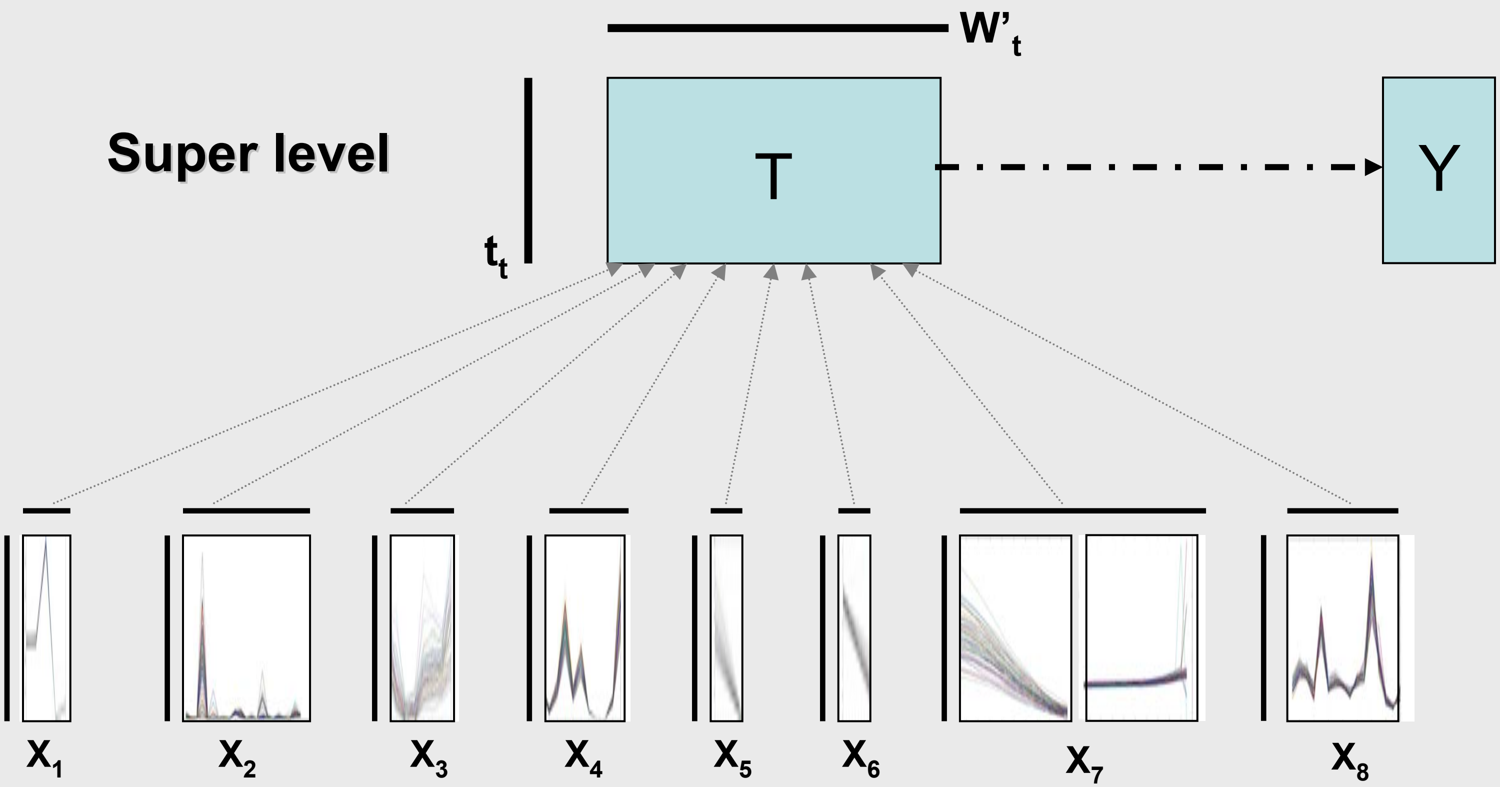
Following measurements are performed at week 35

Sensory Evaluation

- 12 sensory attributes

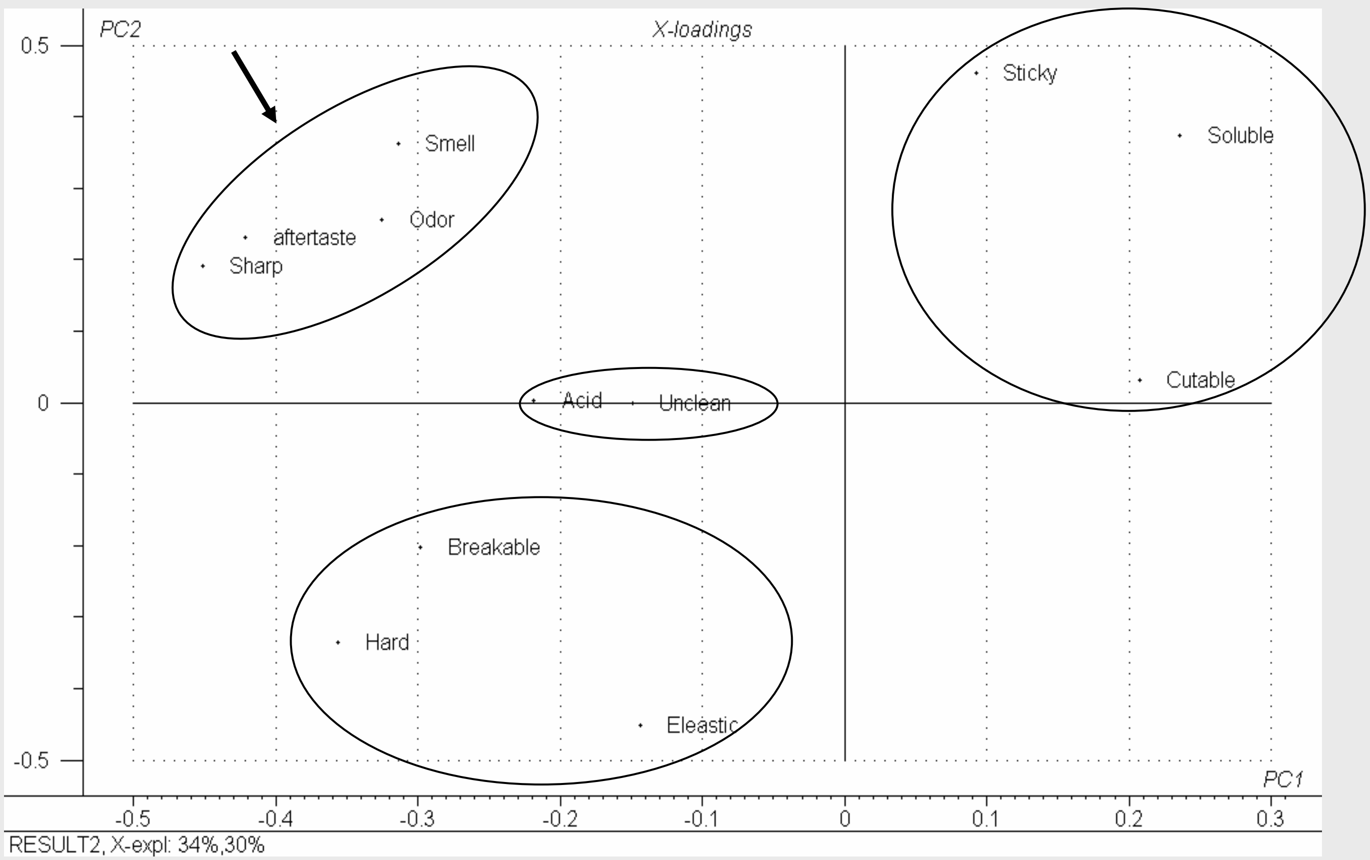
Second example

Multi-block model



Second example

Selection of Y-variables (sensory attributes)



Second example

Multi-block decomposition

<u>Different view-points:</u>

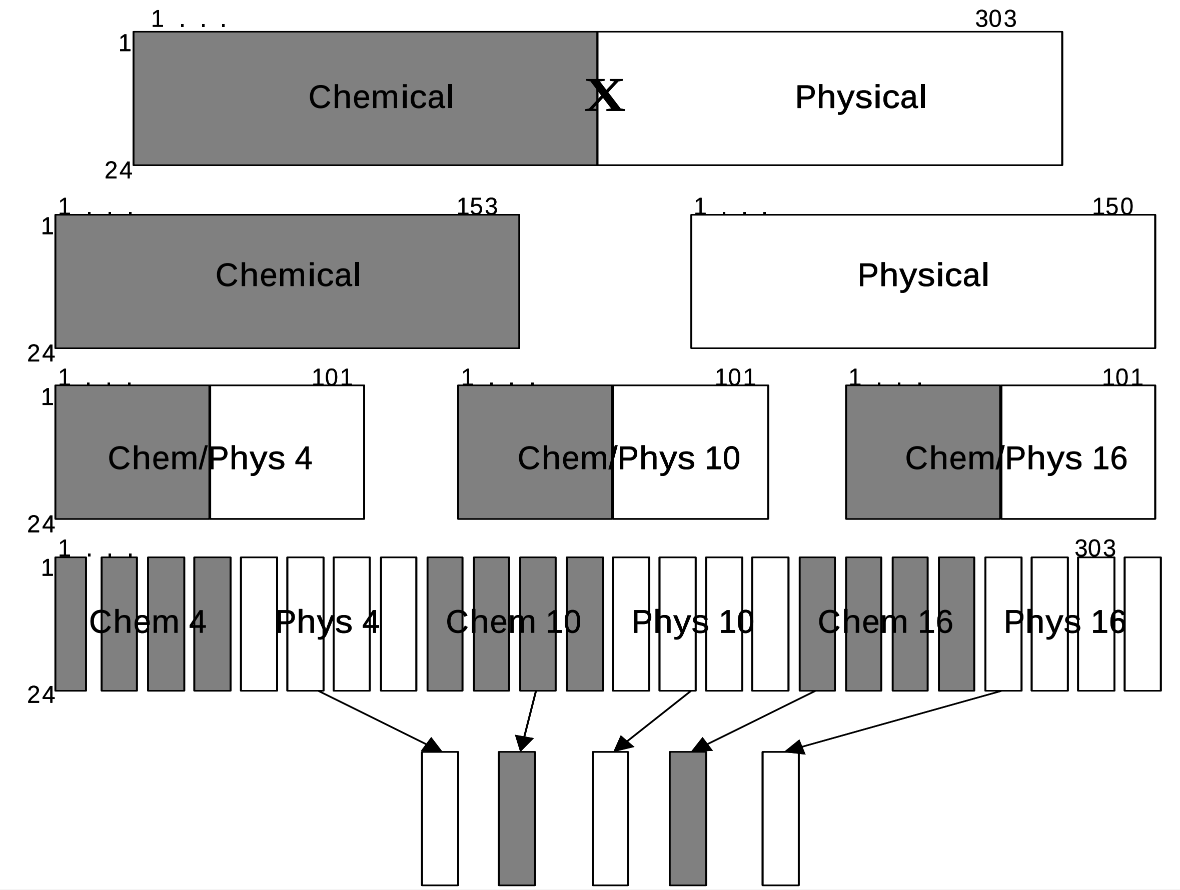
All there is

Nature of the signal

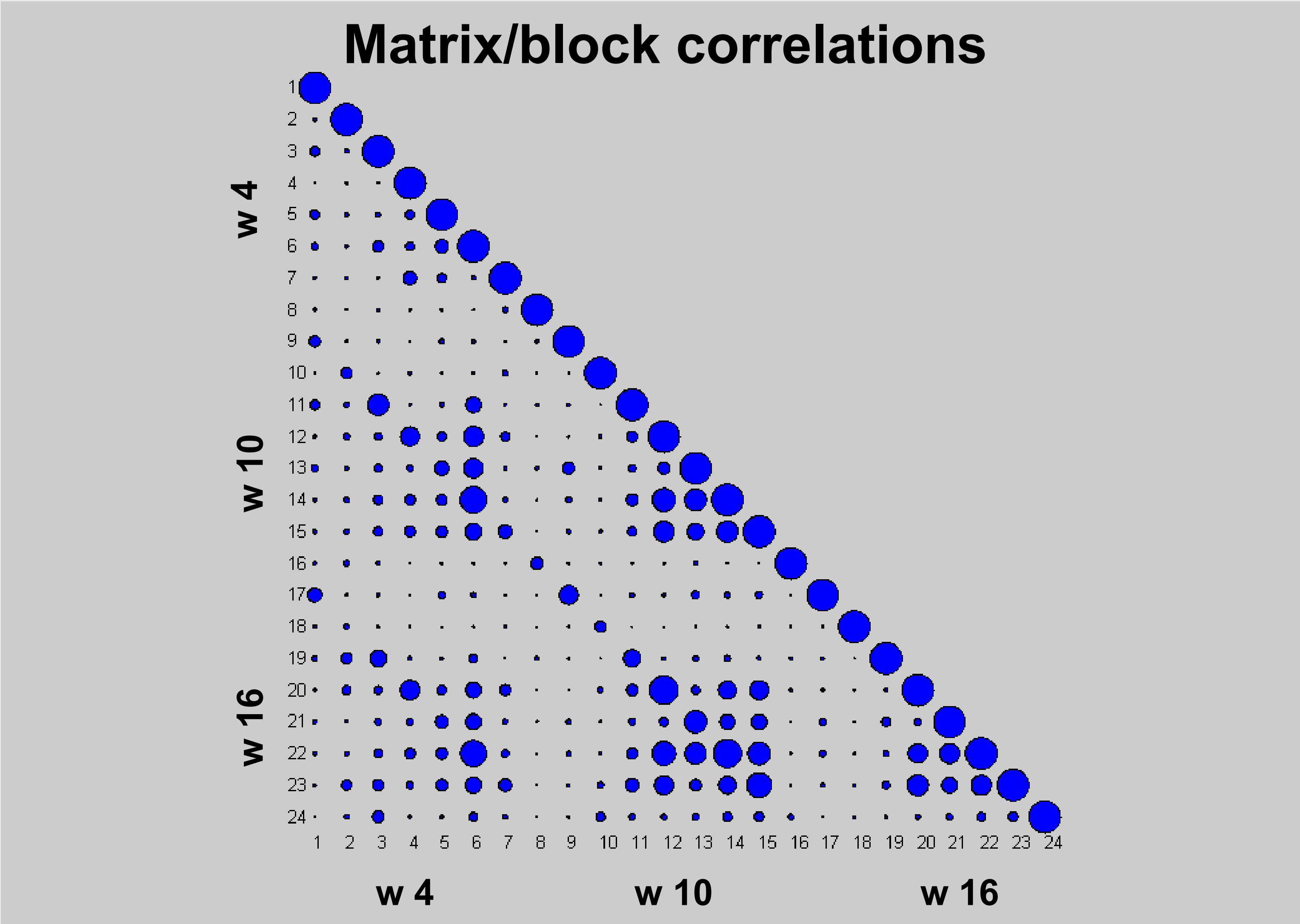
Time

Time/Signal

Selection

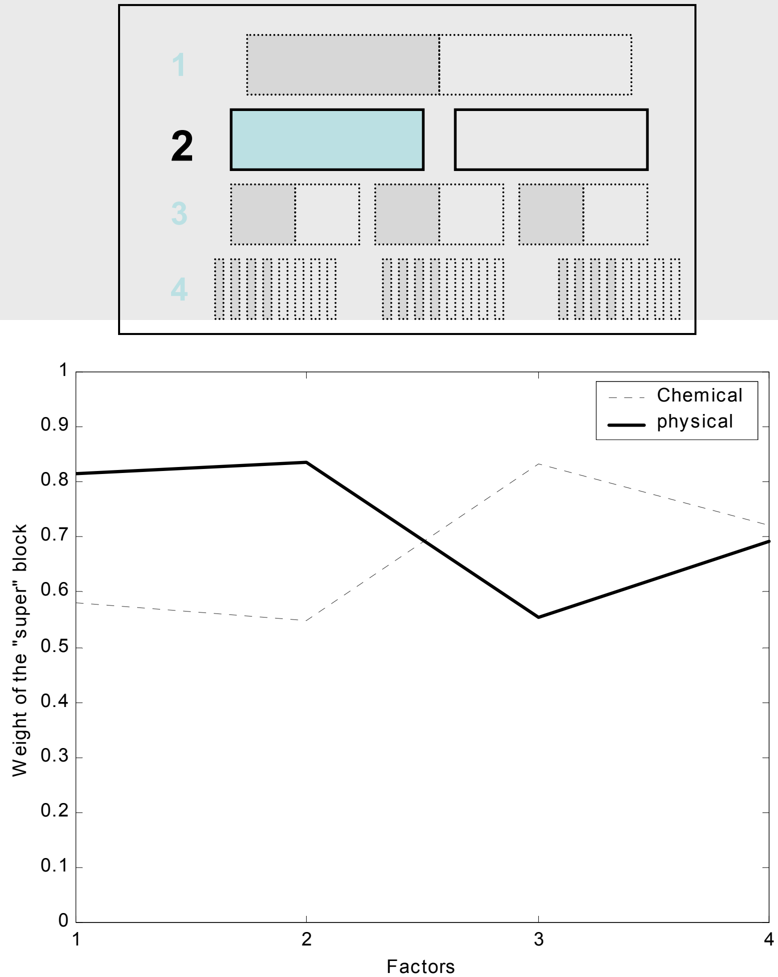


Second example

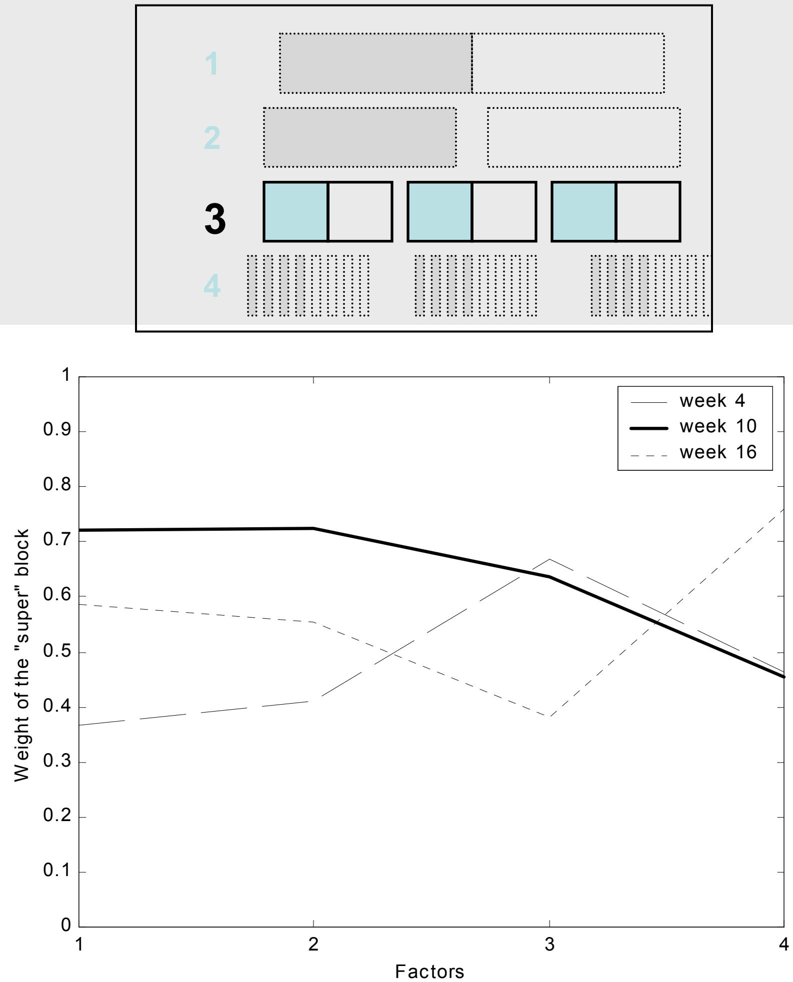


Second example

Y1:decomposition step 2

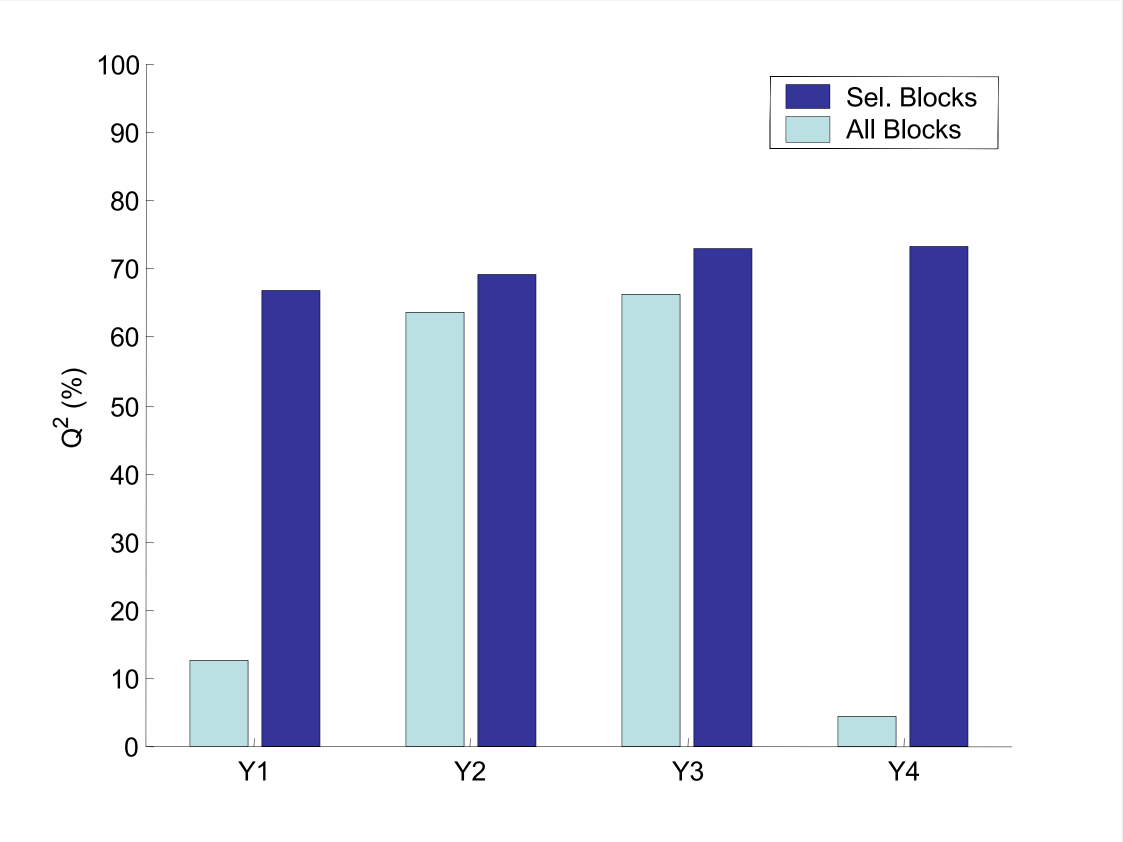


Y1:decomposition step 3



Second example

Optimization in Q^2 for the new models



Y1: Cheese smell

Y2: Cheese taste

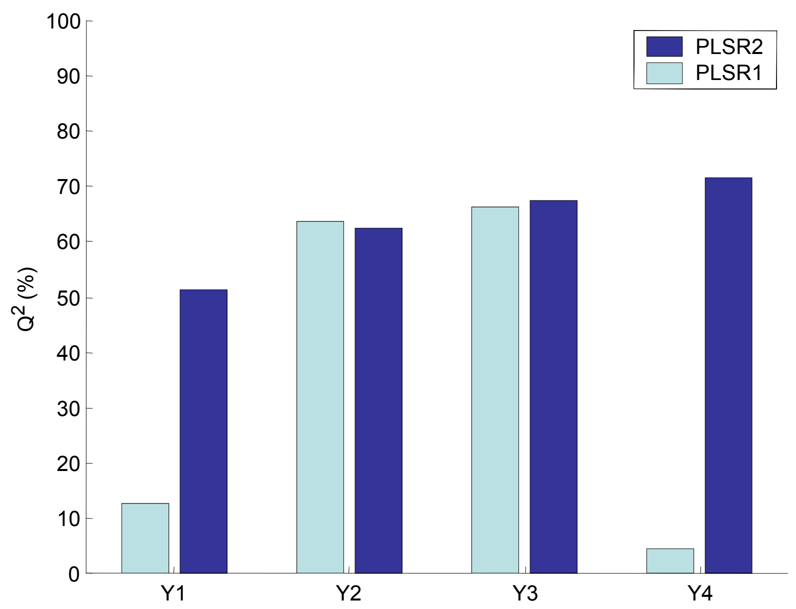
Y3: Sharp taste

Y4: After taste.

Second example

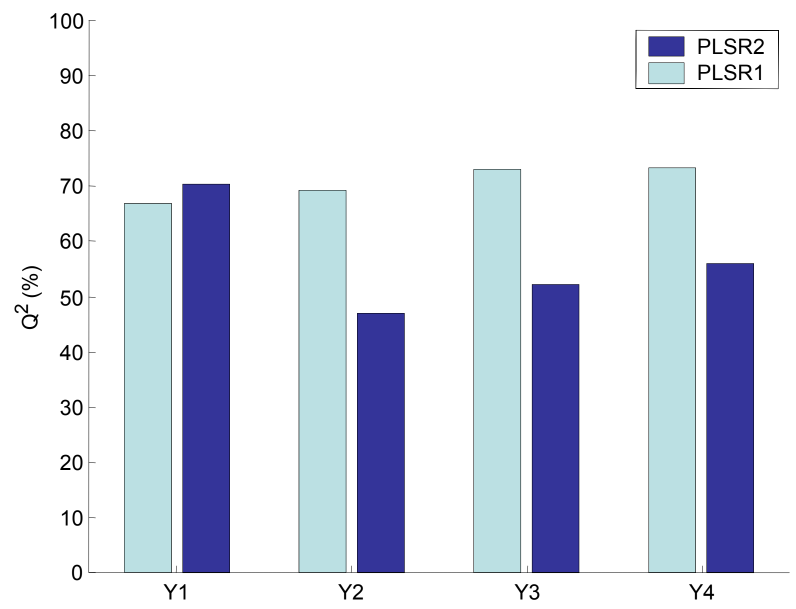
Q^2 for the PLSR1 vs. PLSR2 models

All blocks



7 factor PLSR2

Selected blocks



5 factor PLSR2

Break

(0.00)

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

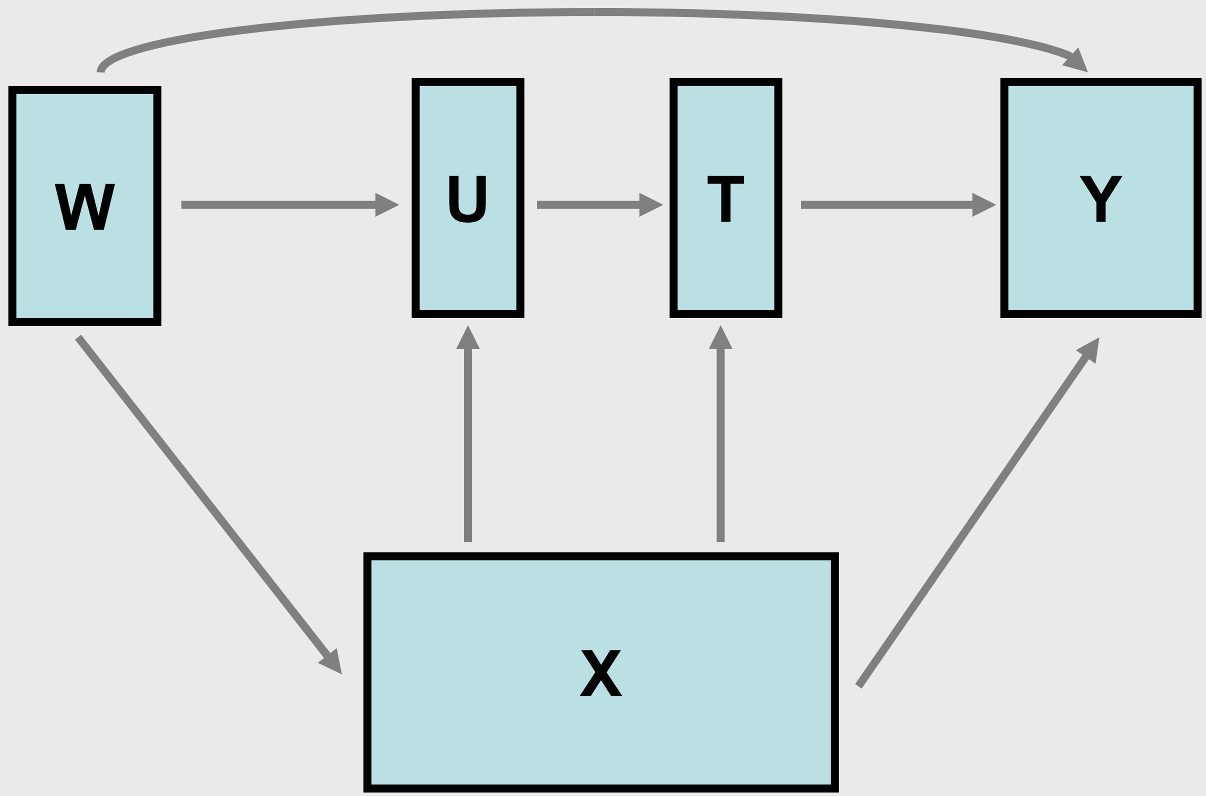
- 1982 H.Wold	path-PLS models
- 1984 S.Wold/Martens/H.Wold	multiblock-PLS
- 1984 Frank/Kowalski	wine, chemical and sensory
- 1988 Wangen/Kowalski	simulations

Origin of PLS like we know it is rooted in Multi-block PLS!!!

Path-PLS
Causality relations

Michel Tenenhaus et.al. *PLS methodology to study relationships between hedonic judgements and product characteristics*
Food Quality and Preference (2004) in press

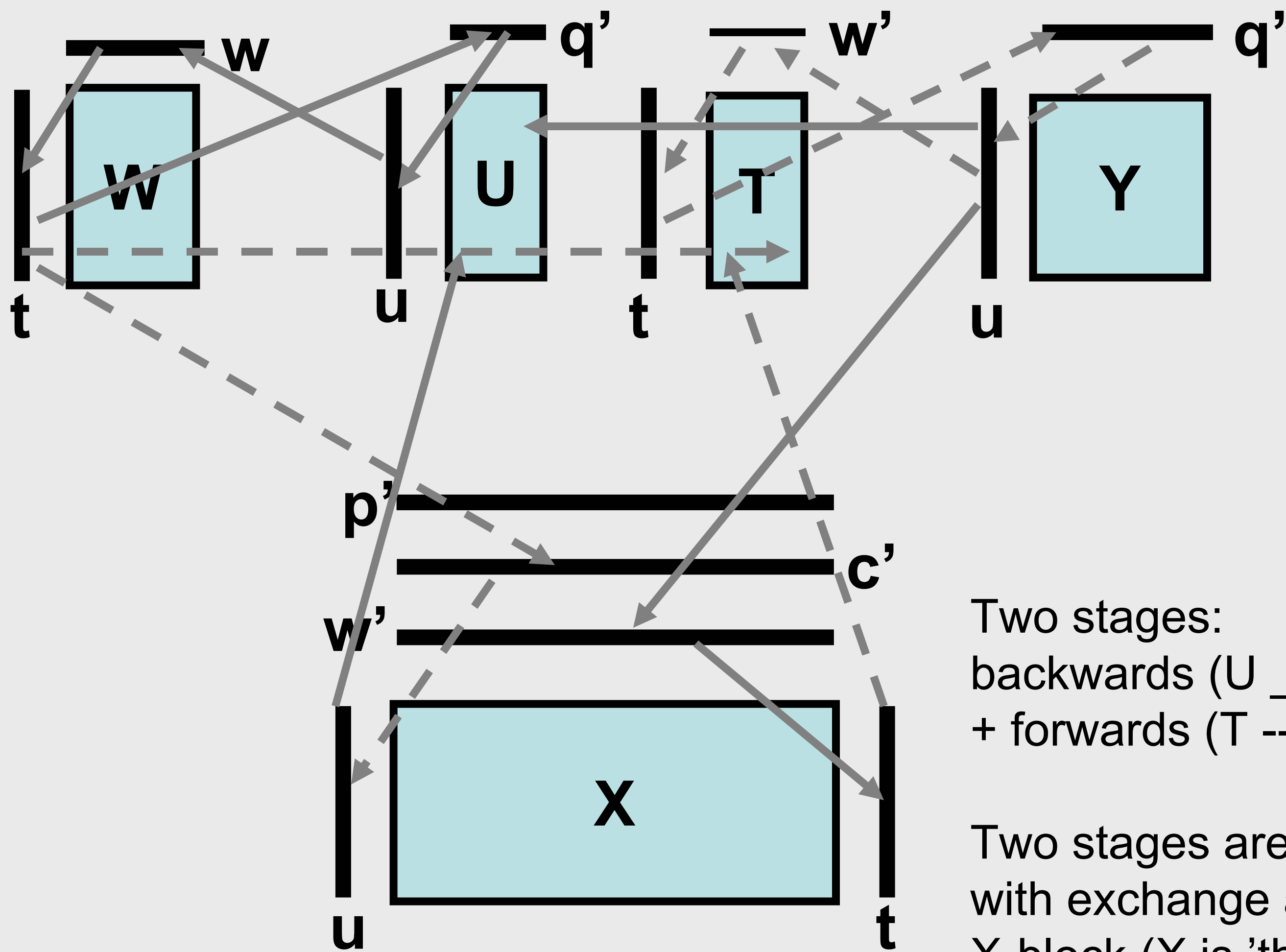
Response to LISREL from social and business sciences
(”one factor PLS per connection”)   Notice that there is a direction!



(2.04)

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’)

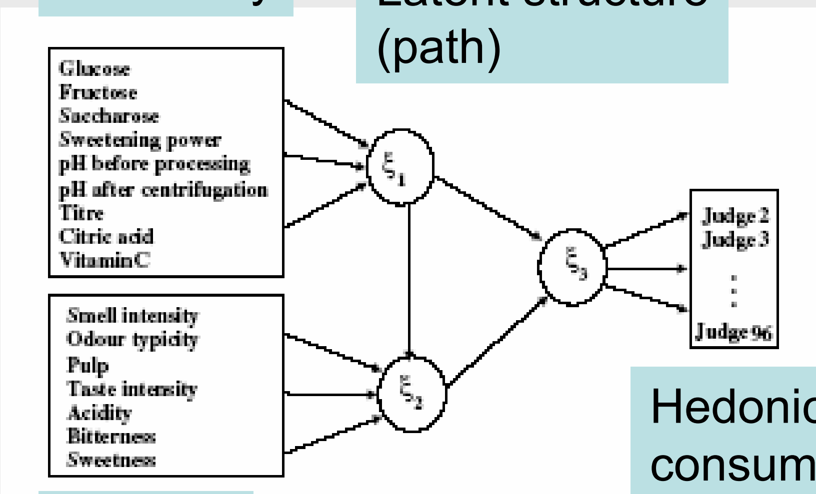
Path-PLS

Causality relations

Evaluation of six orange juices

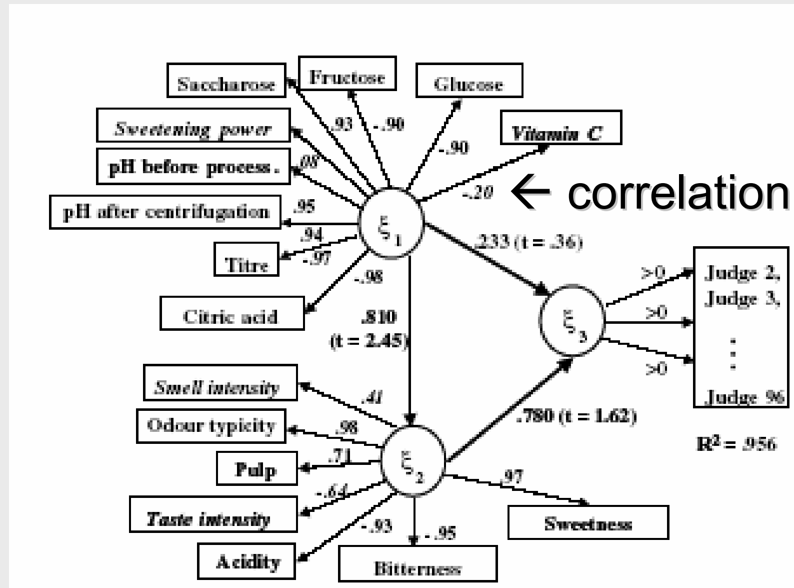
Chemistry

(Imposed) Latent structure (path)



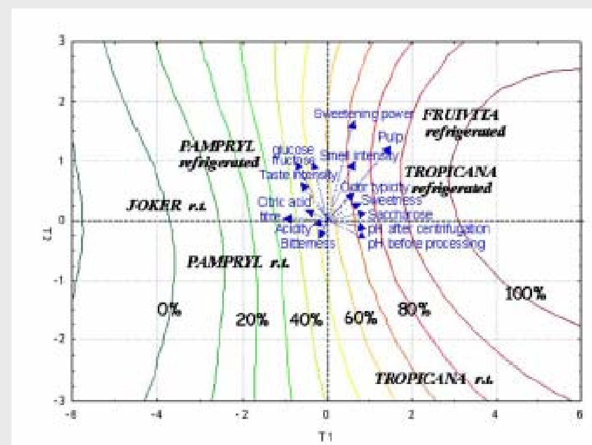
Sensory

Hedonic consumer judgment



← correlation

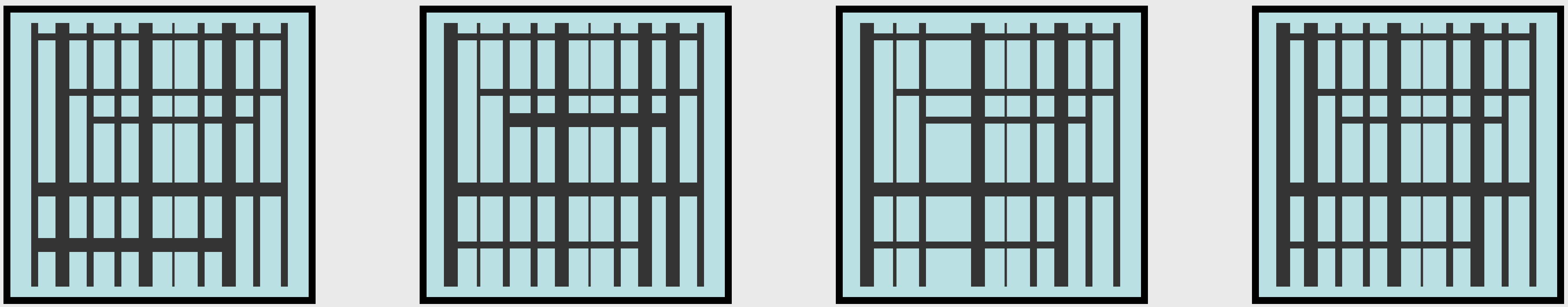
Different objective:
e.g. product optimization
or sales



Multi-block

(2.06)

Same block sizes



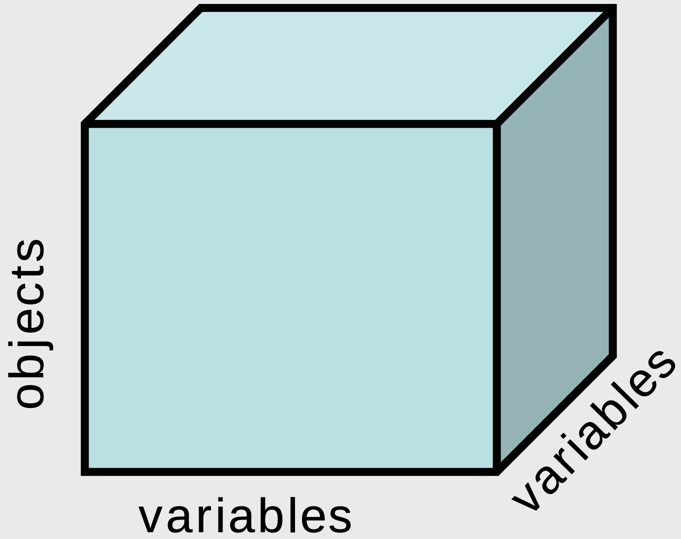
Same information matrix/data-table for different samples

PARAFAC and Tucker

(2.07)

Same block sizes

Rasmus Bro *PARAFAC Tutorial and applications* - Chemometrics and Intelligent Laboratory Systems 38(1997)149-171
Claus A: Andersson et.al. *Analysis of N-dimensional data arrays from fluorescence spectroscopy of an intermediate sugar product* Fresenius Journal of Analytical Chemistry 359(1997)138-142

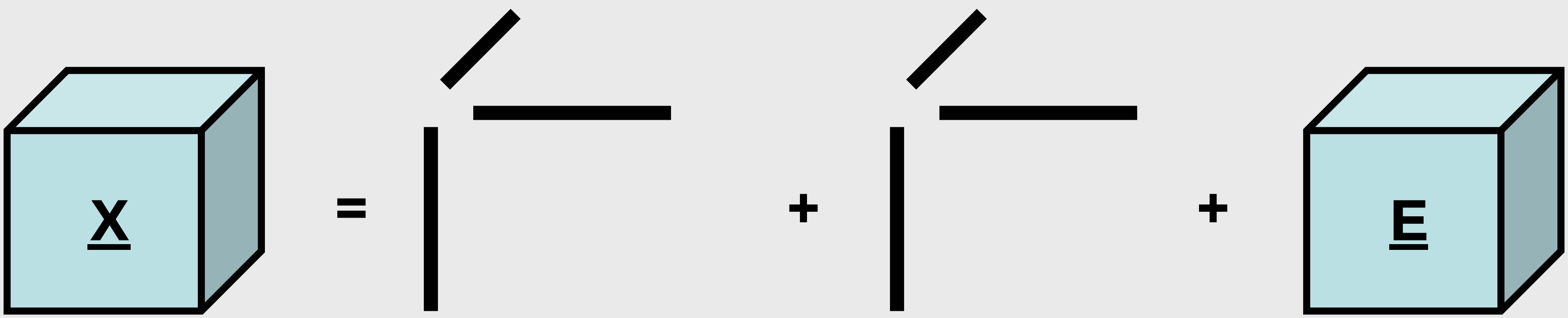


Hyphenated Analytical Chemistry : samples x GC x MS

Sensory Science : products x assessors x attributes

Process Monitoring : batches x process variables x time

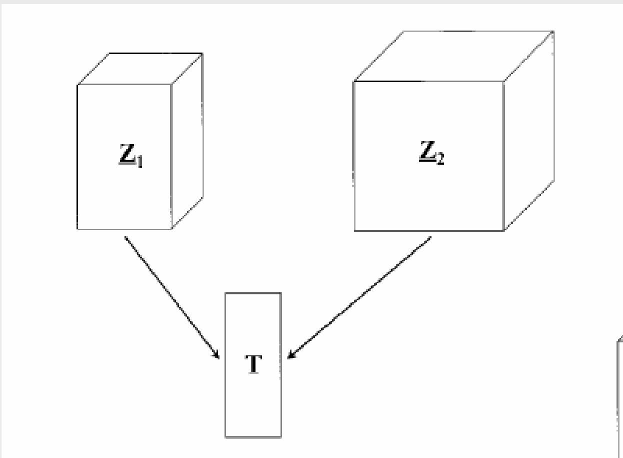
...

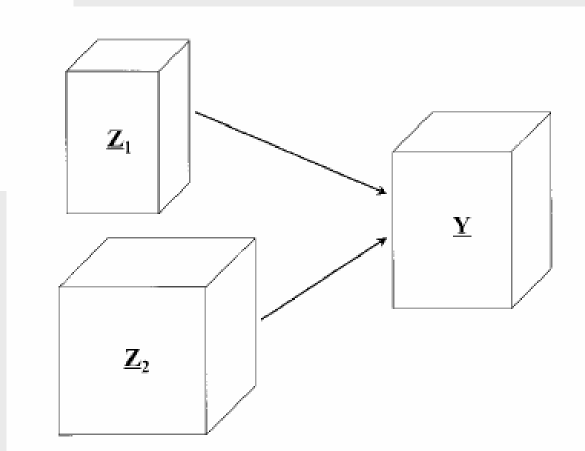


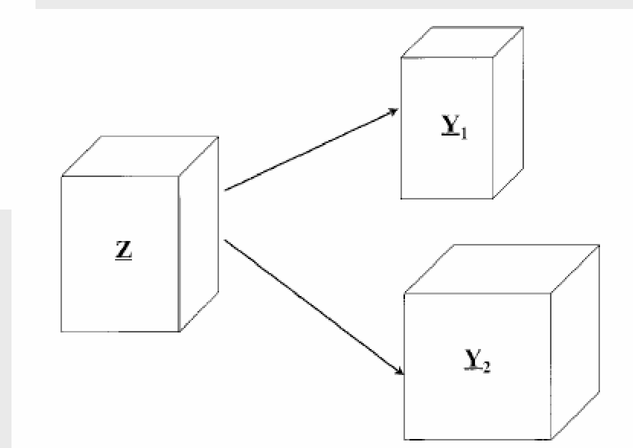
Extensions

Several blocks and tables combined ...

Age Smilde et.al. *Multiway multiblock component and covariates regression models – Journal of* Chemometrics 14(2000)301-311
Age Smilde and Henk Kiers *Multiway covariate regression models Journal of Chemometrics 13(1999)31-48*







Covariate regression

Component-wise versus simultaneous estimation

Age Smilde et.al. *Multiway multiblock component and covariates regression models – Journal of* Chemometrics 14(2000)301-311
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)

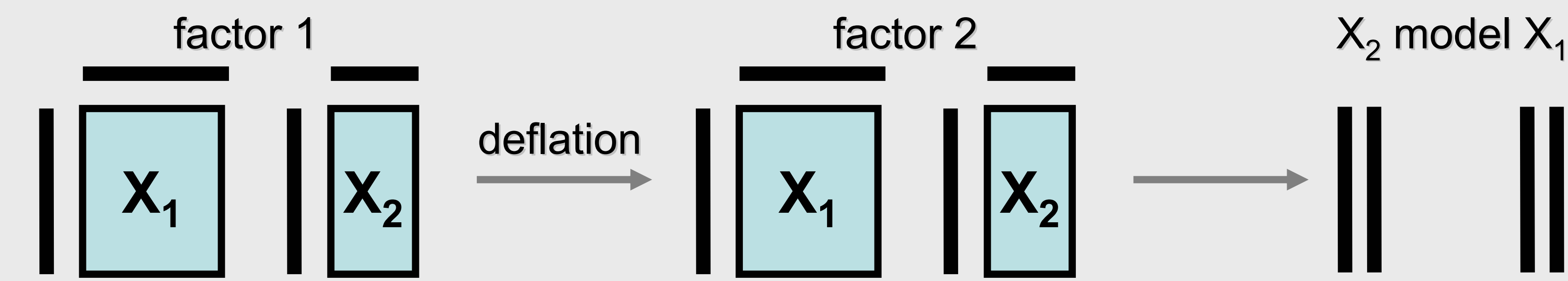
Covariate regression

(2.10)

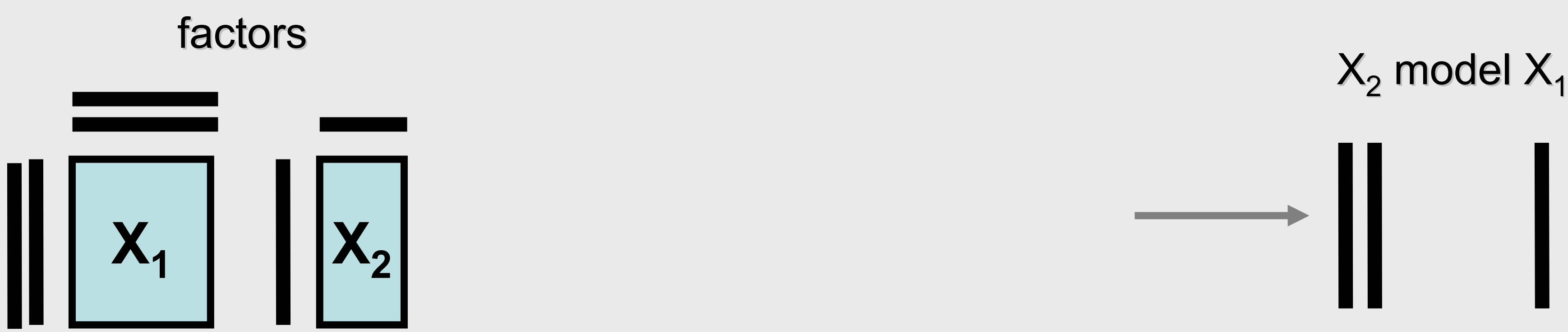
Component-wise versus simultaneous estimation

Age Smilde et.al. *Multiway multiblock component and covariates regression models – Journal of* Chemometrics 14(2000)301-311
Age Smilde and Henk Kiers *Multiway covariate regression models Journal of Chemometrics 13(1999)31-48*

Component-wise (NIPALS)



Simultaneously (ALS) – like Covariates regression



$$\min_{\mathbf{W}}\left[\beta\|\mathbf{X}-\mathbf{XWP}_{\mathrm{X}}^{\mathrm{T}}\| + (1-\beta)\|\mathbf{y}-\mathbf{XWp}_{\mathrm{y}}\|\right] \quad \text{for a given } \beta$$

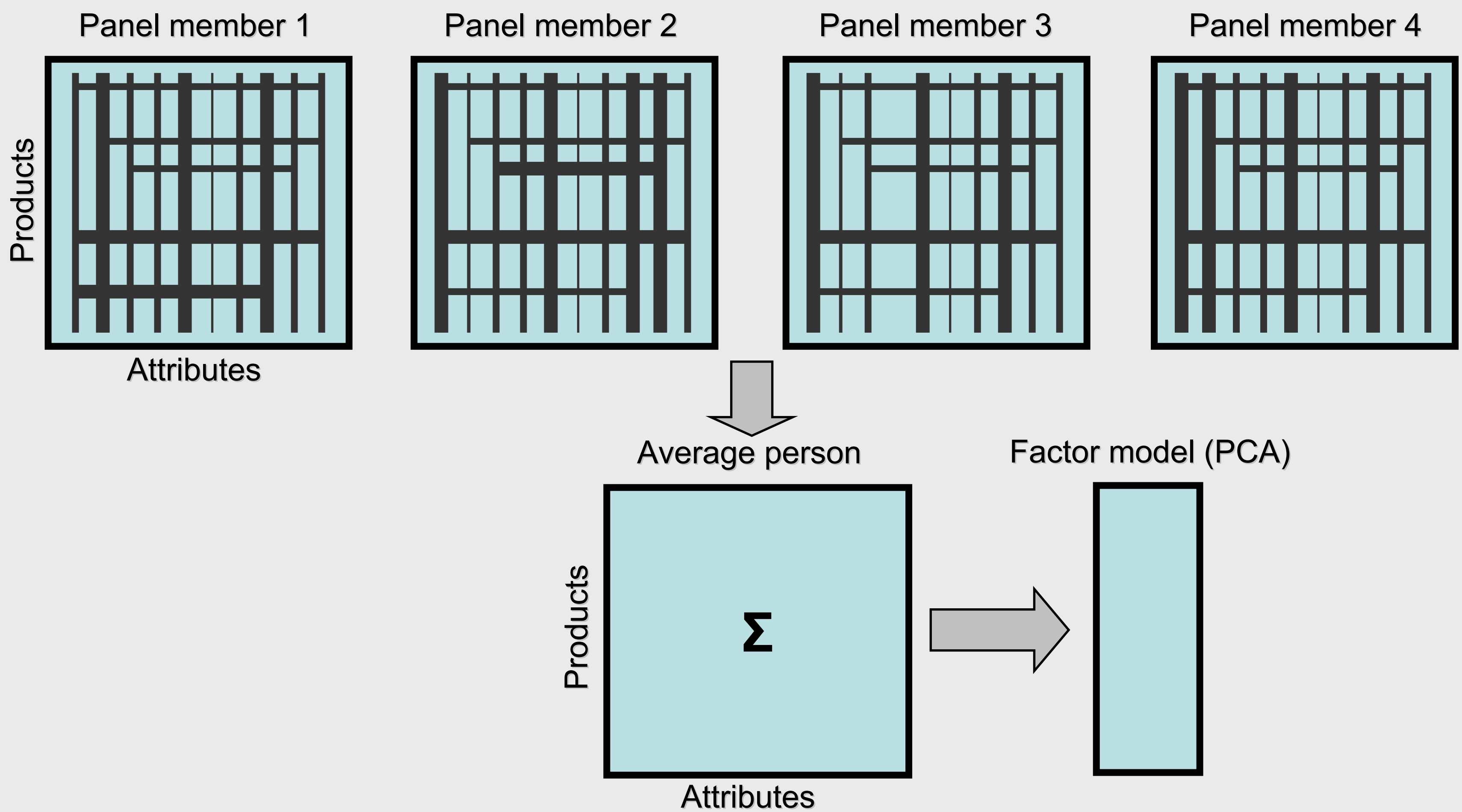
$$0 \le \beta \le 1.$$

Generalized Procrustes Analysis

Same block sizes

(2.11)

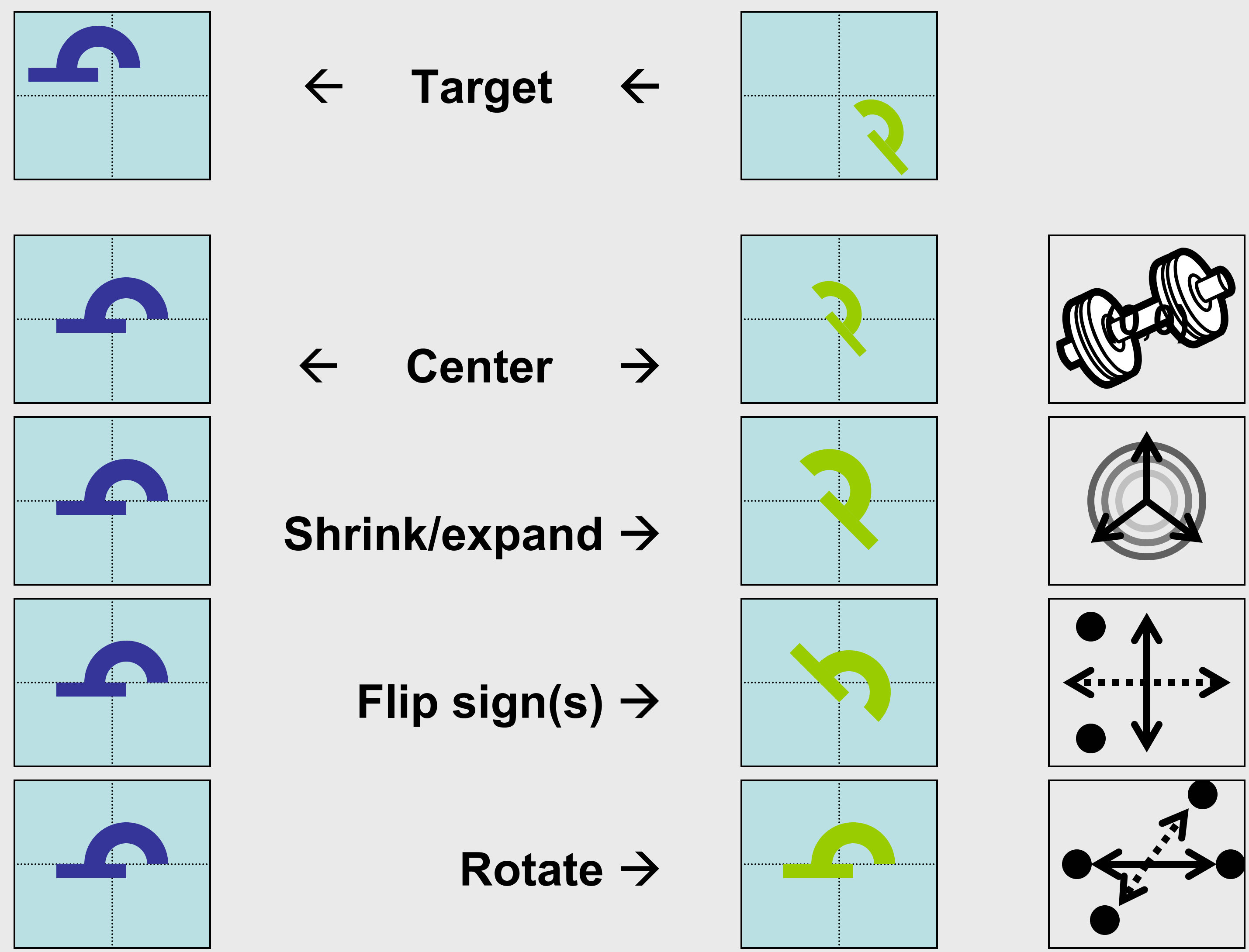
J.C. Gower *Generalized Procrustes Analysis* Psychometrika 40/1(1975)33-51



Generalized Procrustes Analysis

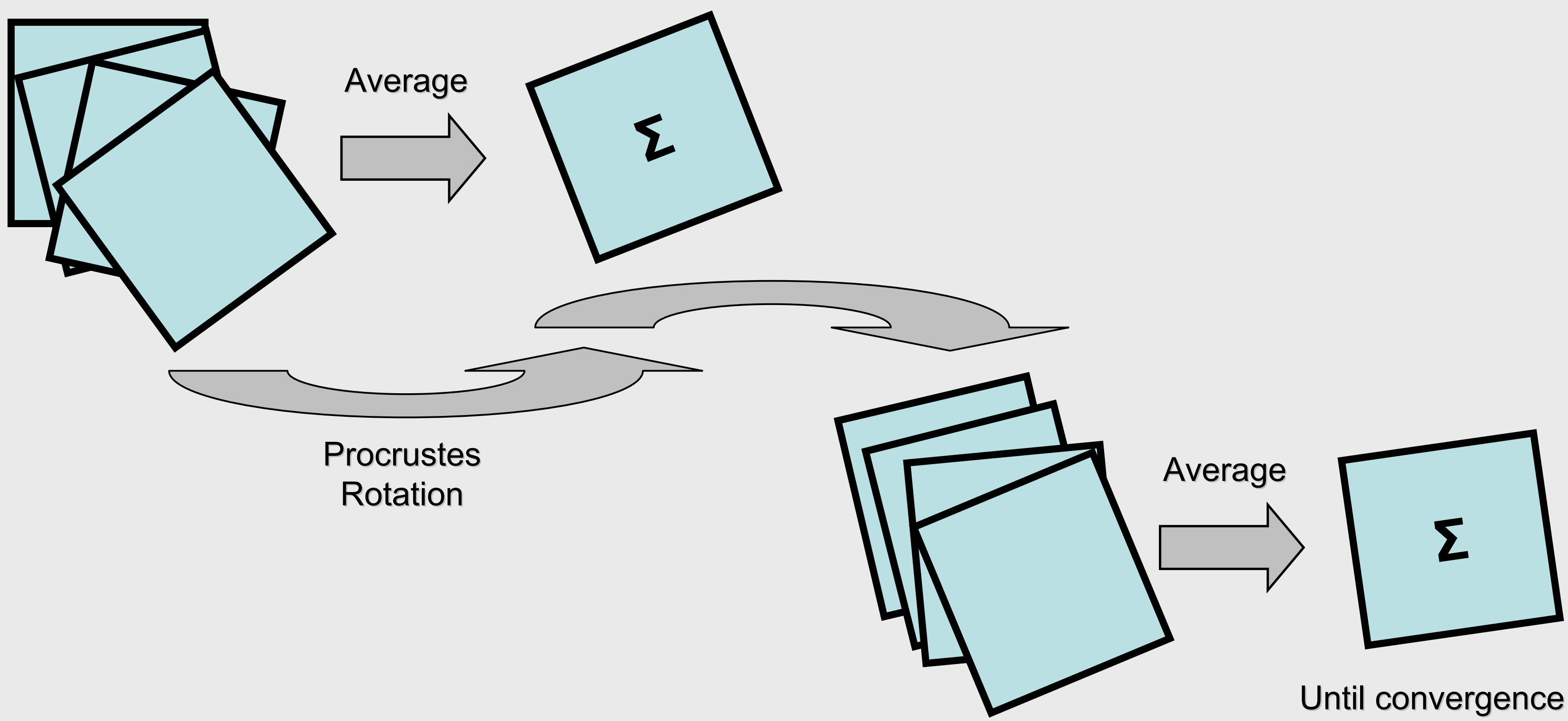
Procrustes rotation

(2.12)



Generalized Procrustes Analysis (2.13)

Same block sizes



Generalized Procrustes Analysis

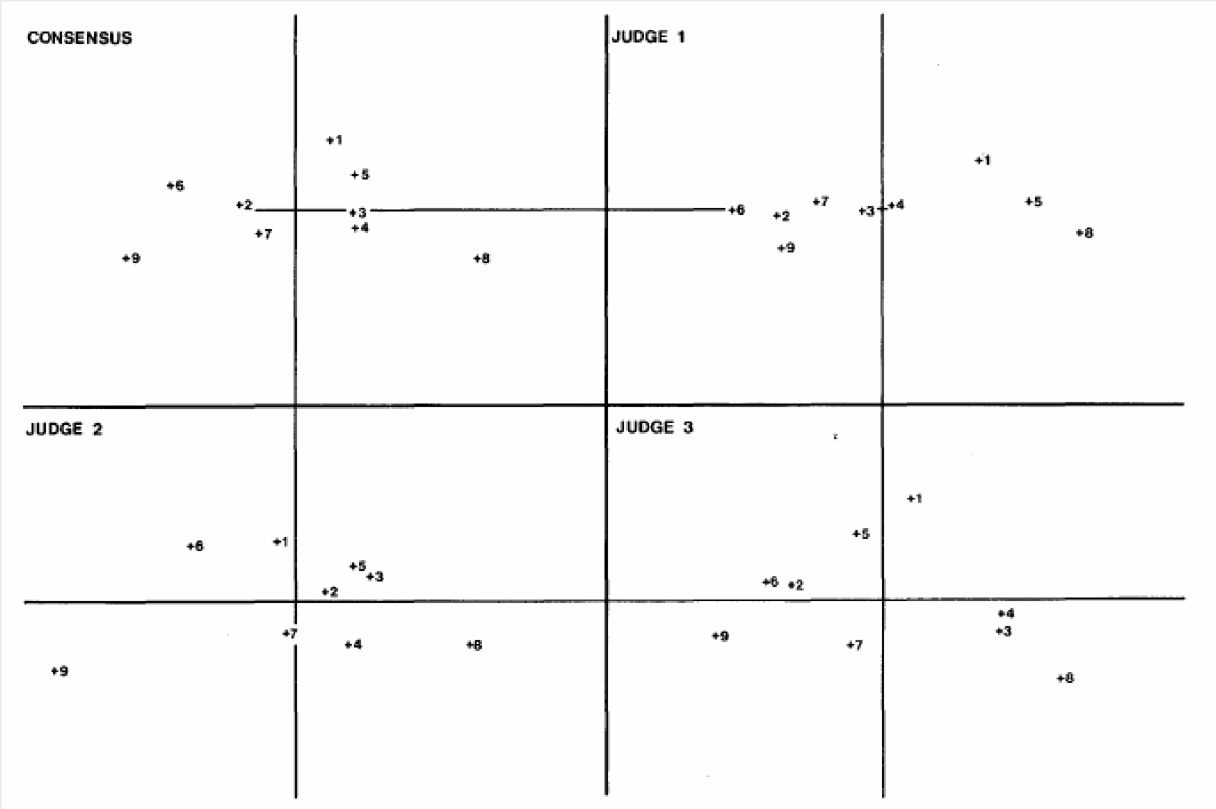
(2.14)

Same block sizes

J.C. Gower *Generalized Procrustes Analysis* Psychometrika 40/1(1975)33-51

Two objectives:

- Get a good average / consensus estimate for the underling phenomena
- Compare individual results with group average / consensus (panel training, homogeneity, etc.)

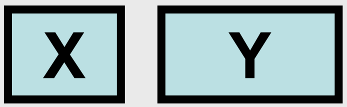


Block correlation (RV) coefficient

(2.15)

Intermediate

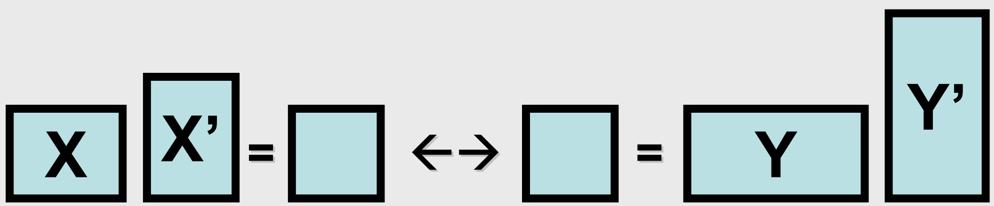
P. Robert and Y. Escouvier *A unifying tool for linear multi-variate statistical methods: the RV-coefficient* **Applied Statistics 25/3(1976)257-265**
J. Ramsay, J. ten Berge and G. Styan *Matrix Correlation* Psychometrika 49/3(1984)403-423



$$RV(\mathbf{X},\mathbf{Y}) = tr(\mathbf{X}.\mathbf{X}^T.\mathbf{Y}.\mathbf{Y}^T) / \sqrt{(tr(\mathbf{X}.\mathbf{X}^T.\mathbf{X}.\mathbf{X}^T).tr(\mathbf{Y}.\mathbf{Y}^T.\mathbf{Y}.\mathbf{Y}^T))}$$

→ RV number between 0 and 1

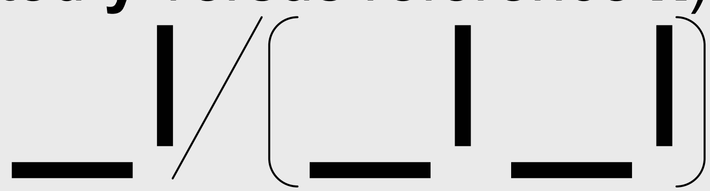
→ RV = 1 means c.**X**.**H** = **Y** or **X** can be rotated into **Y**



tr = trace
= sum of diagonal elements
= sum of eigenvalues for
(semi) positive matrix

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

$$r^2 = (\mathbf{x}^T.\mathbf{y} / \sqrt{(\mathbf{x}^T.\mathbf{x}.\mathbf{y}^T.\mathbf{y})})^2$$



STATIS

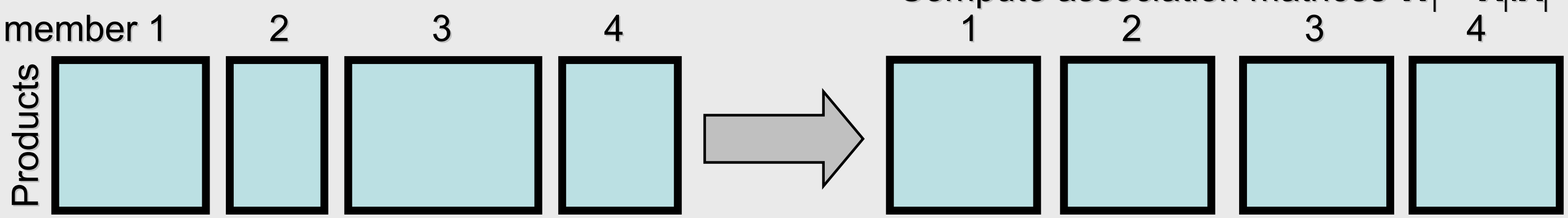
Sensory data

(2.16)

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 $\mathbf{W}_i = \mathbf{X}_i.\mathbf{X}_i^T$

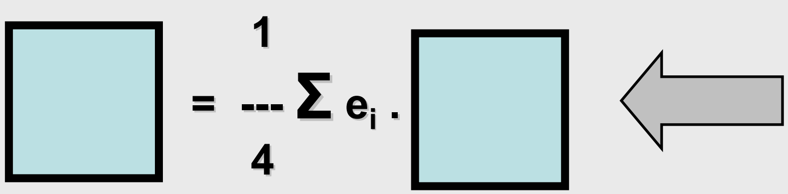


Attributes (e.g. free choice profiling)

Compute correlation matrix for all members and the first eigenvalue

Weights

$$\begin{array}{l} e_1 \leftarrow \\ e_2 \leftarrow \\ e_3 \leftarrow \\ e_4 \leftarrow \end{array} \left| \begin{pmatrix} R(1,1) & R(1,2) & R(1,3) & R(1,4) \\ R(2,1) & R(2,2) & R(2,3) & R(2,4) \\ R(3,1) & R(3,2) & R(3,3) & R(3,4) \\ R(4,1) & R(4,2) & R(4,3) & R(4,4) \end{pmatrix} \right.$$



Average (product) association matrix is weighted average over all panel members (so-called compromise matrix)

Eigenvector is normalized, $\Sigma\, e_i = 1$

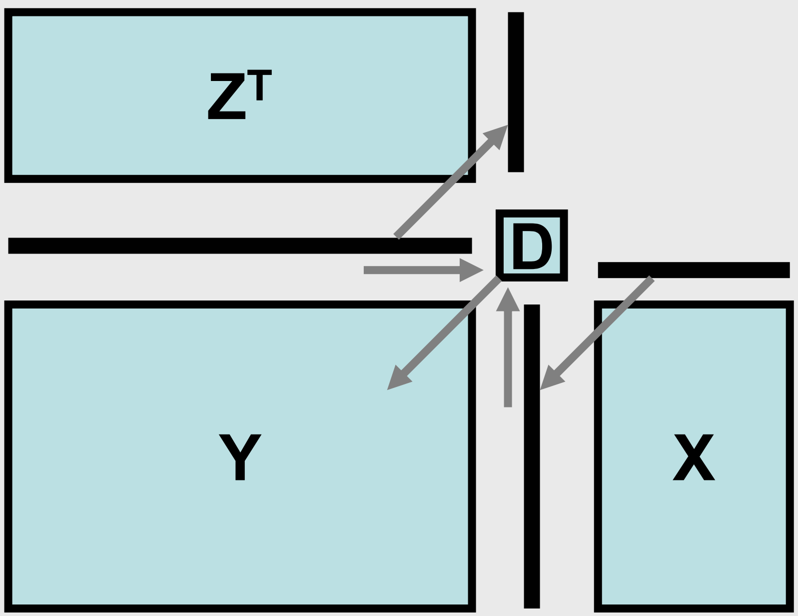
Members close to the average get high scores
Members far off (outliers) get low scores

Myrtille Vivien and Robert Sabatier *Generalized orthogonal multiple co-inertia analysis(-PLS): new multiblock component and regression methods* Journal of Chemometrics 17(2003)287-301

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*
Computational statistics & Data Analysis (2004) in press



Three sources of information:

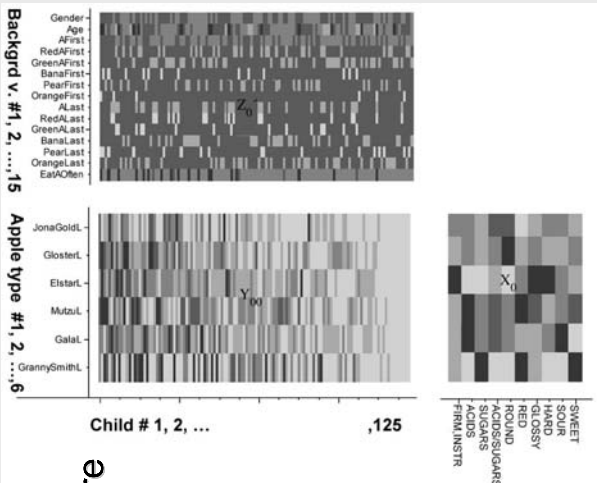
- Z and X both related to Y

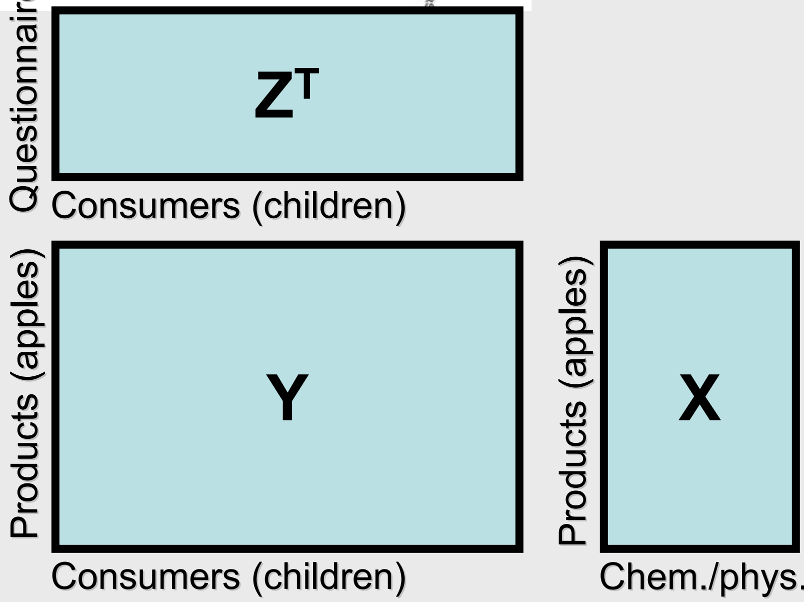
- Z and X no common dimension

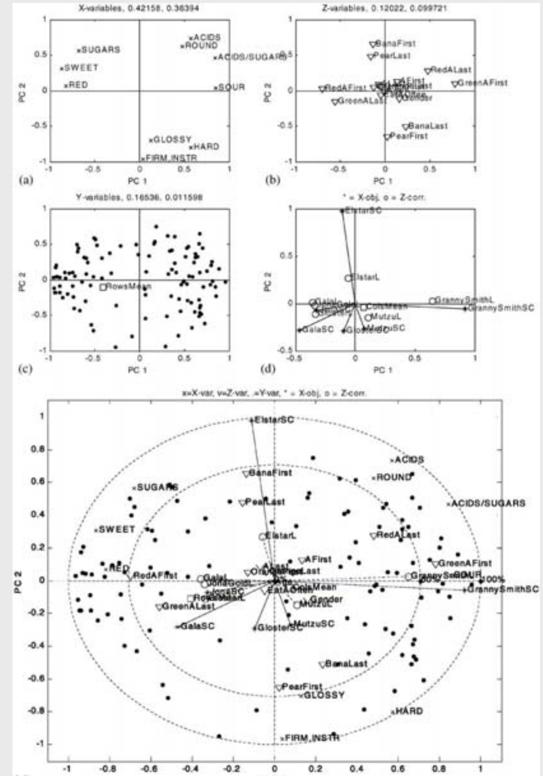
- Bi-linear models for all blocks, in accordance with PLS-concept, modeling of Y guided by X and Z

L-PLS

X, Y and Z data







Many relations (to many?)

Extensions possible (U-PLS, etc.)

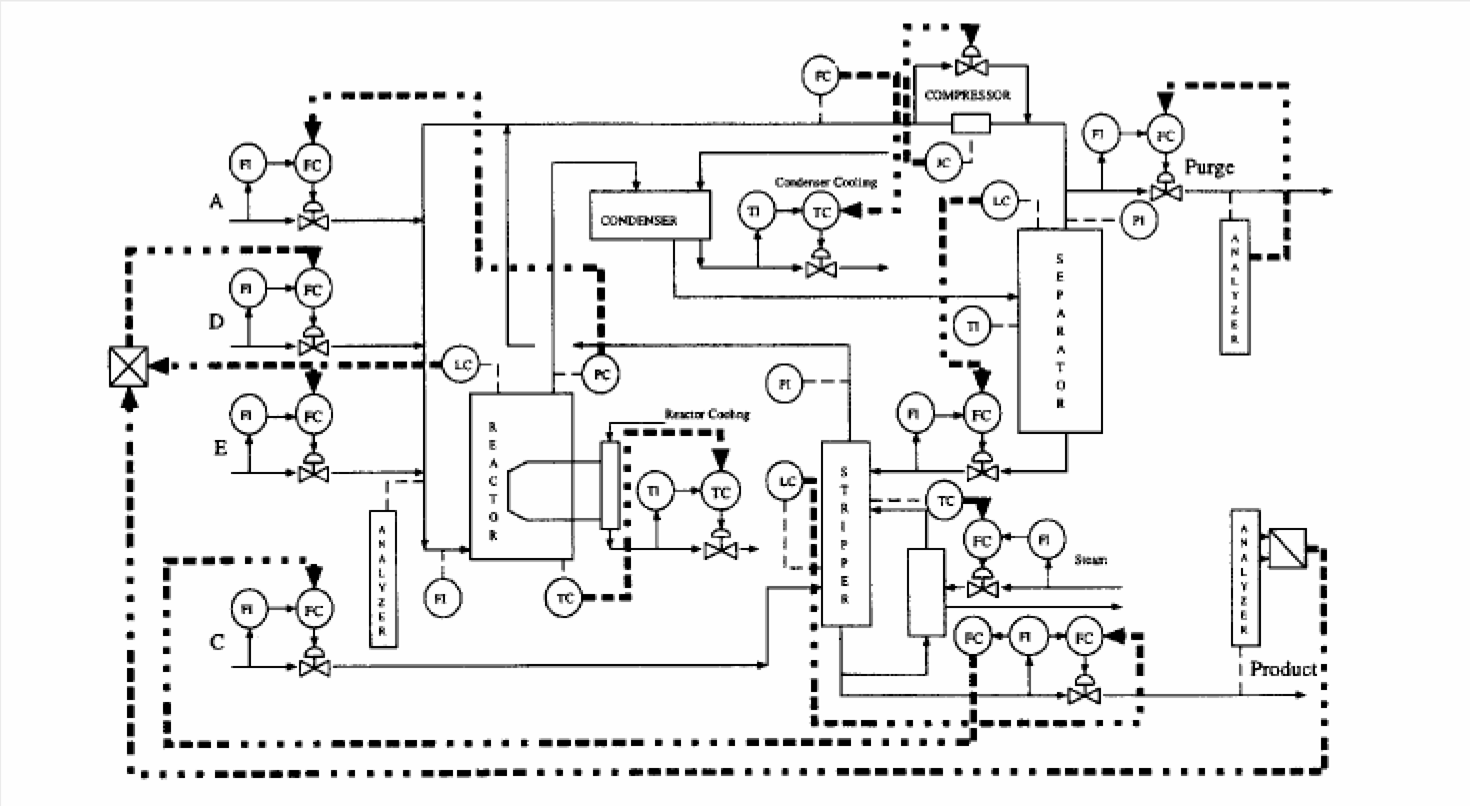
(2.19)

Multi-block PCA and PLS

Process Monitoring and Control

Gang Chen and Thomas MacAvoy *Predictive on-line monitoring of continuous processes* Journal of Process Control 8(1998)409-420

Tennessee Eastman benchmark (famous in control community)
as example in Statistical Process Monitoring by multi-block methods
41 measurements and 12 control variables



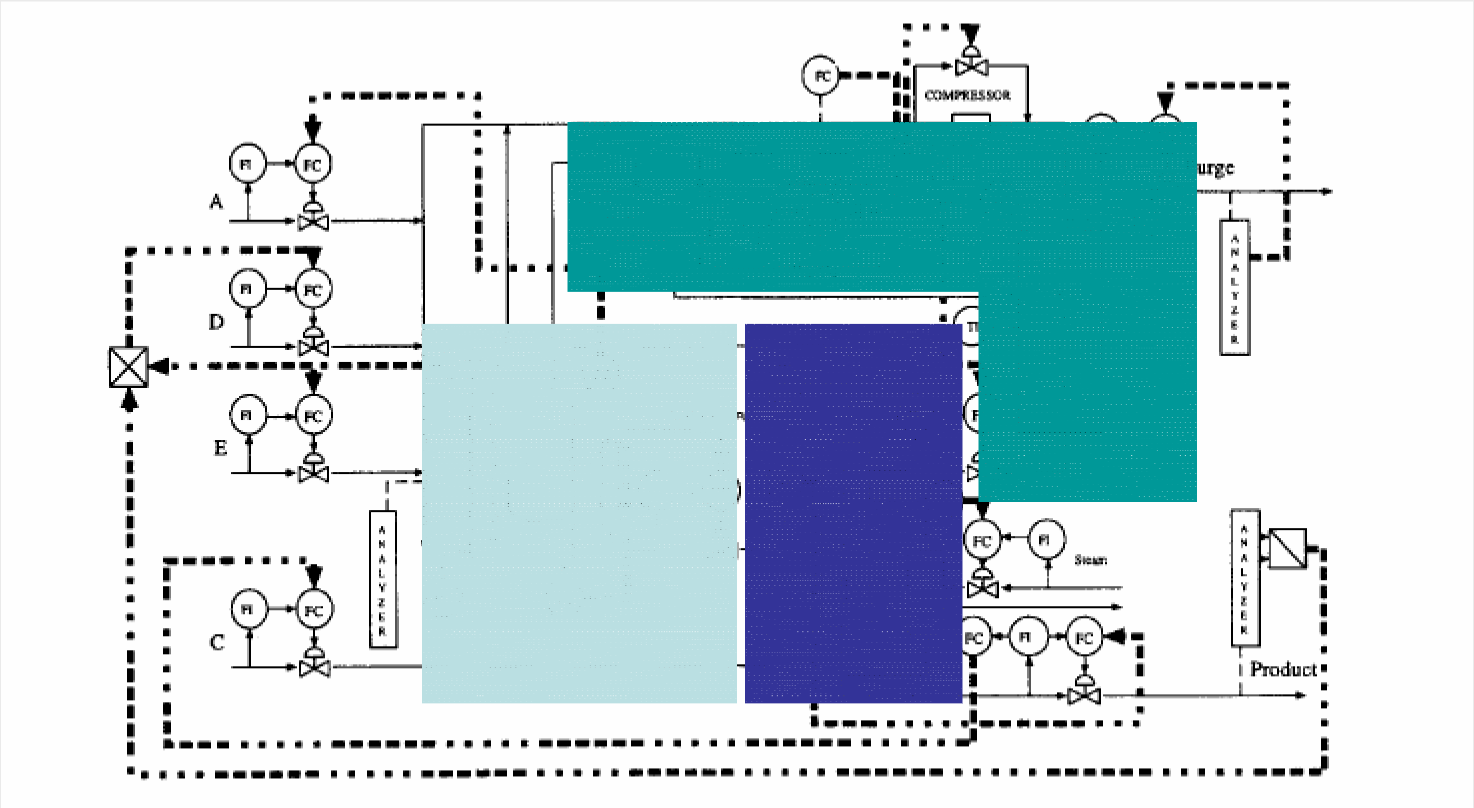
Multi-block PCA and PLS

Process Monitoring and Control

Gang Chen and Thomas MacAvoy *Predictive on-line monitoring of continuous processes* Journal of Process Control 8(1998)409-420

Blocks: conceptually meaningful (for an engineer!)

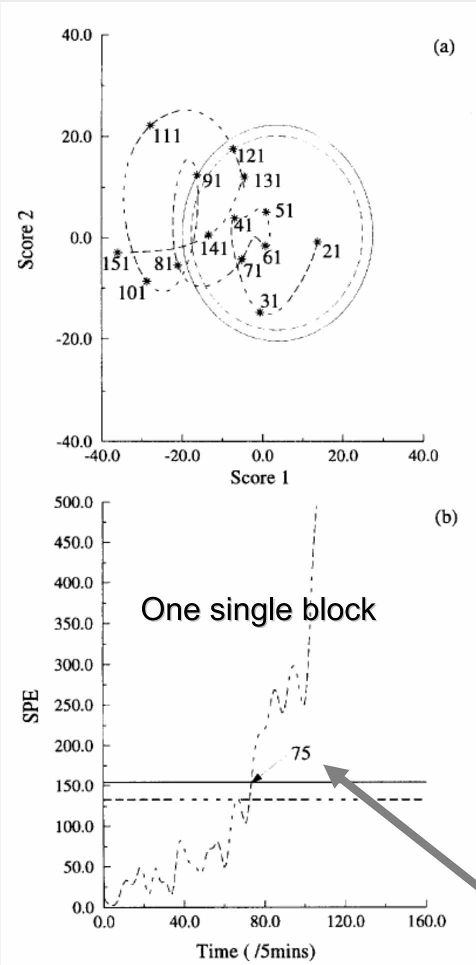
Reactor Stripper Separator



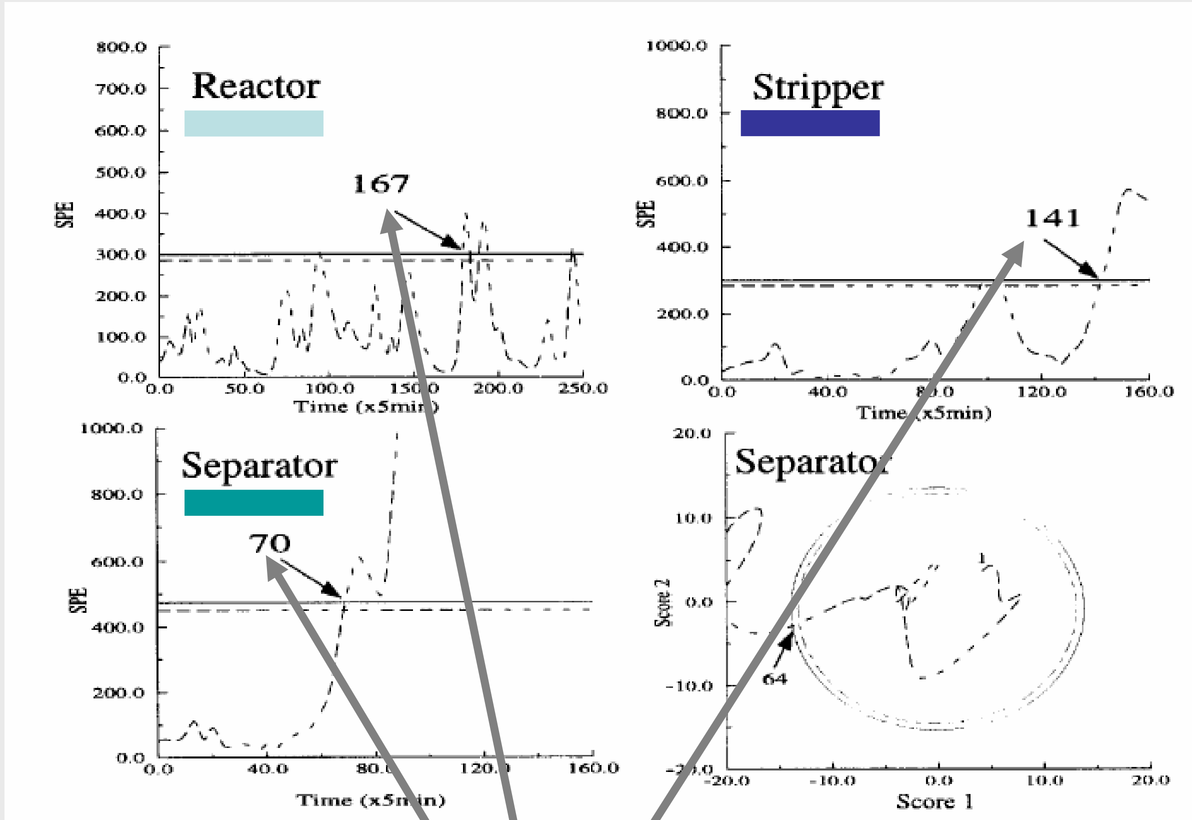
Multi-block PCA and PLS

Process Monitoring and Control

Gang Chen and Thomas MacAvoy *Predictive on-line monitoring of continuous processes* Journal of Process Control 8(1998)409-420



CASE 3 (B composition measurement drift+IDV(8))



Result for multi-block model approach for case 3

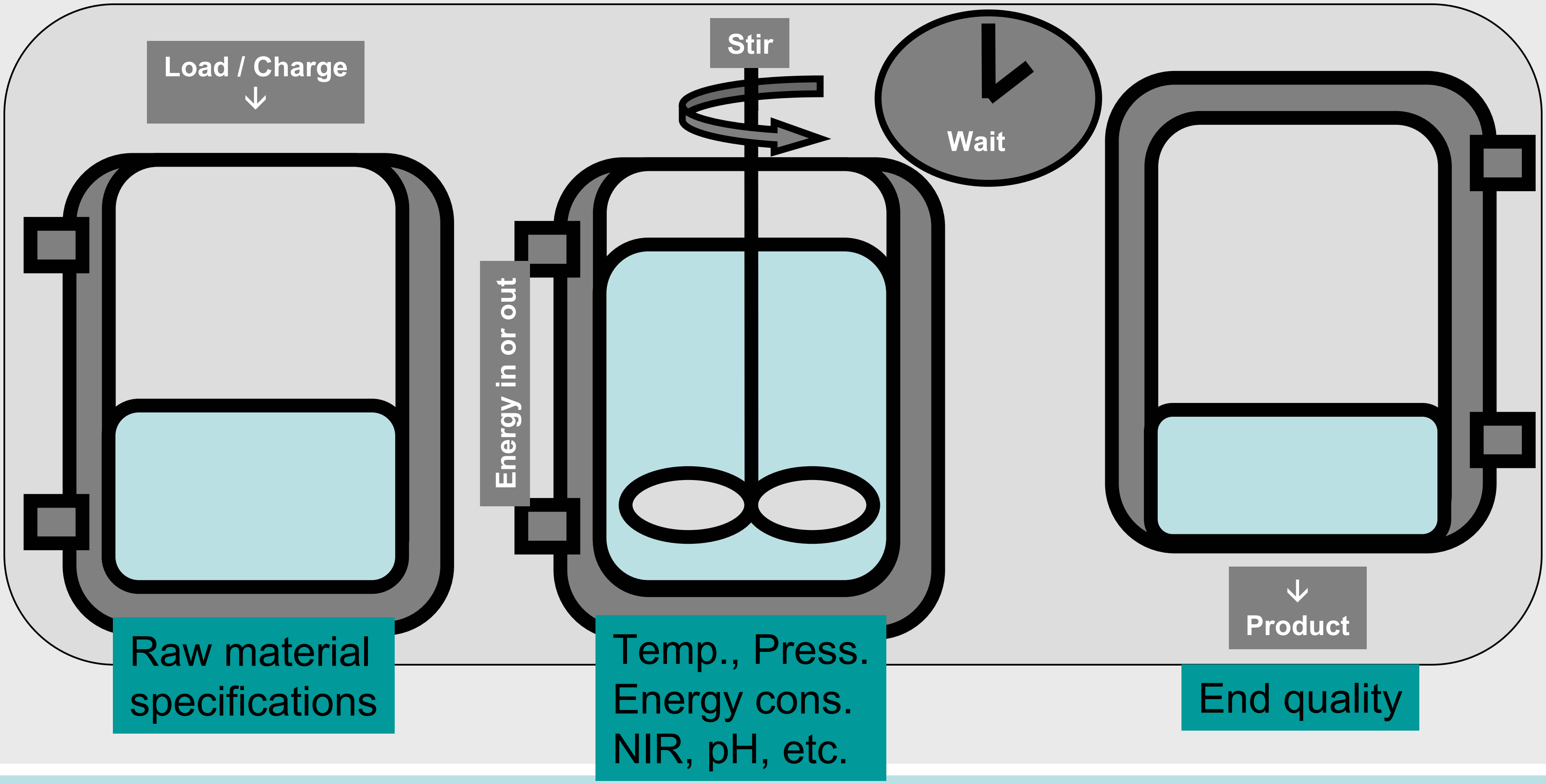
Fault detection:
Overall process / Process divided in blocks

(2.22)

Multi-block PCA and PLS

Process Monitoring and Control

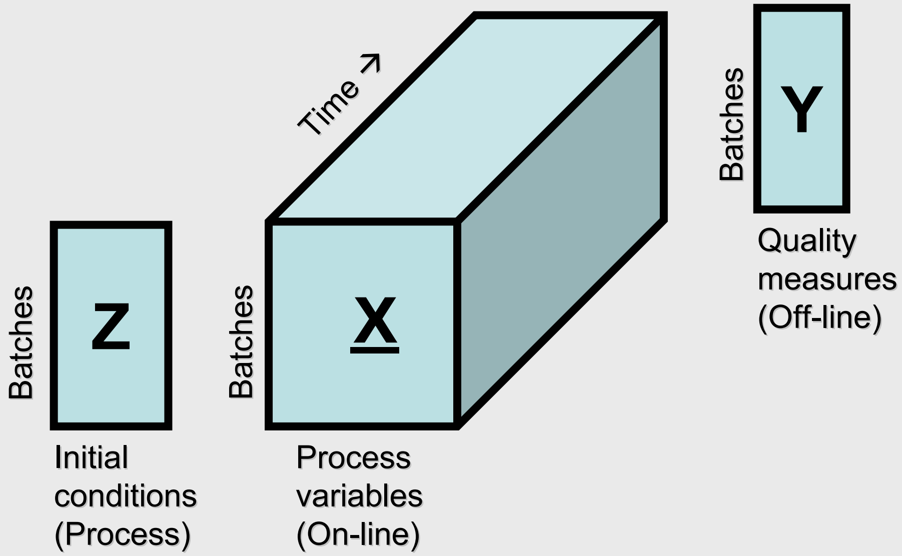
Theodora Kourti et.al. *Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS*
Journal of Process Control 4(1995)277-284
Theodora Kourti and John MacGregor *Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods*
Chemometrics 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



Multi-block PCA and PLS

Process Monitoring and Control

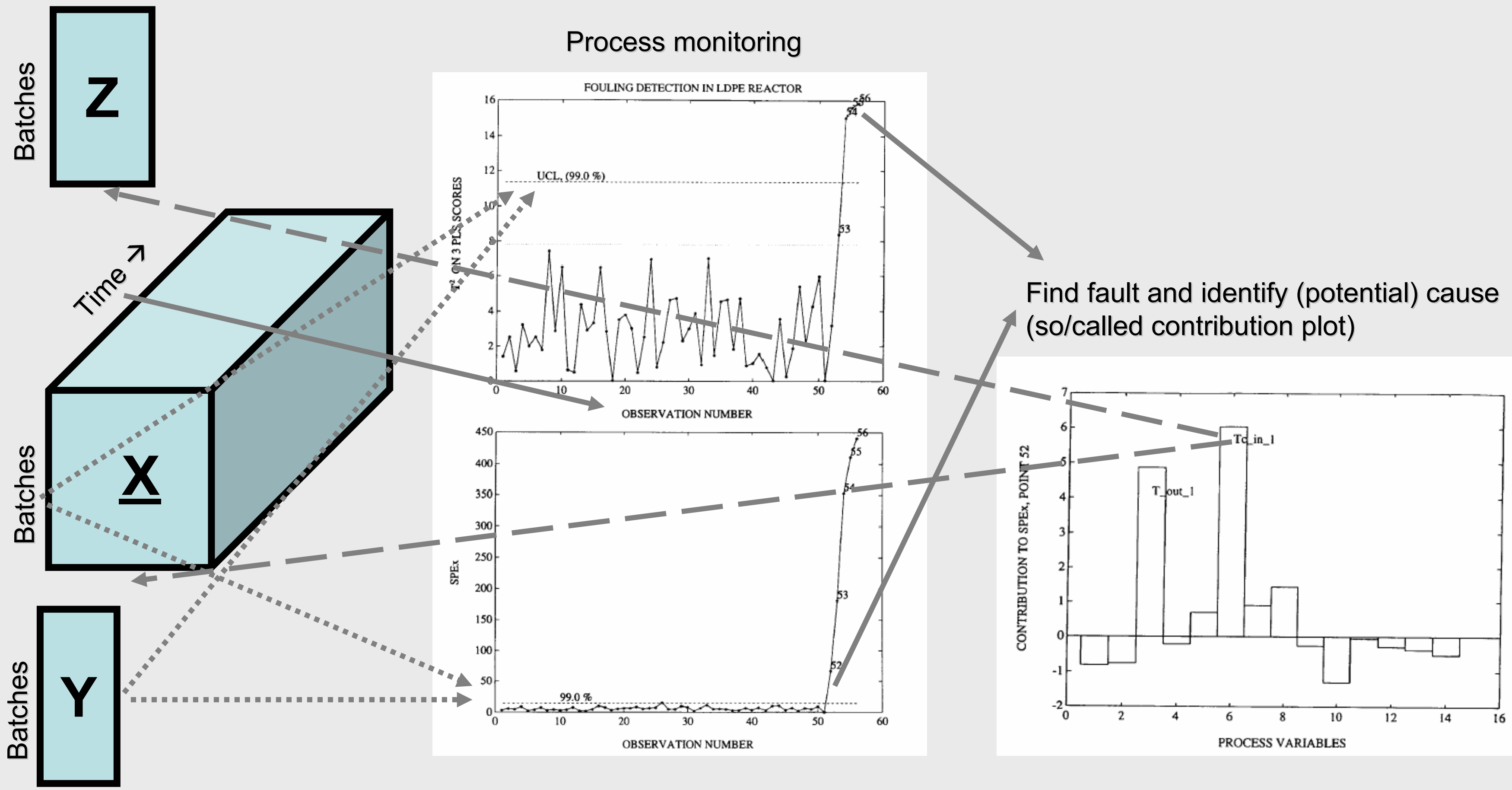
Theodora Kourti et.al. *Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS*
Journal of Process Control 4(1995)277-284
Theodora Kourti and John MacGregor *Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods*
Chemometrics 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*
Journal of Process Control 4(1995)277-284
Theodora Kourti and John MacGregor *Tutorial: Process Analysis, monitoring and diagnosis, using multivariate projection methods*
Chemometrics and Intelligent Laboratory Systems 28(1995)3-21

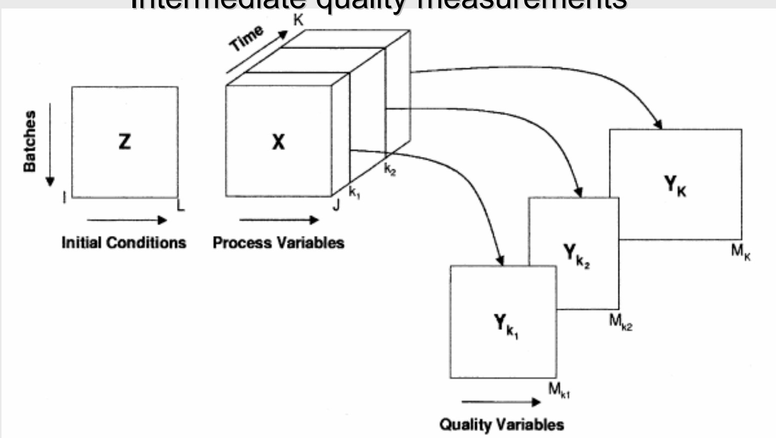


Multi-block PCA and PLS

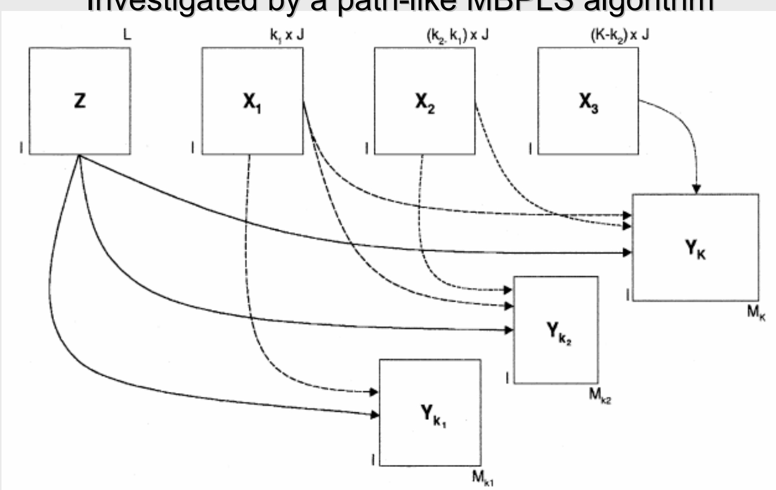
Process Monitoring and Control

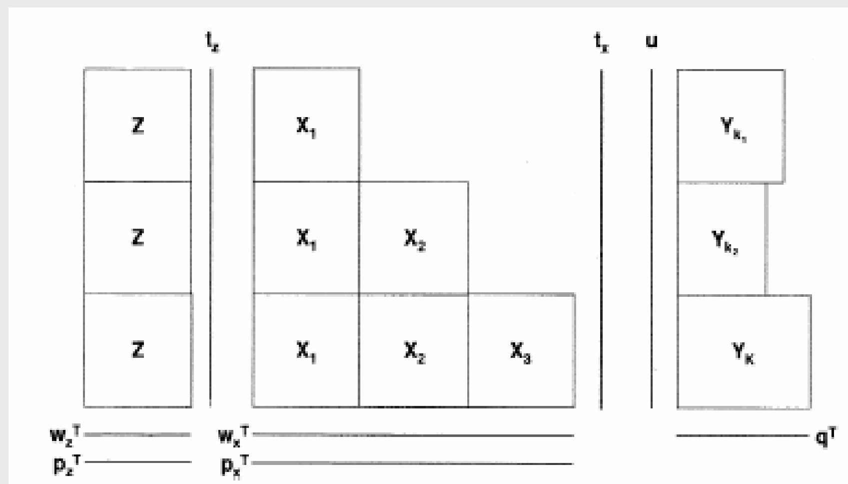
Carl Duchesne and John MacGregor *Multivariate analysis and optimization of process variable trajectories for batch processes*
Chemometrics and Intelligent Laboratory Systems 51(2000)125-137

Intermediate quality measurements



Investigated by a path-like MBPLS algorithm

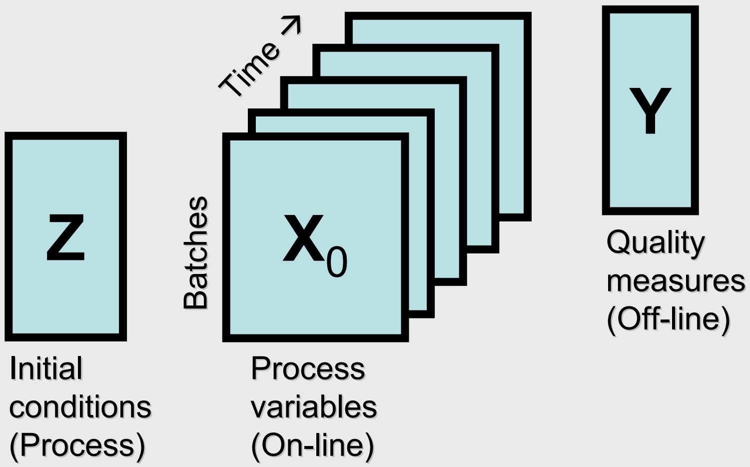


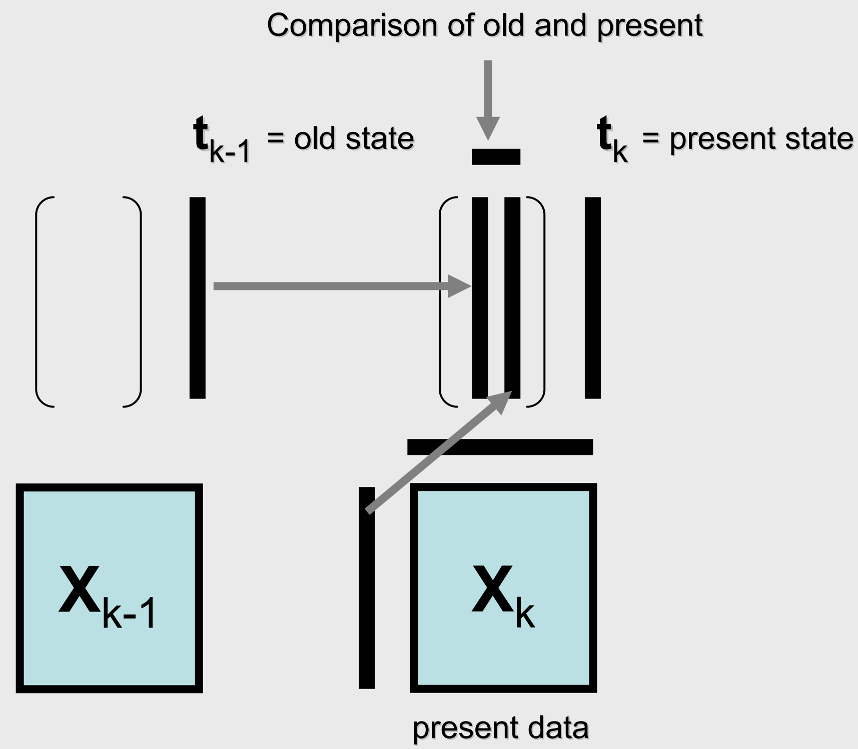


Multi-block PCA and PLS

Process Monitoring and Control

Theodora Kourti et.al. *Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS*
Journal of Process Control 4(1995)277-284
Theodora Kourti 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





EWMA-idea

Multi-block PCA and PLS

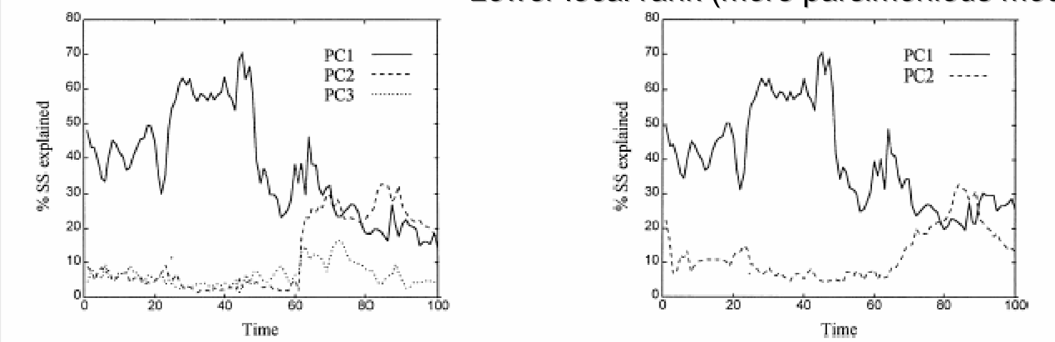
(2.27)

Process Monitoring and Control

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

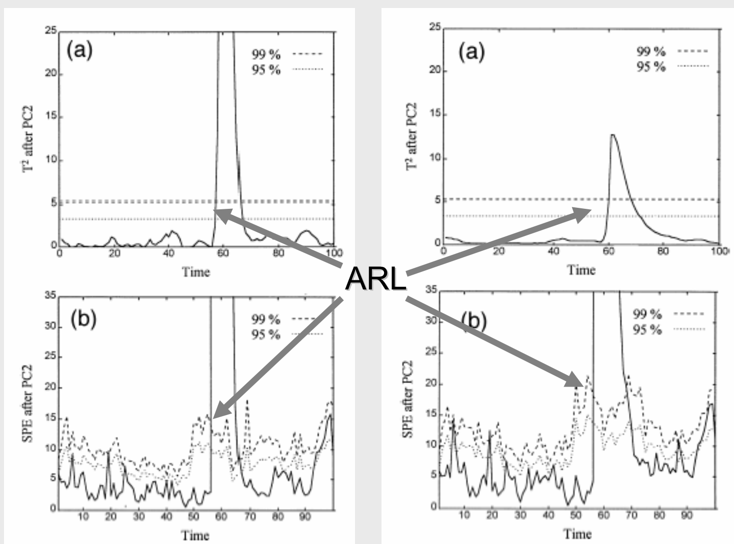
Full set

EWMA
Lower local rank (more parsimonious model)



Weight-ratio
old/new observations
Determines
Average Run Length

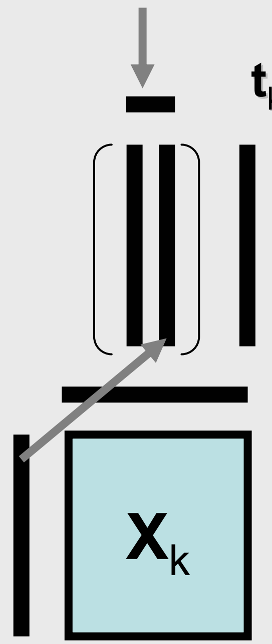
(d = 1.00 and d = 0.33)



d = block weights

Comparison of old and present

$\mathbf{t}_k$ = present state



present data

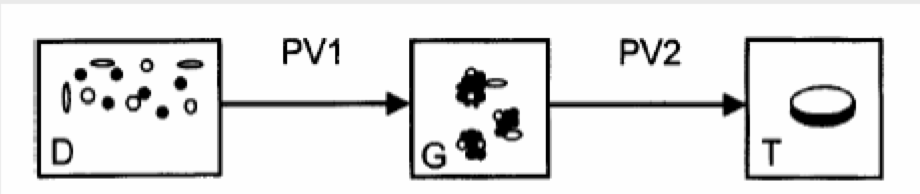
EWMA-idea

Multi-block PCA and PLS

Wet granulation and tablet pressing

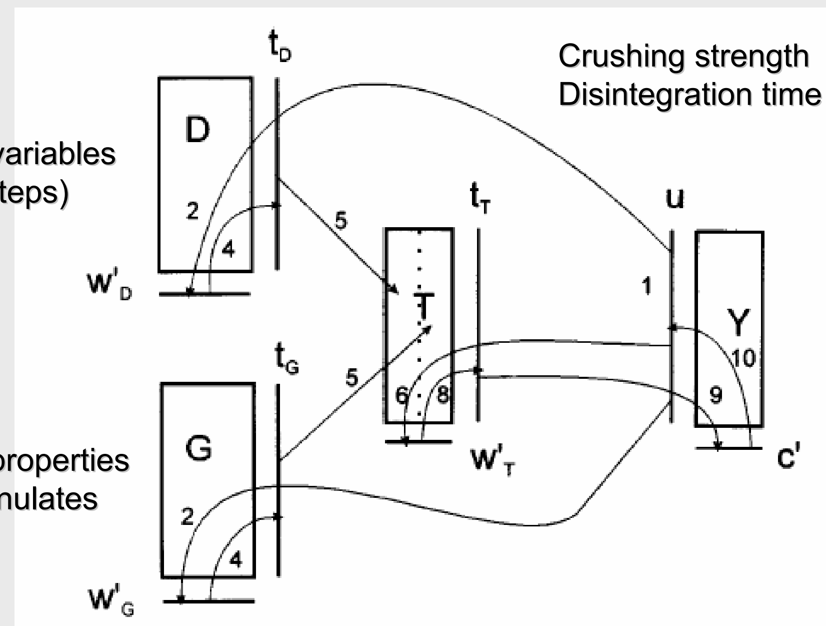
Johan Westerhuis and Pierre Coenegracht *Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with Multiblock partial least squares* Journal of Chemometrics 11(1997)379-392

Two-step process
Mixing tablet pressing tablet quality



Process variables
(of both steps)

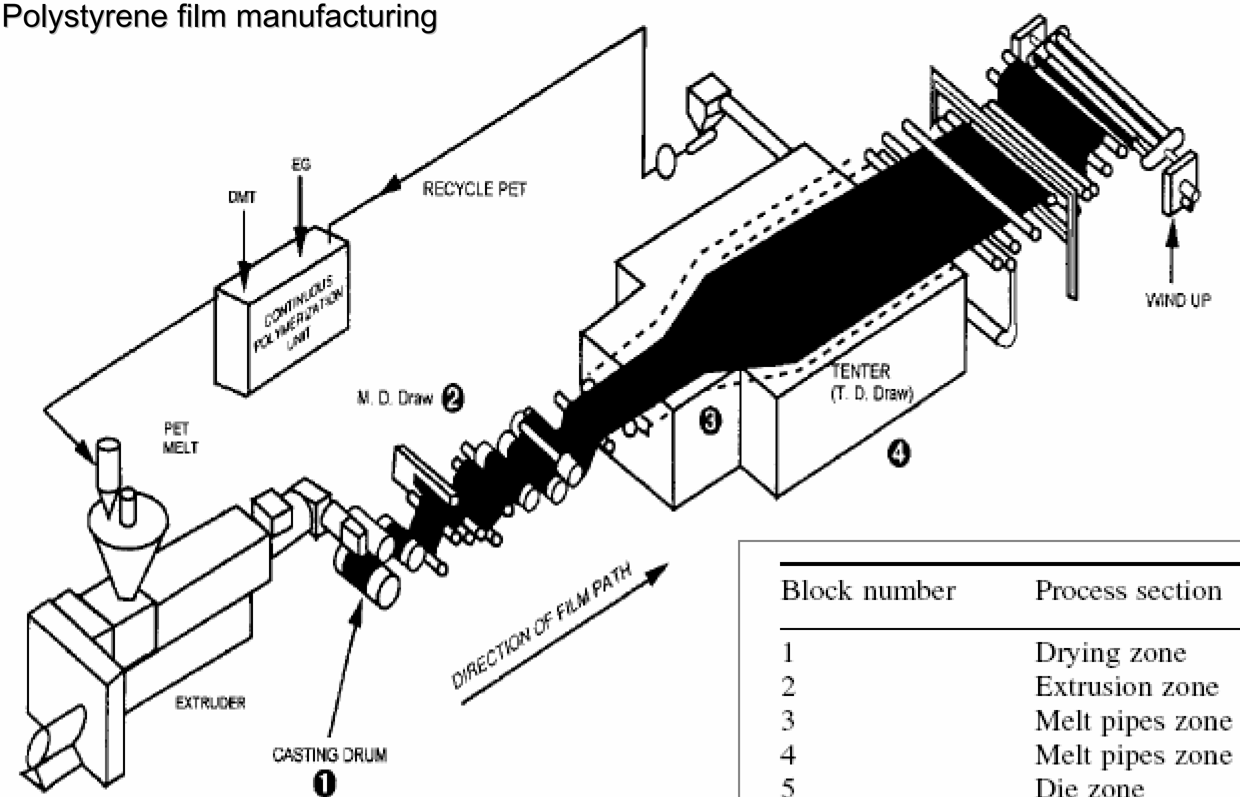
Physical properties
of the granulates



Multi-block PCA and PLS
Decentralized process monitoring

Joe Qin et.al. *On unifying multiblock analysis with application to decentralized process monitoring* Journal of Chemometrics 15(2001)715-742

Polystyrene film manufacturing



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

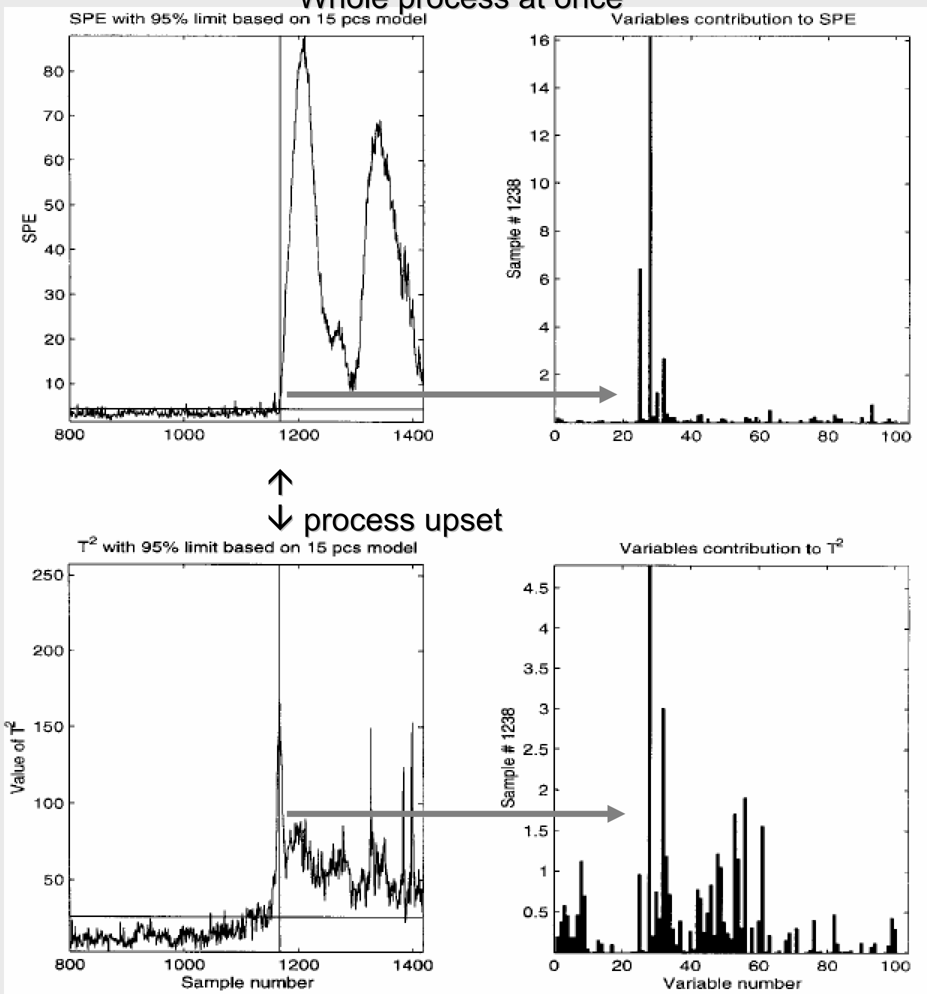
2879 time-frames 103 process variables → 7 blocks

Multi-block PCA and PLS

Decentralized process monitoring

Joe Qin et.al. *On unifying multiblock analysis with application to decentralized process monitoring* Journal of Chemometrics 15(2001)715-742

Whole process at once

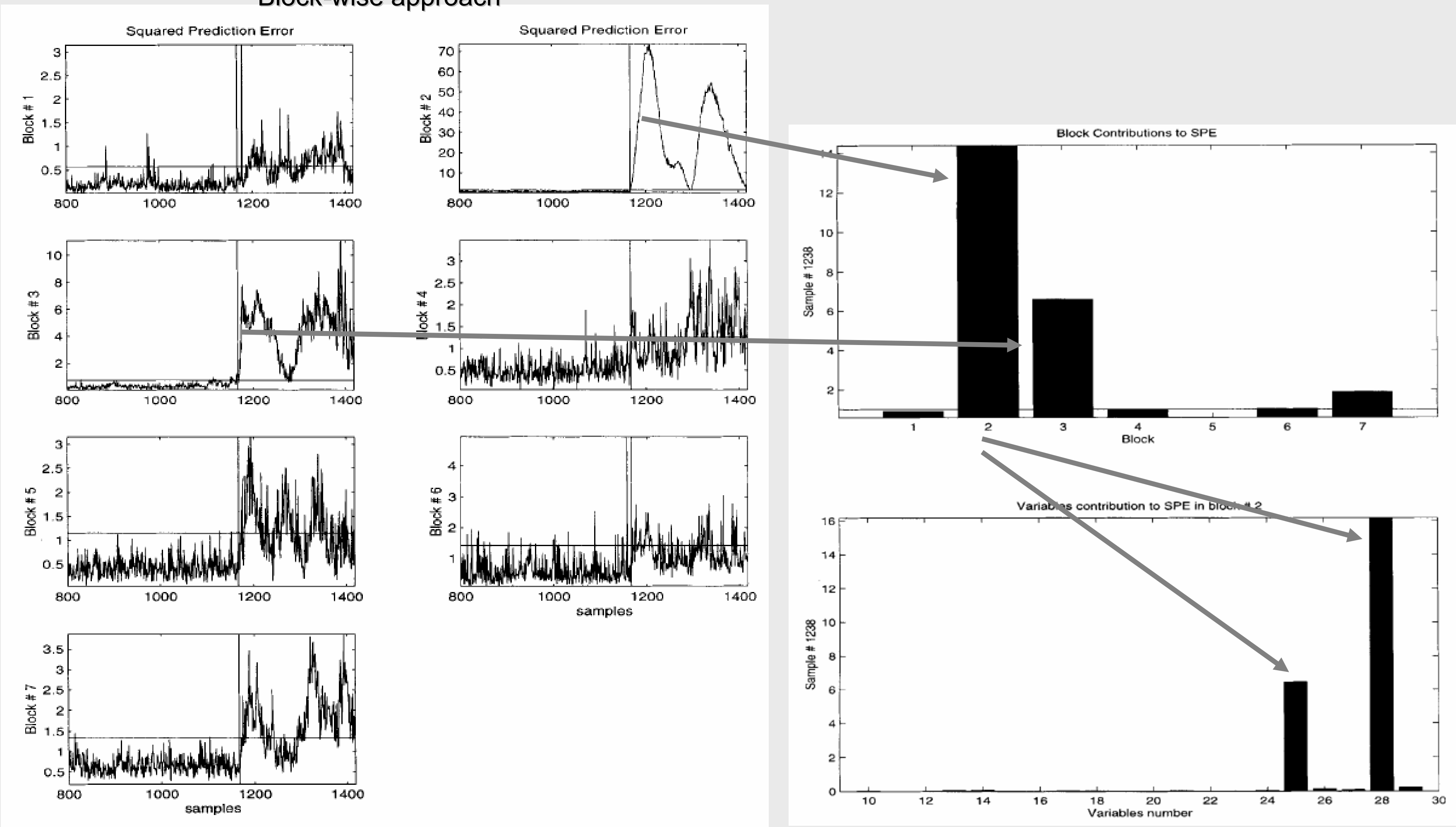


Multi-block PCA and PLS

Decentralized process monitoring

Joe Qin et.al. *On unifying multiblock analysis with application to decentralized process monitoring* Journal of Chemometrics 15(2001)715-742

Block-wise approach



(2.32)

Multi-block (consensus) PCA

The one we will use

Johan Westerhuis et.al. *Analysis of multiblock and Hierarchical PCA and PLS Models* Journal of Chemometrics 12(1998)301-321
Johan Westerhuis and Age Smilde *Short Communication Deflation in multiblock PLS* Journal of Chemometrics 15(2001)485-493
Joe Qin et.al. *On unifying multiblock analysis with application to decentralized process monitoring* Journal of Chemometrics 15(2001)715-742

PCA

Choose $\mathbf{t}$
Loop
- $\mathbf{p} = \mathbf{X}^T.\mathbf{t}/(\mathbf{t}^T.\mathbf{t})$
- $\mathbf{p} \rightarrow ||\mathbf{p}|| = 1$
- $\mathbf{t} = \mathbf{X}.\mathbf{p}$

End
$\mathbf{E} = \mathbf{X} - \mathbf{t}.\mathbf{p}^T$

CPCA

Choose $\mathbf{t}$
Loop
- $\mathbf{p}_b = \mathbf{X}_b{}^T.\mathbf{t}_t/(\mathbf{t}_t{}^T.\mathbf{t}_t)$
- $\mathbf{p}_b \rightarrow ||\mathbf{p}_b|| = 1$
- $\mathbf{t}_b = \mathbf{X}_b.\mathbf{p}_b$
- $\mathbf{T} = [\mathbf{t}_1\ \mathbf{t}_2\ \ldots\ \mathbf{t}_b]$
- $\mathbf{w}_t = \mathbf{T}^T.\mathbf{t}_t/(\mathbf{t}_t{}^T.\mathbf{t}_t)$
- $\mathbf{w}_t \rightarrow ||\mathbf{w}_t|| = 1$
- $\mathbf{t}_t = \mathbf{T}.\mathbf{w}_t$

End
$\mathbf{p}_b = \mathbf{X}_b{}^T.\mathbf{t}_t/(\mathbf{t}_t{}^T.\mathbf{t}_t)$
$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_t.\mathbf{p}_b{}^T$

Relation
(b = block, t = super-level)

$\mathbf{t}_t == \mathbf{t}$

$\mathbf{p}_b = \mathbf{X}_b{}^T.\mathbf{t}_t/(\mathbf{t}_t{}^T.\mathbf{t}_t)$
$\mathbf{p}_b \rightarrow ||\mathbf{p}_b|| = 1$

$\mathbf{t}_b = \mathbf{X}_b.\mathbf{p}_b$
$\mathbf{T} = [\mathbf{t}_1\ \mathbf{t}_2\ \ldots\ \mathbf{t}_b]$

$\mathbf{w}_t = \mathbf{T}^T.\mathbf{t}_t/(\mathbf{t}_t{}^T.\mathbf{t}_t)$

$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_t.\mathbf{p}_b{}^T$

Multi-block PLS

The one we will use

Johan Westerhuis et.al. *Analysis of multiblock and Hierarchical PCA and PLS Models* Journal of Chemometrics 12(1998)301-321
Johan Westerhuis and Age Smilde *Short Communication Deflation in multiblock PLS* Journal of Chemometrics 15(2001)485-493
Joe Qin et.al. *On unifying multiblock analysis with application to decentralized process monitoring* Journal of Chemometrics 15(2001)715-742

PLS

Choose $\mathbf{u}$
Loop
$\mathbf{w} = \mathbf{X}^T.\mathbf{u}/(\mathbf{u}^T.\mathbf{u})$
$\mathbf{w} \rightarrow ||\mathbf{w}|| = 1$
$\mathbf{t} = \mathbf{X}.\mathbf{w}$
$\mathbf{q} = \mathbf{Y}^T.\mathbf{t}/(\mathbf{t}^T.\mathbf{t})$
$\mathbf{u} = \mathbf{Y}.\mathbf{q}$
End
$\mathbf{p} = \mathbf{X}^T.\mathbf{t}/(\mathbf{t}^T.\mathbf{t})$
$\mathbf{E} = \mathbf{X} - \mathbf{t}.\mathbf{p}^T$
$\mathbf{F} = \mathbf{Y} - \mathbf{t}.\mathbf{q}^T$

MBPLS

Choose $\mathbf{u}$
Loop
$\mathbf{w}_b = \mathbf{X}_b^T.\mathbf{u}/(\mathbf{u}^T.\mathbf{u})$
$\mathbf{w}_b \rightarrow ||\mathbf{w}_b|| = 1$
$\mathbf{t}_b = \mathbf{X}_b.\mathbf{w}_b$
$\mathbf{T} = [\mathbf{t}_1\ \mathbf{t}_2\ \ldots\ \mathbf{t}_b]$
$\mathbf{w}_t = \mathbf{T}^T.\mathbf{u}/(\mathbf{u}^T.\mathbf{u})$
$\mathbf{w}_t \rightarrow ||\mathbf{w}_t|| = 1$
$\mathbf{t}_t = \mathbf{T}.\mathbf{w}_t$
$\mathbf{q} = \mathbf{Y}^T.\mathbf{t}_t/(\mathbf{t}_t^T.\mathbf{t}_t)$
$\mathbf{u} = \mathbf{Y}.\mathbf{q}$
End
$\mathbf{p}_b = \mathbf{X}_b^T.\mathbf{t}_t/(\mathbf{t}_t^T.\mathbf{t}_t)$
$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_t.\mathbf{p}_b^T$
$\mathbf{F} = \mathbf{Y} - \mathbf{t}_t.\mathbf{q}^T$

Relation
(b = blcok, t = super-level)

$\mathbf{t}_t == \mathbf{t}$
$\mathbf{u} == \mathbf{u}$

$\mathbf{w}_b = \mathbf{X}_b^T.\mathbf{u}/(\mathbf{u}^T.\mathbf{u})$

$\mathbf{t}_b = \mathbf{X}_b.\mathbf{w}_b$
$\mathbf{T} = [\mathbf{t}_1\ \mathbf{t}_2\ \ldots\ \mathbf{t}_b]$

$\mathbf{w}_t = \mathbf{T}^T.\mathbf{u}/(\mathbf{u}^T.\mathbf{u})$

$\mathbf{E}_b = \mathbf{X}_b - \mathbf{t}_t.\mathbf{p}_b^T$
$\mathbf{F} = \mathbf{Y} - \mathbf{t}_t.\mathbf{q}^T$

Multi-block PCA and PLS

The ones we will use

Johan Westerhuis et.al. *Analysis of multiblock and Hierarchical PCA and PLS Models* Journal of Chemometrics 12(1998)301-321
Johan Westerhuis and Age Smilde *Short Communication Deflation in multiblock PLS* Journal of Chemometrics 15(2001)485-493
Joe Qin et.al. *On unifying multiblock analysis with application to decentralized process monitoring* Journal of Chemometrics 15(2001)715-742

Conclusions:

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

- Use e.g. The Unscrambler or SIMCA output

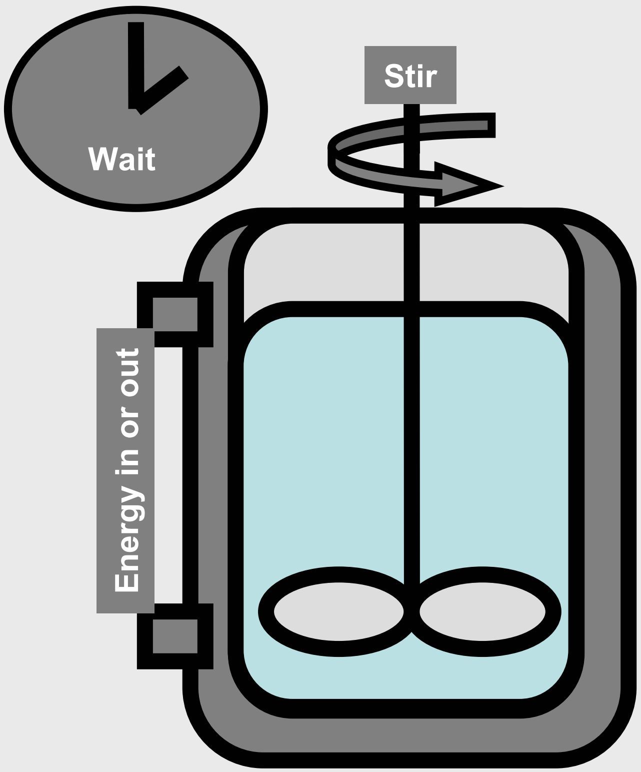
- In this view multi-block methods are not a new “modeling technique” (no better models), but rather a new / alternative way of looking at large data-sets with conceptually meaningful blocks
  New plotting diagnostics at the block level

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

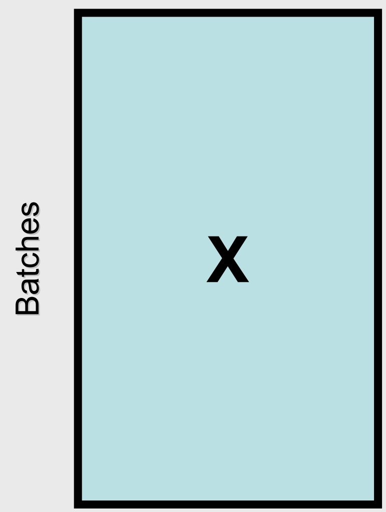
Many names, same result; e.g. Generalized PCA and SUM-PCA
Ph. Casin *A generalization of principal component analysis to K sets of variables* Computational Statistics & Data Analysis 35(2001)417-428
Eduard Derks et.al. *An introduction to Multi-blcok Component Analysis by means of a flavor language case study*
Food Quality and Preference 14(2003)497-506

Multi-block PCA and PLS

Block weighing



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

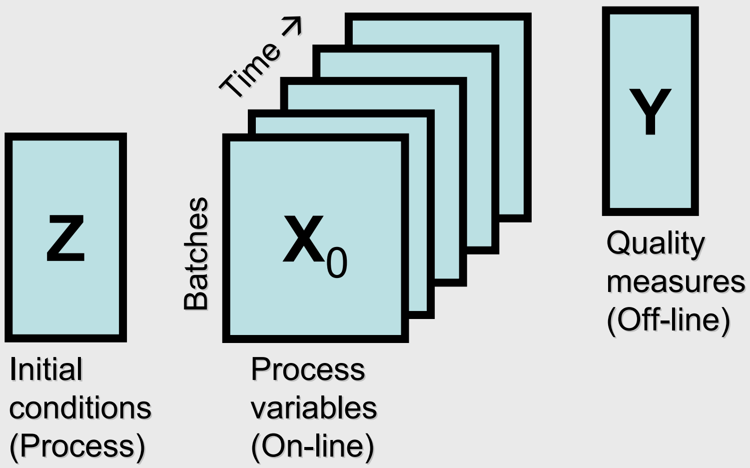
By auto-scaling we give them
equal importance:
center = 0
variance = 1 ($\Sigma x_i^2 / (N-1) = 1$)

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



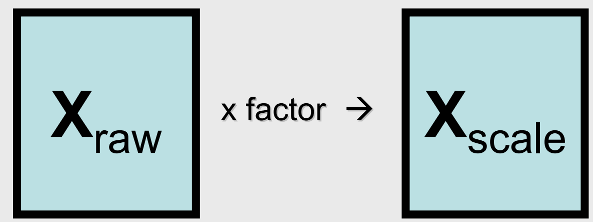
Same holds for multi-block methods:
e.g.
- 1 Input block
- 1 Output block
- 100 Time Frames on the process variables

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.

Block weighting turns out to be a difficult / important issue

We will use sum-of-squares to scale blocks



$\Sigma\ (x_{raw}(i,j)^2 \text{ x factor}) = \text{constant} = \Sigma\ x_{raw}(i,j)^2$

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

Break

(0.00)

NIR temperature effects

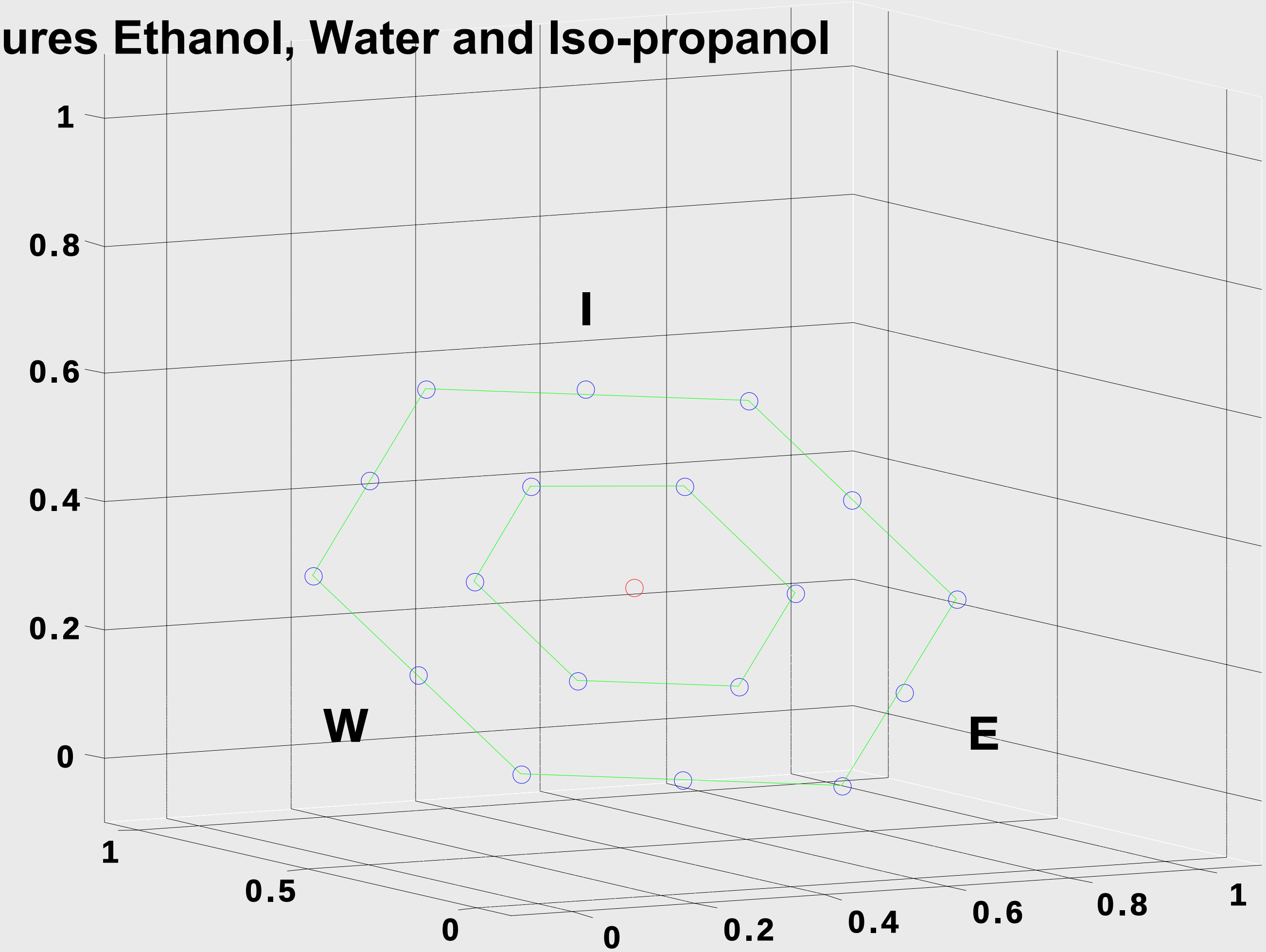
Mixture designs

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)

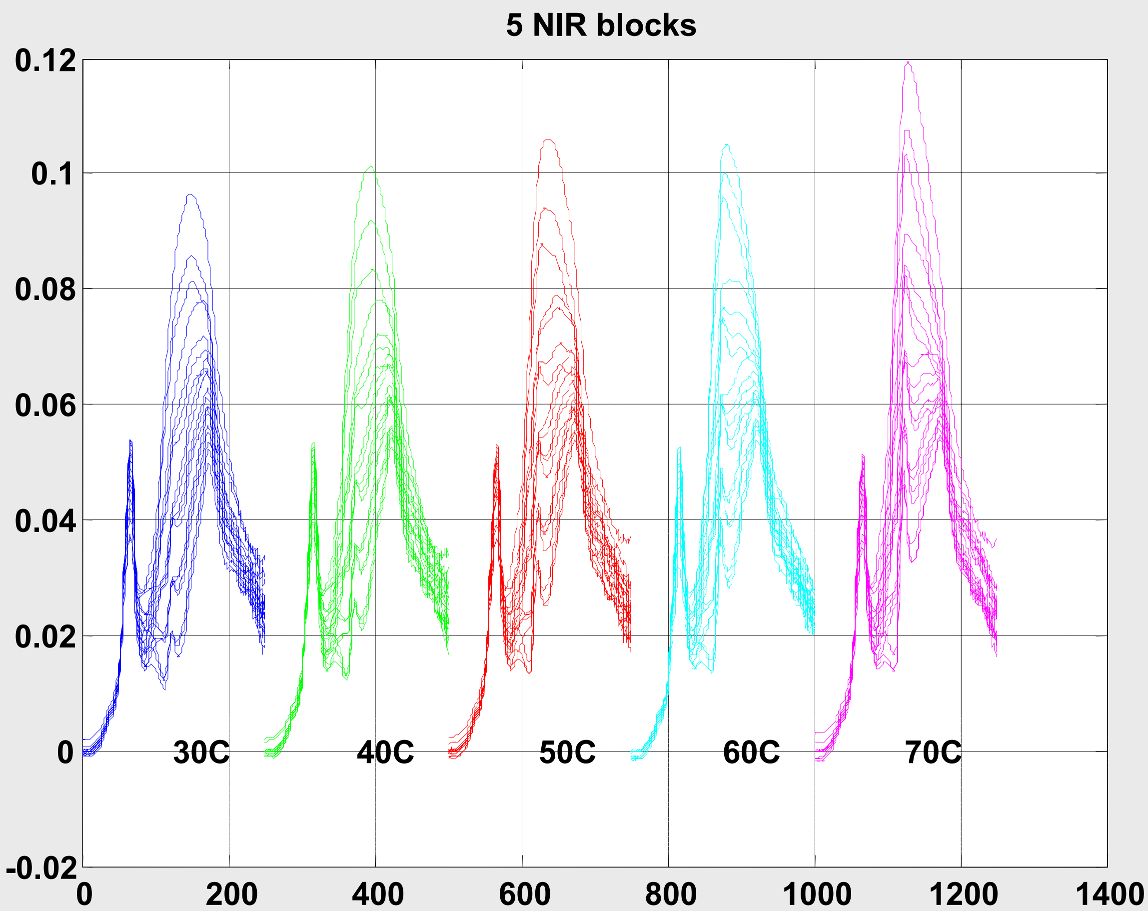
E	**W**	**I**
0.66	0.34	0
0.67	0.16	0.16
0.67	0	0.33
0.50	0.50	0
0.50	0.33	0.17
0.50	0.17	0.33
0.50	0	0.50
0.33	0.67	0
0.33	0.50	0.17
0.33	0.33	0.33
0.32	0.16	0.51
0.34	0	0.66
0.17	0.67	0.17
0.17	0.50	0.33
0.17	0.33	0.50
0.16	0.16	0.67
0	0.67	0.33
0	0.50	0.50
0	0.33	0.67



NIR temperature effects

Mixture designs

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

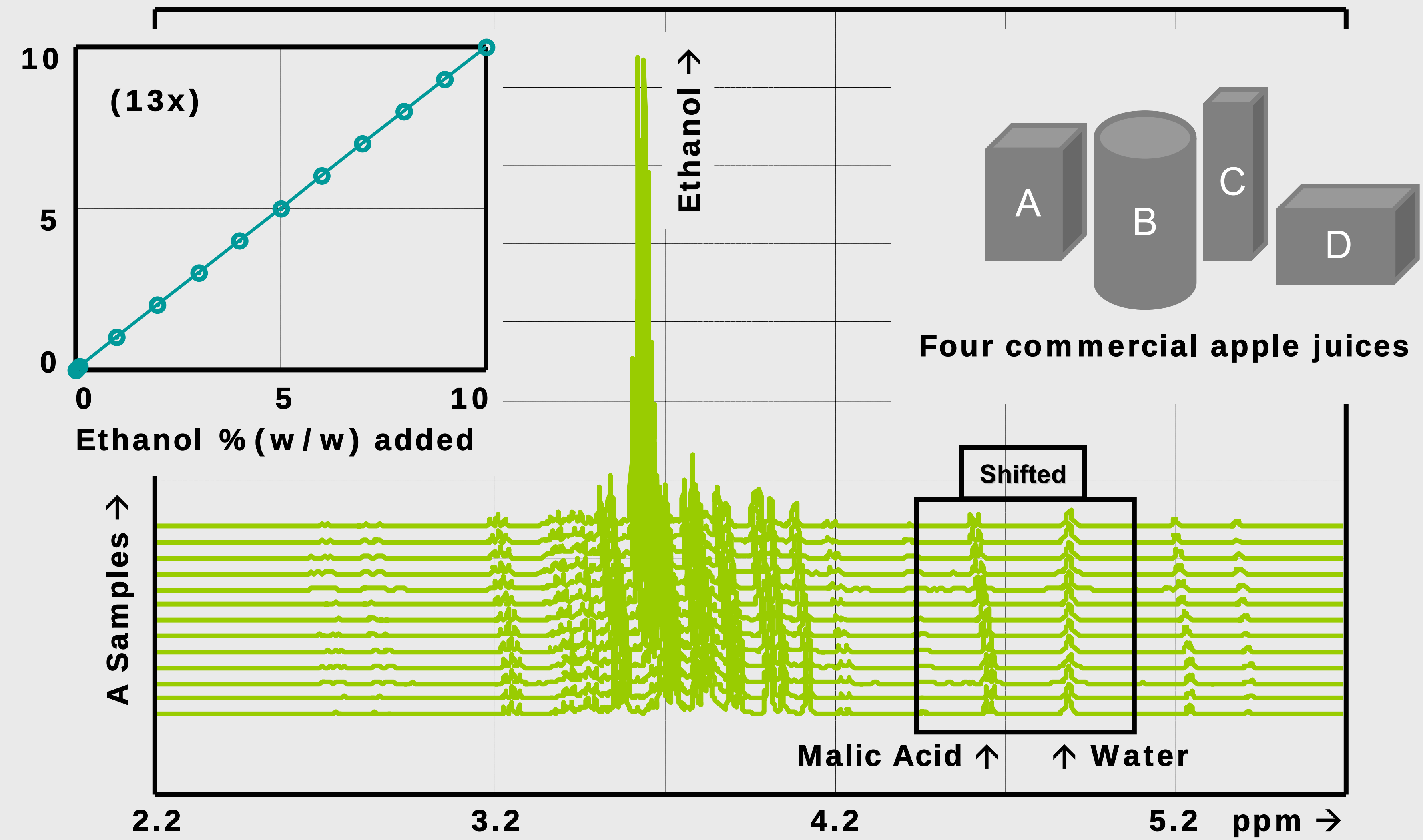
Matlab functions

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

(3.04)

Apple juice

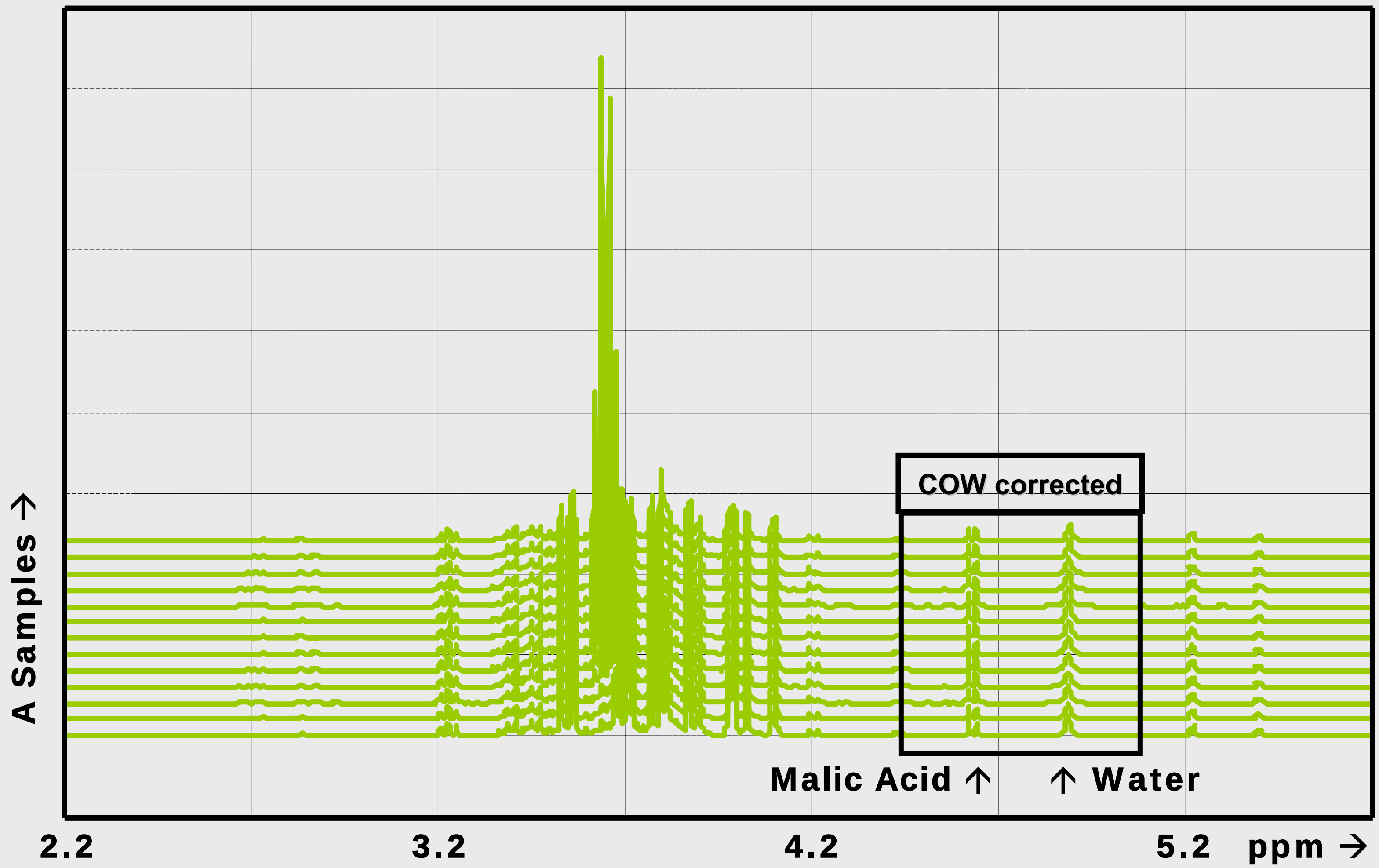


^{1}H-NMR with shifts

Apple juice

(3.05)

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

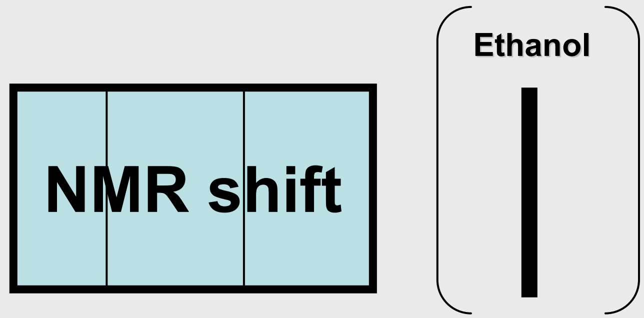


^{1}H-NMR with shifts

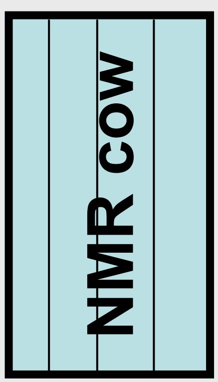
Apple juice

Block schemes

Different regions in spectrum



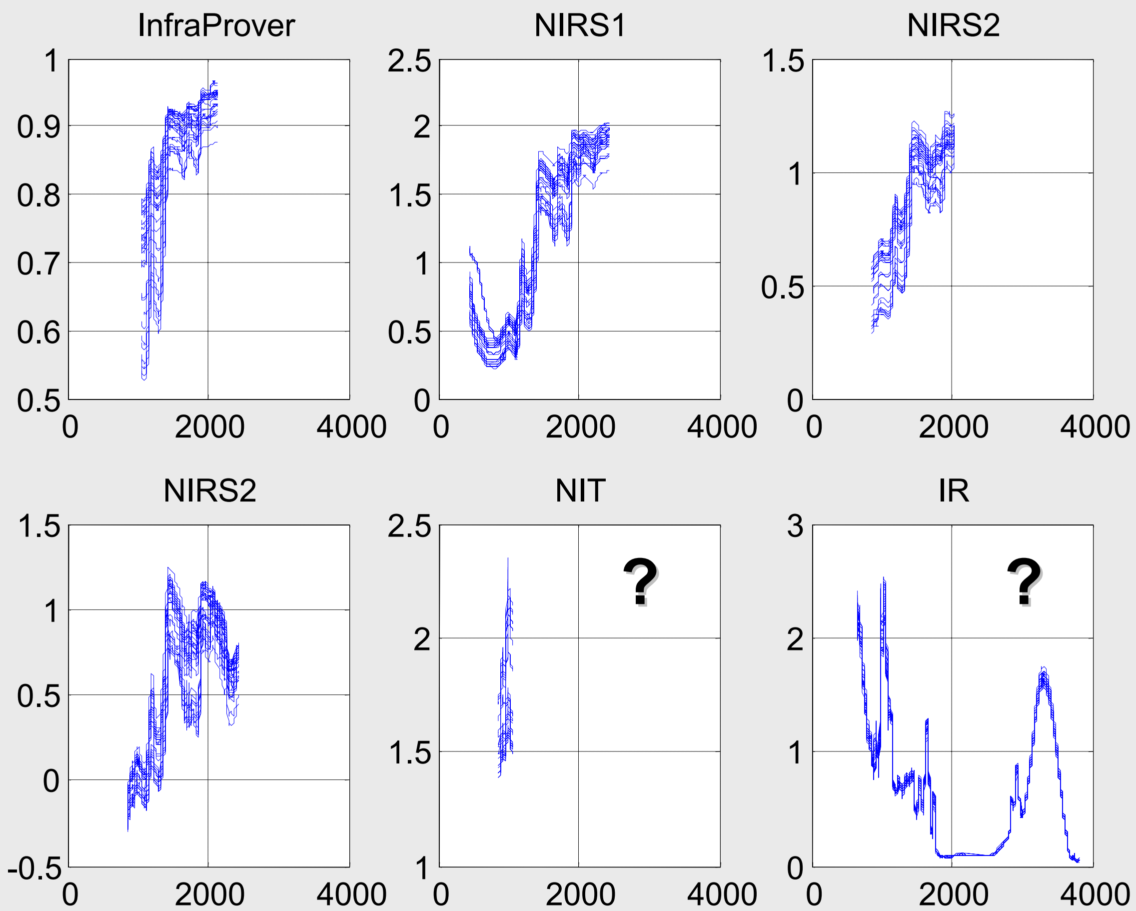
Shifted versus corrected NMR



Different brands
(two are from the manufacturer,
which ones?)

NIR Marzipan
Different instruments

Jakob Christensen, Lars Nørgaard, Hanne Heimdal, Joan Grønkjær Pedersen and Søren Balling Engelsen *Rapid Spectroscopic Analysis of Marzipan - Comparative Instrumentation* Journal of Near Infrared Spectroscopy 12(2004)xxx-xxx



32 marzipan samples
+
Moisture contents
Sugar contents

Measured on 6 instruments

Compare instruments
Block selection