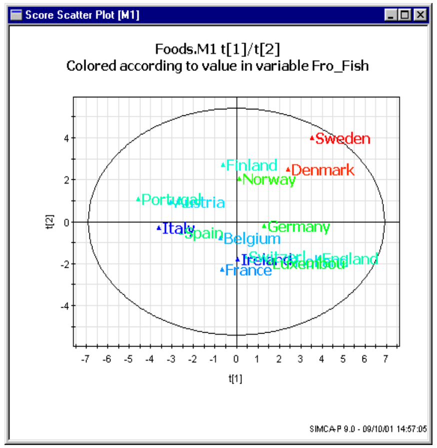
The OPLS method and recent extensions:
-unlocking interpretation in two block modeling and multivariate calibration

Johan Trygg

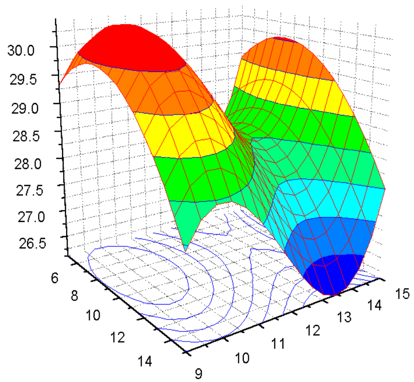
Computational Life Science cluster (CLiC), Umeå University, Sweden
Group leader, Chemometrics & Bioinformatics Umeå Plant Science Centre
Co-director WCN Metabolomics platform

Research group for Chemometrics
Department of Chemistry, Umeå University, Sweden

MVA and DoE



- **Multivariate Analysis**
  – Extract information from multivariate data sets
  – Graphical overview & quantitative modelling
  – Large datasets and databases
  – Reduces complexity



- **Design of Experiments (DoE)**
  – For new experiments and already collected data
  – Maximise the information, minimize experiments
  – Get causality no correlation
    - What are the important factors?
    - What to Optimise
    - Improve Robustness and Quality

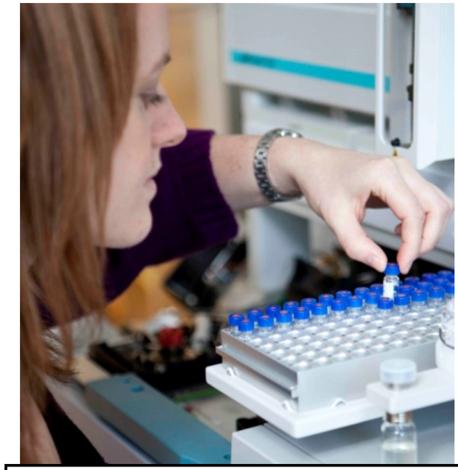
Umeå University/Chemometrics
- Bio-applications

Disease diagnosis, theranostics
- Autoimmune diseases
- Neurodegenerative diseases.
- Cancer

Post surgery:
- Kidney transplant

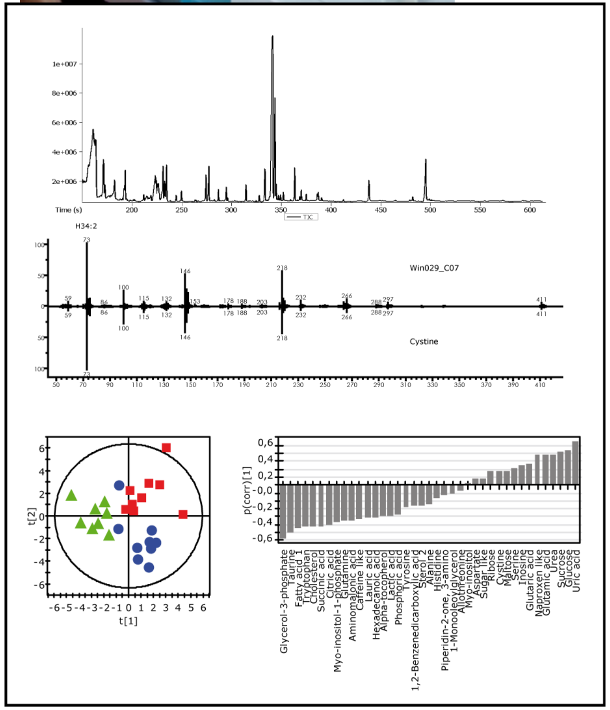
Nutrition:
- Functional foods

Rehabilitation
- Medical imaging of muscle tissue function



Urine test to monitor kidney-transplant rejection

A urine test that diagnoses acute rejection without the need for an invasive biopsy



Multi-disciplinary collaborations
Biotech companies, patient organizations, leading research groups, clinicians, and diagnostics companies

Umeå University/Chemometrics
- PAT/QbD initiative for process understanding

Food and Drug Administration (FDA)

Instead go from product testing to *quality by design*!

Risk minimisation – process understanding

The FDA's initiative is built on the premise that
if manufacturers can demonstrate understanding of their processes, **less risk** of making bad product.

To control the manufacturing process

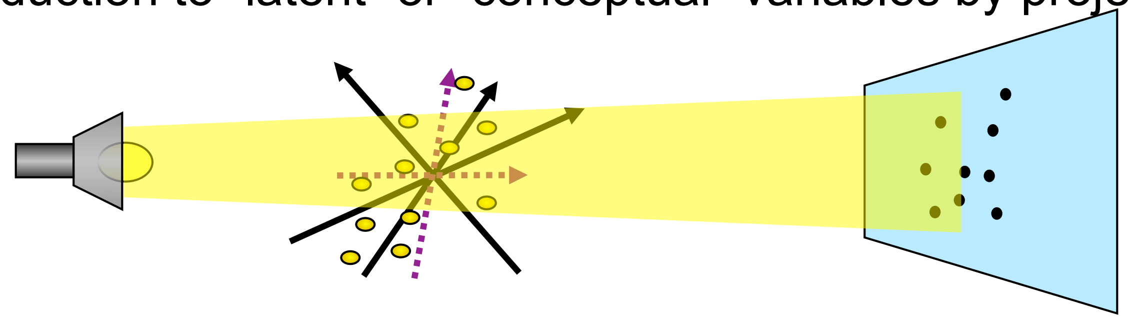
Control = to ensure

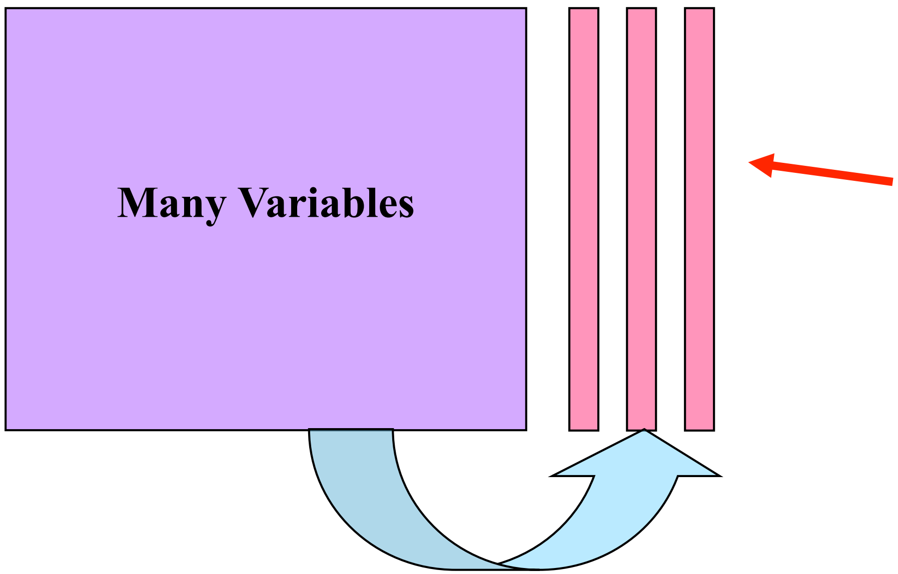
Control ≠ to check

Multivariate data analysis by means of projection methods

PRINCIPAL COMPONENT ANALYSIS

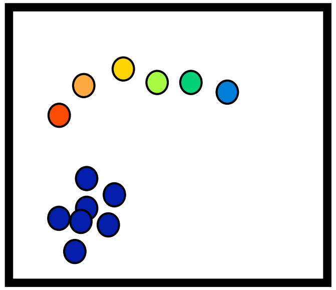
- Data reduction to “latent” or “conceptual” variables by projection

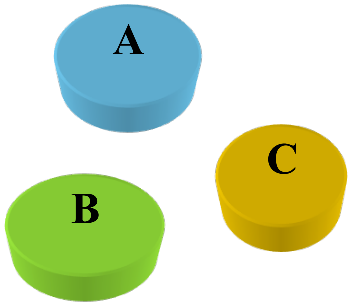


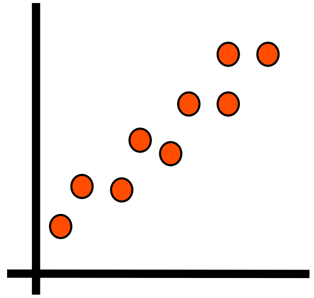


- **Fewer ‘Latent’ Variables**
- **Concise summary of the old**
- **Finds correlations (=systematic variation)**
- **Leaves noise behind (=unsystematic variation)**

3 Basic Data Analysis Questions





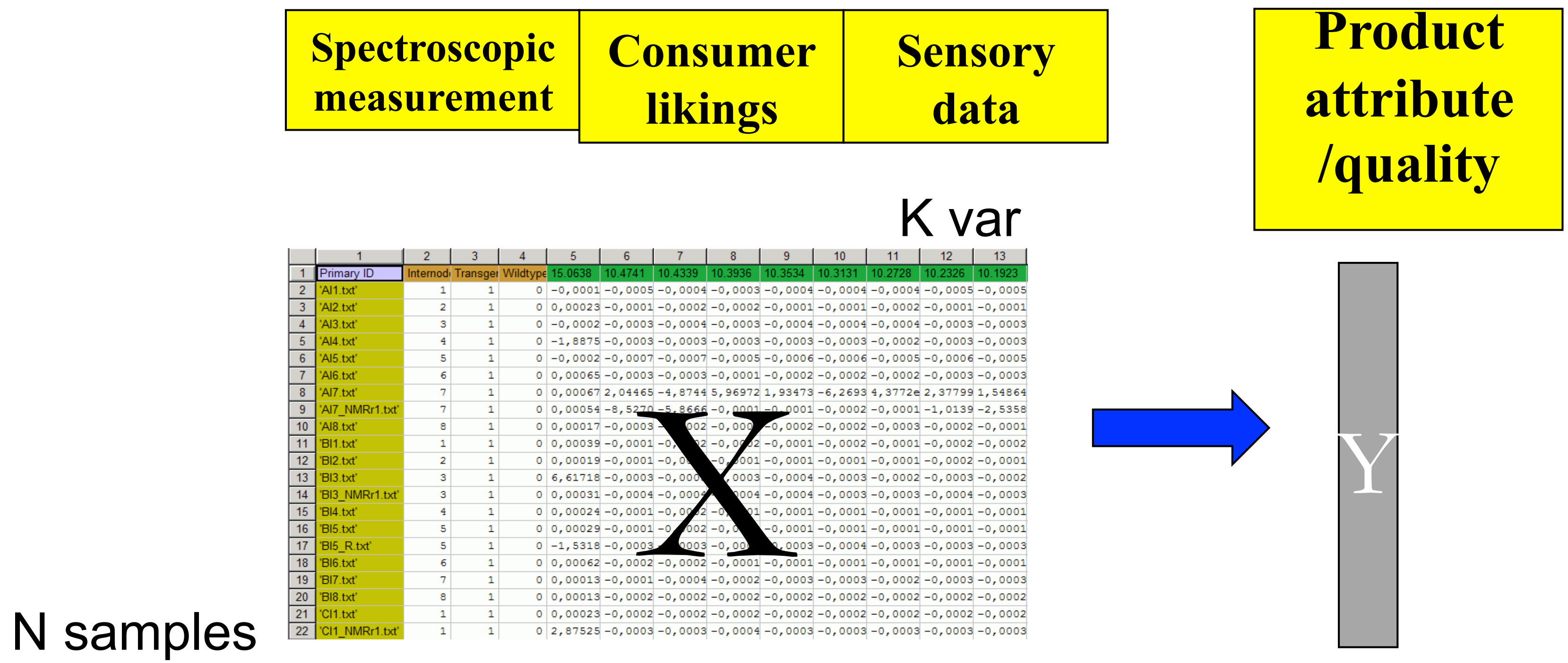


Overview	**Classification**	**Relationship**
Process monitoring Chemical Property Maps Selection of drug candidates Encoding proteins and DNA sequences Assessing biological variation Trends in quality Sensory profiling Competitor Analysis Silicon chip manufacture	Classification of raw materials / foodstuffs Wine authenticity Drug Transport & Toxicity Mechanisms Genomics / Proteomics / Metabonomics i.e. Control / Treated Genetic Modification detection	Drug Activity (QSAR) Detergent design New material properties Calibration models Online NIR – moisture/ particle size / actives Sensory information Wheat Quality Process Quality prediction Batch Modelling
PCA	**SIMCA / PLS-DA / OPLS-DA**	**PLS / OPLS / O2PLS**

Model the relation between two blocks of data

Chemometrics: Multivariate calibration, MC

Samples	Powders, molecules, industrial process samples, plasma, tissue (leaf), …
and sample characterisation	Spectrometers (NIR,UV, IR, NMR, MS), chromatography, chemical descriptors, gene-arrays, metabolites
provide LOTS of data	highly multidimensional (1000’s of variables)

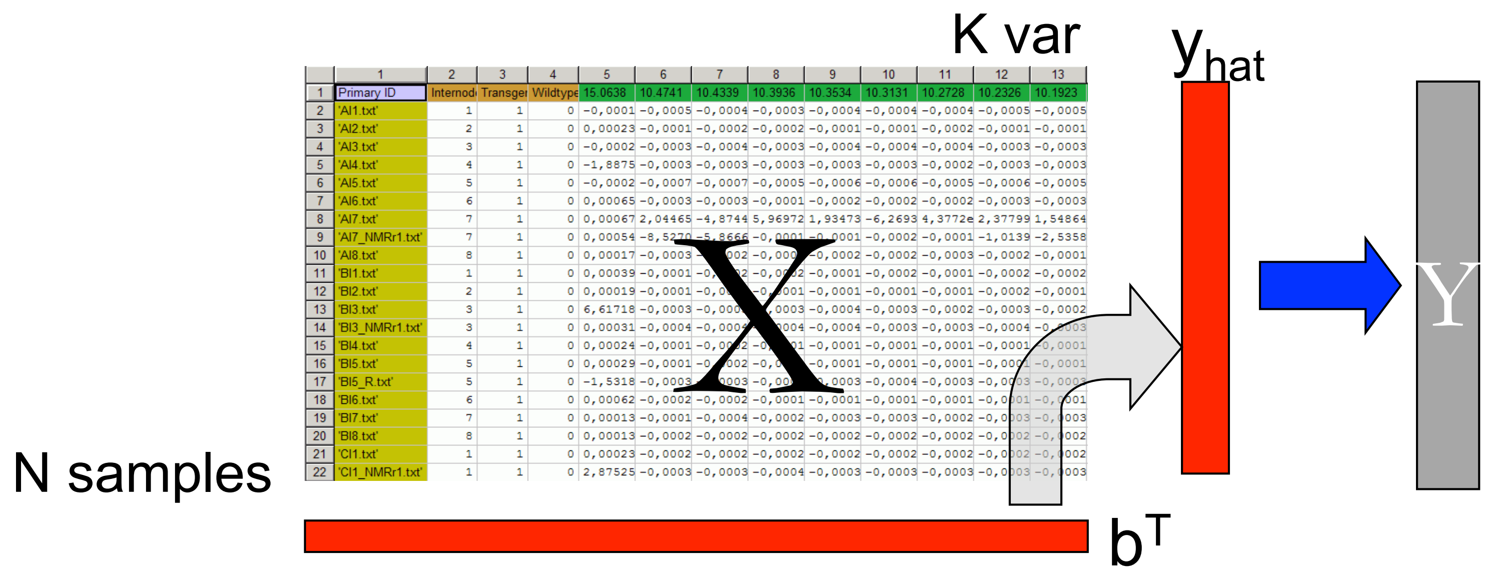


	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Primary ID	Internod	Transger	Wildtype	15.0638	10.4741	10.4339	10.3936	10.3534	10.3131	10.2728	10.2326	10.1923
2	'AI1.txt'	1	1	0	-0,0001	-0,0005	-0,0004	-0,0003	-0,0004	-0,0004	-0,0004	-0,0005	-0,0005
3	'AI2.txt'	2	1	0	0,00023	-0,0001	-0,0002	-0,0002	-0,0001	-0,0001	-0,0002	-0,0001	-0,0001
4	'AI3.txt'	3	1	0	-0,0002	-0,0003	-0,0004	-0,0003	-0,0004	-0,0004	-0,0004	-0,0003	-0,0003
5	'AI4.txt'	4	1	0	-1,8875	-0,0003	-0,0003	-0,0003	-0,0003	-0,0003	-0,0002	-0,0003	-0,0003
6	'AI5.txt'	5	1	0	-0,0002	-0,0007	-0,0007	-0,0005	-0,0006	-0,0006	-0,0005	-0,0006	-0,0005
7	'AI6.txt'	6	1	0	0,00065	-0,0003	-0,0003	-0,0001	-0,0002	-0,0002	-0,0002	-0,0003	-0,0003
8	'AI7.txt'	7	1	0	0,00067	2,04465	-4,8744	5,96972	1,93473	-6,2693	4,3772e	2,37799	1,54864
9	'AI7_NMRr1.txt'	7	1	0	0,00054	-8,5270	-5,8666	-0,0001	-0,0001	-0,0002	-0,0001	-1,0139	-2,5358
10	'AI8.txt'	8	1	0	0,00017	-0,0003	[illegible]	[illegible]	-0,0002	-0,0002	-0,0003	-0,0002	-0,0001
11	'BI1.txt'	1	1	0	0,00039	-0,0001	[illegible]	[illegible]	-0,0001	-0,0002	-0,0001	-0,0002	-0,0002
12	'BI2.txt'	2	1	0	0,00019	-0,0001	[illegible]	[illegible]	-0,0001	-0,0001	-0,0001	-0,0002	-0,0001
13	'BI3.txt'	3	1	0	6,61718	-0,0003	[illegible]	[illegible]	-0,0004	-0,0003	-0,0002	-0,0003	-0,0002
14	'BI3_NMRr1.txt'	3	1	0	0,00031	-0,0004	-0,0004	[illegible]	-0,0004	-0,0003	-0,0003	-0,0004	-0,0003
15	'BI4.txt'	4	1	0	0,00024	-0,0001	[illegible]	[illegible]	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
16	'BI5.txt'	5	1	0	0,00029	-0,0001	[illegible]	[illegible]	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
17	'BI5_R.txt'	5	1	0	-1,5318	-0,0003	[illegible]	[illegible]	[illegible]	-0,0004	-0,0003	-0,0003	-0,0003
18	'BI6.txt'	6	1	0	0,00062	-0,0002	-0,0002	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
19	'BI7.txt'	7	1	0	0,00013	-0,0001	-0,0004	-0,0002	-0,0003	-0,0003	-0,0002	-0,0003	-0,0003
20	'BI8.txt'	8	1	0	0,00013	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002
21	'CI1.txt'	1	1	0	0,00023	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002
22	'CI1_NMRr1.txt'	1	1	0	2,87525	-0,0003	-0,0003	-0,0004	-0,0003	-0,0003	-0,0003	-0,0003	-0,0003

Model the relation between two blocks of data

Chemometrics: Multivariate calibration, MC

Samples	Powders, molecules, industrial process samples, plasma, tissue (leaf), …
and sample characterisation	Spectrometers (NIR,UV, IR, NMR, MS), chromatography, chemical descriptors, gene-arrays, metabolites
provide LOTS of data	highly multidimensional (1000's of variables)



Linear prediction model: $y = Xb + f$

Objective: Provide good fit to estimate y, and good predictions for future samples

Focus: How to solve for b?

Many different methods to choose from

Linear methods

Full rank methods

- Multiple Linear Regression (MLR)
- Stepwise MLR
- Ridge Regression
- Canonical correlation

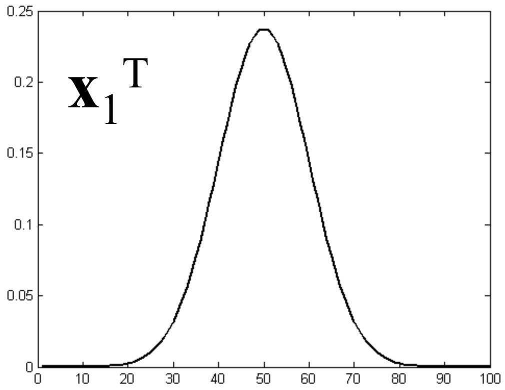
Latent variable regression (LVR) methods

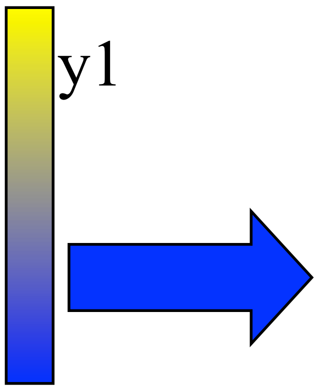
- Principal Component Regression (PCR)
- **Partial Least Squares (PLS)**
- ***Orthogonal Projections to Latent Structures (OPLS)***

Non-Linear methods

- Neural Networks (NN)
- Support Vector Machines (SVM)
- Regression trees

Example: One component system

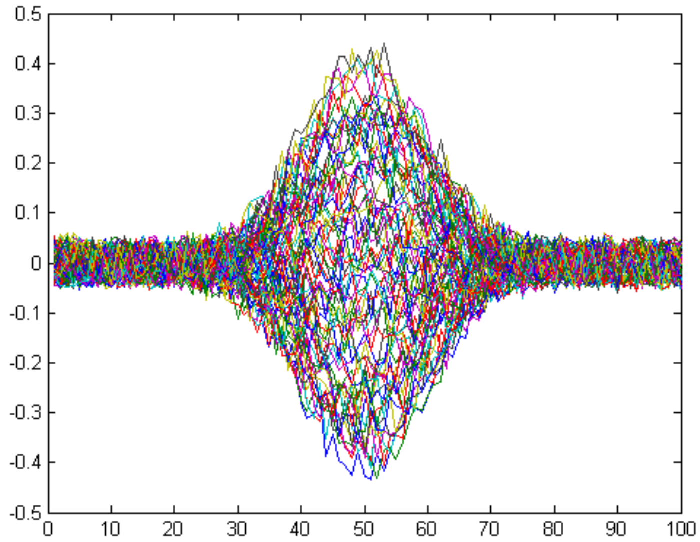




Spectral profile of
Y-predictive component

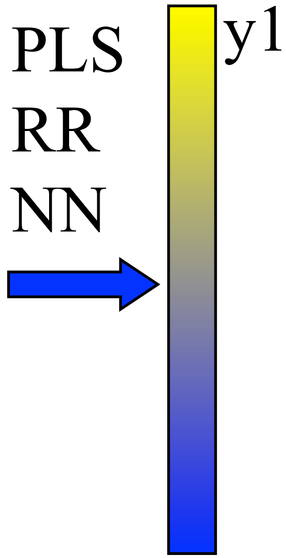
X matrix

100



70

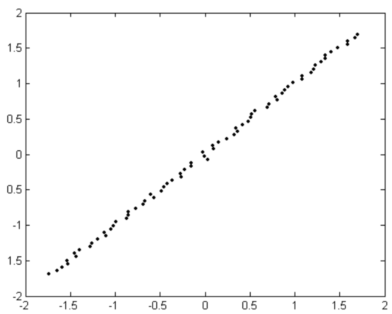
PLS
RR
NN



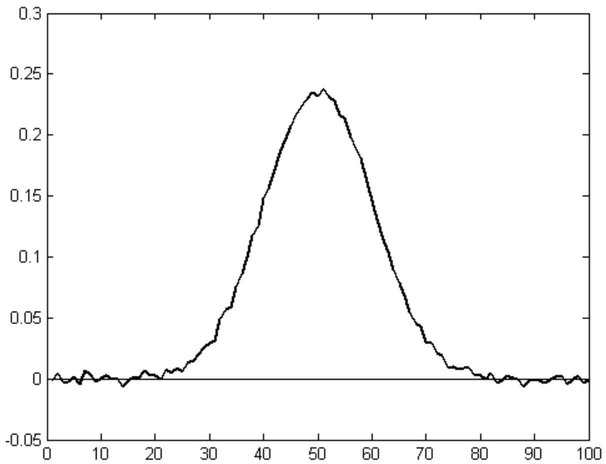
$$\mathbf{X} = \mathbf{y}_1 \mathbf{x}_1^T + \mathbf{E}$$

Example: One component model

PLS regression

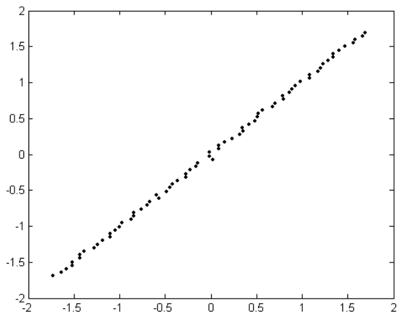


Observed vs Predicted

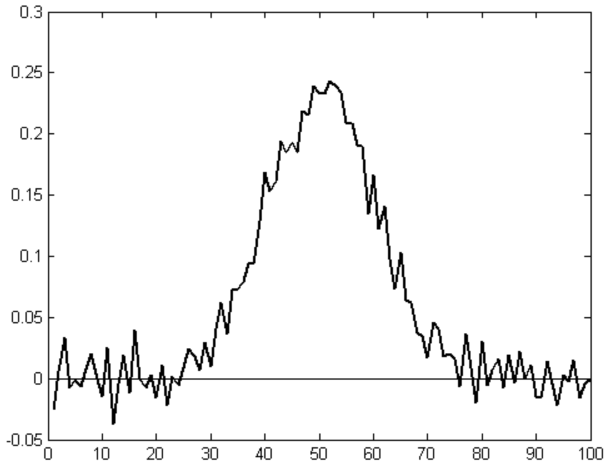


b coefficients

Ridge Regression

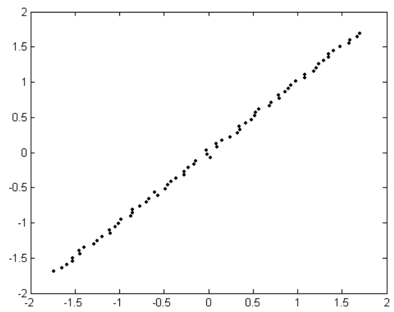


Observed vs Predicted

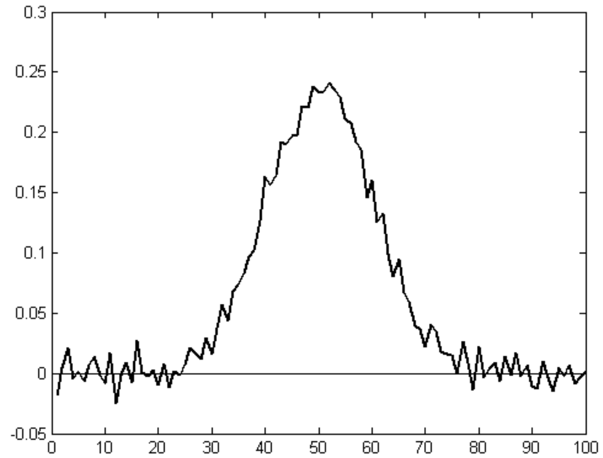


b coefficients

Linear Neural Net



Observed vs Predicted



b coefficients

Svante Wold's, Harald Martens et al... problem!

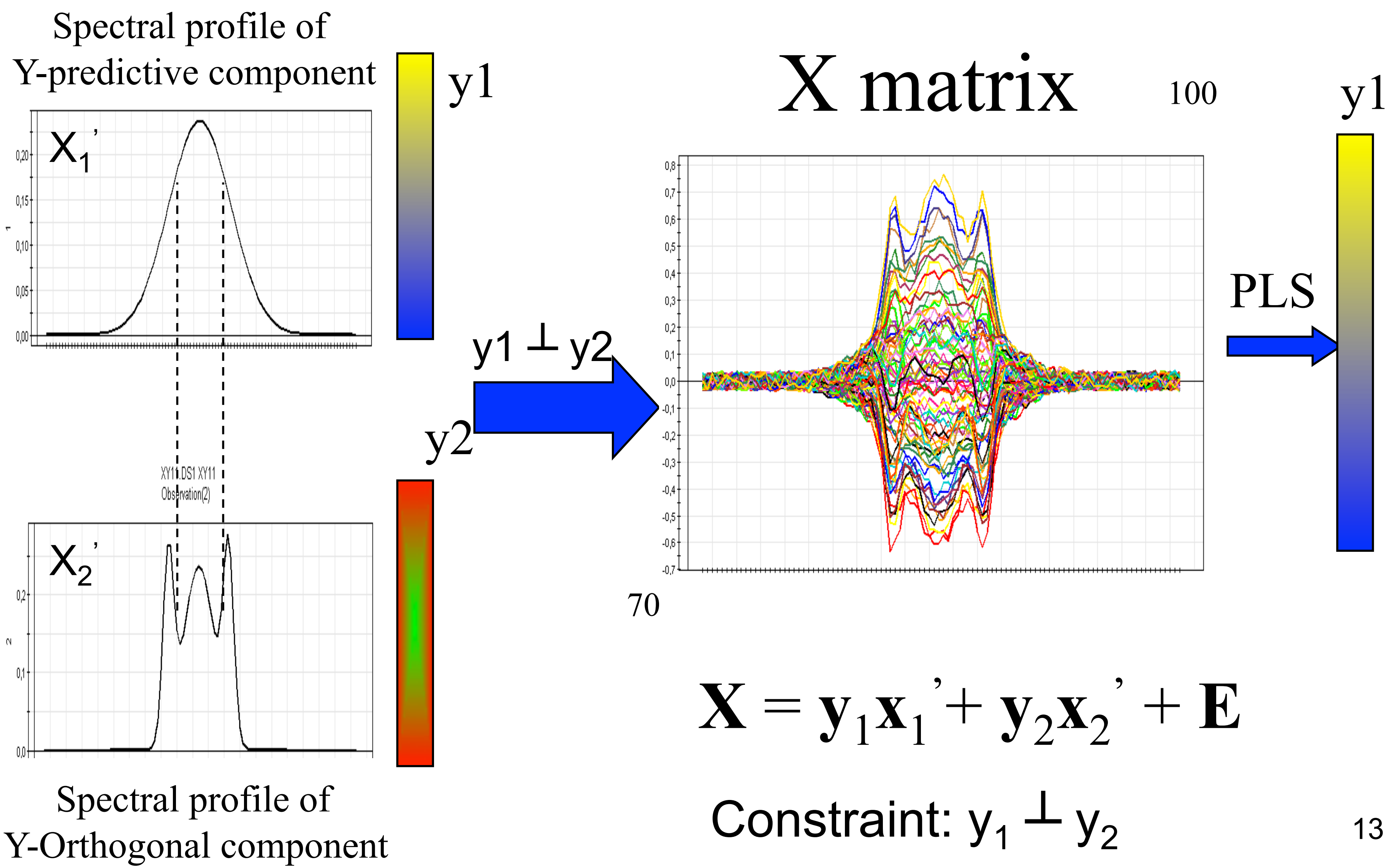
Sensory/Consumer/Chemical / biological data are complex

- **Single PLS component is simply not enough!!**

- **Lots of unknown systematic variation – mostly due to poor knowledge…**
  - strong dietary, environmental, hormonal variations, etc…
  - Experimental variation, sampling, instrumental variation
  - Input material varies with supplier

- **Measured signal is the sum of many contributing factors**
  - Human urine sample (e.g. genetics, diet, gender, age, stress, disease)
  - Pharmaceutical tablet formulation (e.g. binders, fillers, active drug, lubricant)
  - Plant biotech / Pulp & paper (e.g. wood species, cellulose & lignin content, water, age)
  - In QSAR the molecular descriptor profile is a function of its chemical and biological property/activity/function

Example: Two component system (overlap)

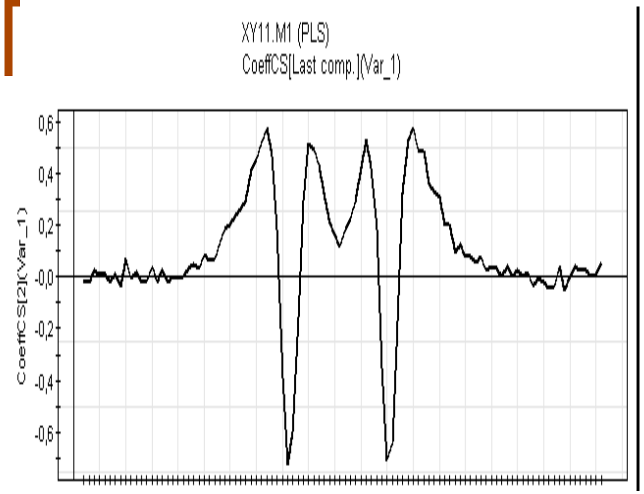


$$\mathbf{X} = \mathbf{y}_1\mathbf{x}_1' + \mathbf{y}_2\mathbf{x}_2' + \mathbf{E}$$

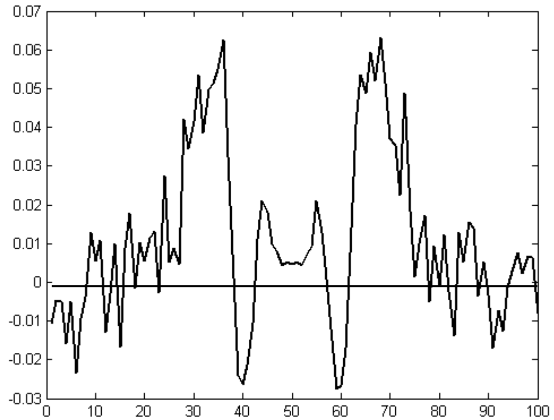
Constraint: $y_1 \perp y_2$

Example: Two component system (overlap)
Model interpretation by regression coefficient profile

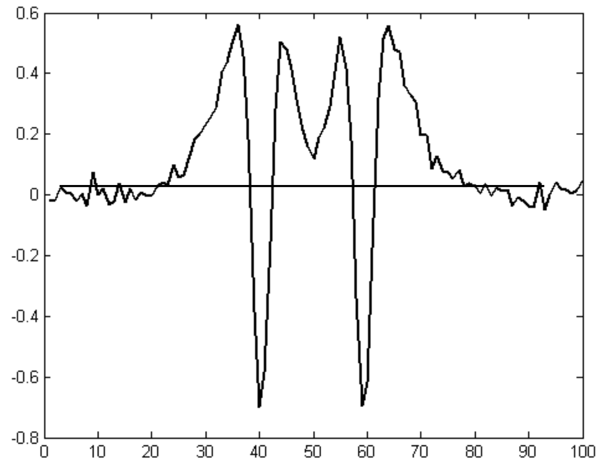
PLS regression | **Ridge Regression** | **Linear Neural Net**



Coefficients



Coefficients



Coefficients

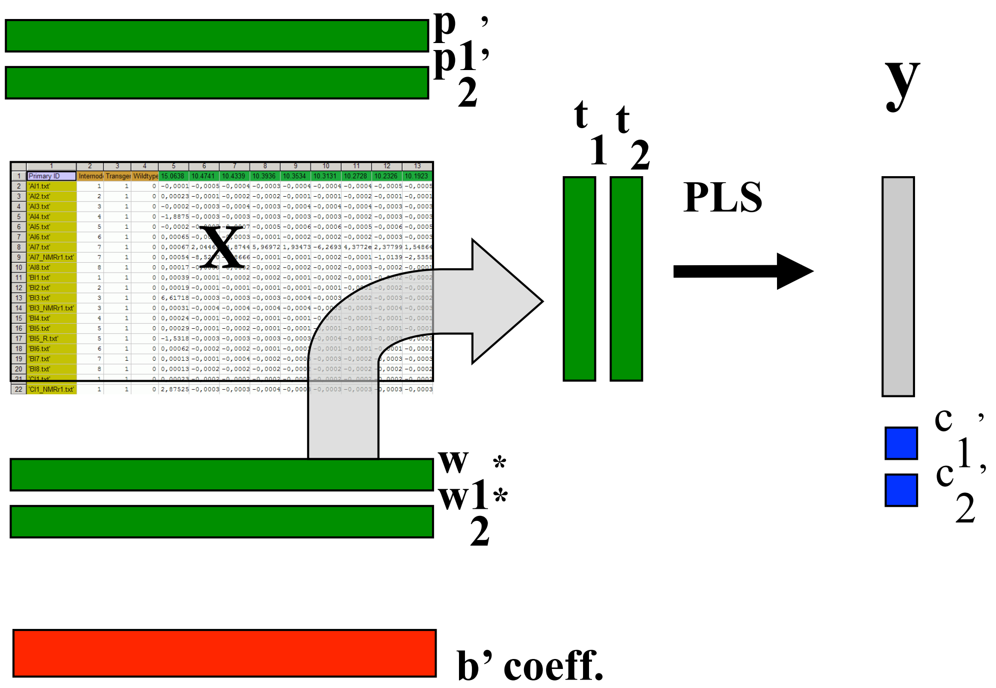
Negative dips observed in regression coefficients, but we have only positive correlation in the data?
→ Disturbance is due to Y-Orthogonal variation

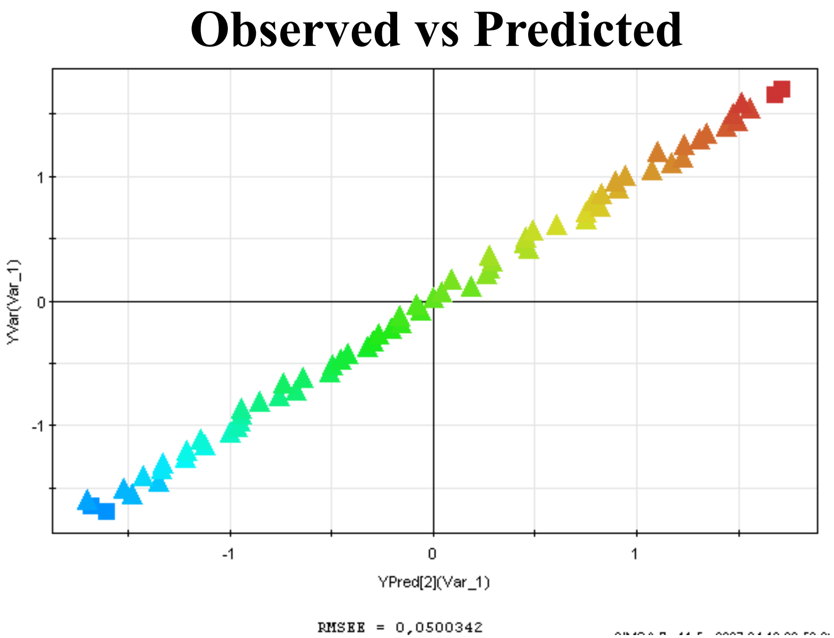
Wold, Martens and colleagues PLS NIPALS (1982)

$$X = TP' + E$$
$$y = Tc' + f$$

Ability to handle unknown variation in X

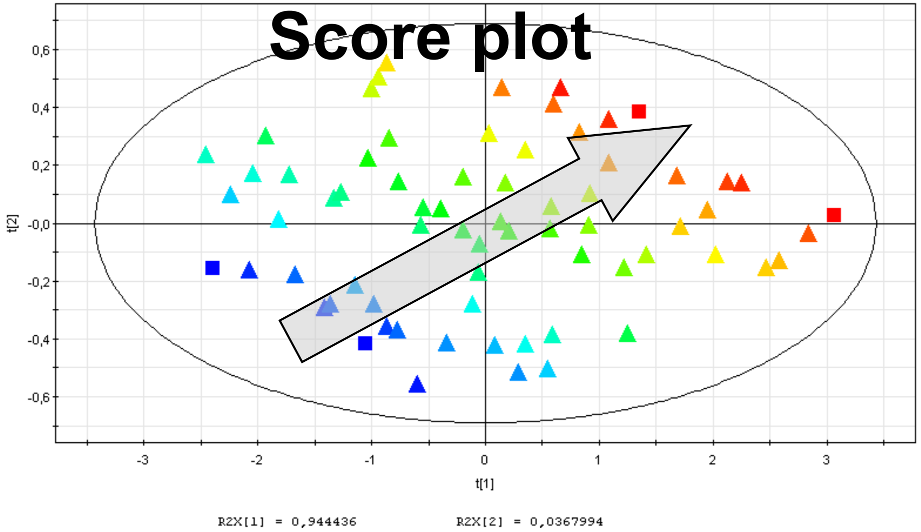
- **Success story of PLS and chemometrics**

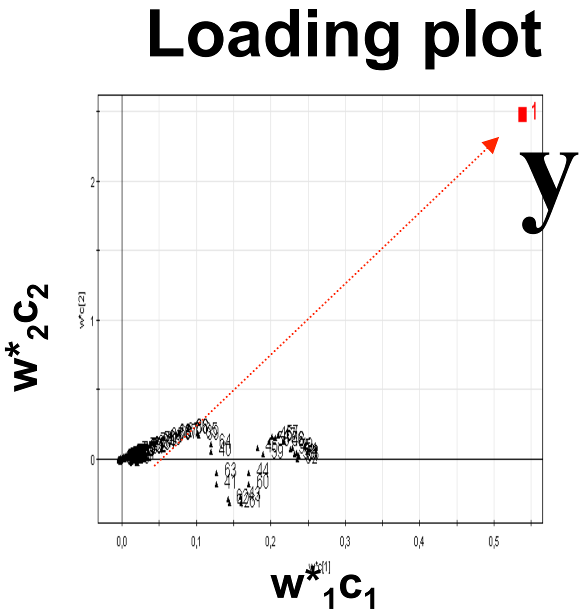




What about interpretation of PLS model

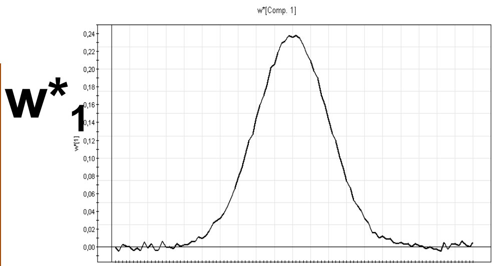
Example: Single-Y, two component system

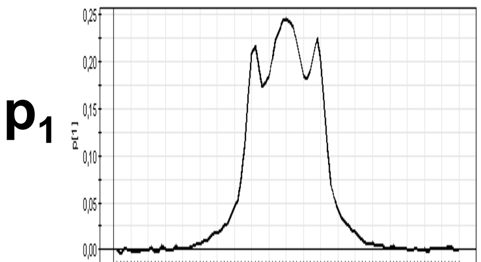


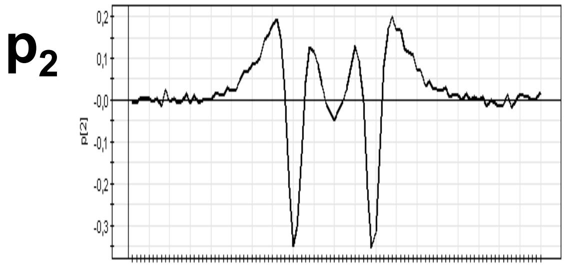


R2Xcomp1=94% variation

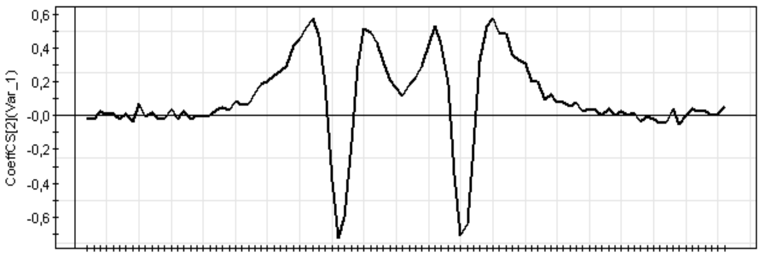
R2Xcomp2= 3.7% variation







Regression coefficients, b

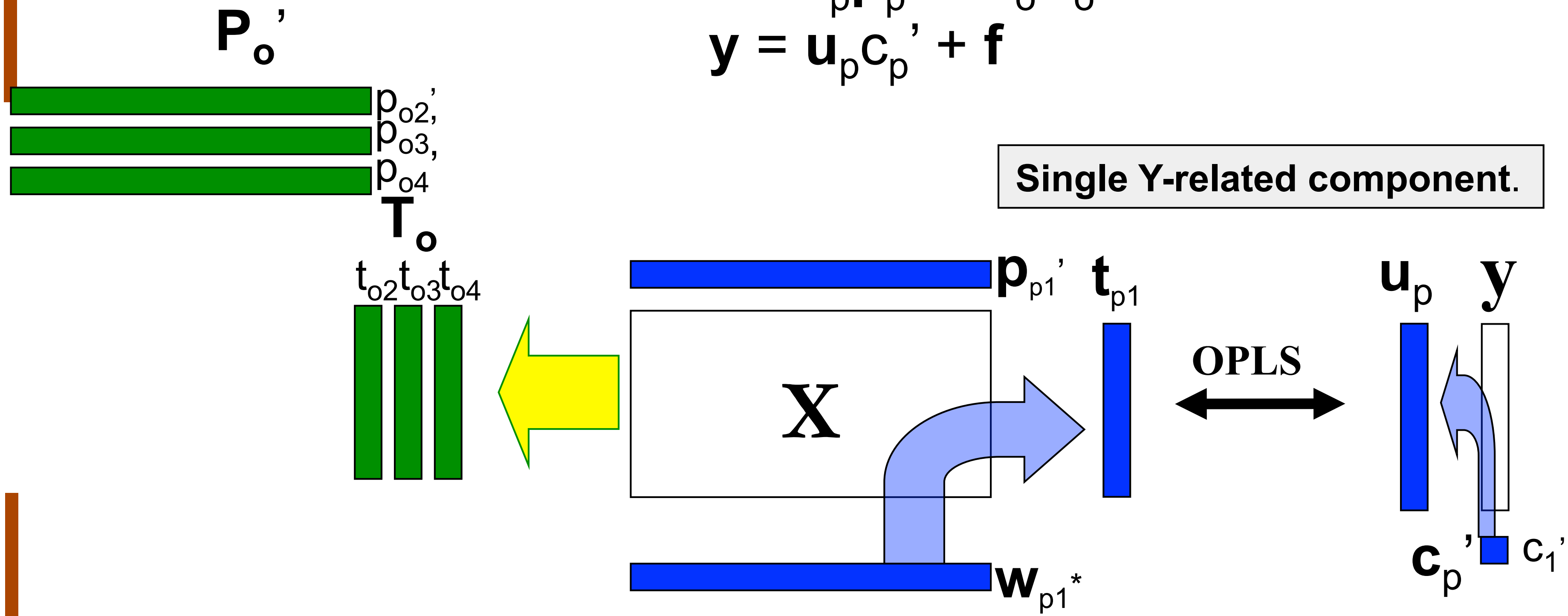


OPLS method -Trygg & Wold (2002)

Trygg J, Wold S. J. Chemometr., 2002; **16**: 119-128

$$\mathbf{X} = \mathbf{t}_p\mathbf{p}_p' + \mathbf{T}_o\mathbf{P}_o + \mathbf{E}$$

$$\mathbf{y} = \mathbf{u}_p c_p' + \mathbf{f}$$



The concept of Orthogonal variation

- Not all systematic variation in the X-block is related to the Y-block

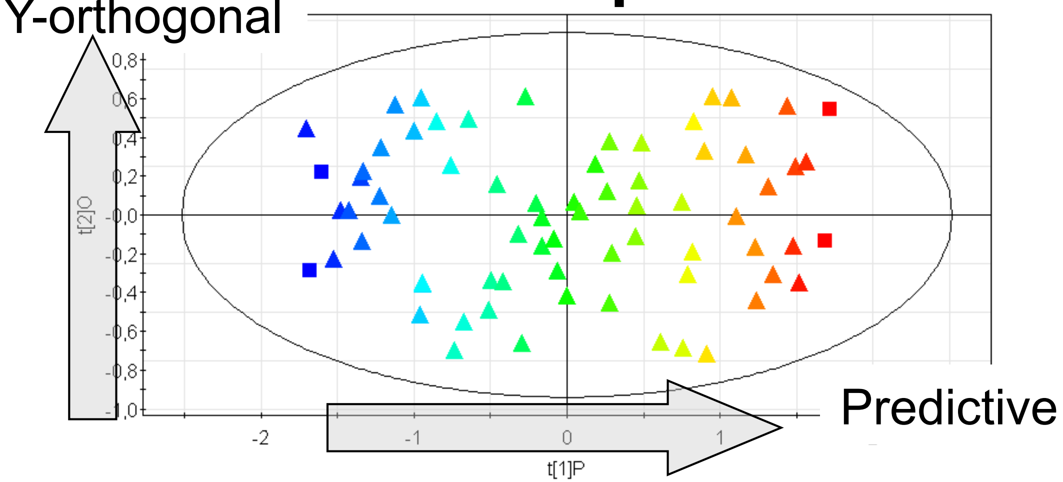
- The new ‘O’-methods, OPLS and O2PLS, are able to divide the systematic X-variation in two parts:
  - What in X is related to Y; Y-predictive variation
  - What in X is not related to Y; Y-orthogonal variation

- This Y-orthogonal variation is important information for the total understanding of the studied system or process
  - Gender
  - Drift
  - Unknown interferents
  - Sampling / Experimental problems
  - Non-linearities
  - Within class variation

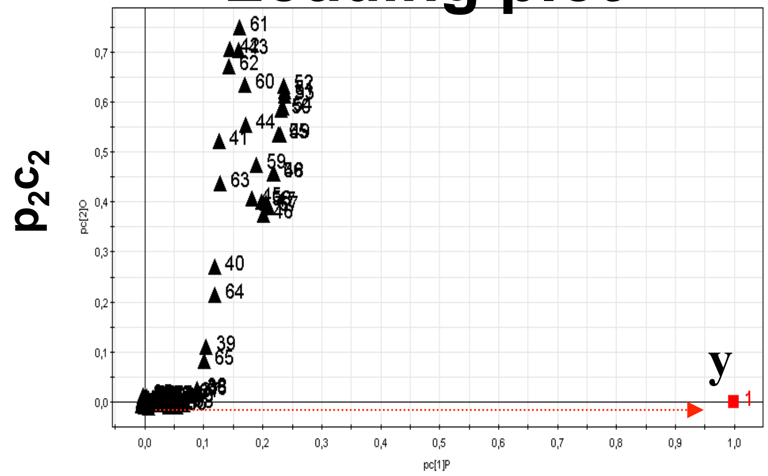
OPLS model

Example: Single-Y, two component system

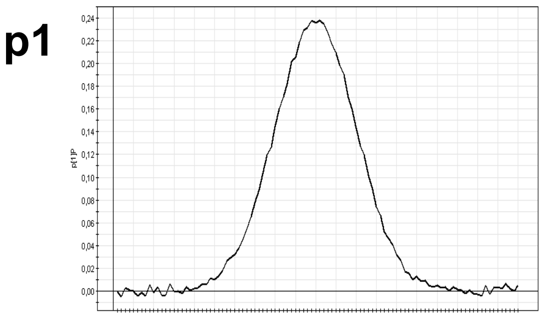
Scores plot



Loading plot

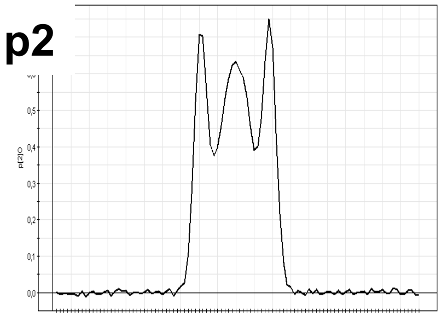


49% variation

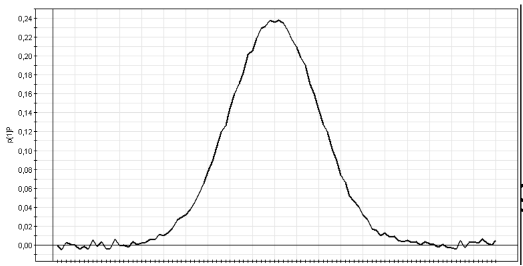


Predictive profile

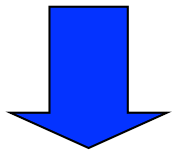
49% variation



Y-orthogonal profile



Predictive profile



Some theoretical properties of OPLS

<u>Objective function:</u>
Find t_o that maximizes overlap with predictive score matrix T

$$\max(\mathbf{T}^{\mathrm{T}}\mathbf{t}_{\mathrm{osc}})^2 = \max(\mathbf{T}^{\mathrm{T}}\mathbf{X}\mathbf{w}_{\mathrm{osc}})^2$$

$$\|\mathbf{w}_{\mathrm{osc}}\| = 1, \qquad \mathbf{Y}^{\mathrm{T}}\mathbf{t}_{\mathrm{osc}} = \mathbf{0}$$

Solution (PLS principle)

$$\mathrm{PCA}(\mathbf{Y}^{\mathrm{T}}\mathbf{X}) = \mathbf{C}\mathbf{W}^{\mathrm{T}}$$

$$\mathrm{E}_{\mathrm{XY}} = \mathbf{X} - \mathbf{T}\mathbf{W}^{\mathrm{T}} \quad \text{where } \mathbf{T} = \mathbf{X}\mathbf{W}$$

$$(\mathbf{E}_{\mathrm{XY}}^{\mathrm{T}}\mathbf{T}\mathbf{T}^{\mathrm{T}}\mathbf{E}_{\mathrm{XY}})\mathbf{w}_{\mathrm{osc}} = \lambda\mathbf{w}_{\mathrm{osc}}$$

Benefits of OPLS modeling

- ✓ **Model diagnostics:**
  – $R^2(XY)$: How much variation in X is correlated to Y, and vice versa?
  – $R2(X_{yo})$: How much is not correlated to Y? (to X?)

- ✓ **Model interpretation**
  – More focussed components (plots) & easier interpretation
    - Predictive components ($\mathbf{T}_p\mathbf{P}_p^T$)
    - Y-orthogonal components ($\mathbf{T}_o\mathbf{P}_o^T$)
  – Pure profile estimation

- ✓ **Model (prediction):**
  – Understand & correct for faults/mistakes found in Y-orthogonal components
  – e.g. experimental, sampling

- **Multi-block modeling** (X←→Y)
  – Integrate, compare and filter multiple data tables

Preference mapping

understanding the descriptive sensory attributes that relate to consumer preferences

Product attributes / Sensory attributes

consumer likings

K var

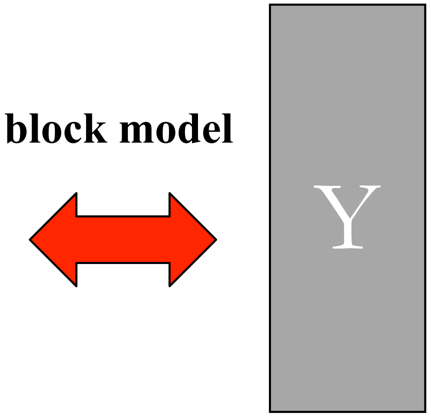
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Primary ID	Internod	Transgei	Wildtype	15.0638	10.4741	10.4339	10.3936	10.3534	10.3131	10.2728	10.2326	10.1923
2	'AI1.txt'	1	1	0	-0,0001	-0,0005	-0,0004	-0,0003	-0,0004	-0,0004	-0,0004	-0,0005	-0,0005
3	'AI2.txt'	2	1	0	0,00023	-0,0001	-0,0002	-0,0002	-0,0001	-0,0001	-0,0002	-0,0001	-0,0001
4	'AI3.txt'	3	1	0	-0,0002	-0,0003	-0,0004	-0,0003	-0,0004	-0,0004	-0,0004	-0,0003	-0,0003
5	'AI4.txt'	4	1	0	-1,8875	-0,0003	-0,0003	-0,0003	-0,0003	-0,0003	-0,0002	-0,0003	-0,0003
6	'AI5.txt'	5	1	0	-0,0002	-0,0007	-0,0007	-0,0005	-0,0006	-0,0006	-0,0005	-0,0006	-0,0005
7	'AI6.txt'	6	1	0	0,00065	-0,0003	-0,0003	-0,0001	-0,0002	-0,0002	-0,0002	-0,0003	-0,0003
8	'AI7.txt'	7	1	0	0,00067	2,04465	-4,8744	5,96972	1,93473	-6,2693	4,3772e	2,37799	1,54864
9	'AI7_NMRr1.txt'	7	1	0	0,00054	-8,5270	-5,8666	-0,0001	-0,0001	-0,0002	-0,0001	-1,0139	-2,5358
10	'AI8.txt'	8	1	0	0,00017	-0,0003	[illegible]	[illegible]	-0,0002	-0,0002	-0,0003	-0,0002	-0,0001
11	'BI1.txt'	1	1	0	0,00039	-0,0001	[illegible]	-0,0002	-0,0001	-0,0002	-0,0001	-0,0002	-0,0002
12	'BI2.txt'	2	1	0	0,00019	-0,0001	[illegible]	-0,0001	-0,0001	-0,0001	-0,0001	-0,0002	-0,0001
13	'BI3.txt'	3	1	0	6,61718	-0,0003	[illegible]	[illegible]	-0,0004	-0,0003	-0,0002	-0,0003	-0,0002
14	'BI3_NMRr1.txt'	3	1	0	0,00031	-0,0004	-0,0004	[illegible]	-0,0004	-0,0003	-0,0003	-0,0004	-0,0003
15	'BI4.txt'	4	1	0	0,00024	-0,0001	-0,0002	[illegible]	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
16	'BI5.txt'	5	1	0	0,00029	-0,0001	-0,0002	[illegible]	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
17	'BI5_R.txt'	5	1	0	-1,5318	-0,0003	[illegible]	[illegible]	[illegible]	-0,0004	-0,0003	-0,0003	-0,0003
18	'BI6.txt'	6	1	0	0,00062	-0,0002	-0,0002	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001	-0,0001
19	'BI7.txt'	7	1	0	0,00013	-0,0001	-0,0004	-0,0002	-0,0003	-0,0003	-0,0002	-0,0003	-0,0003
20	'BI8.txt'	8	1	0	0,00013	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002
21	'CI1.txt'	1	1	0	0,00023	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002	-0,0002
22	'CI1_NMRr1.txt'	1	1	0	2,87525	-0,0003	-0,0003	-0,0004	-0,0003	-0,0003	-0,0003	-0,0003	-0,0003

X

Products

N samples

Two block model



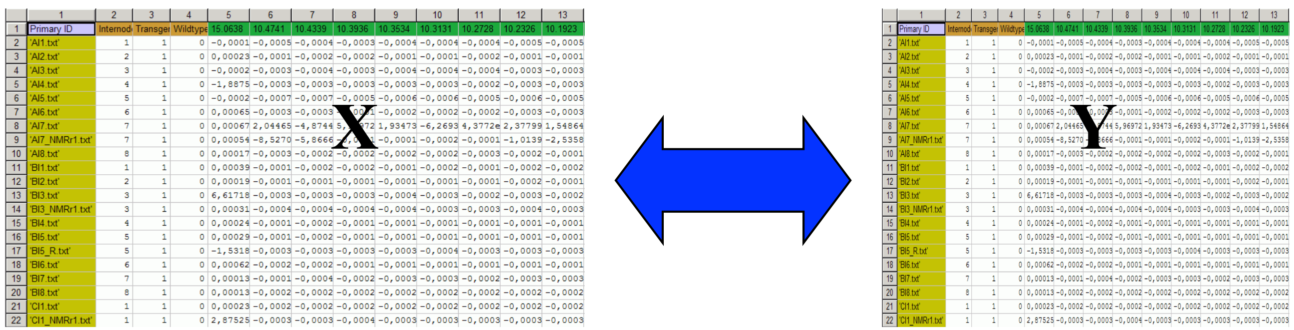
M var

Multi-block modeling

- Similar to multivariate calibration (**X→Y**), but extended to *both directions*
- Compare & Integrate **X** and **Y** in terms of….
  - Experimental conditions, Process step, Time (drift), Instruments, Replication, Pre-treatments, and so on…
- Understand…
  - Overlap? What is jointly related?
  - What is unique for X, for Y?

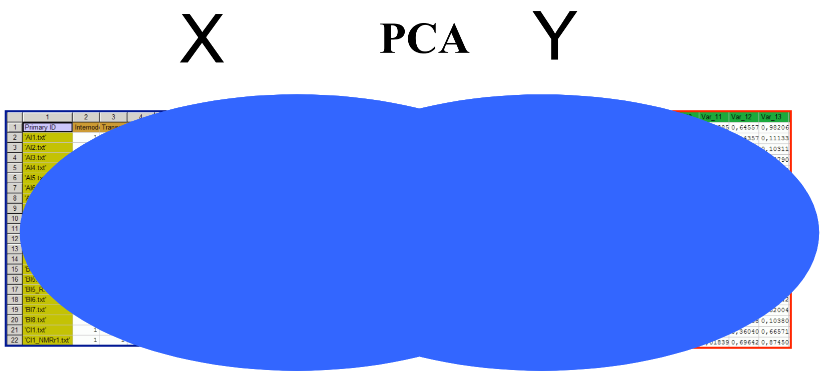
Product attributes / Sensory attributes

consumer likings



Current methods lack proper multi-block structure

- ✓ **They mix all the different variations**
- ✓ **and/or… uni-directional models (X→Y)**
- ✓ **and/or… focus on correlations only**

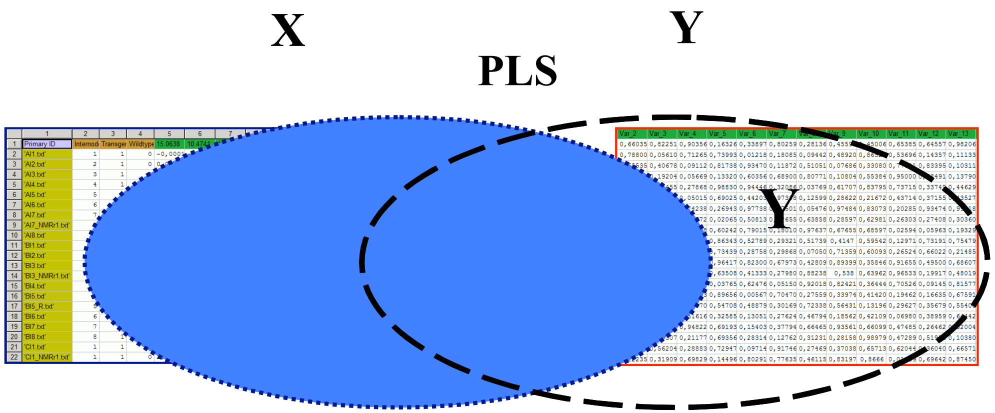


a.) Single-block model (e.g. PCA)

- Mixes all variation together
- No relation between blocks

b.) Two-block model (e.g. PLS, PCR, MLR, RR, SVM, CC)

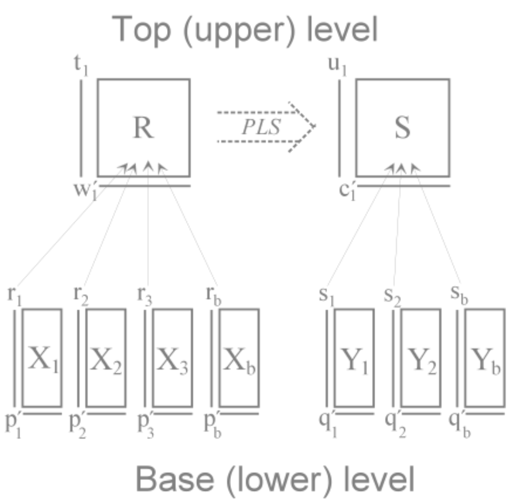
- Uni-directional (X→Y), only X-block is modelled
- Mixes all variations in X block together,
- Non-overlapping variation of no interest
    - **Sometimes, the most interesting!!**



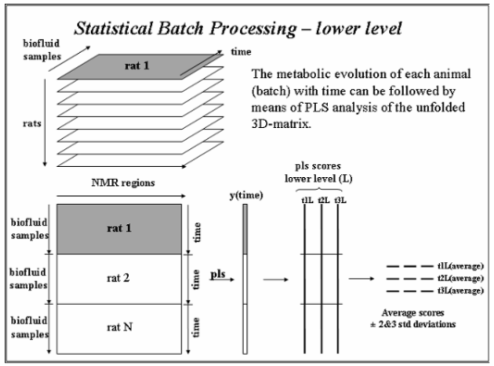
…including
Multi-block extensions of PCA/PLS modeling

c.) Hierarchical modeling
d.) Batch modeling
e.) BIF-PLS, L-PLSR (more?)

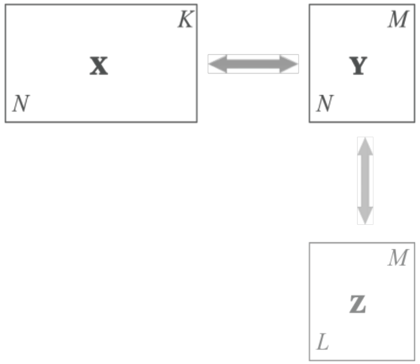
Hierarchical modeling



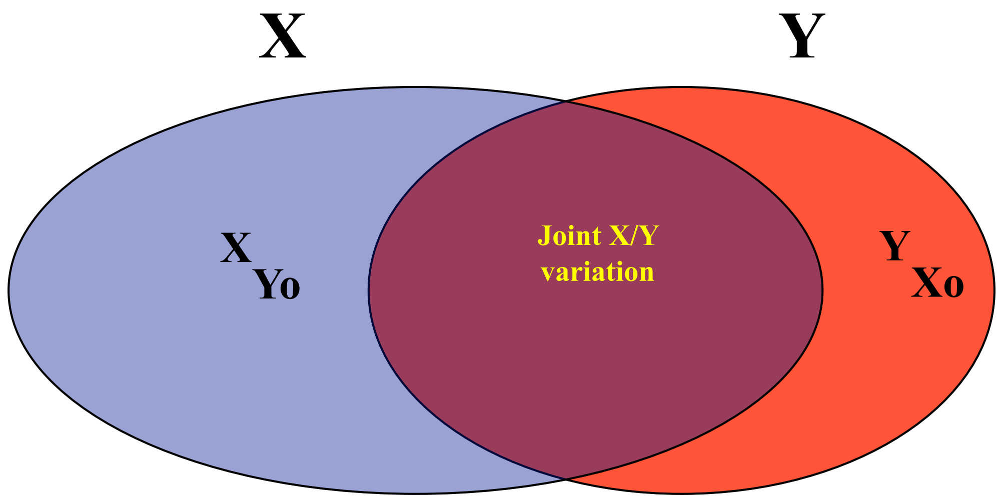
Batch modeling



Three block modeling
(BIF-PLS, L-PLSR)



O2PLS for multi-block modeling

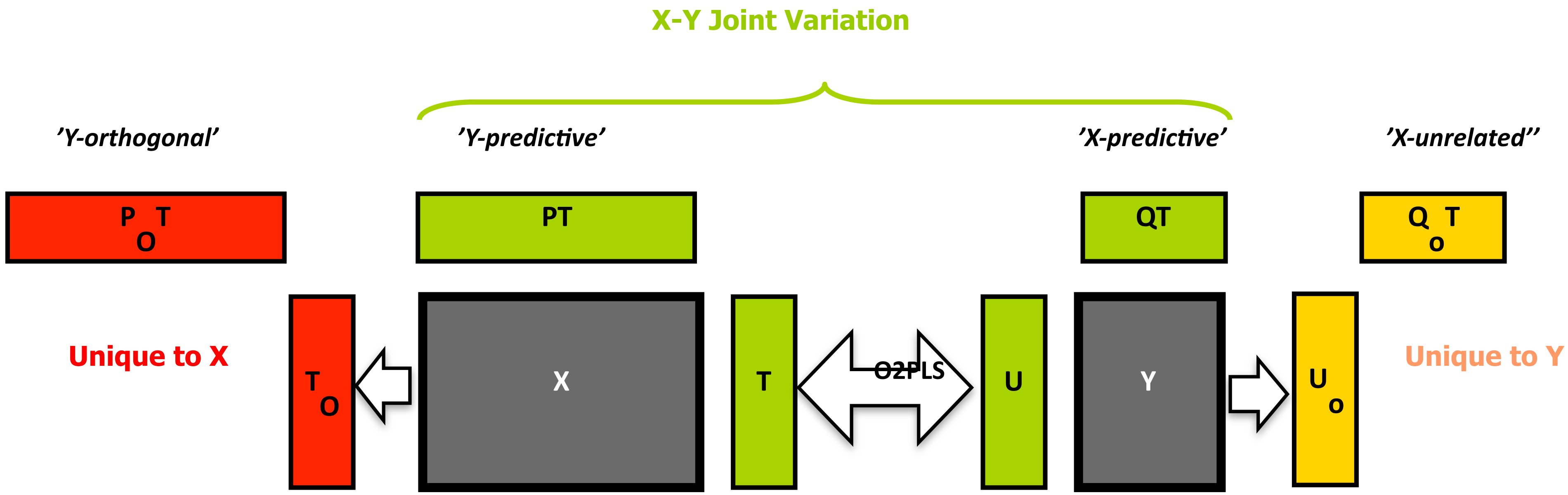


- Does not mix all different variations together
  - *Separates 'between block' (Y-related) variation from 'within block' (Y-orthogonal) variation*

- Only uses joint X/Y variation for modeling Y from X (and vice versa)
  - Main reason for differences in how we interpret a PLS model (W,W*, P, B)

The O2PLS model structure

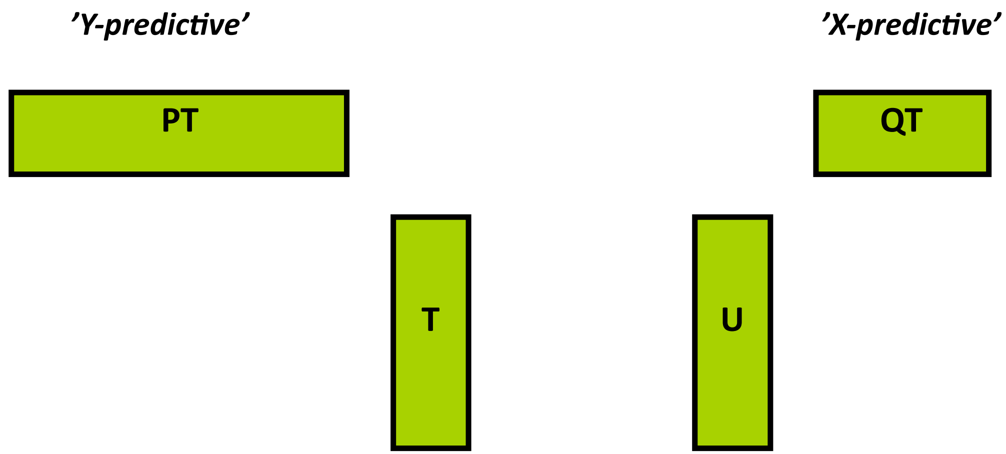
- Separate model for joined and orthogonal variation
- Model of **X**: $\mathbf{X} = \mathbf{T}_p\mathbf{P}_p^T + \mathbf{T}_o\mathbf{P}_o^T + \mathbf{E}$
- Model of **Y**: $\mathbf{Y} = \mathbf{U}_p\mathbf{Q}_p^T + \mathbf{U}_o\mathbf{Q}_o^T + \mathbf{F}$
- O2PLS components number for **X-Y** joined variation obtained by PCA of their co-variance matrix ($\mathbf{X}^T\mathbf{Y}$)



Trygg J, Wold S, O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter JOURNAL OF CHEMOMETRICS 17 (1): 53-64 JAN 2003

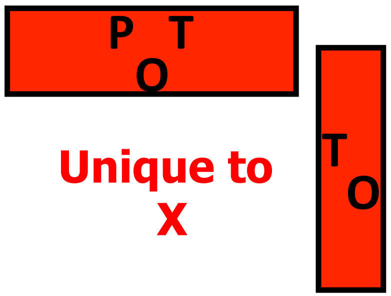
O2PLS - Interpretation

- **What is the overlapping variation between X↔Y?**



O2PLS - Interpretation

- **Which variation is only found in X?**

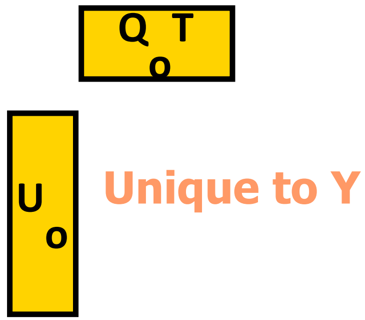


- What is seen in the orthogonal vectors, X⊥Y?
  - Systematic variation in **X** that is orthogonal between **X** and **Y**
  - Look at $\mathbf{T}_o$ and corresponding $\mathbf{P}_o$

O2PLS - Interpretation

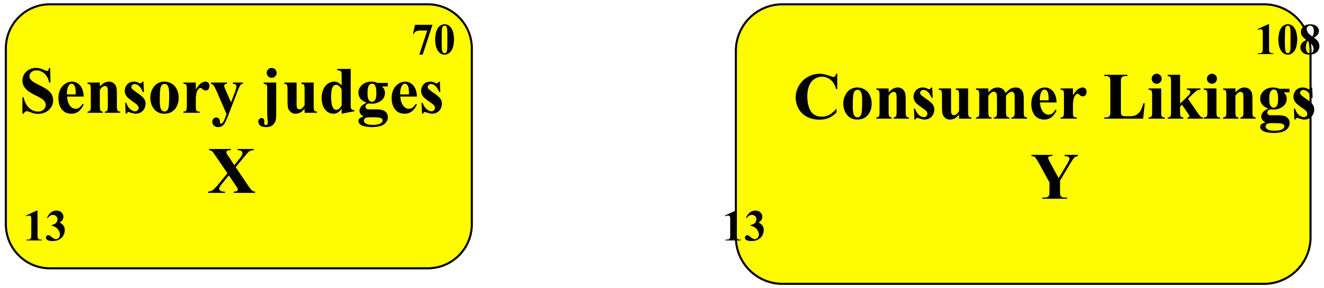
- **Which variation is only found in Y?**

- What is seen in the unrelated vectors, $Y \perp X$?
  - Systematic variation in **<u>Y</u>** that is uncorrelated between **X** and **Y**
  - Look at $\mathbf{U}_o$ and corresponding $\mathbf{Q}_o$



O2PLS modeling of Apples
Preference mapping

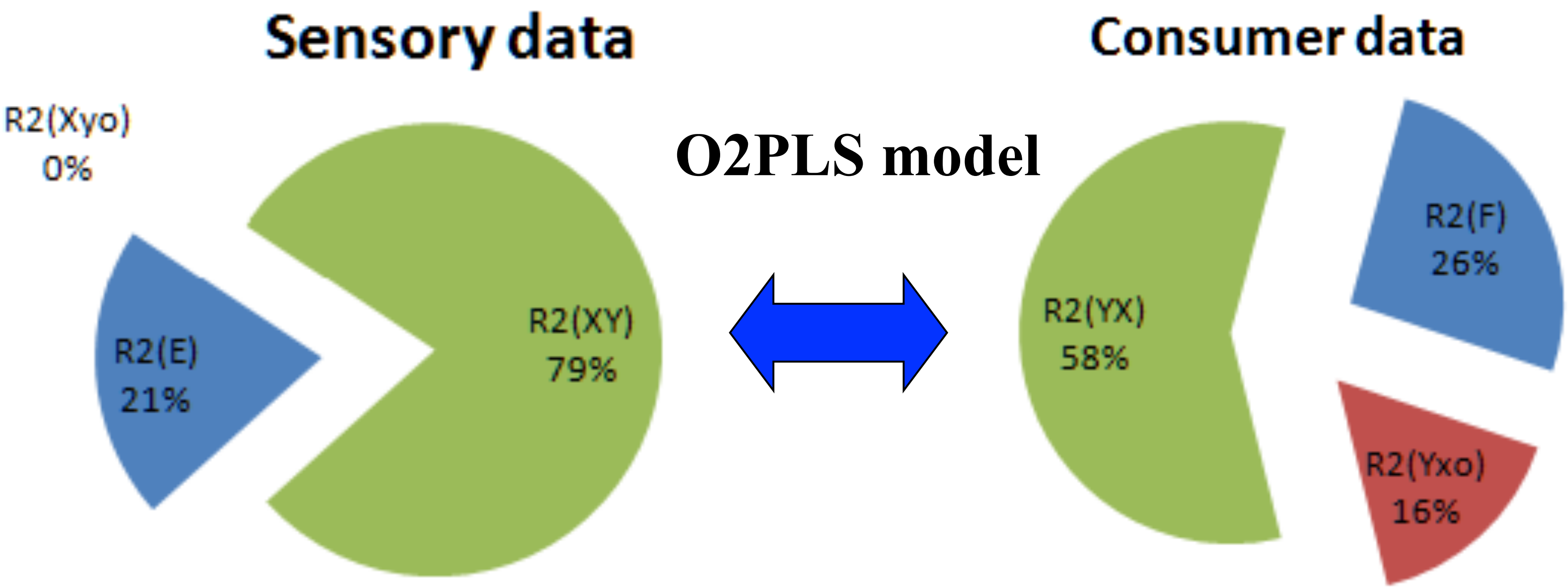
- Sensory and preference data for a set of 13 apples
  - 70 sensory attributes (X-variables); panel averages across 12 judges
  - 108 comsumer likings (Y-variables), expressed on a nine-grade scale
  - Original reference [MacFie, H., *et al.*, 1999].



- Group formation among sensory attributes
  - 1_ is a "First Bite" attribute,
  - E_ is an "External Appearance" attribute,
  - EA_ is "External Aroma”,
  - A_ is "Astringent aftertaste”,
  - F_ is "Flavor",
  - I_ is "Internal Appearance“,
  - T_ is "Texture"

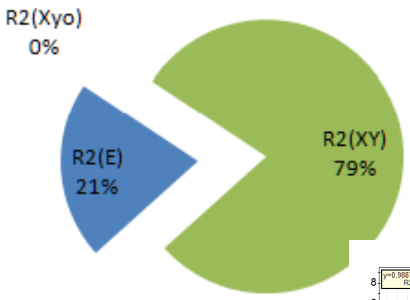
O2PLS modeling of Apples
Preference mapping

- O2PLS analysis results in a 4 + 0 + 2 model
  - Joint information; $R^2X = 0.79$; $R^2Y = 0.56$
  - Y-orthogonal (uncorrelated to the judges); ---
  - X-orthogonal (uncorrelated to the consumers); $R^2Y_{X\text{-orth}} = 0.16$

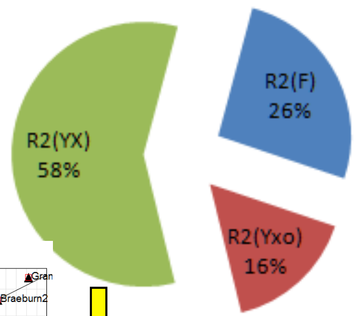


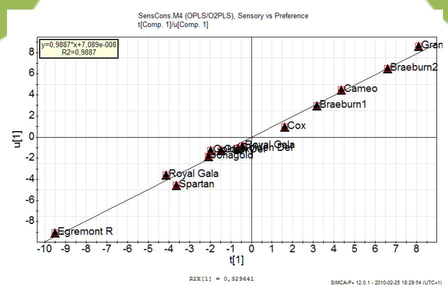
O2PLS modeling in Preference mapping
Joint variation (overlapping)

Sensory data



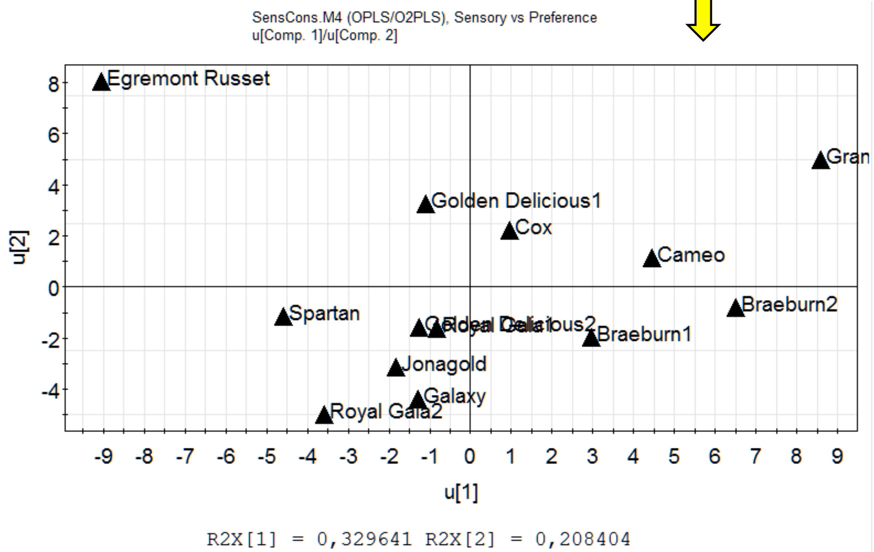
Consumer data





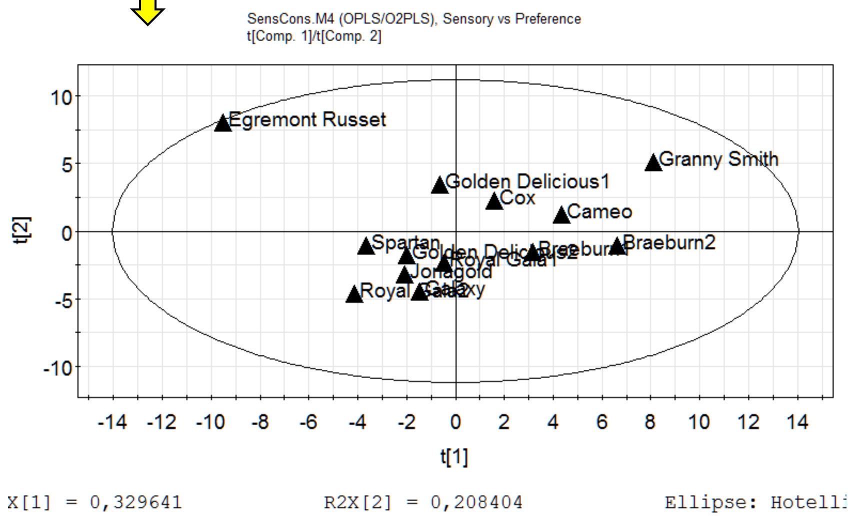
Sensory data plot

Score plot (t_1-t_2)



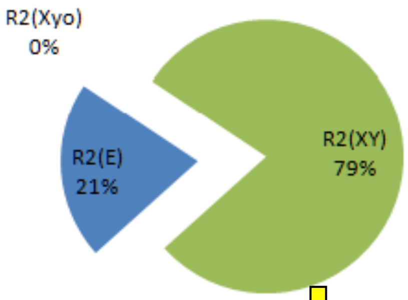
Consumer data plot

Score plot (t_1-t_2)

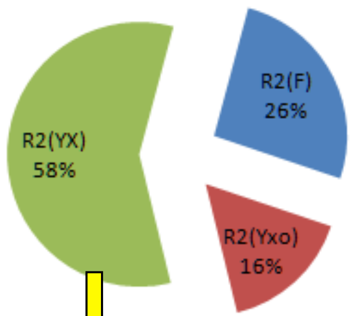


O2PLS modeling in Preference mapping
Joint variation (overlapping)

Sensory data

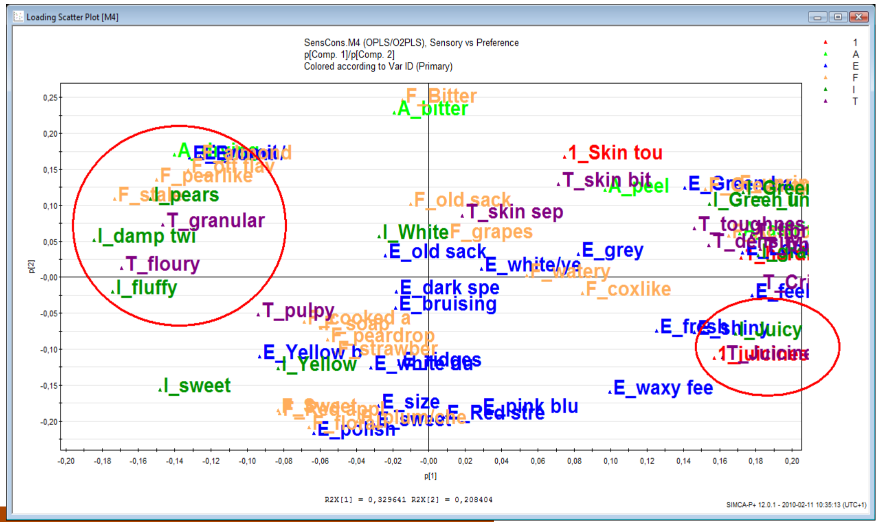


Consumer data



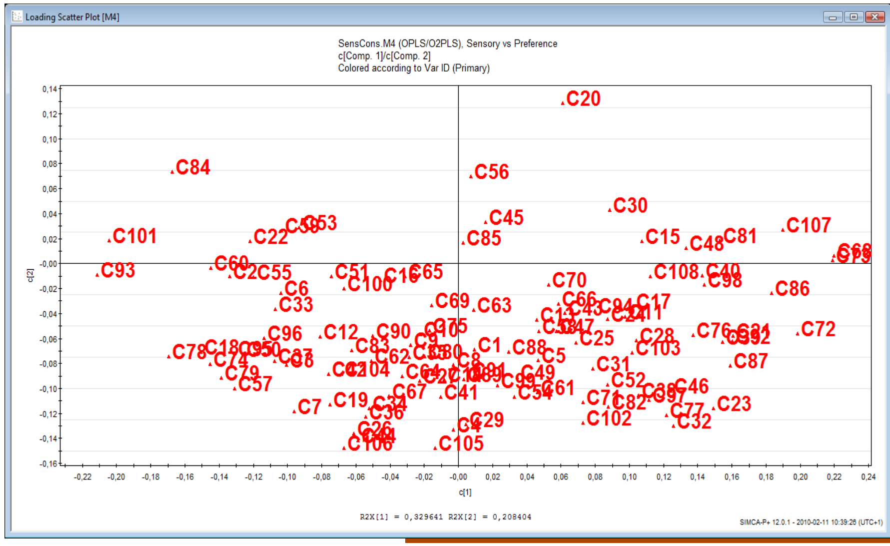
Joint variation (79%)

Loading plot (p_1-p_2)



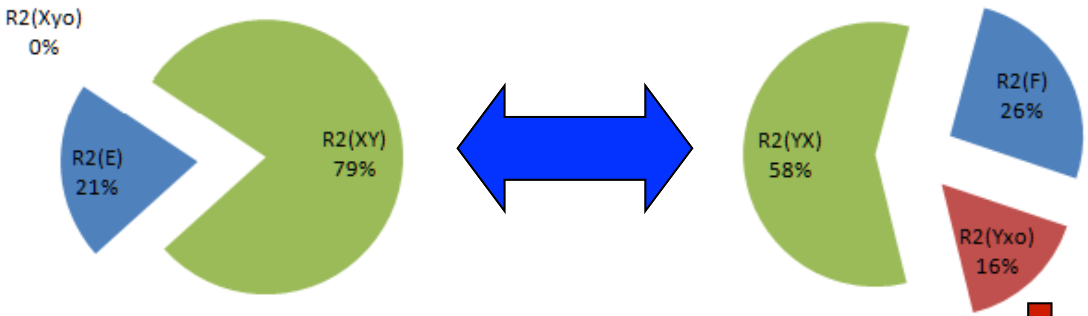
Joint variation (58%)

Loading plot (q_1-q_2)



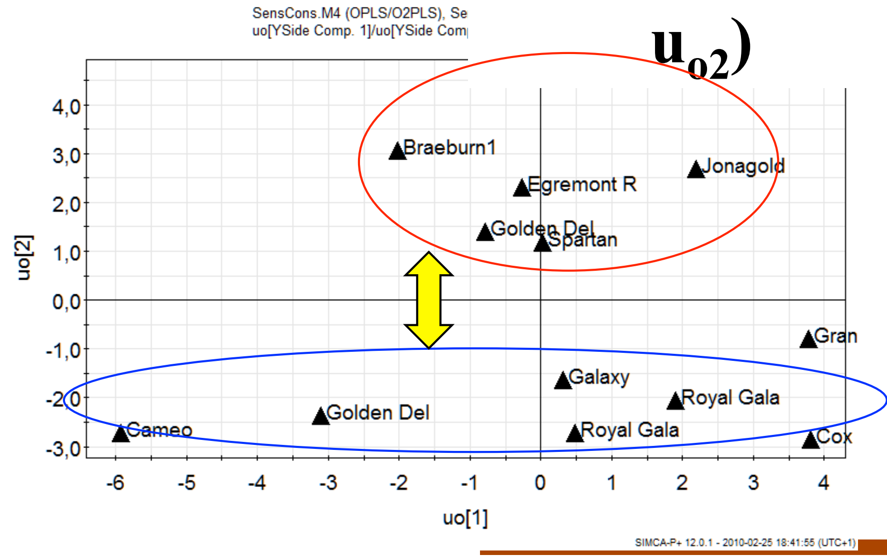
O2PLS modeling in Preference mapping
Unique variation in Y (uncorrelated)

Sensory data **Consumer data**



Unique variation in Y

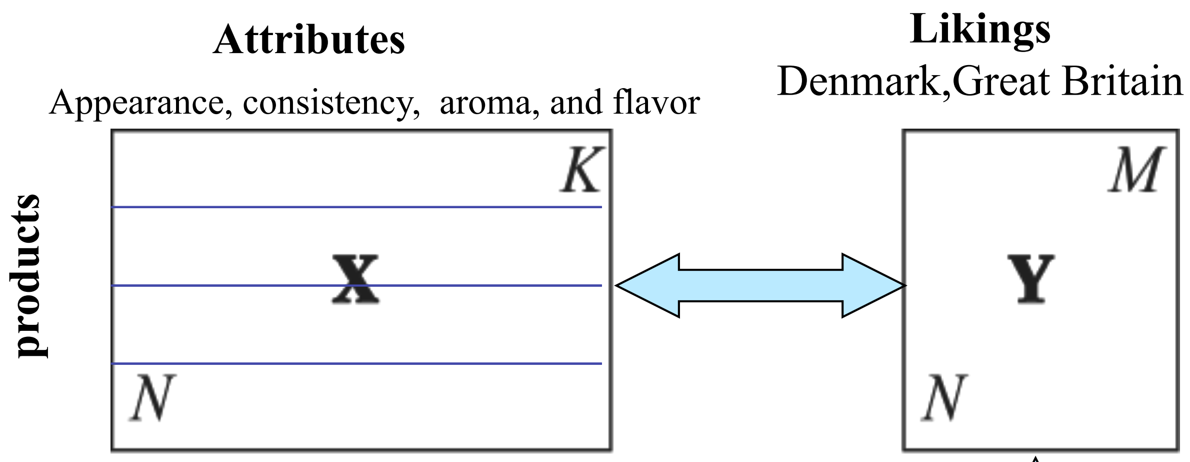
Apples plot
(Score plot u_{o1}-u_{o2})



Unique variation in Y (16%)

Consumers discriminate
Between two groups of apples
Not picked up by Sensory data

Preference mapping using 3 block extension of O2PLS

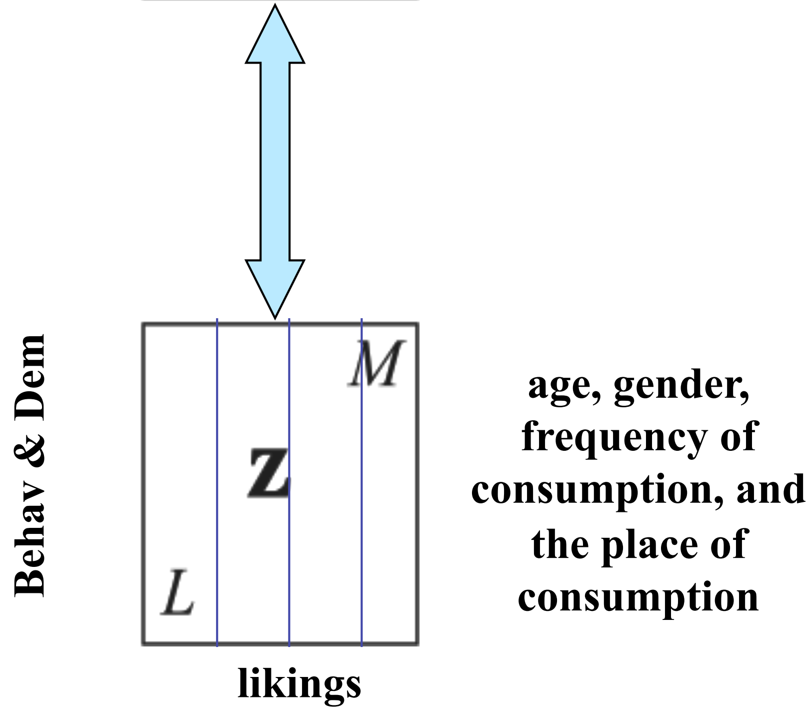


N = 8 products

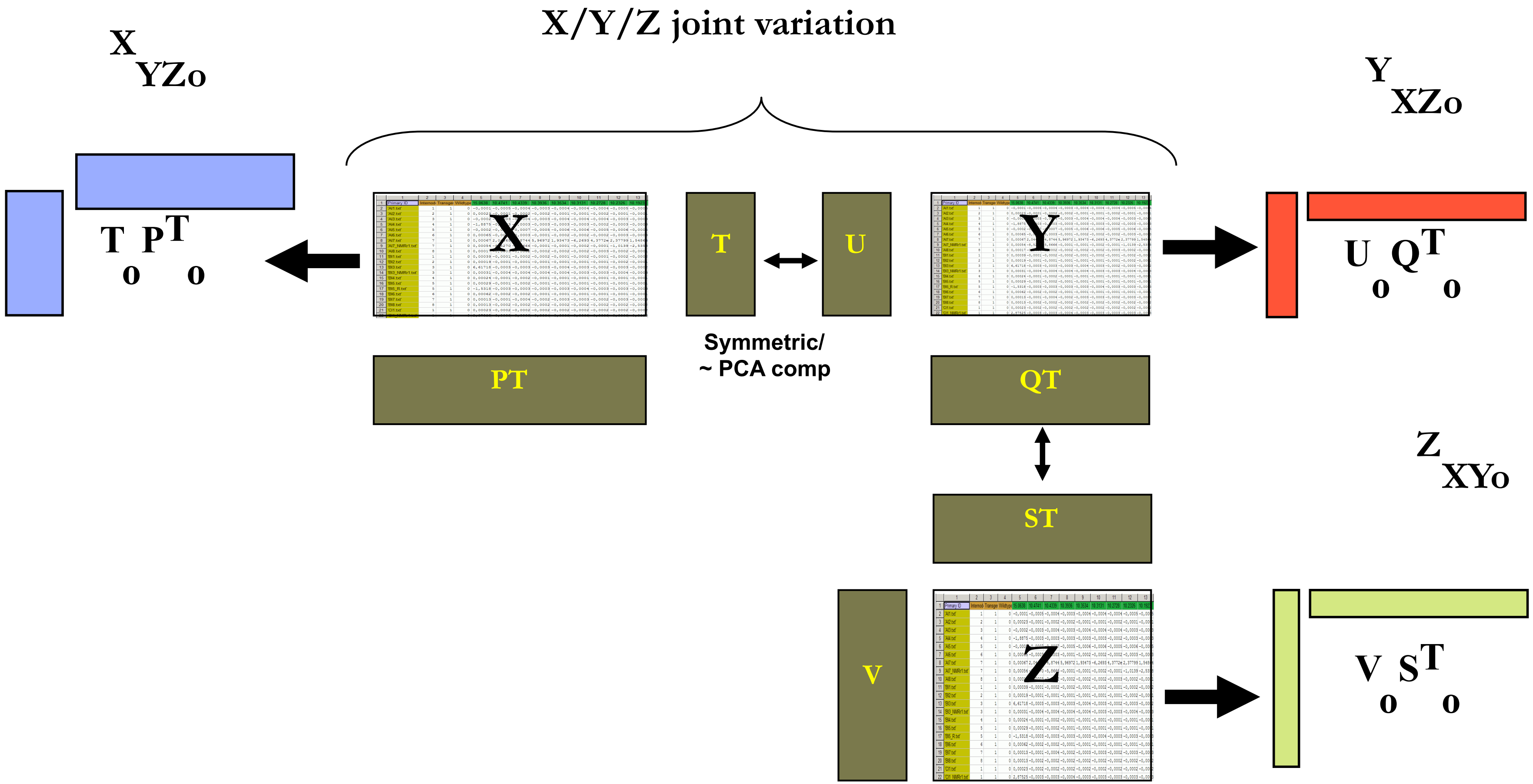
K = 29 sensory attributes (X)

M = 156 consumer likings (Y)

L = 8 behavioural and demographic data (Z)



Structure of the extended O2PLS model



Concluding remarks

- OPLS and O2PLS are the only methods that have a multi-focus
  - Analyzes the correlation
  - analyzes the Orthogonal variation i.e. the uncorrelated, unique structure in a comprehensive way.
- Predictive component
  - What is related between blocks
- Uncorrelated structure, orthogonal variation
  - What is NOT related between blocks

Acknowledgements

Swedish Foundation for Strategic Research (SSF)
Swedish Research Council
FORMAS FuncFibre
Knut & Alice Wallenberg Foundation
GlaxoSmithKline
AstraZeneca
MKS Umetrics

Chemometrics group , Umeå University
M. Bylesjö
S. Wiklund
H. Stenlund
P. Jonsson
H. Antti
M. Hedenström
M. Sjöström

Umeå Plant Science Center, Umeå Univ
T. Moritz
A. Johansson
D. Eriksson
A. Sjödin
B Sundberg
G. Sandberg

Bergen University
Olav Kvalheim

Acure Pharma .
T. Lundstedt

Umetrics AB, Sweden
E. Johansson
S. Wold
L. Eriksson

AstraZeneca
O. Berntsson
M Josefsson
S Folestad
J Gottfries

GlaxoSmithKline
R Escott
C Airau
T Thurston

SweTreeTechnologies
M Hertzberg
K Johansson
A Karlsson