SARTORIUS

Simplifying Progress

OPLS® in Process Modeling

Lennart Eriksson, Ph.D., Assoc. Prof.
Senior Lecturer and Principal Data Scientist

Born in Data Analytics

- Company founded in 1987 by Professor Svante Wold, in Umeå, Sweden
  - Originator of Chemometrics and the SIMCA® Methodology
- Patented technologies in Design of Experiments and Multivariate Data Analysis
- We help our customers bring high-quality products to market faster
- Part of Sartorius Stedim Biotech since April 2017
- Products like MODDE®, SIMCA® and SIMCA®-online
- Global strength with local presence

Business Growth Through the Entire Product Lifecycle

Active Dashboard
Interactive performance insight

Control Advisor
Avoid problems before they arise

MODDE®
Get it right from the start

SIMCA®
Turn data into growth

SIMCA®-online
Ensured manufacturing success

Umetrics® Suite

Education/training and consulting
Transferring expertise to you

Upcoming Webinars (https://www.sartorius.com/en/company/exhibition-conferences)

UPCOMING EVENTS | PAST EVENTS | CALENDAR | SCHEDULE +

Search

STANDARD
Design of Experiments (DOE) for the Beginner
TUE, JAN 26, 2021 03:00 PM - 04:00 PM CET

STANDARD
Multivariate Data Analysis (MVDA) for the Beginner
THU, JAN 28, 2021 03:00 PM - 04:00 PM CET

STANDARD
Lean-and-clean DOE using One-click analysis
TUE, FEB 16, 2021 03:00 PM - 04:00 PM CET

STANDARD
OPLS® in process modeling
THU, FEB 18, 2021 03:00 PM - 04:00 PM CET

STANDARD
Robust optimization made easy
TUE, MAR 2, 2021 03:00 PM - 04:00 PM CET

STANDARD
Analyzing batch process data, a step-by-step guide
THU, MAR 4, 2021 03:00 PM - 04:00 PM CET

STANDARD
From Design of Experiments to Design Space Estimation
TUE, MAR 23, 2021 03:00 PM - 04:00 PM CET

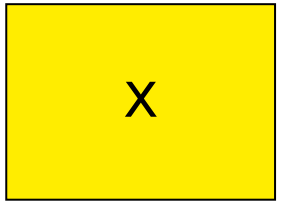
STANDARD
Multiblock Orthogonal Component Analysis (MOCA) – A Novel Tool for Data Integration
THU, MAR 25, 2021 03:00 PM - 04:00 PM CET

Multivariate Process Data

- Monitoring a process
  - Early warning of disturbances
  - Diagnostics - finding ”assignable causes”

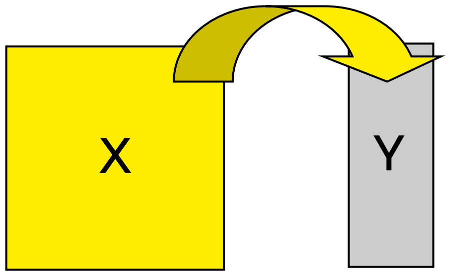
- Modelling a process output
  - Monitor Quality of final product

- Modelling Batch Processes
  - Majority of industrial processes
  - More complex analysis



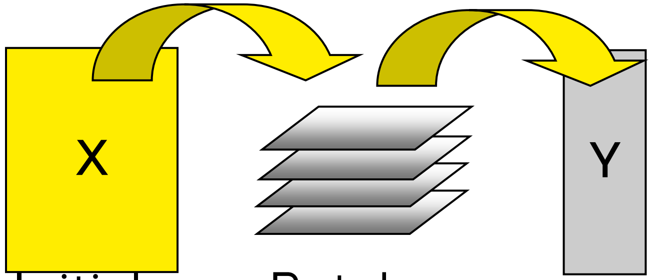
PCA

The process



PLS
OPLS

The process Quality



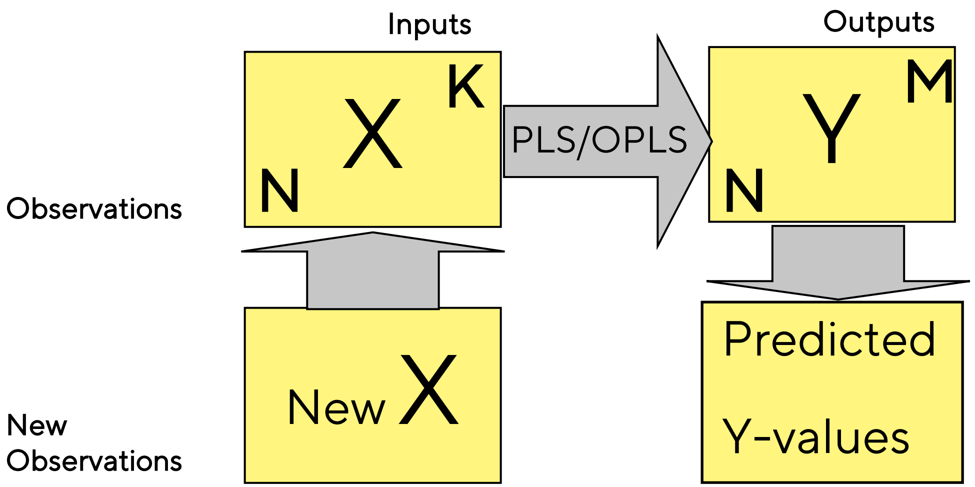
PCA
PLS
OPLS

Initial Conds
Batch Evolution
Quality

Schematic View of the Regression Problem

- Find relationships between sets of multivariate data X and Y

X - Inputs:
- Chemical measurements
- in Situ data
- Spectra
- Process data
- Chemical Composition data

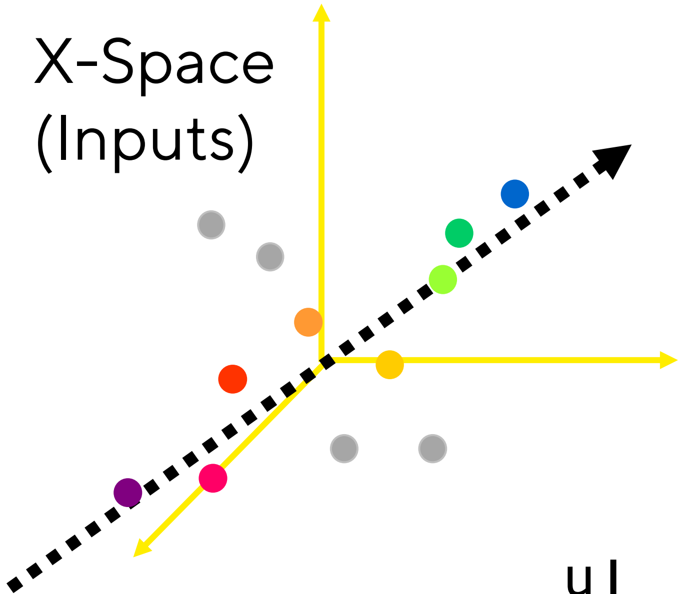


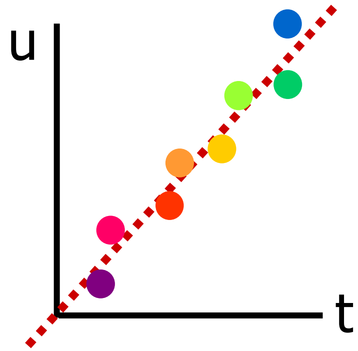
Y - Outputs:
- Yield
- Purity
- Quality
- Metrology
- Biological data

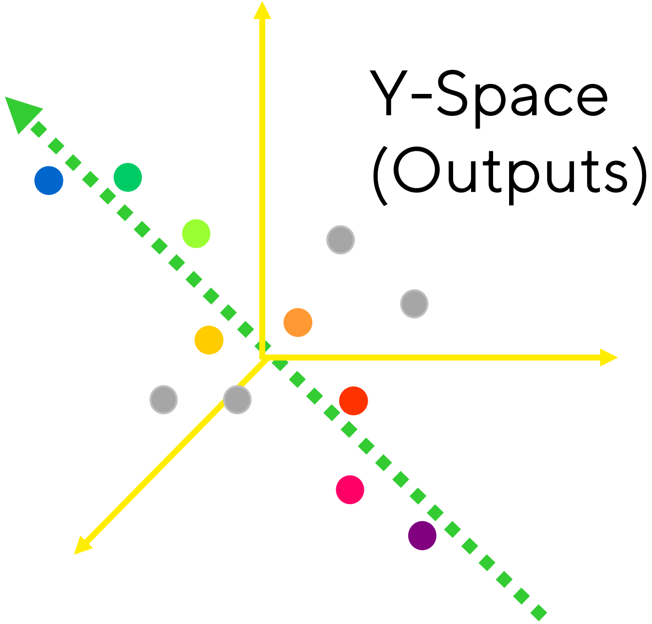
Introduction to PLS/OPLS

- Relationships between two blocks of data (often called X and Y) can be explored by regression extensions of PCA, i.e. PLS and OPLS
- PLS (= Partial Least Squares) was originated around 1975
  - Refined around 1982-83
- An improvement called OPLS (= Orthogonal PLS) was presented 2002
  - OPLS offers enhanced model interpretation
  - PLS and OPLS models for single-Y are identical wrt to prediction

Projection-Based Regression Modeling





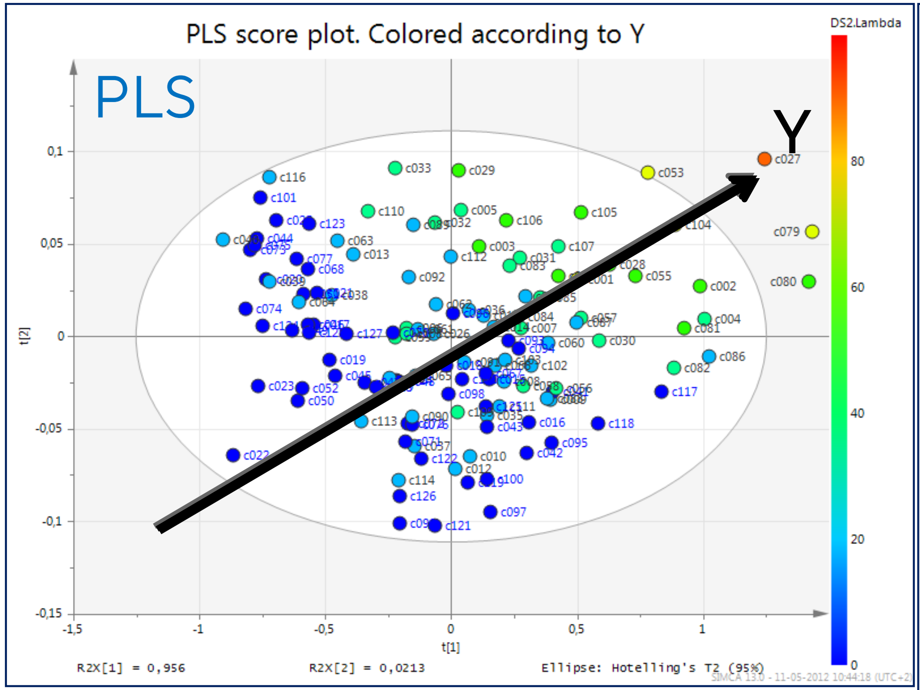


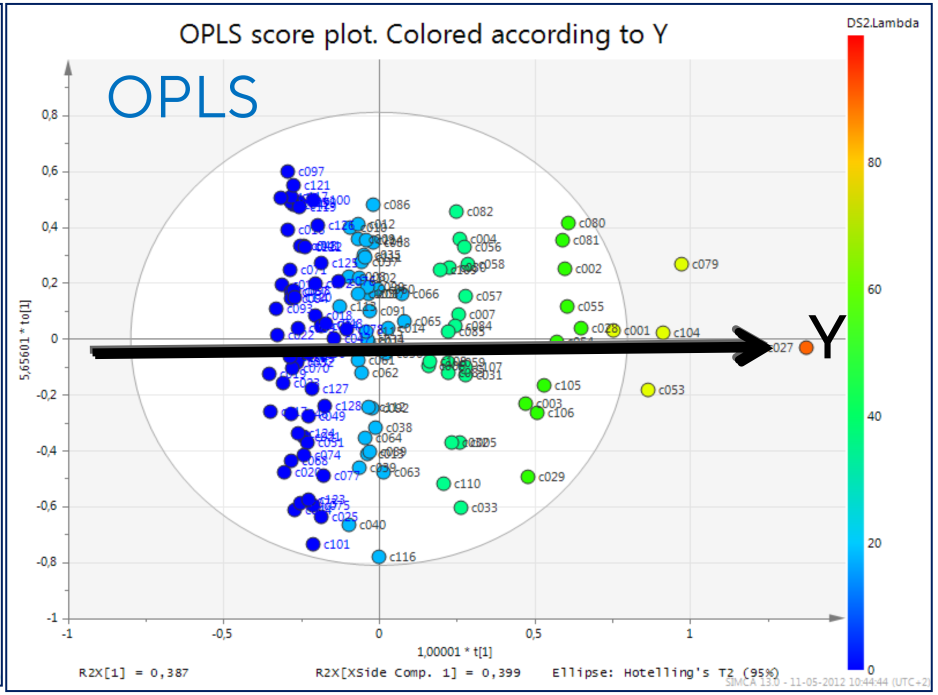
Finds directions (here: colors) in X which are predictive of directions in Y

PLS works well when colored structures dominate; when gray structures dominate OPLS is needed to filter it out

Orthogonal-PLS (OPLS)

- A rotation or "transformation" of the PLS solution
  - OPLS and PLS models with single-Y and same number of components are equivalent; same predictive power

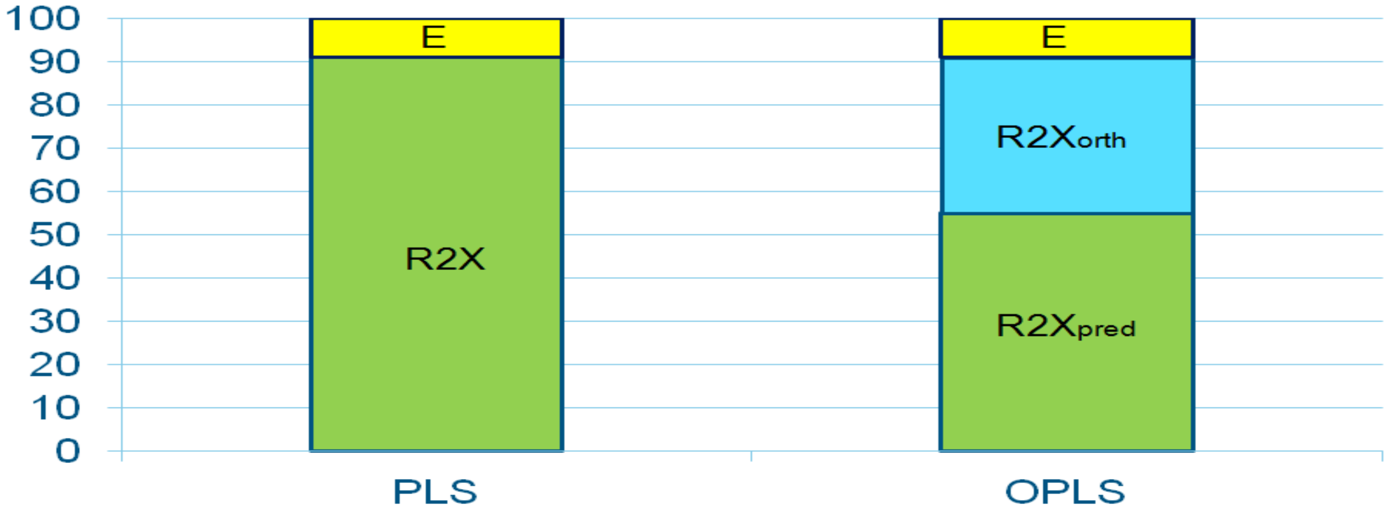




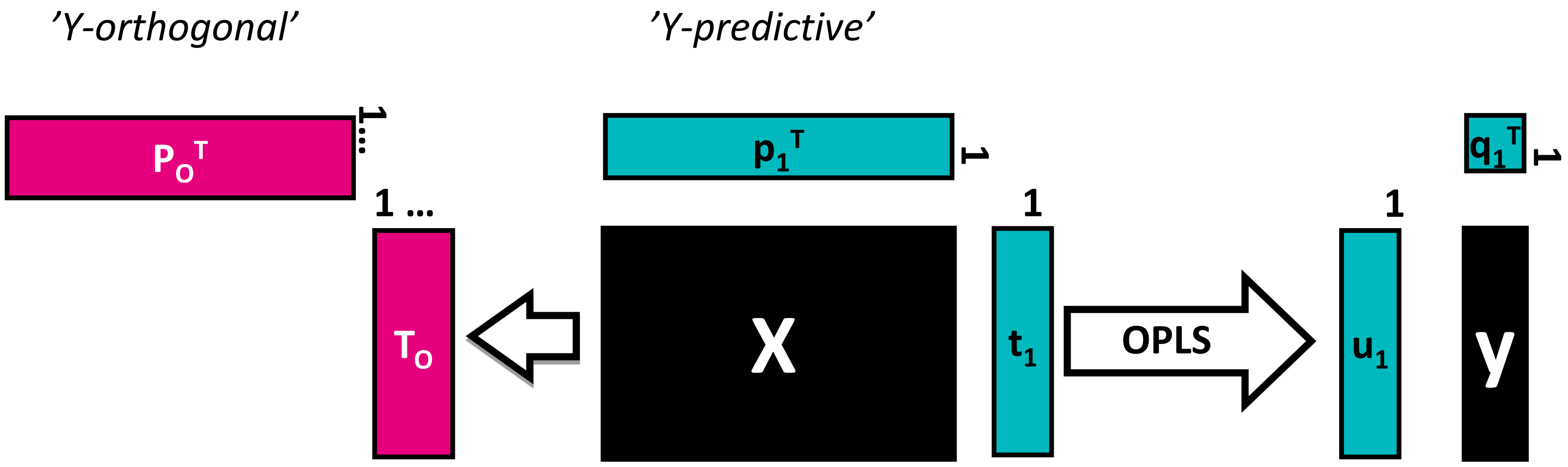
SARTORIUS

OPLS Terminology

- PLS divides the variability in the X-matrix in two parts, the systematic variability and the residual variability.
- OPLS further splits the systematic variability, R2X, in two parts, the part that is correlated (predictive) to Y and the part that is uncorrelated (orthogonal) to Y.



OPLS Model Structure (Single-Y)



OPLS Model

$$X = t_1 p_1^T + T_O P_O^T + E$$

$$Y = t_1 q^T_1 + F$$

Example: Binary Powder I

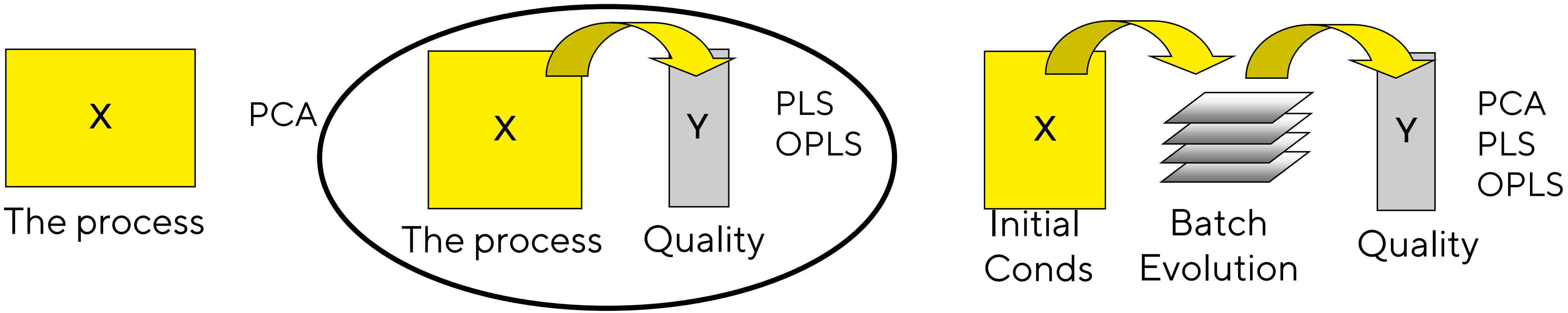
- Mixing two powders with similar particle sizes which represents a common situation in pharmaceutical production
  - The two model powders are lactose and salicylic acid.

- Sample mixtures were prepared using salicylic acid/lactose mixtures in the range 45/55% (w/w) to 55/45% (w/w).

- Main reference: O Berntsson, LG Danielsson, MO Johansson, and S Folestad, Quantitative determination of content in binary powder mixtures using diffuse reflectance near infrared spectrometry and multivariate data analysis. Analytica Chimica Acta, 419 (2000) 45-54.

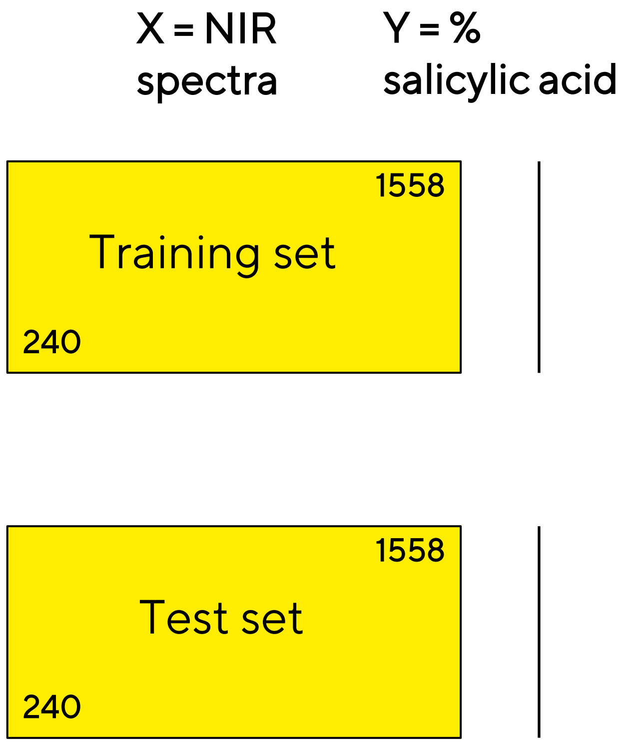
Multivariate Process Data

- Monitoring a process
  - Early warning of disturbances
  - Diagnostics - finding ”assignable causes”
- Modelling a process output
  - Monitor Quality of final product
- Modelling Batch Processes
  - Majority of industrial processes
  - More complex analysis



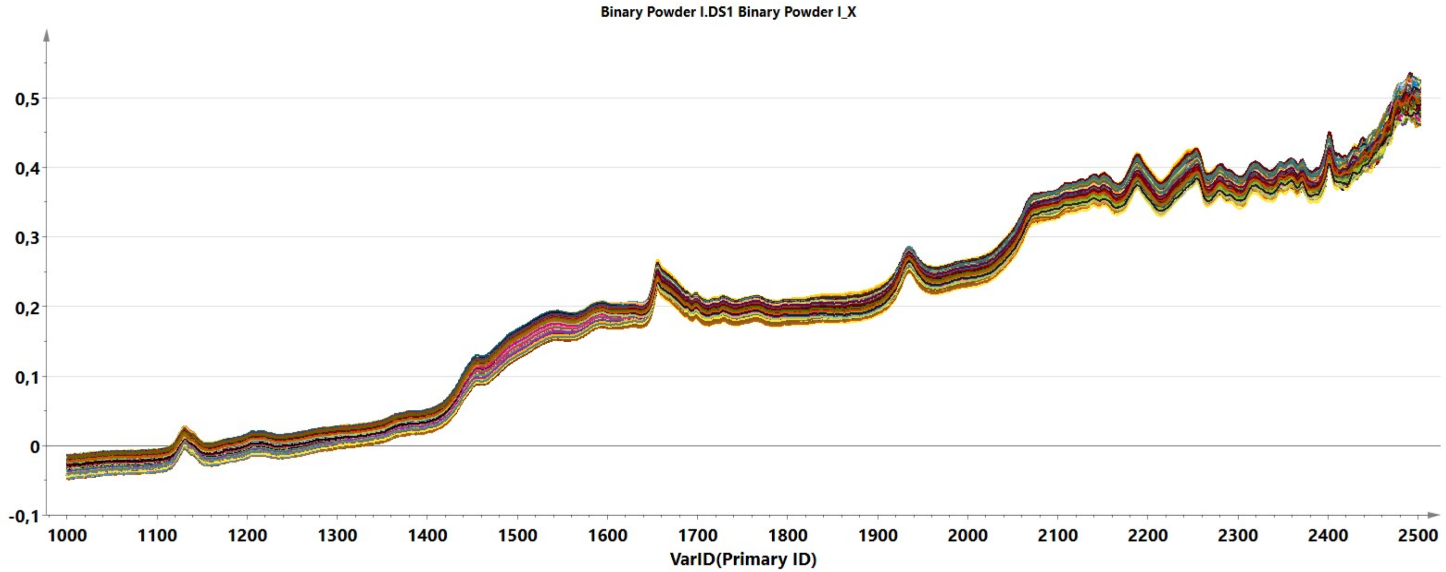
Data

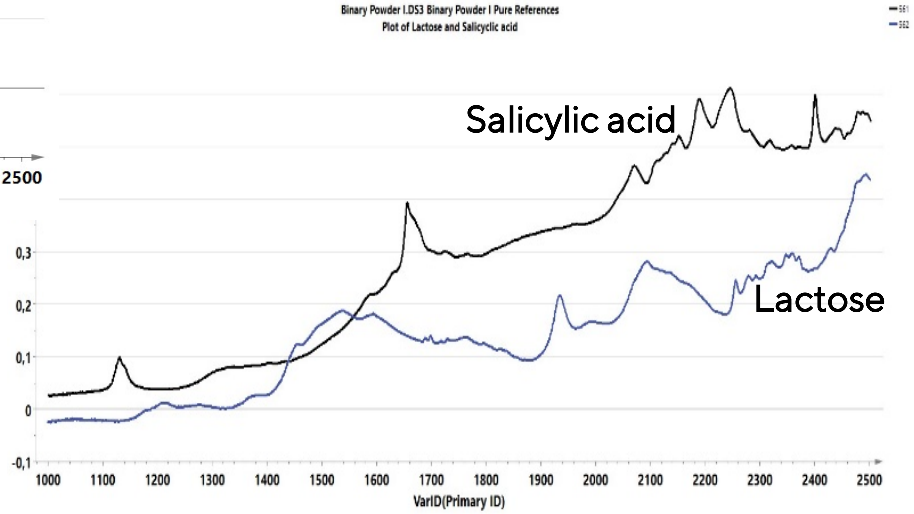
- The data consists of 12 batches of powders with salicylic acid content ranging from 45% to 55%.
  - 6 batches used as training set and 6 batches used as test set
- For each batch, 40 NIR spectra were acquired by holding the fibre-optic probe against the powder surface in 40 randomly chosen positions.
  - 12*40 = 480 observations; divided 240/240
- NIR spectra were obtained as log (1/R) in the 1000-2500nm range giving a total of 1558 X-variables.



SARTORIUS

Spectra of Training Set Samples





SARTORIUS

OPLS Model

- The PLS and OPLS models have identical performance
- However, the additional insight provided by the OPLS model is that only 22% of the variation in the NIR data is connected to the variation in the levels of salicyclic acid
- 77% of NIR variance systematic but not predictive to Y

Binary Powder I - M4

Workset... Options... Title: OPLS odd batches

Type: OPLS-Class(1) Observations (N)=240, variables (K)=1559 (X=1558, Y=1)

Component	R2X	R2X(cum)	Eigenvalue	R2	R2(cum)	Q2	Limit	Q2(cum)	R2Y	R2Y(cum)	EigenvalueY	Significance
Model		0,989			0,972			0,956		1		
Predictive		0,22			0,972			0,956		1		
P1	0,22	0,22	52,8	0,972	0,972	0,956	0,01	0,956	1	1	1	R1
Orthogonal in X(OP...		0,769			0							
O1	0,6...	0,699	168	0	0							R1
O2	0,0...	0,743	10,4	0	0							R1
O3	0,0...	0,753	2,52	0	0							R1
O4	0,0...	0,761	1,78	0	0							R1
O5	0,0...	0,765	0,959	0	0							R1
O6	0,0...	0,769	0,945	0	0							R1

Binary Powder I - M2

Workset... Options... Title: PLS odd batches

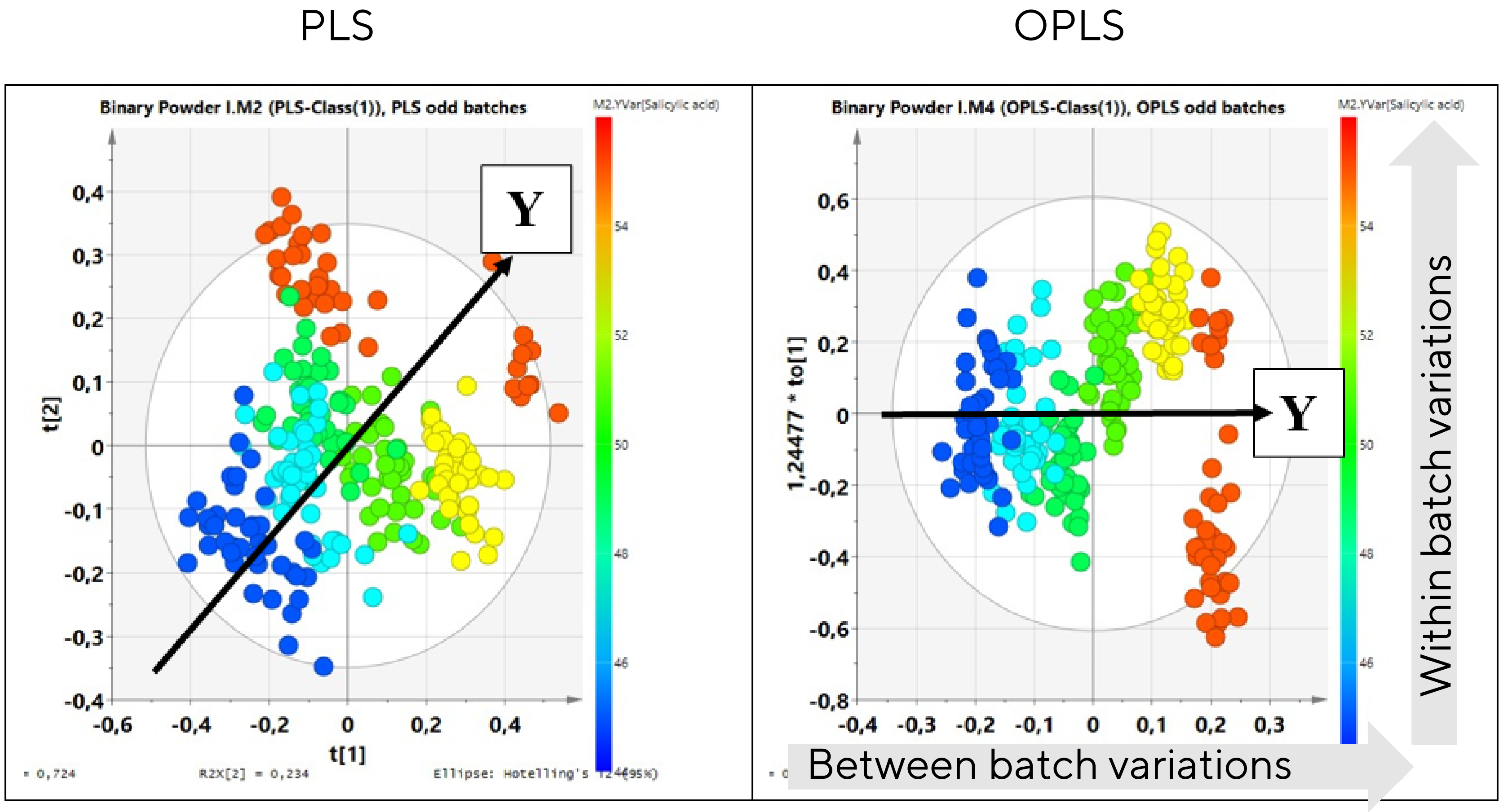
Type: PLS-Class(1) Observations (N)=240, variables (K)=1559 (X=1558, Y=1)

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)	Significance	Iterations
0	Cent.									
1	0,724	0,724	174	0,425	0,425	0,422	0	0,422	R1	1
2	0,234	0,958	56,2	0,406	0,831	0,705	0	0,829	R1	1
3	0,0218	0,98	5,22	0,0698	0,901	0,402	0	0,898	R1	1
4	0,0026	0,983	0,624	0,0311	0,932	0,195	0	0,918	R1	1
5	0,004...	0,987	1,03	0,0127	0,945	0,155	0	0,931	R1	1
6	0,001...	0,988	0,279	0,011[illegible]	0,957	0,09...	0	0,937	R1	1
7	0,000...	0,989	0,106	0,015[illegible]	0,972	0,121	0	0,945	R1	1

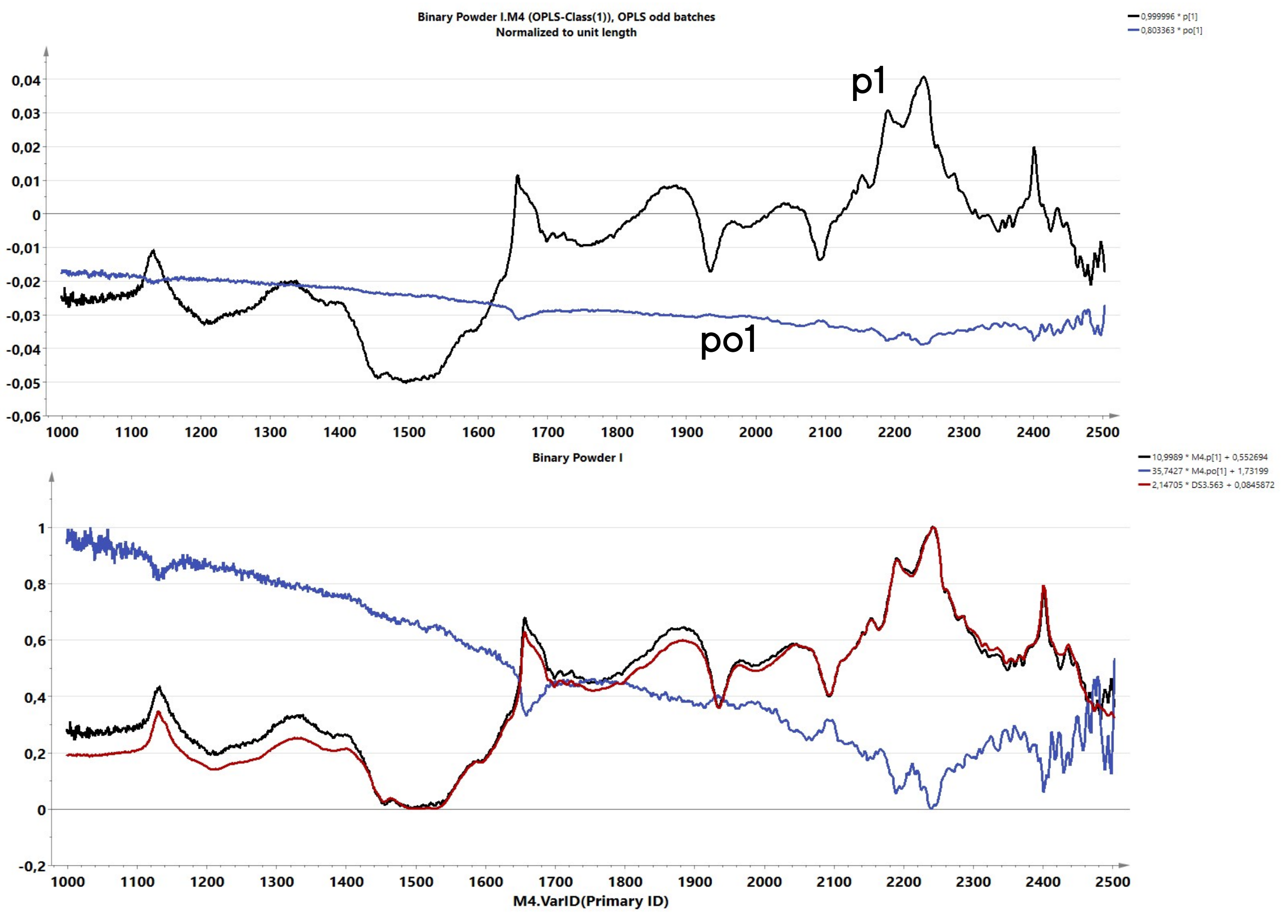
SARTORIUS

Rotation of Projection Towards Y

-



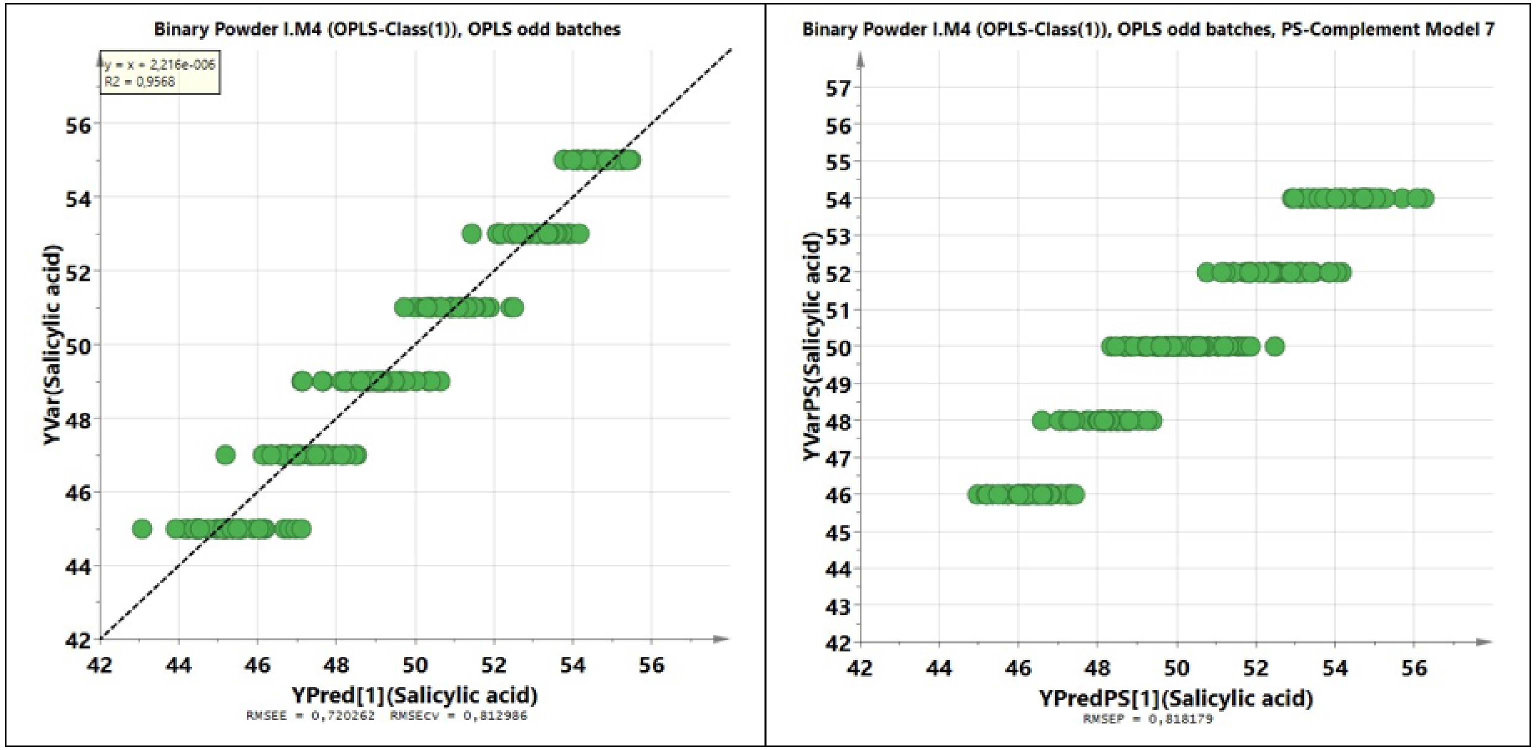
Interpretation of Predictive and Orthogonal Components (No Mixing)



SARTORIUS

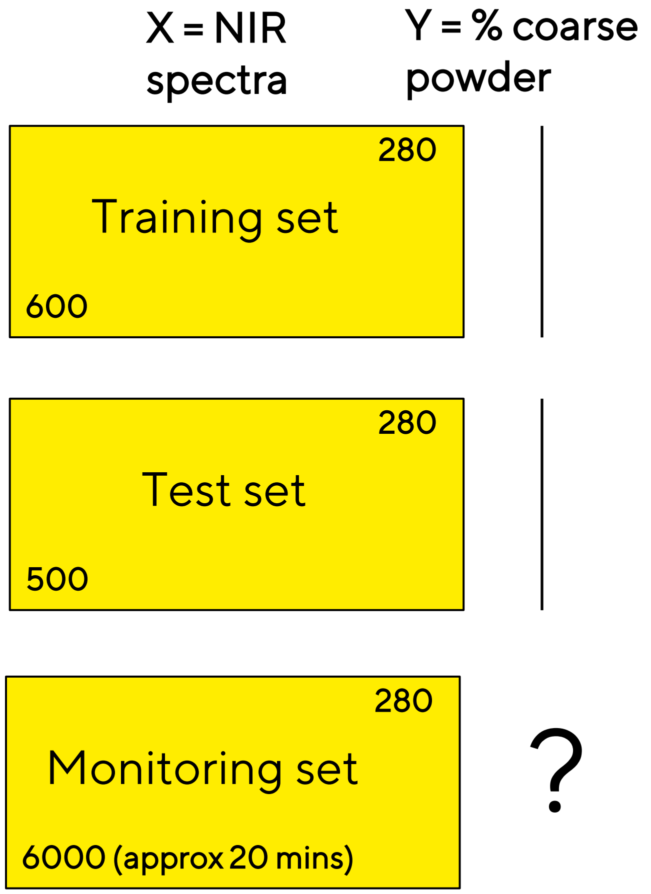
Prediction For Test Set

- Good agreement between fit and prediction measures
- RMSEE = 0.72, RMSEcv = 0.81, RMSEP = 0.82.



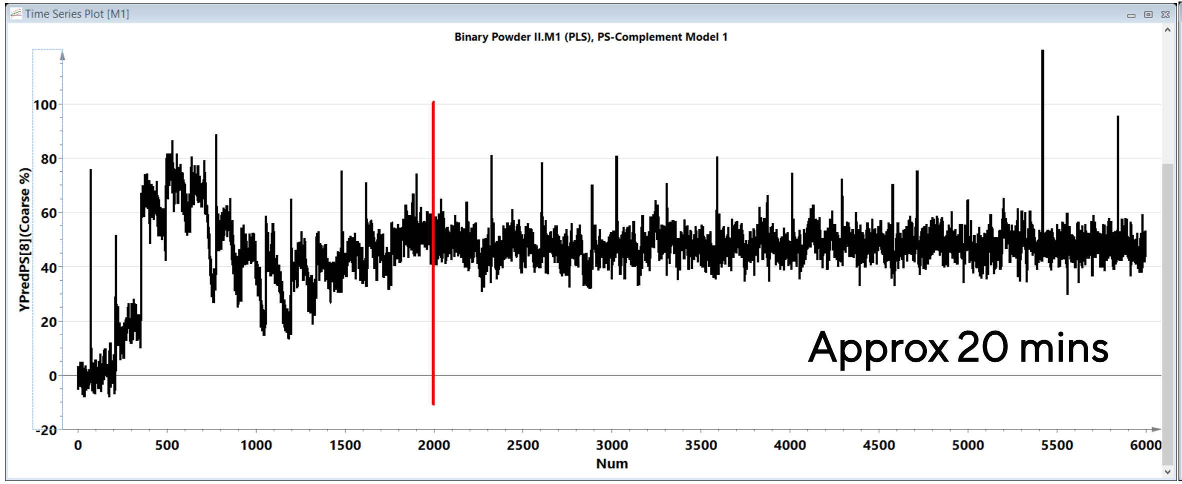
Example: Binary Powder II

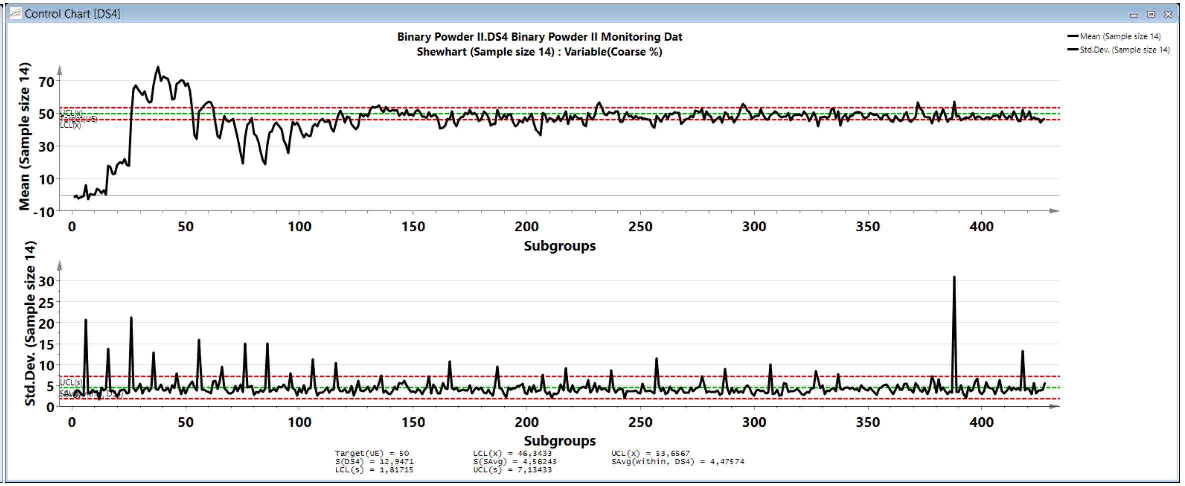
- Mixing two powders with dissimilar particle sizes
  - 0 - 100% coarse powder
- Training set 6*100 = 600 spectra
- Test set 5*100 = 500 spectra
- Monitoring set = 6000 spectra
- Spectral range: 1082-2025 nm
  (➔ 280 variables)
- Scenario: In the monitoring experiment, equal amounts of the coarse and fine powders were loaded into a vertical cone mixer with the coarse powder on top.
- Question: When is mixing complete?



When Is Mixing Complete?

- Mixing complete about 1/3 into the sampling period
- Effect of orbiting screw seen every 140th spectrum





Batch Example: Hydrogenation Reaction

- PLS and OPLS are available for BEM and BLM
- PLS maximizes covariance between X and Y; risk is that undesired X-variability is picked up by BEM in terms of more components (more control charts to monitor)
- OPLS divides systematic SS of X in two parts, pred and orth; this may affect trajectory estimations and confidence intervals

Analytica Chimica Acta 857 (2015) 28–38

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

ELSEVIER

OPLS in batch monitoring – Opens up new opportunities

CrossMark

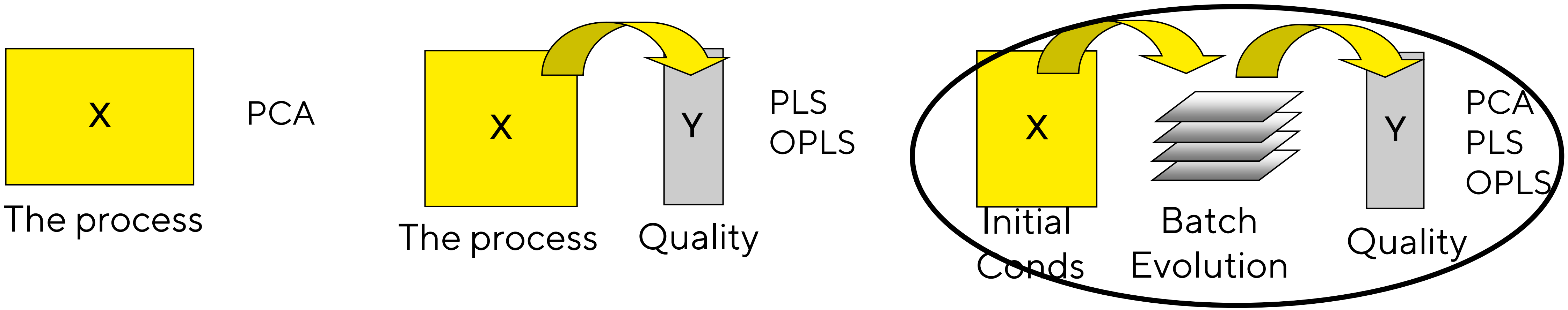
Nabil Souihi [a], Anders Lindegren [b], Lennart Eriksson [b], Johan Trygg [a,*]

[a] Computational Life Science Cluster (CLiC), Department of Chemistry, Umeå University, Sweden
[b] MKS Umetrics, Umeå, Sweden

SARTORIUS

Multivariate Process Data

- Monitoring a process
  - Early warning of disturbances
  - Diagnostics - finding "assignable causes"
- Modelling a process output
  - Monitor Quality of final product
- Modelling Batch Processes
  - Majority of industrial processes
  - More complex analysis



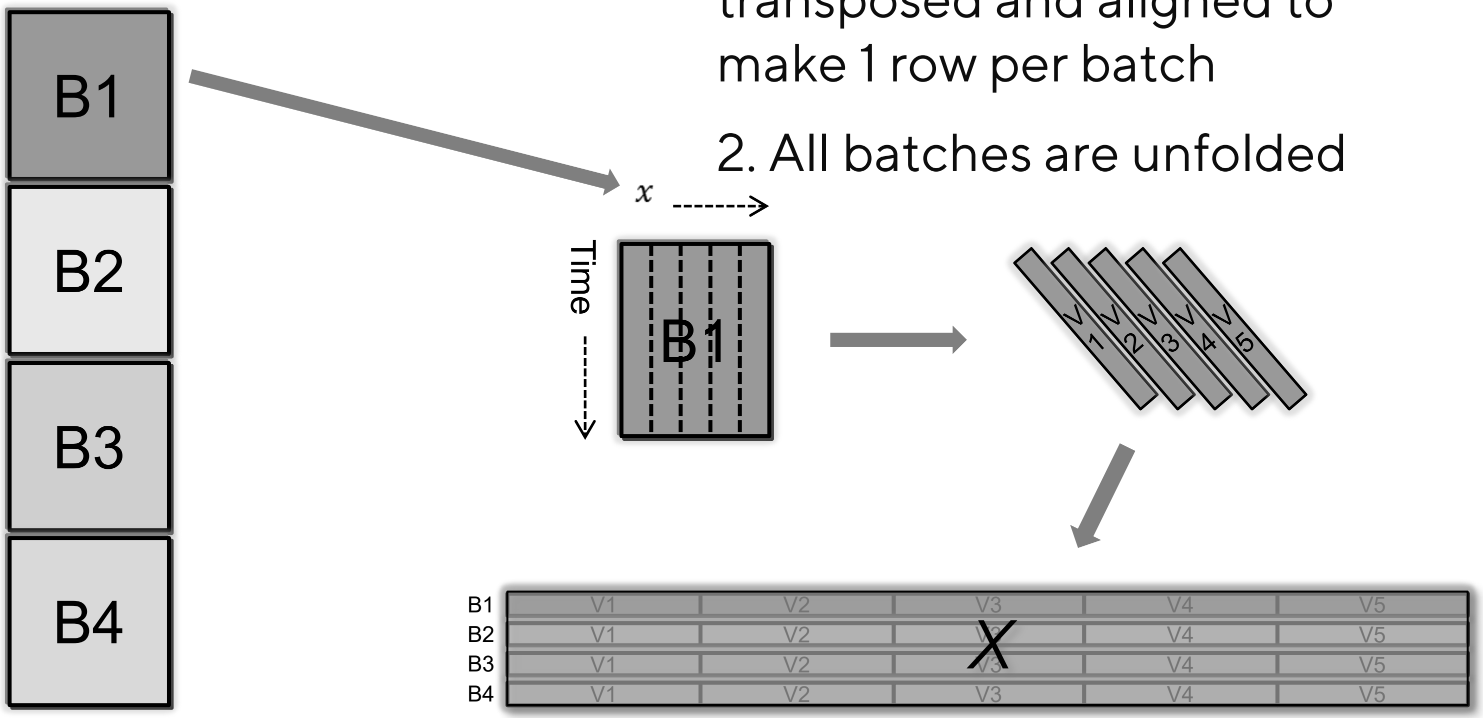
Batch Data; BEM and BLM

- Batch data has time dependency
  - A table of data is generated for each batch
  - Variables measured over time

Batch-wise Unfolding

1. Each variable is transposed and aligned to make 1 row per batch

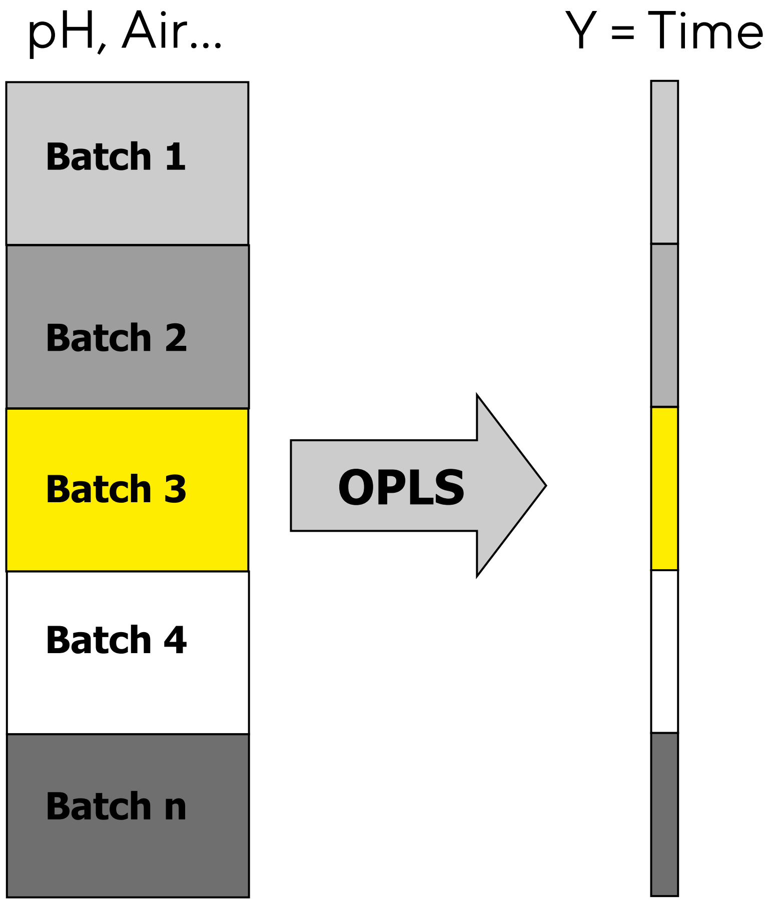
2. All batches are unfolded



SARTORIUS

The Batch Evolution Model (BEM)

- Time (or maturity) is used as a Y-variable to give the model a direction
- Maturity need not be time. It could be say, for example, be Ethanol in beer brewing



Example: Hydrogenation Reaction

- 6 centerpoint batches, NOC batches, used for model training
- 5 additional batches used for model testing
  - 3 corners of DOE (one corner missing)
  - 2 PD, process disturbances, batches (process upsets induced)
- 87 variables, 80 spectral (UV 200-300 nm, 1st derivative) and 7 process variables
  - reactor temp, reactor pressure, gas feed, jacket-in temp, jacket-out temp, flow rate of oil, and stirrer speed

J. Gabrielsson, H. Jonsson, J. Trygg, C. Airiau, B. Schmidt, R. Escott, AIChE J. 52 (2006) 3164–3172.

Example: Hydrogenation Reaction (Nitrobenzene to Aniline)

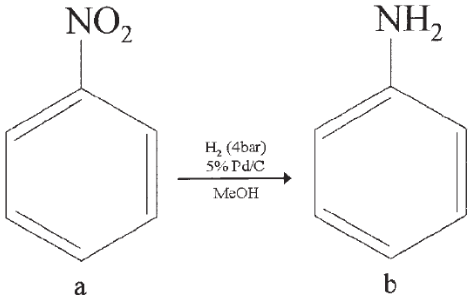


Figure 2. Reaction scheme for the conversion of Nitrobenzene (a) to Aniline (b).

DOE

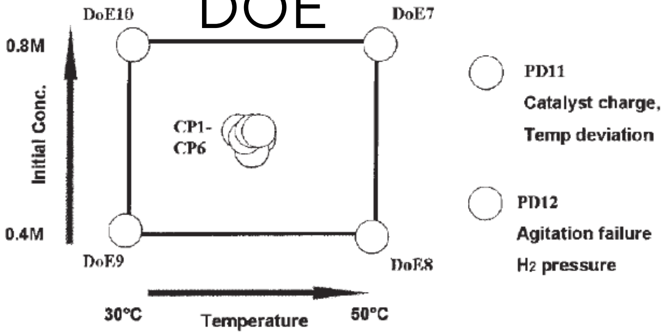


Figure 1. Illustration of the 2^2 full factorial design that was implemented in the study.

X data
- 1st der UV data (200-300 nm)
- Process data

1. Reactor temperature,
2. Reactor pressure
3. gas feed

4-7 jacket temp and stirrer speed

Table 1. Factor Settings for the 2^2 Full Factorial Experimental Design for the Conversion of Nitrobenzene to Aniline

Batch Name (abbreviation)	Reaction Temperature (°C)	Initial Conc. (mol/L)	Duration (min)
Center point 1 (CP1)	40	0.6	45.7
Center point 2 (CP2)	40	0.6	39.1
Center point 3 (CP3)	40	0.6	37.8
Center point 4 (CP4)	40	0.6	37.9
Center point 5 (CP5)	40	0.6	38.9
Center point 6 (CP6)	40	0.6	37.3
Experiment 7 (DoE7)	50	0.8	24.1
Experiment 8 (DoE8)	50	0.4	—
Experiment 9 (DoE9)	30	0.4	66.1
Experiment 10 (DoE10)	30	0.8	45.4
Process deviation 11 (PD11)	40	0.6	47.6
Process deviation 12 (PD12)	40	0.6	33.9

Included also are six center points and two batches with introduced process deviations (explained in detail in Table 2). The resulting duration of each batch is given in min.

SARTORIUS

BEM Results

- 3 PLS components
- 1+2 OPLS components
  - R2Xpred = 65%
  - R2Xorth = 19%,
  - Xres = 16%

Hydrogenation_SIMCA 14 - M1

Workset... Options... Title: Untitled

Type: PLS Observations (N)=1188, variables (K)=92 (X=91, Y=1), included batches: 6

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)	Significance	Iterations
0	Cent.									
1	0,669	0,669	60,9	0,93	0,93	0,93	0	0,93	RB1	1
2	0,0578	0,727	5,26	0,0227	0,953	0,325	0	0,953	RB1	1
3	0,111	0,838	10,1	0,003..	0,956	0,06...	0	0,956	RB1	1

Hydrogenation_SIMCA 14 - M2

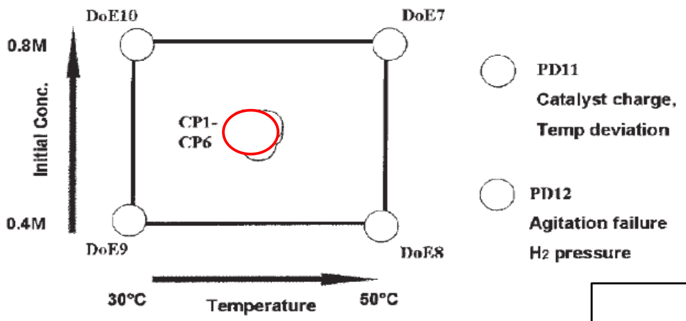
Workset... Options... Title: Untitled

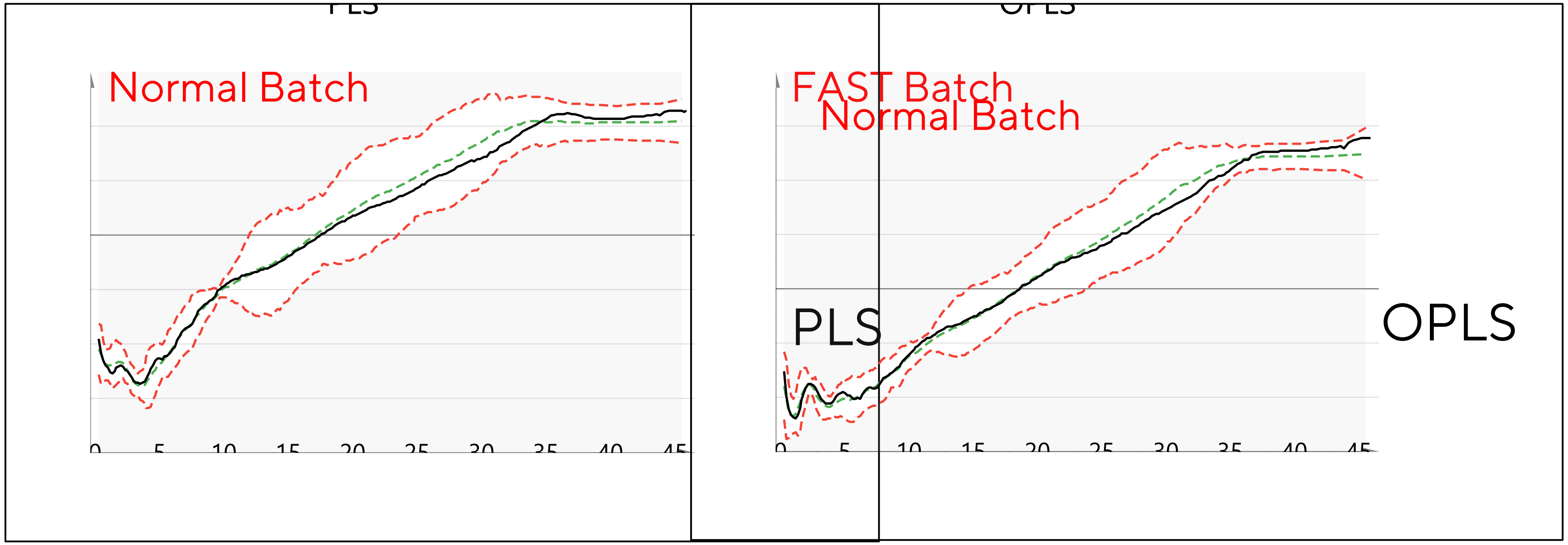
Type: OPLS Observations (N)=1188, variables (K)=92 (X=91, Y=1), included batches: 6

Component	R2X	R2X(cum)	Eigenvalue	R2	R2(cum)	Q2	Limit	Q2(cum)	R2Y	R2Y(cum)	EigenvalueY	Significance
Model		0,838			0,956			0,956		1		
Predictive		0,65			0,956			0,956		1		
P1	0,65	0,65	59,2	0,956	0,956	0,956	0,01	0,956	1	1	1	R1
Orthogonal in X(OP...		0,188			0							
O1	0,0...	0,0747	6,8	0	0							R1
O2	0,1...	0,188	10,3	0	0							NS

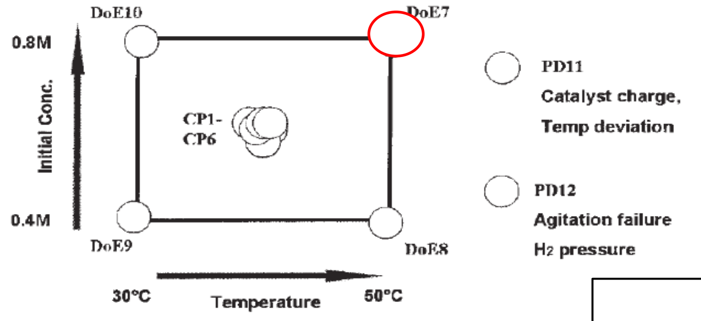
SARTORIUS

Control Charts Representing NOC





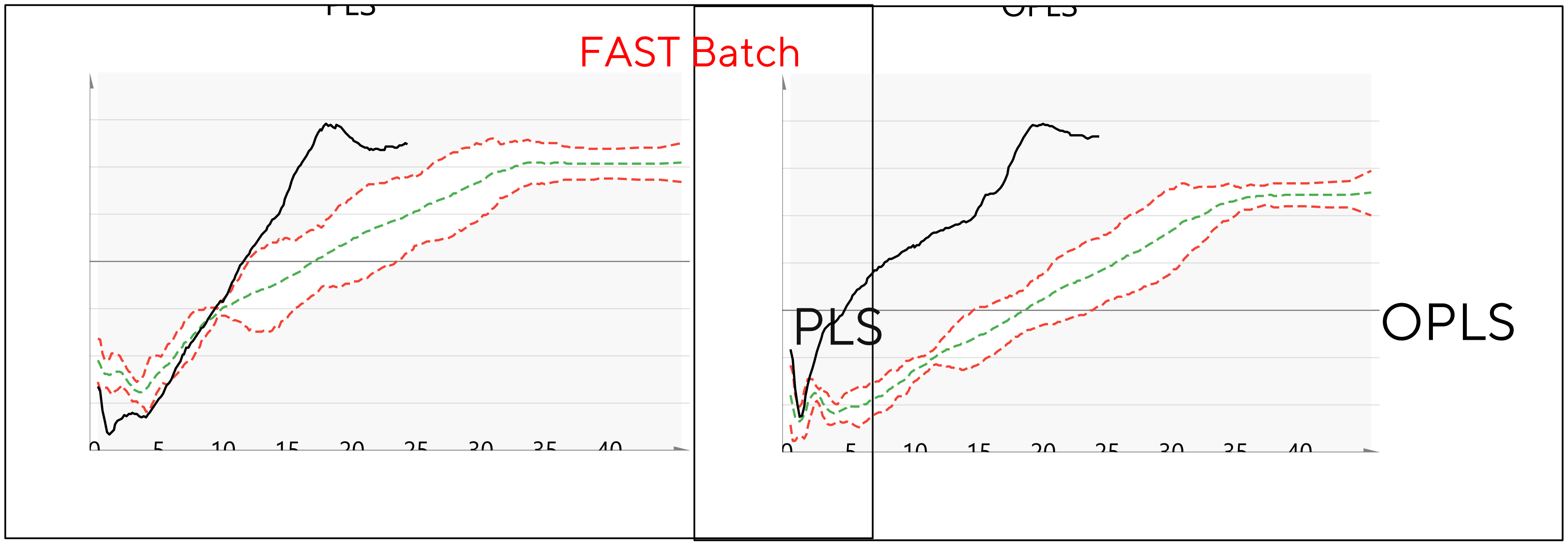
Prediction of Batch DOE7 (tps1) – High Temp / High Initial Conc.



Reaction fast and high amount of product

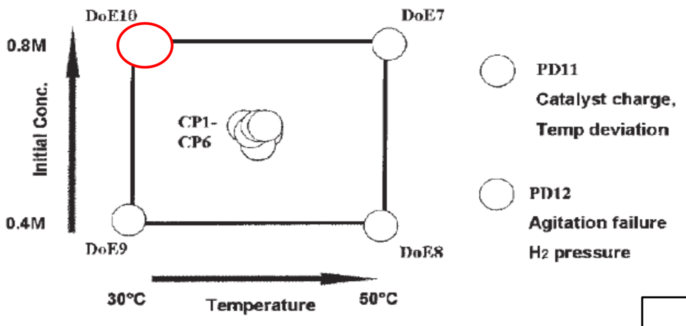
OPLS consistently shows this in score plot

PLS not as sharp and capable



SARTORIUS

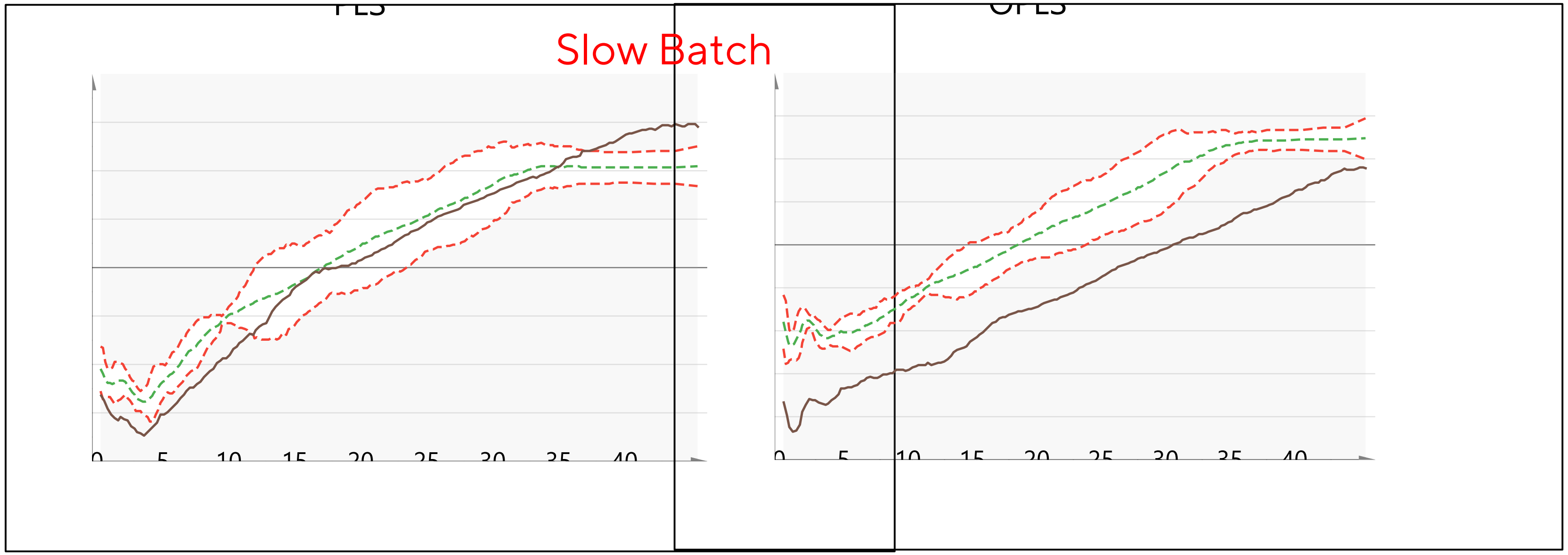
Prediction of Batch DOE10 (tps1) – Low Temp / High Initial Conc.



Reaction slow and low amount of product

OPLS consistently shows this in score plot

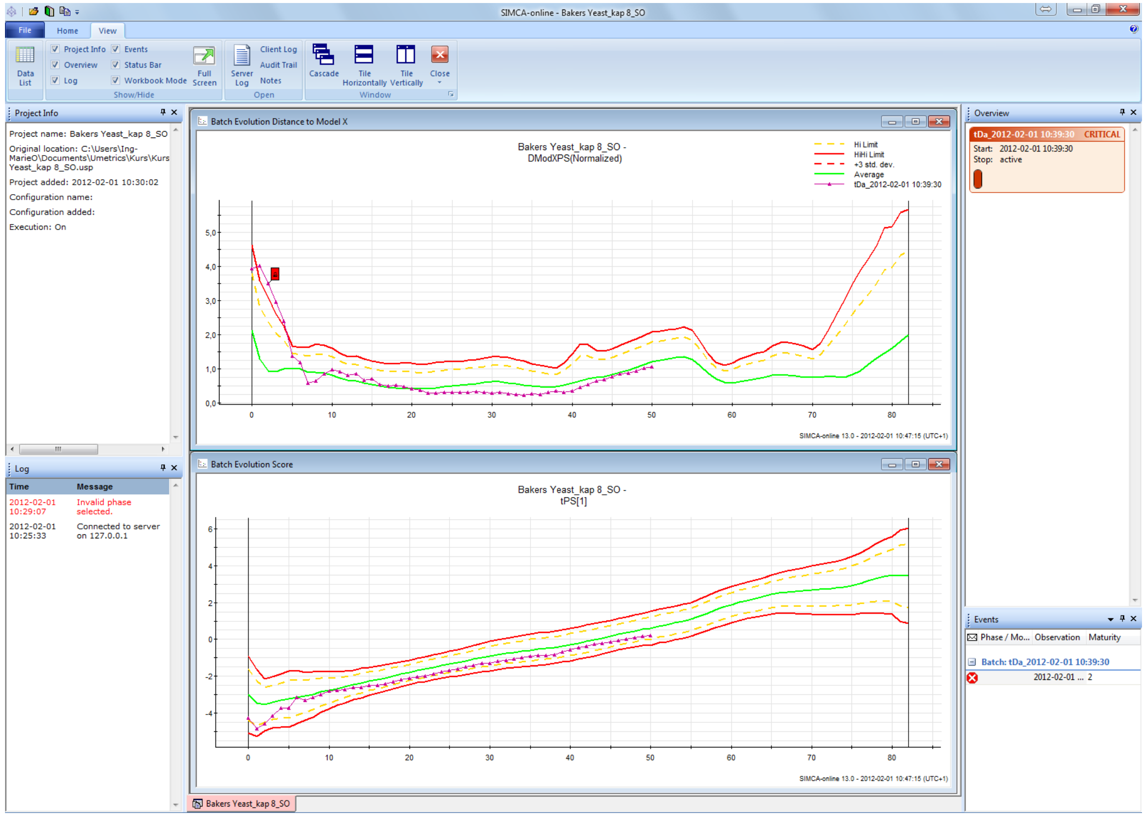
PLS not as sharp and capable



SARTORIUS

SIMCA®-online

- Once the model is created and validated it is ready for online deployment



Demo

Conclusions OPLS

- A single-Y PLS and a single-Y OPLS model with the same number of components are mathematically equivalent.
- Predictions, residuals, R2, R2 per variable, DModX, etc, are the same for PLS and OPLS.
- Loadings and scores are different for PLS compared to OPLS.
- Interpretation is easier with OPLS, since the user has a clearer understanding of what the different components mean

Upcoming Webinars (https://www.sartorius.com/en/company/exhibition-conferences)

UPCOMING EVENTS PAST EVENTS CALENDAR SCHEDULE

Search

STANDARD
Design of Experiments (DOE) for the Beginner
TUE, JAN 26, 2021 03:00 PM - 04:00 PM CET

STANDARD
Multivariate Data Analysis (MVDA) for the Beginner
THU, JAN 28, 2021 03:00 PM - 04:00 PM CET

STANDARD
Lean-and-clean DOE using One-click analysis
TUE, FEB 16, 2021 03:00 PM - 04:00 PM CET

STANDARD
OPLS® in process modeling
THU, FEB 18, 2021 03:00 PM - 04:00 PM CET

STANDARD
Robust optimization made easy
TUE, MAR 2, 2021 03:00 PM - 04:00 PM CET

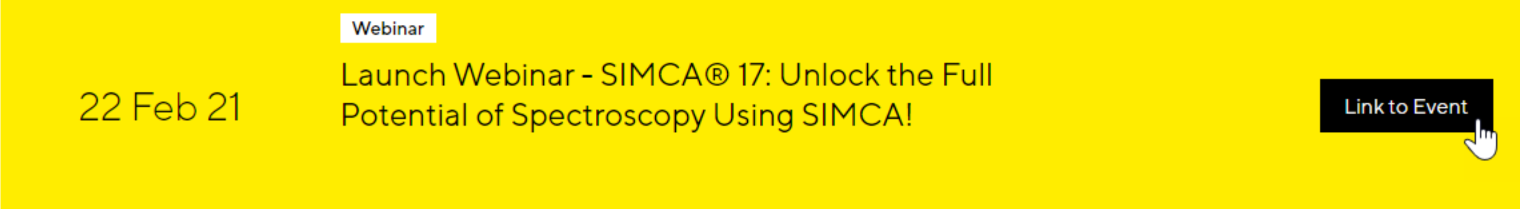
STANDARD
Analyzing batch process data, a step-by-step guide
THU, MAR 4, 2021 03:00 PM - 04:00 PM CET

STANDARD
From Design of Experiments to Design Space Estimation
TUE, MAR 23, 2021 03:00 PM - 04:00 PM CET

STANDARD
Multiblock Orthogonal Component Analysis (MOCA) – A Novel Tool for Data Integration
THU, MAR 25, 2021 03:00 PM - 04:00 PM CET

Launch Webinar – SIMCA® 17

- https://www.sartorius.com/en/company/exhibition-conferences



SARTORIUS