Boehringer Ingelheim






2018 PI WORLD LIFE SCIENCES
A CASE STUDY:
BIOPHARMACEUTICAL DATA ANALYTICS
USING ASSET FRAMEWORK AND EVENT FRAMES FOR MVDA

WHAT IF WE COULD

In biopharmaceutical vaccine manufacturing; what if we:

Created *data-driven* *performance-based* objectives?

What if we:

Aligned our process control strategies to compliance *and* business performance objectives

What would it look like?

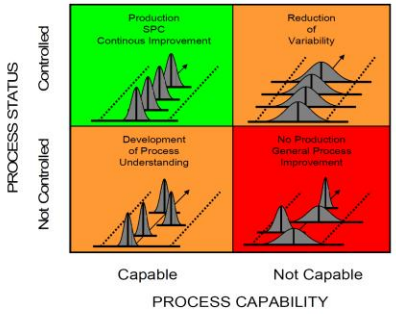
- reductions in production variances…
- business intelligence dashboards (Cycle-times, KPI's, Cost of Goods, Absorptions)…
- mvda characterization (golden batch approach)…
- batch evolution trajectories and prescriptive process controls…
- criticality classifications based on these statistical properties.

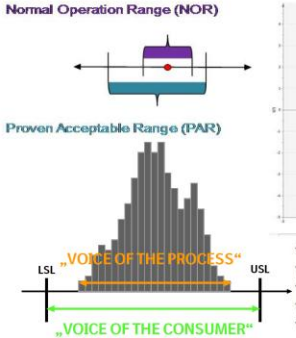
How can we:

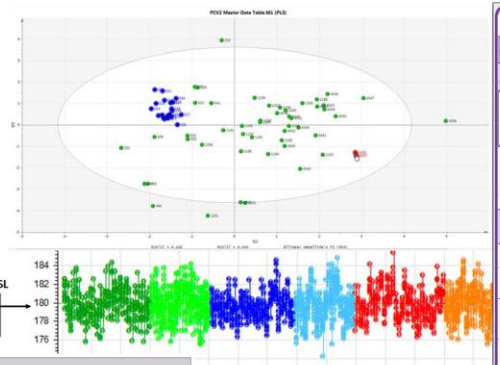
Use Middleware databases and OSI PI Asset Framework , Event Frames, and a series of analytical engines (statistics software and visualizations) to enable it?

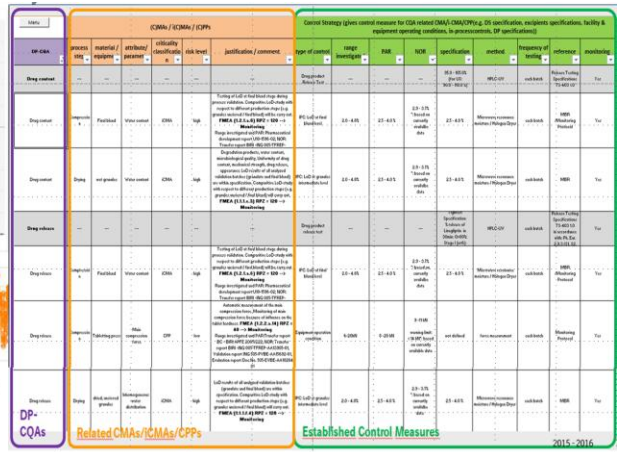
PRIMARY OBJECTIVE: Regulatory Compliance

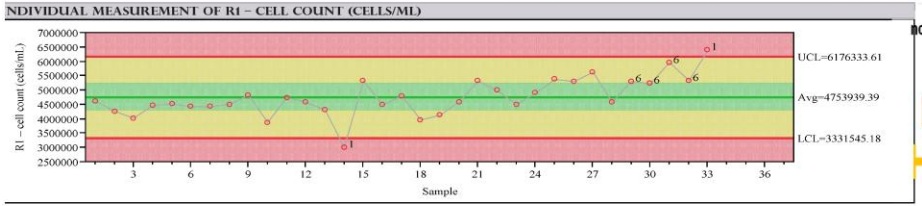
Statisitcal Process Behavior, Control and Capability

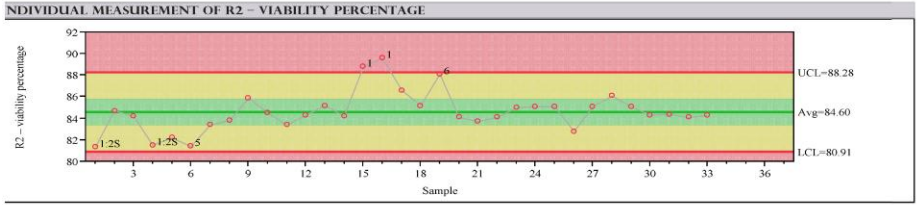


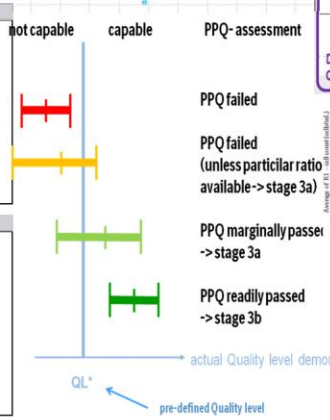


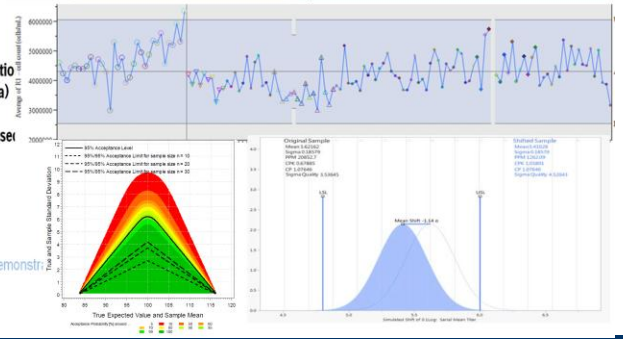






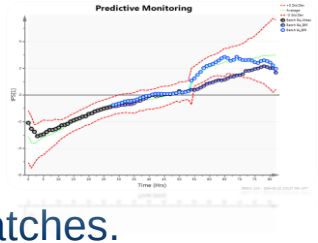






ENHANCED OBJECTIVES

Model functional relationships in
Fx {Organism Biochemistry &Processing Environment &Analytical Technology}



Multivariate statistical characterization and real time monitoring

- Model production variances (high vs. low potency batches)
- fault detection ID and in-batch parametric adjustment
- Prescriptive process control to target high potency trajectory

Value to the business

- Optimized and consistent Yield, less destruction, fewer low performing batches.
- Process predictability, characterize and prescribe controls for efficient use of materials, lower production costs better resource utilizations, cycle times, RFT

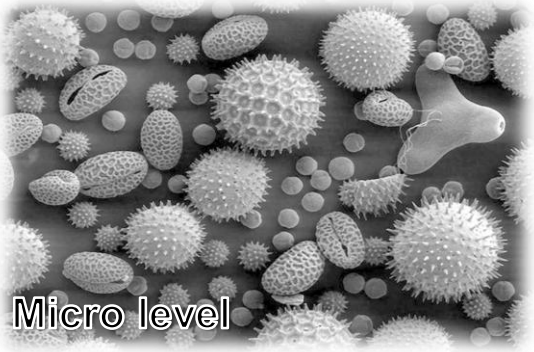
WHAT IF WE COULD

what if we could aggregate process and result data in batch events

SCADA tag (time resolved vectors, process batch)

… and batch conditions (critical parameters and material attributes)

with … response data (targets, potencies, yields)!!



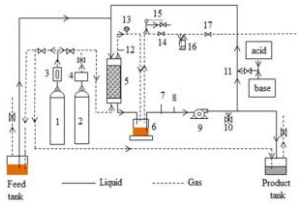
Architecture:

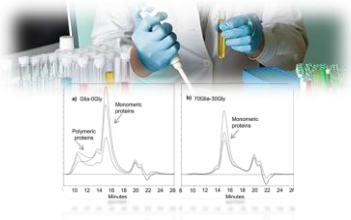
a middleware data-layer (ETL Warehouse and OSI PI) with batch and result contextualization

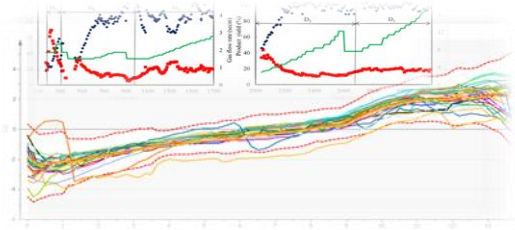
Programs:

1) Asset Framework combines tag/point data from SCADA with attribute and responses from SAP

2) ISA 88 ‘Unit Operation and Phase’ Event Frames Parent (Batch Event) to Child (Phase Steps)





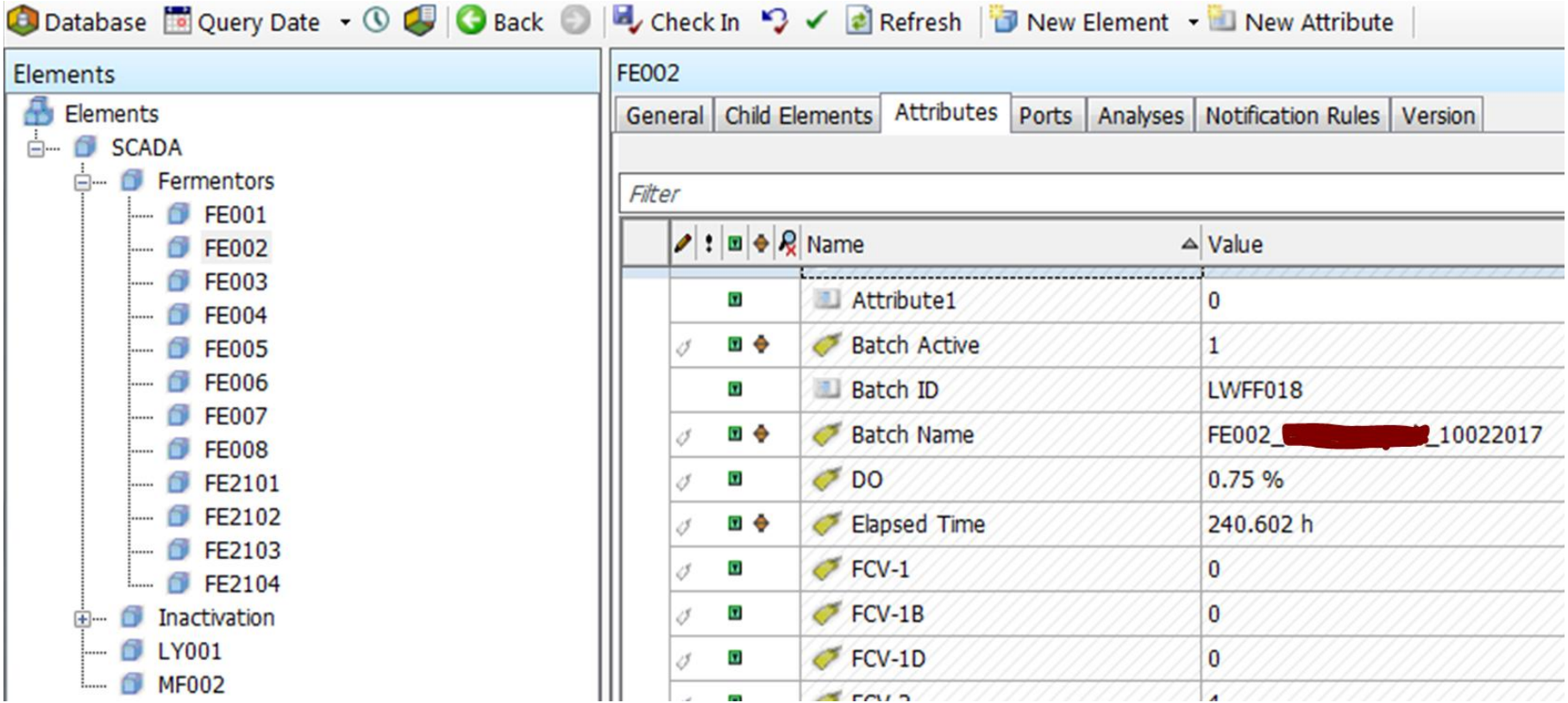


HOW WOULD WE CREATE:

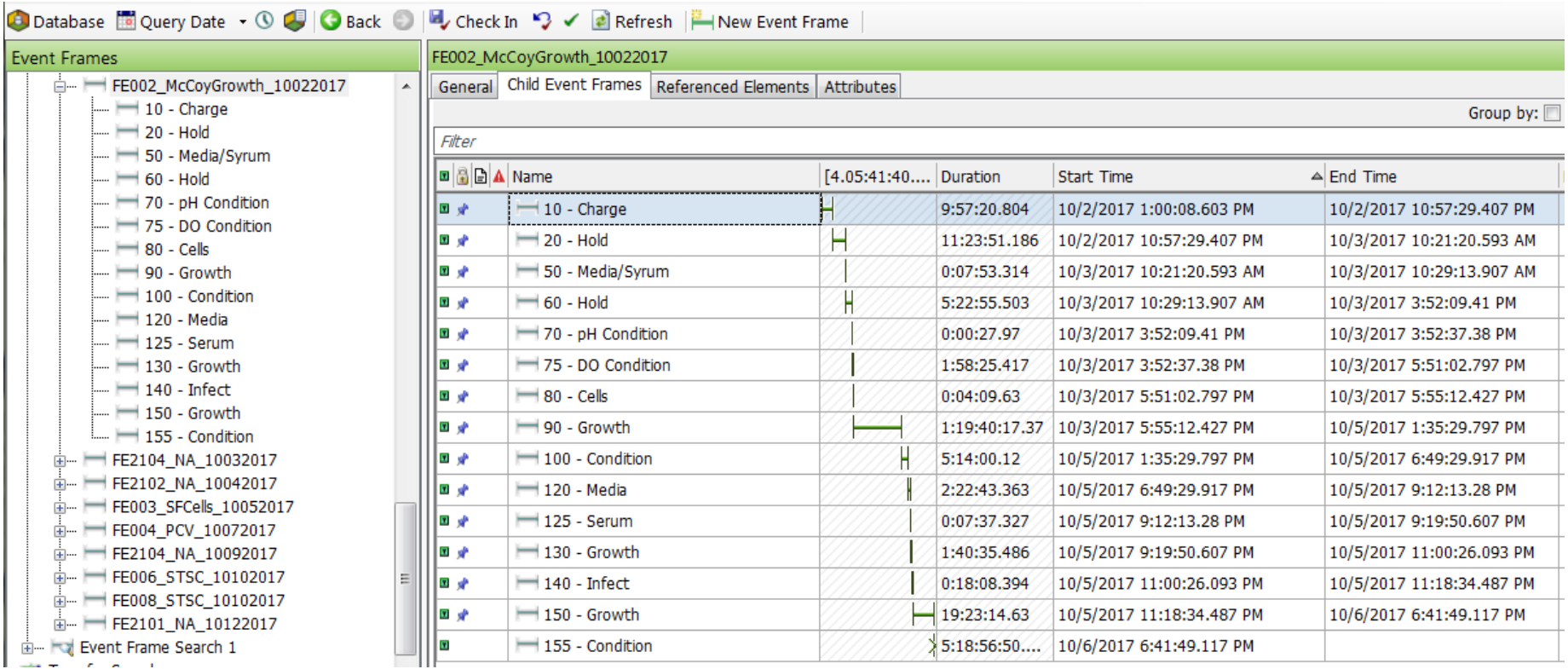
"Batch-contextualized" data without MES and EBR :

Objective:	Current Status
DATA AGGREGATION	Triggers programmed into PLC using standard ISA 88 language to differentiate Batch ID, Unit Operations/Procedures, and Process Stages/Phases. Facilitate wrapping of time-resolved vector data for analysis and comparison.
ASSET FRAMEWORK	Using Primary Elements (Fermenters) from each manufacturing suite, all associated tags are built as PI points and given data type. Once complete for an element, class-based templates are created to roll out Asset Framework for all equipment types according to this class. These templates ensure consistent configuration and provide role specific data flow to event framework.
EVENT FRAMEWORK	PI Event Framework utilizes Asset Framework and available attribute data (SAP, QM, Offline) and wraps the data into Parent-Child Events. This accomplishes acquisition of analysis ready data sets and visualization. Event frames are also integrated to SIMCA Online for Multivariate Data Analysis (MVDA).

OSI PI: ASSET FRAMEWORK



OSI PI: EVENT FRAMES



Name	[4.05:41:40....	Duration	Start Time	End Time
10 - Charge		9:57:20.804	10/2/2017 1:00:08.603 PM	10/2/2017 10:57:29.407 PM
20 - Hold		11:23:51.186	10/2/2017 10:57:29.407 PM	10/3/2017 10:21:20.593 AM
50 - Media/Syrum		0:07:53.314	10/3/2017 10:21:20.593 AM	10/3/2017 10:29:13.907 AM
60 - Hold		5:22:55.503	10/3/2017 10:29:13.907 AM	10/3/2017 3:52:09.41 PM
70 - pH Condition		0:00:27.97	10/3/2017 3:52:09.41 PM	10/3/2017 3:52:37.38 PM
75 - DO Condition		1:58:25.417	10/3/2017 3:52:37.38 PM	10/3/2017 5:51:02.797 PM
80 - Cells		0:04:09.63	10/3/2017 5:51:02.797 PM	10/3/2017 5:55:12.427 PM
90 - Growth		1:19:40:17.37	10/3/2017 5:55:12.427 PM	10/5/2017 1:35:29.797 PM
100 - Condition		5:14:00.12	10/5/2017 1:35:29.797 PM	10/5/2017 6:49:29.917 PM
120 - Media		2:22:43.363	10/5/2017 6:49:29.917 PM	10/5/2017 9:12:13.28 PM
125 - Serum		0:07:37.327	10/5/2017 9:12:13.28 PM	10/5/2017 9:19:50.607 PM
130 - Growth		1:40:35.486	10/5/2017 9:19:50.607 PM	10/5/2017 11:00:26.093 PM
140 - Infect		0:18:08.394	10/5/2017 11:00:26.093 PM	10/5/2017 11:18:34.487 PM
150 - Growth		19:23:14.63	10/5/2017 11:18:34.487 PM	10/6/2017 6:41:49.117 PM
155 - Condition		5:18:56:50....	10/6/2017 6:41:49.117 PM	

BACKGROUND AND STATISTICS
BIOLOGICAL NICHE = HIGH VARIANCE COMPONENTS

Keeping it simple: It's about directional covariance relative to target response.

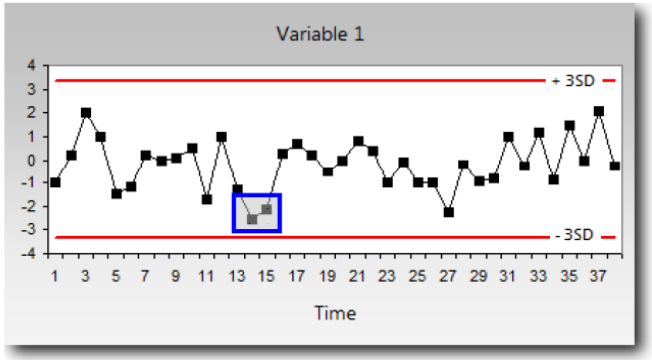
In the process, the goal is to explain the maximum amount of variance in the original data with the fewest number of components.

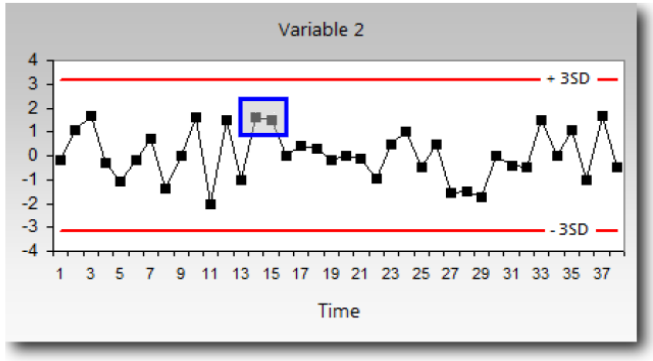
$$\sigma^2_{Total} = \sigma^2_{x^1} + \sigma^2_{x_2} + \sigma^2_{x3} + \sigma^2_{x4} + \in_{total\ error}$$

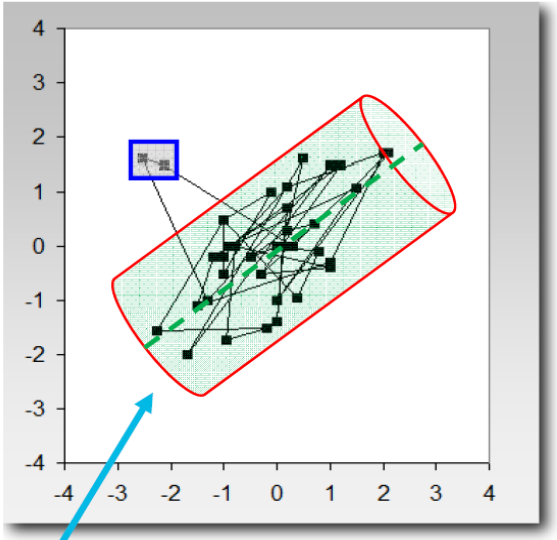
MULTIVARIATE PREDICTION AND CONTROLS

The goal is to explain the maximum amount of variance in univariate and multivariate space.

As we begin to characterize these variance components we begin to better understand our processes and how to create prescriptive process controls.







Multivariate Control Limits detect unusual points

$$\sigma^2_{Total} = \sigma^2_{x1} + \sigma^2_{x2} + \sigma^2_{x3} + \sigma^2_{x4} + \in_{total\ error}$$

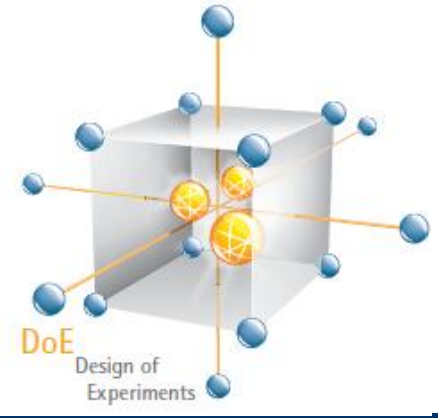
SIMCA & SIMCA ONLINE

Multivariate Umetrics Suite

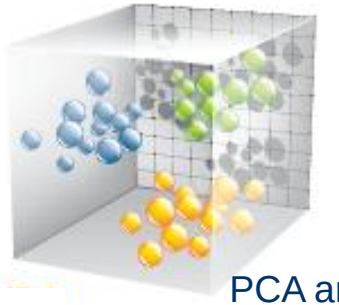
Now acquired by Sartorius Stedim

Industry Leader in MVDA: PCA, PLS, OPLS, and Online BEM

MODDE

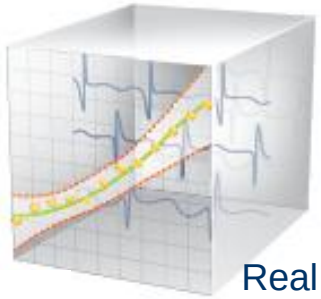


SIMCA



PCA and PLS

SIMCA-Online



Real Time Batch Evolution

SIMCA-online

Sartorius Stedim Data Analytics

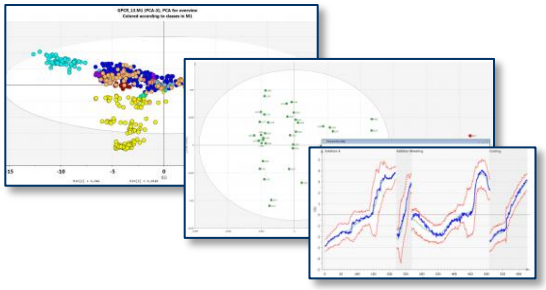
At your fingertips:

- Remote monitoring
- Multivariate predictive monitoring
- Root cause analysis
- Automatic corrective recommendation
- Real-time process control

Analytics in Biologics Manufacturing

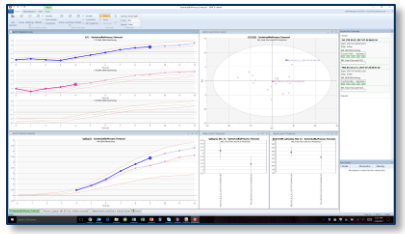
Process Characterization

- Model desired process operation in **SIMCA**
- Use these models to quantify process variation and impact on quality



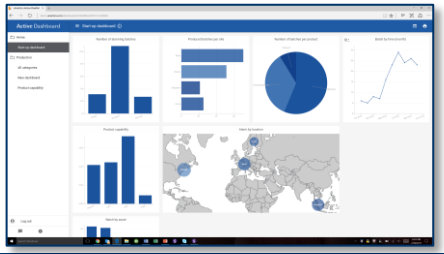
Process Monitoring

- Compare current and future production to desired
- Real-time monitoring provided in **SIMCA-online**



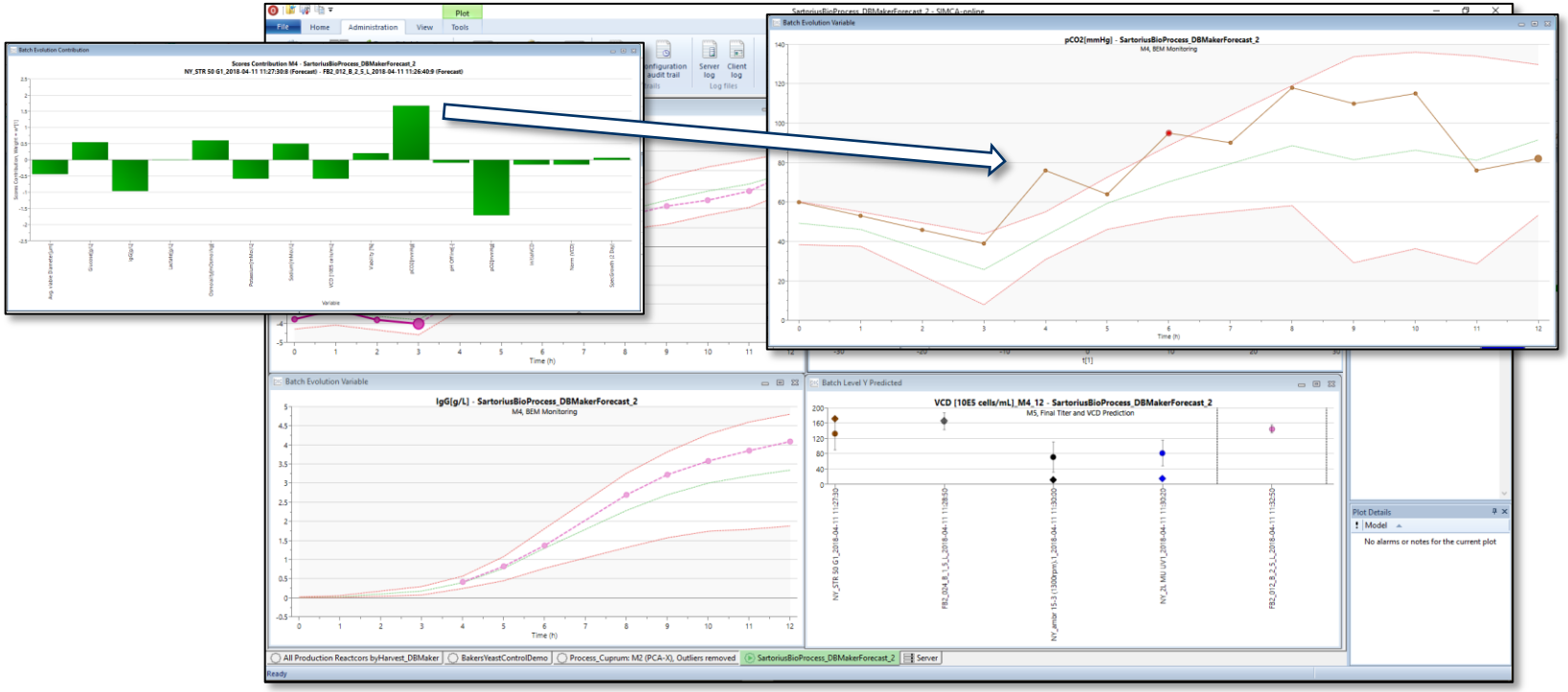
Enterprise Visualization

- Enterprise level overview of variability in process and product quality in **Active Dashboard**
- Connect variability in the process to resulting quality



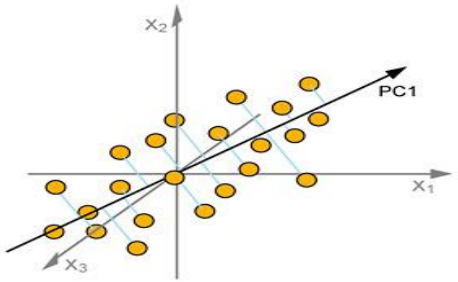
SIMCA-online

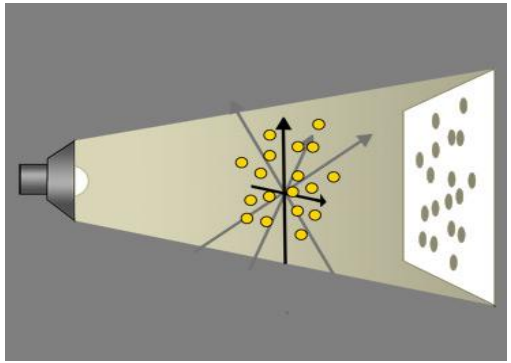
Sartorius Stedim Data Analytics

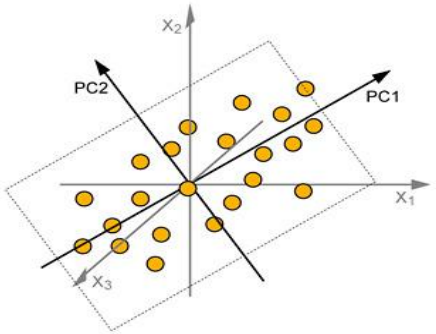


MVDA - PCA & PLS

PCA is a good method for displaying spacial relationships or affinities between observations. Score plots display the spread of the data-points according to the Covariance Matrix







Relationships in data are explained as eigenvalues or eigenvectors which summarize the ‘directional variance’ between observations. Loadings plots are used to summarize coefficients relative to a response.

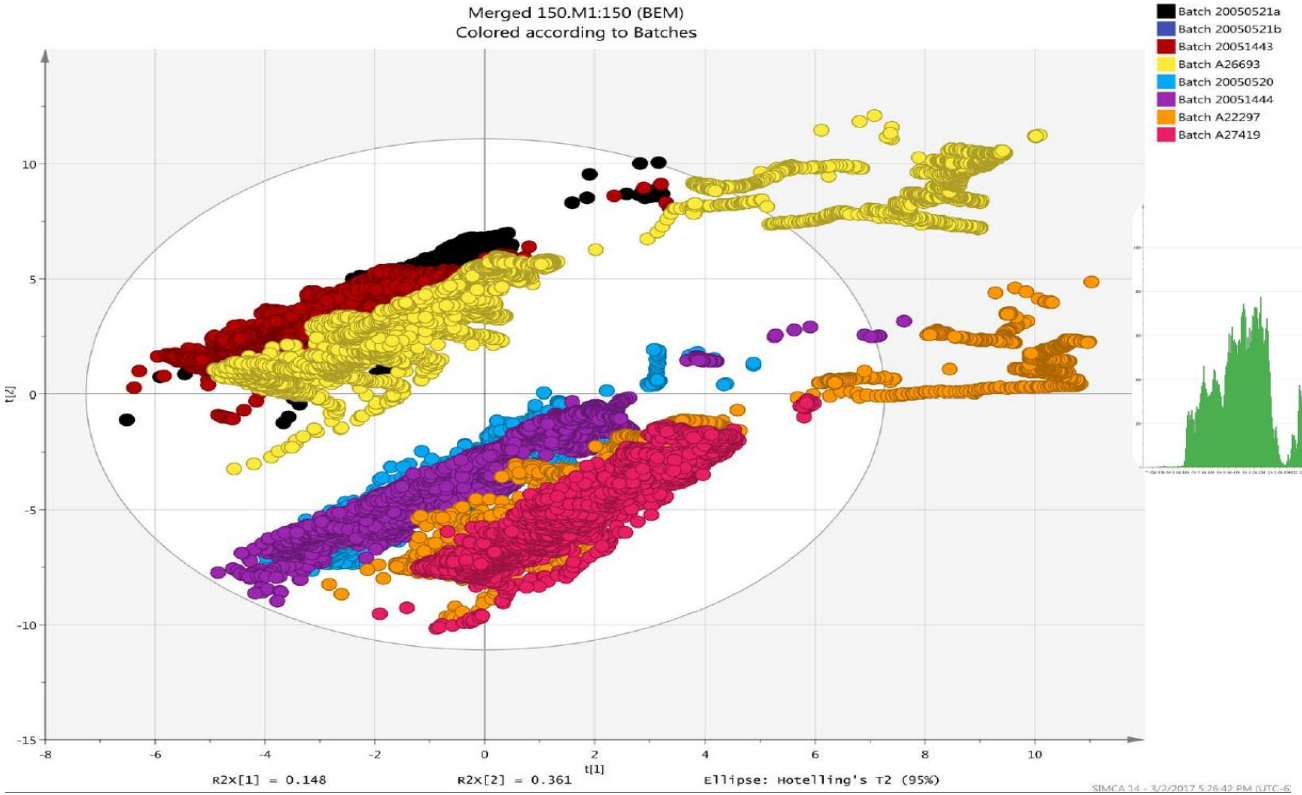
WHY PCA?
WHAT IS IT?

WHY PCA?
WHY ROTATE? - WHAT IS IT?

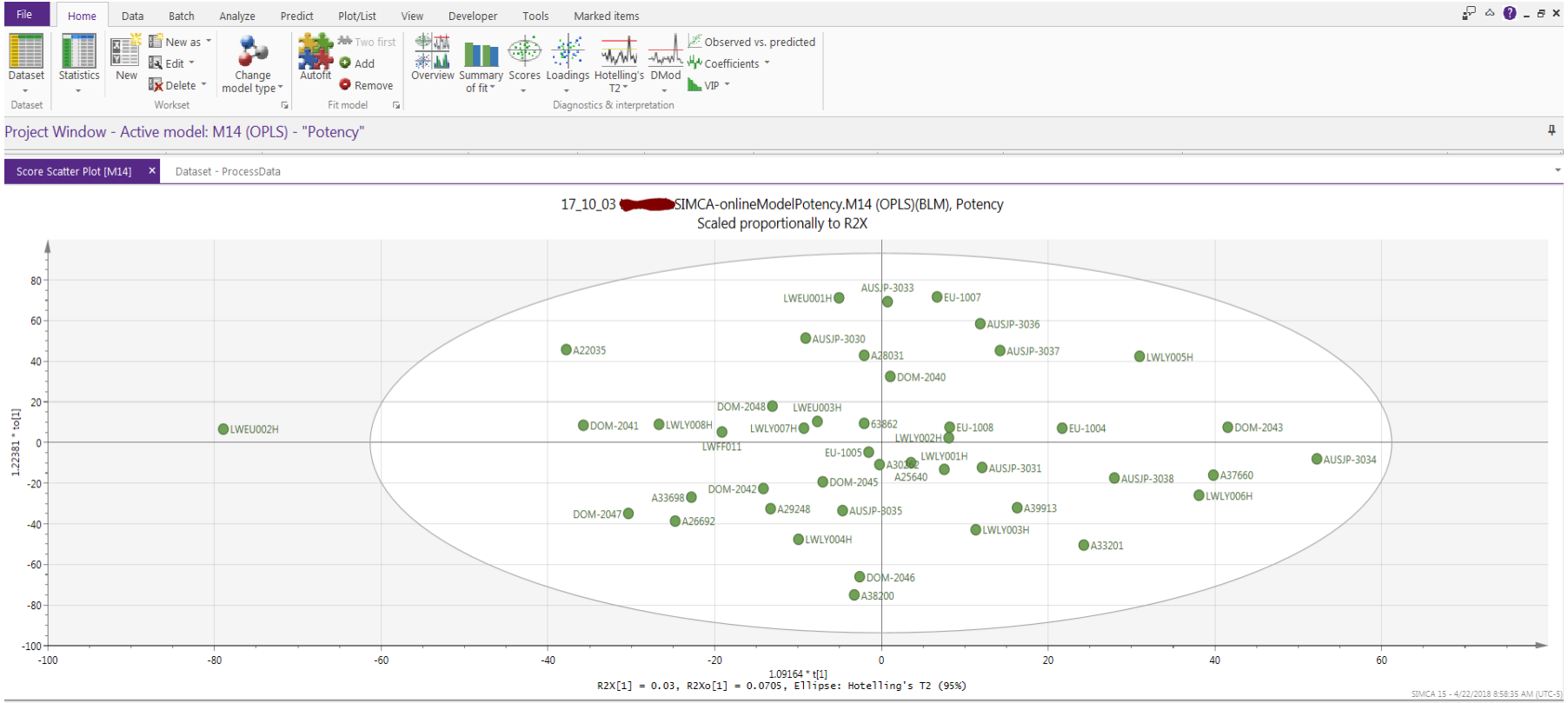
PCA SCORES PLOT – ANTIGEN GROWTH PHASE

PCA scores clearly show two distinct groups, or populations from Antigen Growth Data.

These distinctions typically illuminate major differences in asset framework e.g equipment and instrumentation

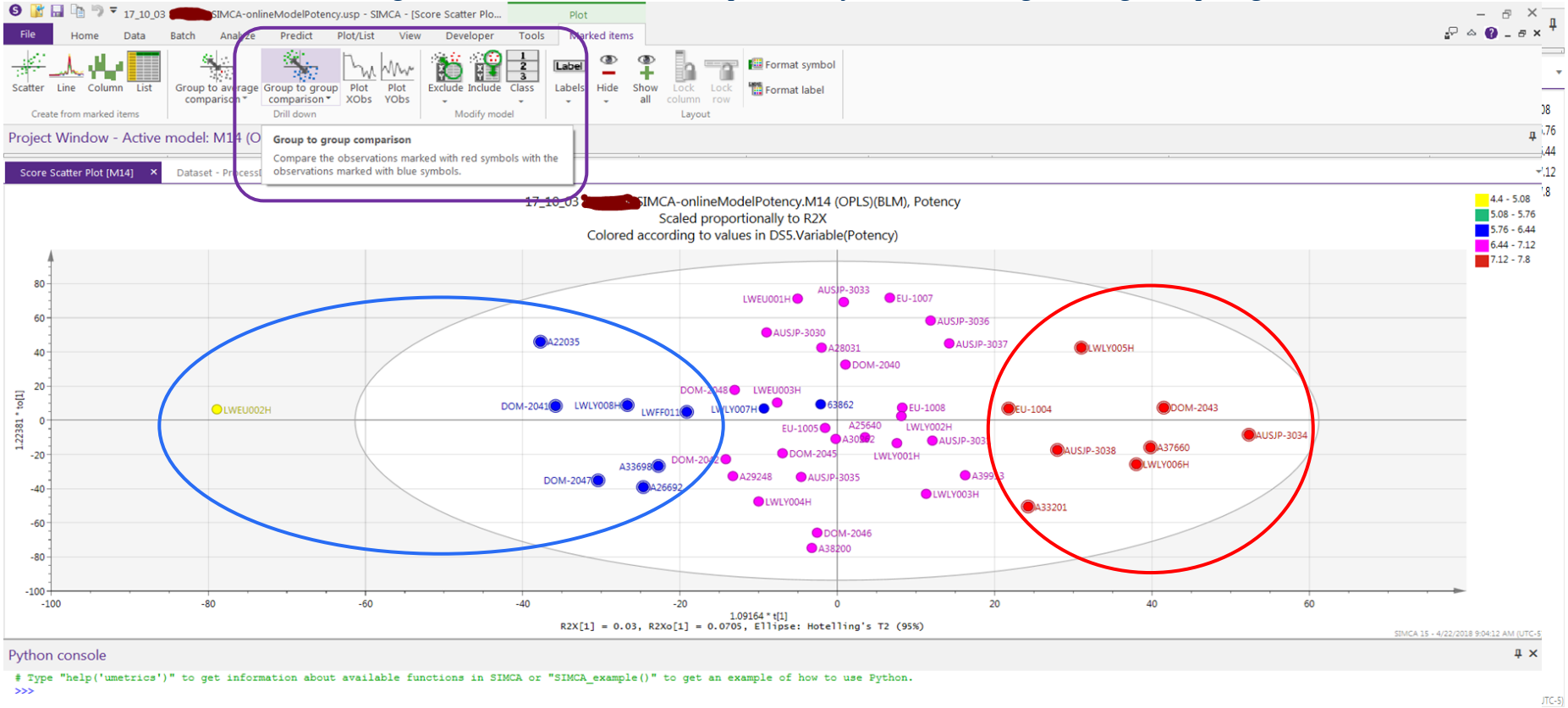


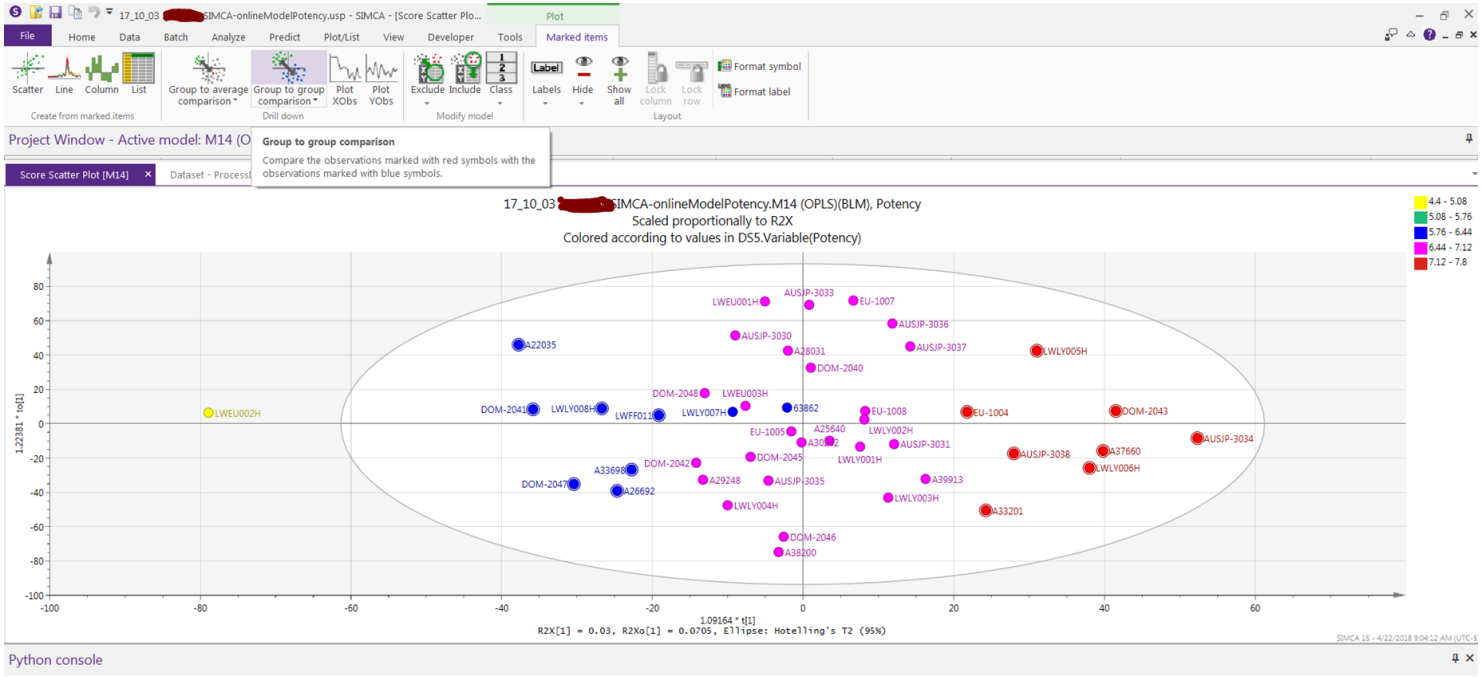
OPLS SCORE PLOT: POTENCY AS TARGET



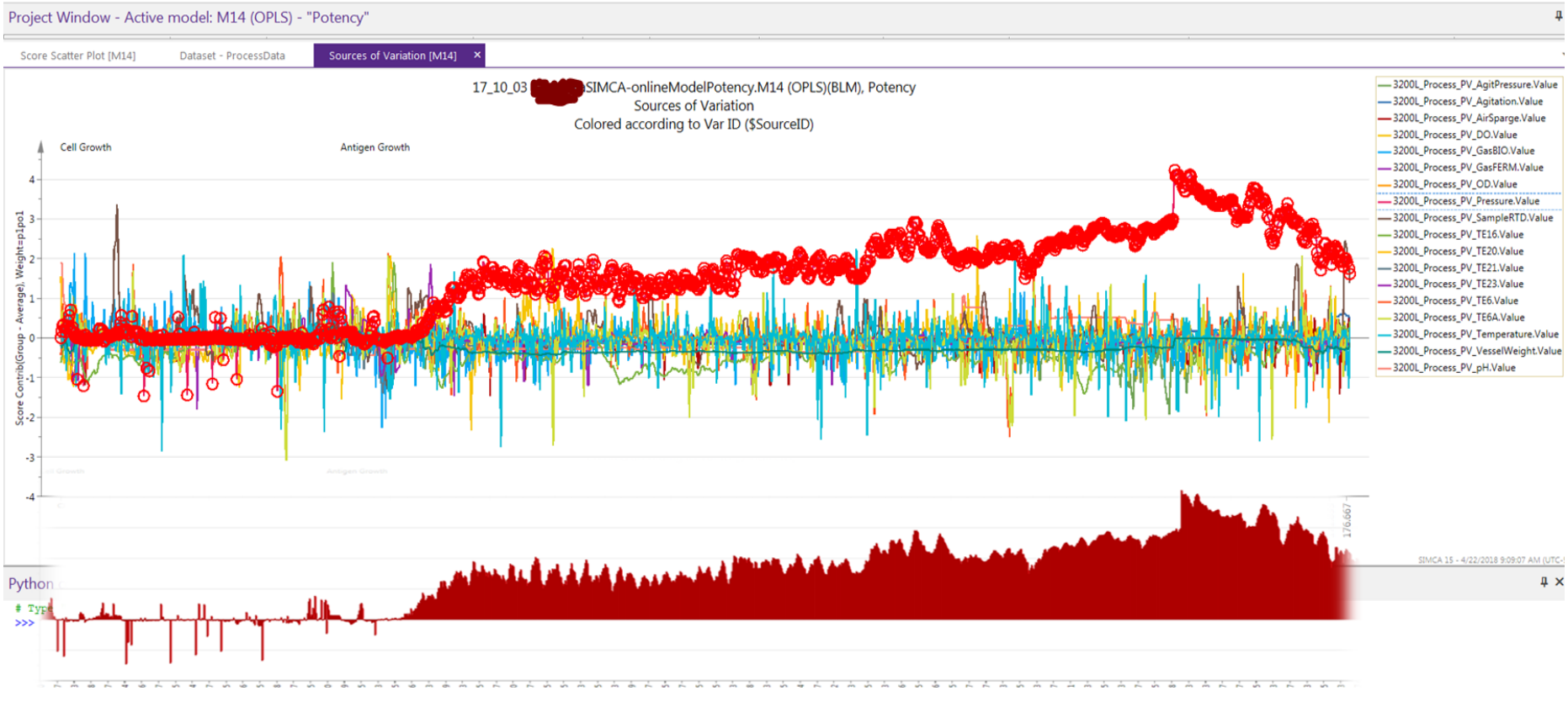
COLORED BY POTENCY

Notice the left to right distribution in potency – looking for groupings




17_10_03 [redacted]SIMCA-onlineModelPotency.usp - SIMCA - [Score Scatter Plo...
Plot
File
Home
Data
Batch
Analyze
Predict
Plot/List
View
Developer
Tools
Marked items
Scatter
Line
Column
List
Create from marked items
Group to average comparison
Group to group comparison
Plot XObs
Plot YObs
Drill down
Exclude
Include
Class
Modify model
Labels
Hide
Show all
Lock column
Lock row
Layout
Format symbol
Format label
Project Window - Active model: M14 (O
Group to group comparison
Compare the observations marked with red symbols with the observations marked with blue symbols.
Score Scatter Plot [M14]
Dataset - Process
17_10_03 [redacted]SIMCA-onlineModelPotency.M14 (OPLS)(BLM), Potency
Scaled proportionally to R2X
Colored according to values in DS5.Variable(Potency)
4.4 - 5.08
5.08 - 5.76
5.76 - 6.44
6.44 - 7.12
7.12 - 7.8
1.22381 * to[1]
1.09164 * t[1]
R2X[1] = 0.03, R2Xo[1] = 0.0705, Ellipse: Hotelling's T2 (95%)
LWEU001H
AUSJP-3033
EU-1007
AUSJP-3036
AUSJP-3030
A22035
A28031
AUSJP-3037
LWLY005H
DOM-2040
DOM-2048
LWEU003H
DOM-2041
LWLY008H
LWFF011H
LWLY007H
63862
EU-1008
EU-1004
DOM-2043
LWEU002H
EU-1005
A25640
LWLY002H
AUSJP-3034
AUSJP-3031
A37660
AUSJP-3038
DOM-2042
DOM-2045
LWLY001H
LWLY006H
A33698
DOM-2047
A29248
AUSJP-3035
A39913
A26692
LWLY003H
LWLY004H
A33201
DOM-2046
A38200
Python console
Type "help('umetrics')" to get information about available functions in SIMCA or "SIMCA_example()" to get an example of how to use Python.
>>>

“ONE OF THESE THINGS IS NOT LIKE THE OTHER(S)”



EXPERT TEAM CONCEPT:

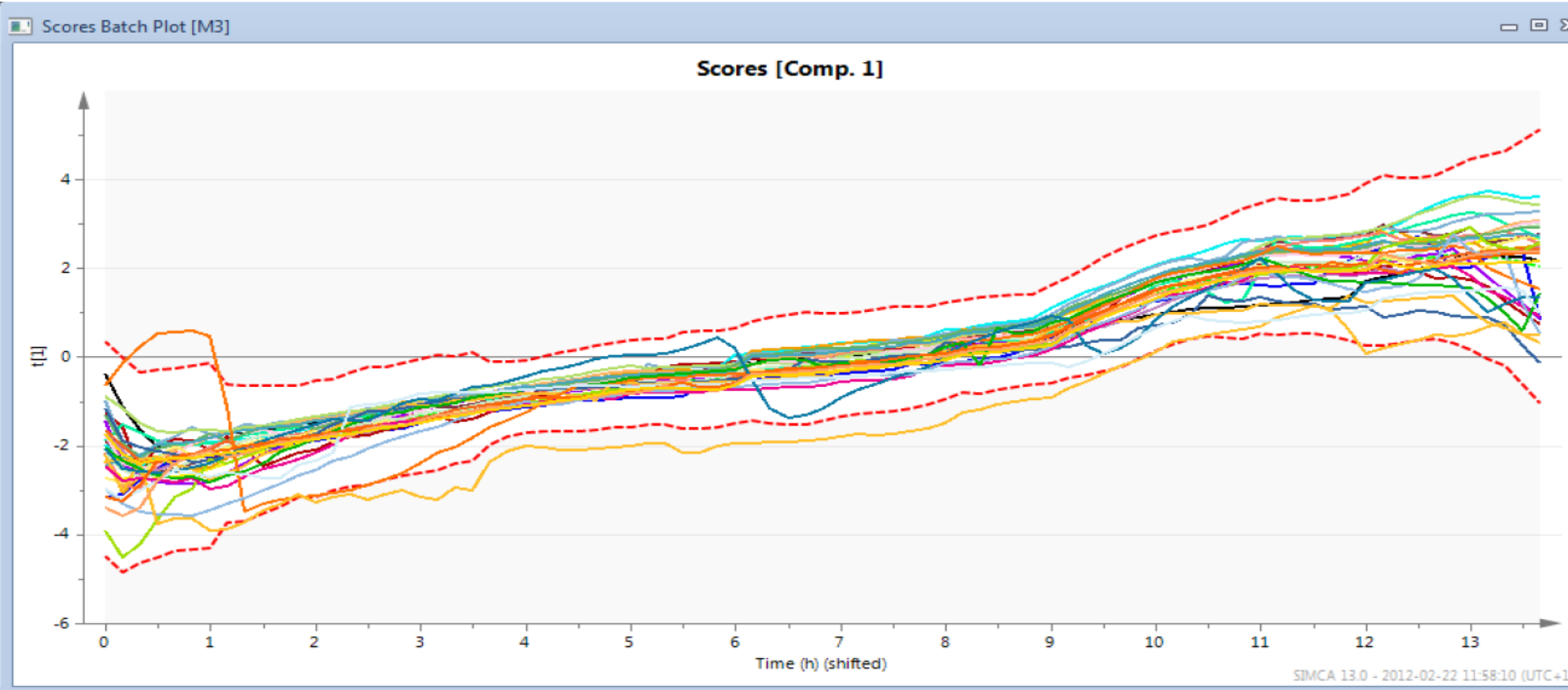
Focused on visualizations, model interpretations, risk assessment, control strategy, and transfer of value propositions into savings

Data Analytics, BioProcess Science, Validation, QA, Production, Engineering, QC, Supply Chain, Sourcing

EXECUTIVE SUMMARY:

Business Impact	Costs	Returns
Resources	Data Architects and Data Modelers (advanced technical skills) Data Management – middleware data warehousing and interface OSI PI and SIMCA Online software costs Program Development	IndStandardCompliance – 100% in CSS and CPV – ROI savings in maximizing profitibility by reducing productionVariances and waste/rework, savings in returning Biostatistics Contracts back in-house.
Status	**Current**	**Future**
Challenges	Data Management PAT Technical Drivers – Real time Potency Metrics	Industrie 4.0 & Technology Shift Data Maturity and Smart/IOT systems
Value Return ROI – batch savings 2-4 batches year for active models	4-6 active models Value Propositions in ET Meetings Time and Project Management	Translate Value Propositions Into Hard Savings

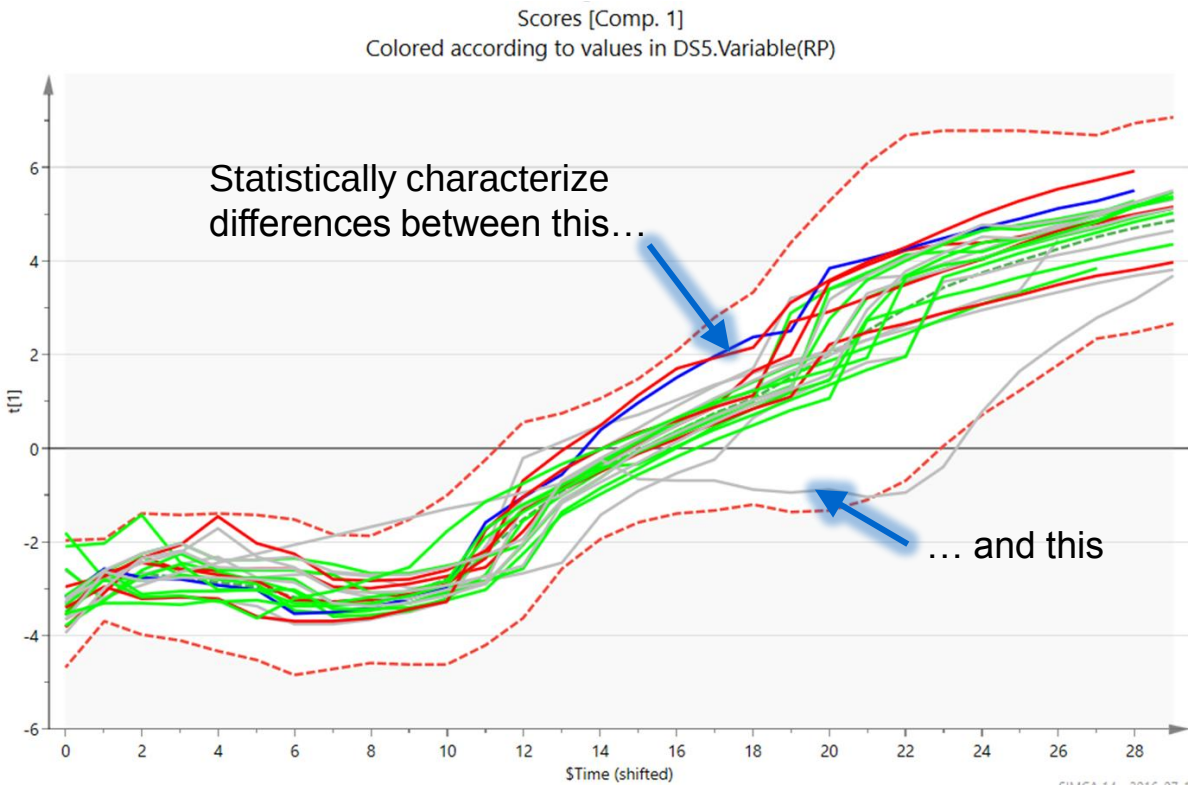
BATCH EVOLUTION MODEL
EACH LINE = 1 BATCH: MULTIVARIATE CONTROL CHART



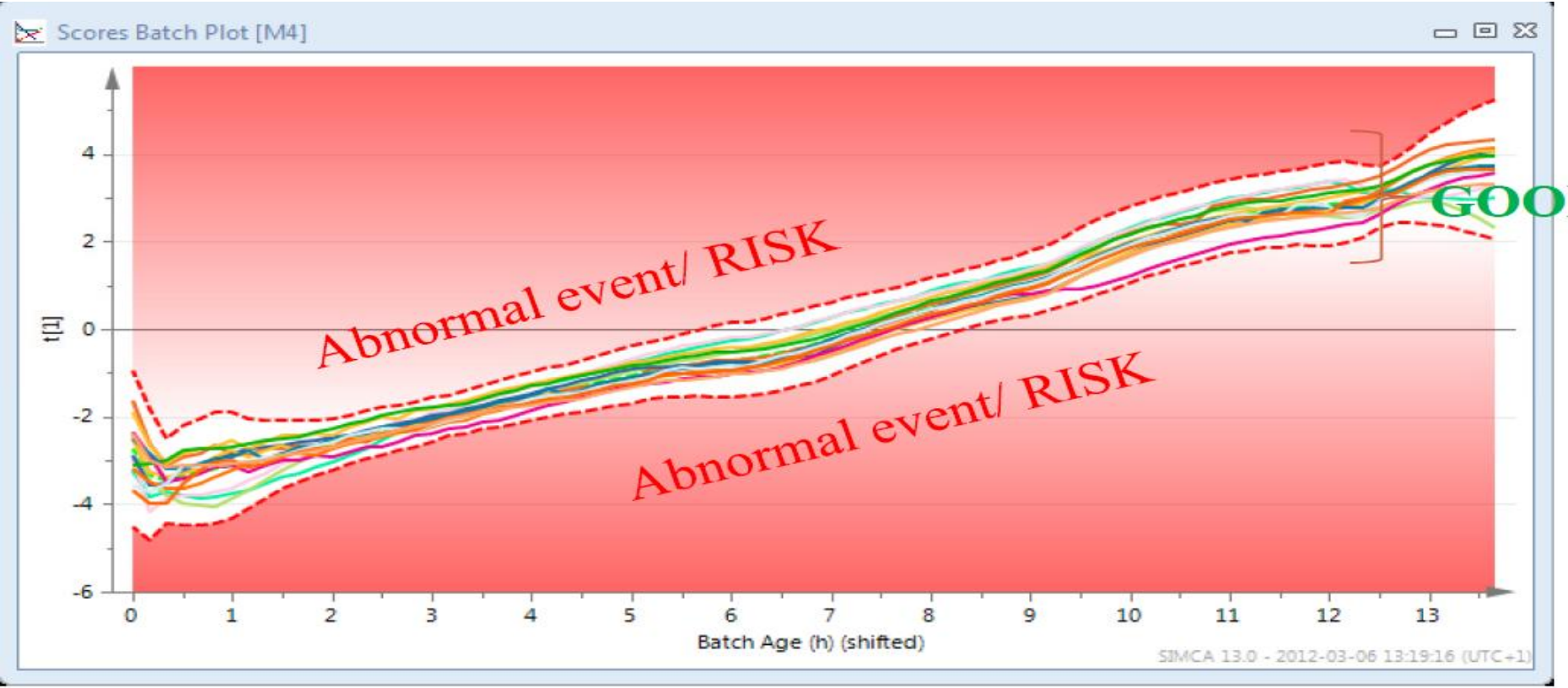
OBJECTIVE OF BATCH EVOLUTION MODELS

Batch Evolution Models (BEM) use time shifted score plots for Antigen Growth.

Process trajectory is currently being monitored on 3 large scale products in

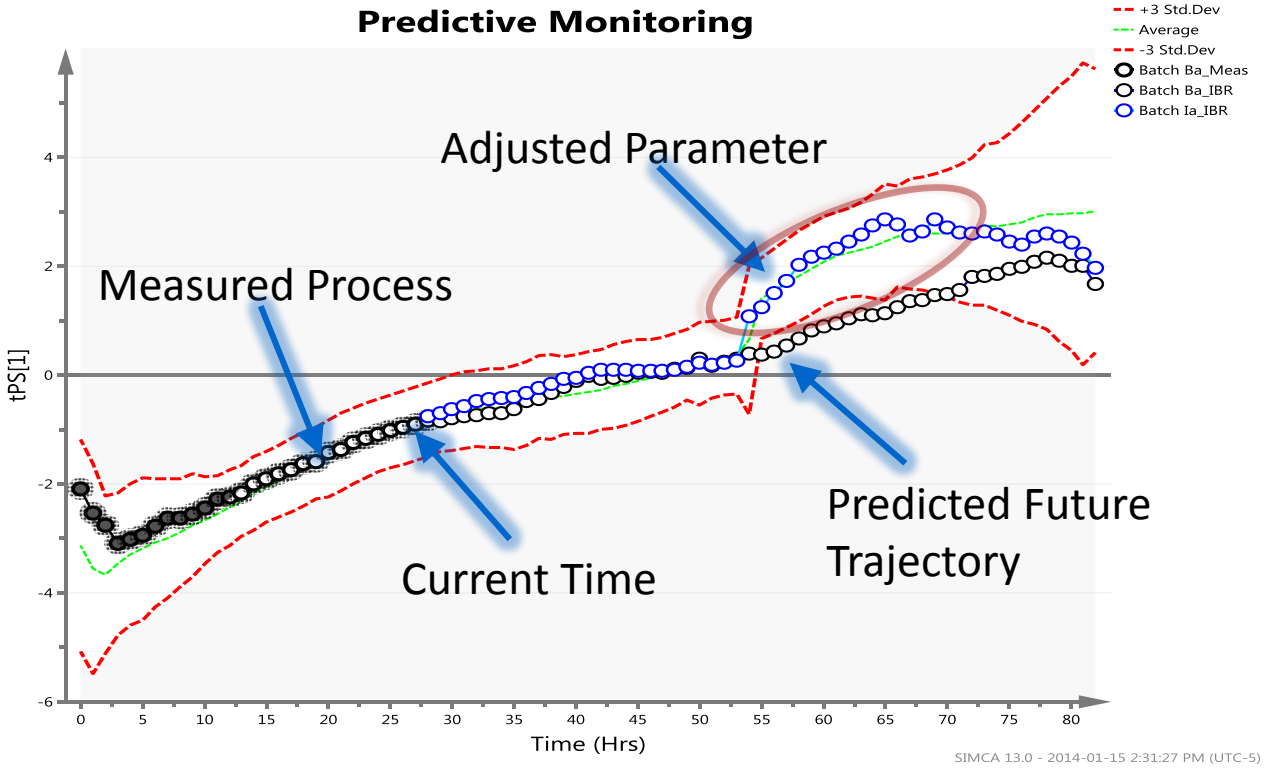


OBJECTIVE:


Scores Batch Plot [M4]
Abnormal event/ RISK
Abnormal event/ RISK
GOOD
t[1]
Batch Age (h) (shifted)
SIMCA 13.0 - 2012-03-06 13:19:16 (UTC+1)

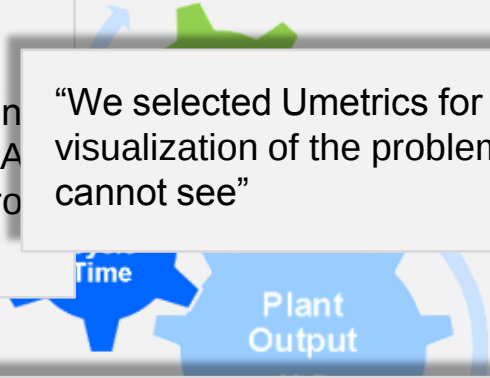
BATCH ADJUSTMENTS

detect and ***mitigate*** process deviations ***before*** they happen



VALUE GENERATION

Process characterization of historical data



Process Monitoring

"[SIMCA-online] has increased o
time fault detection capabilities.
improved process and equipmer
knowledge and understanding. It

Manufacturing Risk

"We achieved significant decrease in
manufacturing risk due to the SIMCA
online system, and shifted culture fro
being reactive to proactive"

"We selected Umetrics for their visualization of the problems we cannot see"

"We selected Umetrics for their visualization of the problems we cannot see"

Model Predictive Control

Real-time optimization of feed profile and other important process parameters. resulted in increase in cell density of 23%

GMP

"Our foundation for monitoring in GMP facilities is built on the SIMCA and SIMCA-online software from Umetrics"

Amgen

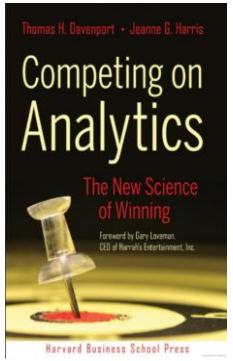
CIO 100 Award 2013

- “The new automated analytics system **saves time** and allows the team to **focus on the trends** instead of gathering and charting data. In one case, a yield issue was rectified in **less than a day** thanks to the new system; with Amgen’s old setup, it would not have become apparent for weeks.”
- “In one recent example, Amgen identified the cause of a cell culture problem about a month earlier than it otherwise might have. For that biologic product, making the fix early -and not losing that month- **saved $2.4 million**, she says. Depending on the product, manufacturing a lot can cost more than $1 million”, CIO Diana McKenzie says. **"This is Amgen's taming of big data**.“
- “Using statistical analysis of data points collected in **real time** during the process, Amgen identifies "weak signals" that could indicate brewing problems in the manufacturing cycle. Scientists then delve deeper, taking corrective steps if necessary. The analytics system includes virtualized data warehousing tools from Denodo, multivariate analysis tools from **Umetrics** and various software modules from SAP.”

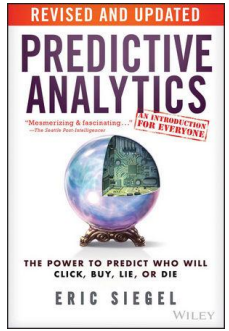
- Diana McKenzie, CIO

LITERATURE

1. Davenport,Harris: Competing on Analytics. Harvard Press



2. Siegel, Eric:Predictive Analytics. Wiley.



1. https://www.forbes.com/sites/bernardmarr/2016/03/01/the-8-hr-analytics-every-manager-should-know-about/2/#55758a7670a7

LITERATURE
OPERATIONS:

Ündey, S. T., Van Sprang, E., Streefland, M., Ramaker, H.-J., L.A., V. d., Beuvery, E., et al. (2007). Manufacturing Vaccines: An Illustration of Using PAT Tools for Controlling the Cultivation of Bordetella pertussis. Quality Engineering , 19, 373–384.	 Manufacturing Vaccines _ An Il...
Ündey, C., Ertunc, T., Mistretta, B., & Loose, S. (2010). Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control. Journal of Process Control , 20:, 1009–1018.	 Applied advanced process analy...
Weiner, M. O. (2010). Building Process Understanding forVaccine Manufacturing Using Data Mining. Quality Engineering , 22, 157–168.	 Building Process Understanding...

I N T R O D U C T I O N :

Will A. Penland *will.penland@boehringer-ingelheim.com*
MSPA, Northwestern University
Principal Specialist, Data Analytics, Boehringer Ingelheim Animal Health

- 12 Years at BI Animal Health
  - Small molecule assay development (contract – stability indication, LC)
  - Pharmaceutical stability lead and QC/LIMS SPC lead
  - Process Validation lead (SM/BIO)
  - Data Analytics: Compliance statistics, predictive/prescriptive Data Science

Present: Data Driven actionable insights. Atypical trends. Golden batch trajectory. Predictive and Prescriptive Process Control

Future: Help the organization reduce inherent chaos associated with Operations Intelligence data sciences: production variances, #bigData, technology shift and Industrie 4.0 landscapes.

I N T R O D U C T I O N :

- Chris McCready,
  Lead Data Scientist, Sartorius Stedim Data Analytics

  - 4 years in process industries in advanced automation and systems design
  - 14 years in health science data analytics

    - Multivariate data analysis
    - Process monitoring
    - R&D into manufacturing analytics, optimization and control

B.A.Sc. Chemical Engineering – University of Waterloo, Canada
M.Eng. Chemical Engineering – McMaster University, Canada

THANK YOU!

Boehringer Ingelheim GmbH

Binger Straße 173, 55216 Ingelheim am Rhein, Germany
Phone +49 (0) 6132 770
Fax +49 (0) 6132 720
info@boehringer-ingelheim.de
www.boehringer-ingelheim.com