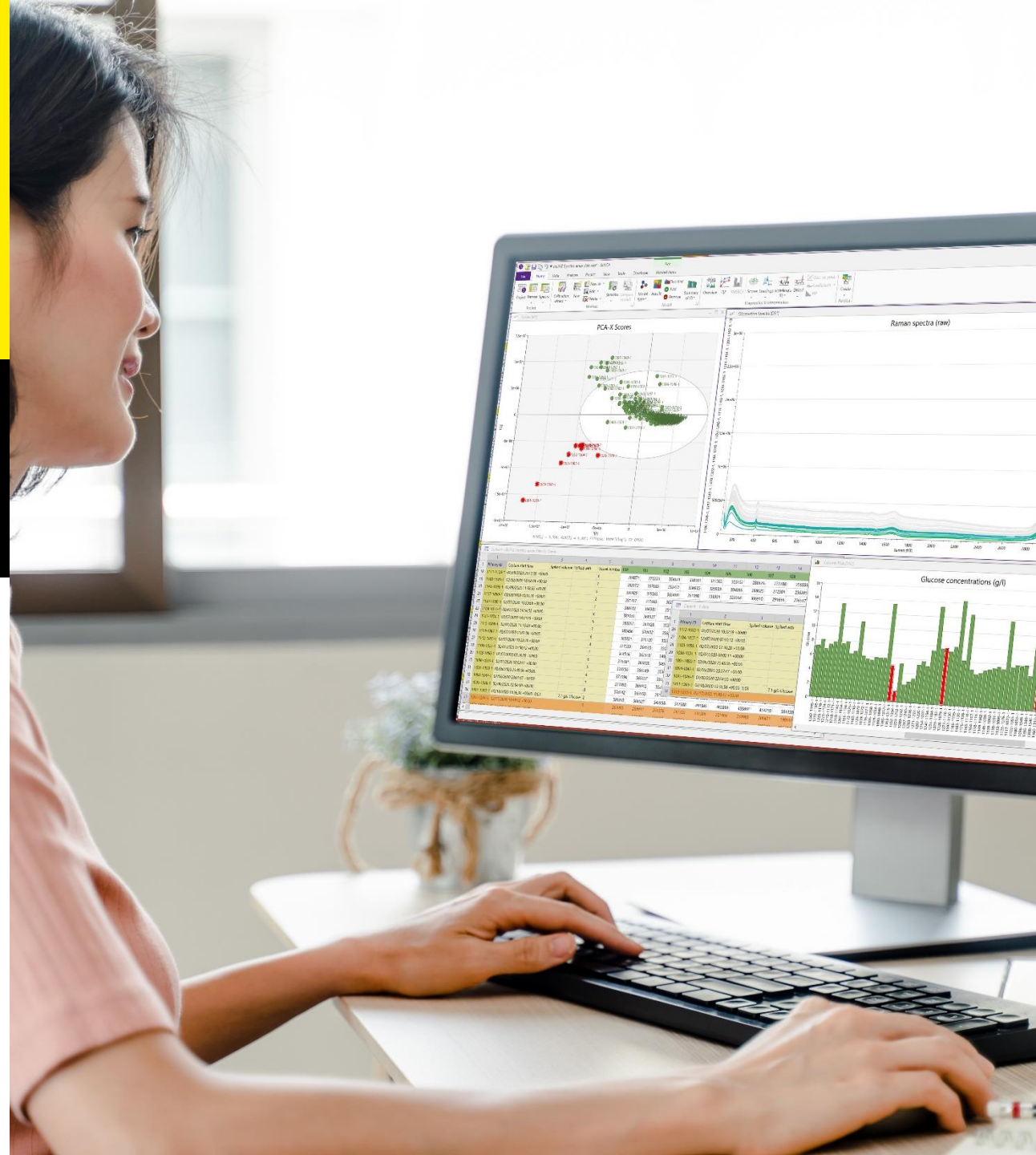


# SARTORIUS

## Simplifying Progress

### SIMCA<sup>®</sup> 17 – What's New?

Release February 9, 2021



# SIMCA® 17 Overview

- Umetrics® suite SIMCA® is focused on delivering a full data analysis experience, from data organization to data based decision making supported by multivariate models.
- SIMCA® 17 is focused on improving spectroscopy data analysis by providing tools to make multivariate calibration easy and reliable
- With SIMCA® 17 you get a complete solution for preprocessing your spectral data to enhance prediction performance
  - Extended library of preprocessing algorithms in an easy-to-use wizard
- Calibration wizard providing instant model quality visualization after each preprocessing choice helping you chose the best prediction model

# SIMCA® 17 Highlights

- Spectroscopy Project
  - Set, or generate Spectral ID at import
  - New spectra visualization
  - Extended library of preprocessing algorithms
  - Preprocessing wizard
  - Calibration wizard for 1 Y models
- Batch Data and Database Import
  - Possibility to control Batch alignment
  - BCC from dataset
  - SimApi:s available at installation and database interface improvements
- Plot Interactivity
  - Color by Rank – new option for better vector coloring
  - Reverse and log axis control in properties pane
  - Color BCC by batch condition vectors
- Other Improvements
  - Multiple file import for CSV, TXT and DIF files
  - Import configuration template
  - Python news and additions
  - Dataset properties
  - Dataset merge to keep un-matched content
  - New bias vectors (MBE)
  - SIMCA® 16 compatibility
  - Performance

# Learning What's New in SIMCA® 17

- In the following slides you will get an overview of the changes and additions made in SIMCA® 17
- For more details on how to use the features, please watch the what's new videos that you can reach from the start page of SIMCA® 17
- Please also check out more videos on SIMCA® and other Umetrics® suite products by looking up Sartorius Data Analytics on YouTube



Simplifying Progress

## SIMCA® 17 What's New – Spectroscopy Project

SARTORIUS

# Spectroscopy Project – Spectral ID

- A new integrated Spectroscopy project
  - Requires a Spectral ID vector
  - X-data (spectra) and Y-data are split in separate datasets
- Set your Spectral ID at import
  - Spectral ID will be used as X-axis in plots
  - Triggers access to the Calibration wizard
- Select your spectral ID, or type in a custom name
  - Generate your own numerical vector if none exist
- Center (ctr) scaling is set as default for spectral data

The screenshot shows the SIMCA software interface with the 'Specify spectral ID' dialog box open. The dialog is positioned over a data table. The 'Specify spectral ID' dialog has the following options:

- Wavelength
- Wavenumber
- Raman shift (selected)
- Custom spectral ID

Below the dialog, an 'Add value range' dialog is open, showing the following settings:

- Start value: 3600
- Step: -1
- Range: 3600 to 200
- Spectral ID name: Wavelength

The background data table is as follows:

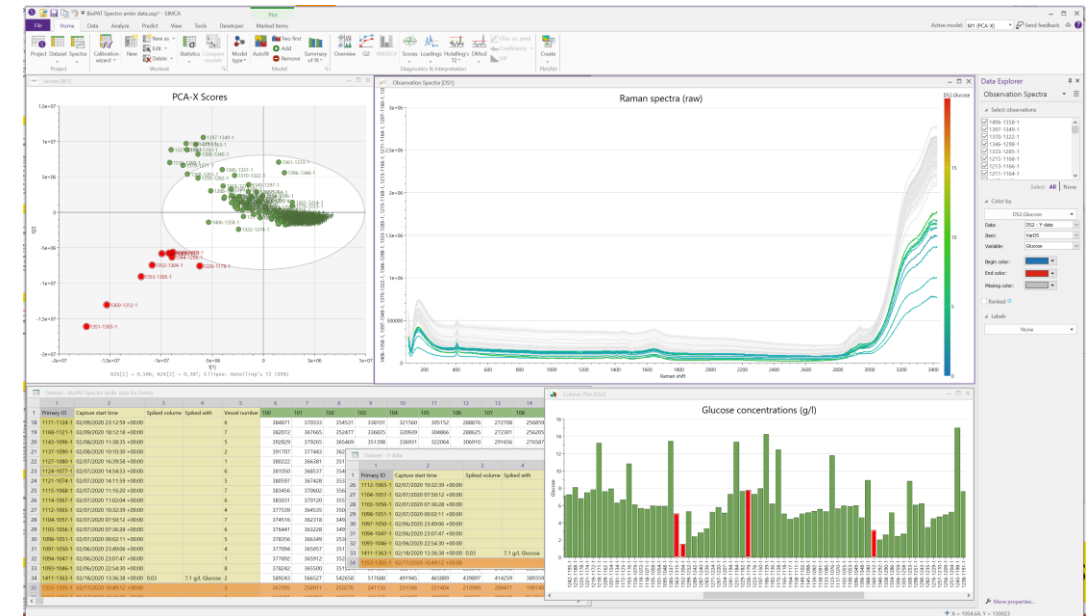
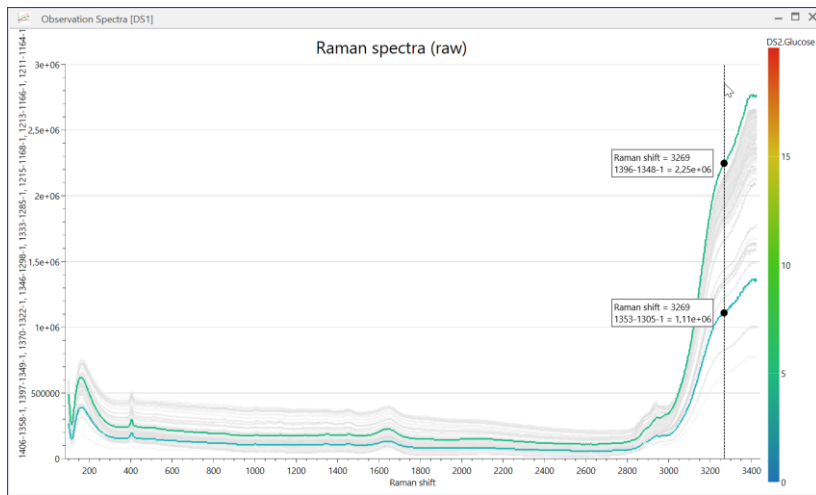
	Primary ID	2		6
Spectral ID	Raman shift	360		3597 3596
2	c001	0,11138		0,7833 0,10744
3	c002	0,09863		0,0553 0,101227
4	c003	0,117		0,0787 0,122423
5	c004	0,0823067	0,08015	0,0784567 0,07686 0,0756433
6	c005	0,102087	0,1015	
7	c006	0,0850233	0,08475	
8	c007	0,09031	0,09209	
9	c008	0,0917233	0,090	
10	c009	0,0716167	0,07133	
11	c010	0,0920633	0,09361	
12	c011	0,0872233	0,08881	



# Spectroscopy Project – New Spectra Visualization

- New selection tool in spectra plot
  - Spectra selection instead of variable selection
  - ctrl click to select more than one spectrum
  - Tooltip follow selected spectra for easy comparison

- SIMCA<sup>®</sup> recognizes spectra as observations and links the plot selections accordingly



# Preprocessing Library Additions

## SIMCA® 16 preprocessing

- Smoothing
  - *Savitzky-Golay*
  - *EWMA (left filter)*
  - *Wavelet denoising*
- Normalization
  - *SNV*
- Baseline correction
  - *Row center*
- Other (Spectra enhancement)
  - *MSC*
  - *Wavelet compression*
  - *Derivatives (Savitzky-Golay) 1<sup>st</sup>-3<sup>rd</sup>*
  - *OSC*

## SIMCA® 17 additions

- Smoothing
  - Moving window
    - average and median
  - EWMA (symmetric filter)
  - AsLS
- Normalization
  - peak height
  - peak area
- Baseline correction
  - Offset, linear baseline, AsLS
- Other (Spectra enhancement)
  - Derivatives (Savitzky-Golay) 4<sup>th</sup>



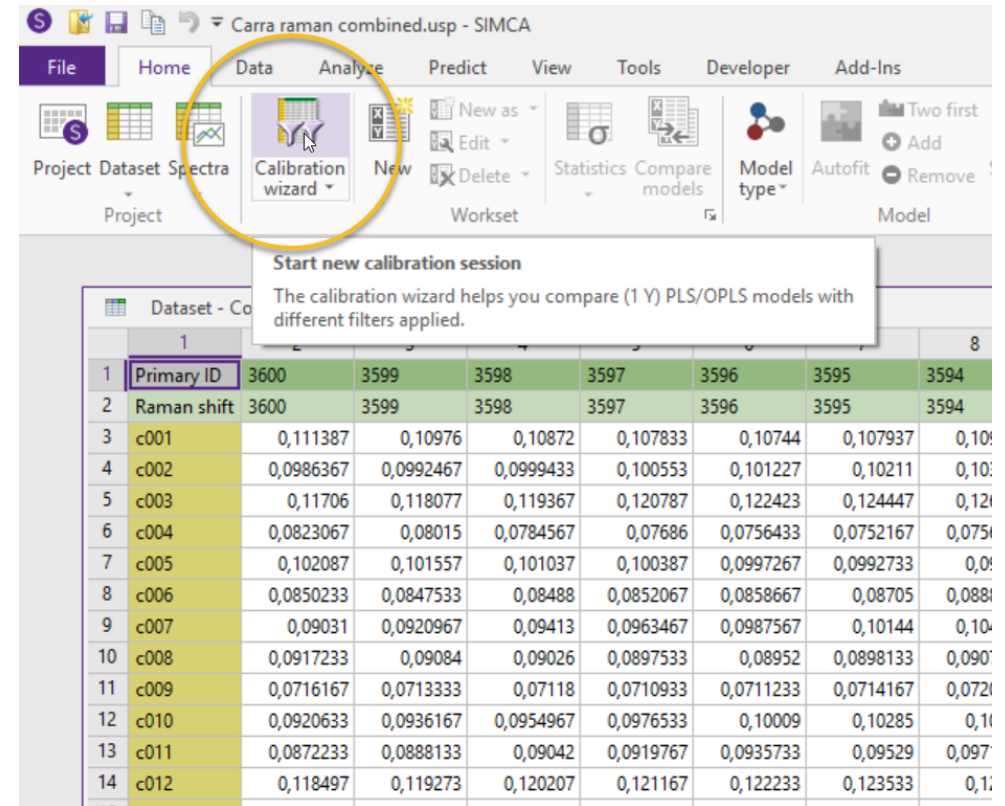
# Preprocessing Wizard

- Extensive selection of spectral preprocessing
  - Individual or chained filters
- Interactive, graphical, wizard for filter settings
  - Direct visualization of filter transformation effect
- Python based preprocessing filters are easily added to any of the categories in the wizard
- All filters are automatically applied to predictions
  - Also in SIMCA<sup>®</sup>-Q and SIMCA<sup>®</sup>-online

The screenshot displays the 'Preprocessing' wizard interface. At the top, a progress bar shows the current step is 'Filters'. Below this, a grid of filter categories is visible: Smoothing (Savitzky-Golay, EWMA, WDS, Moving window, AsLS smoothing), Baseline correction (Row-center, Offset, Linear, AsLS correction), Normalization (SNV, Peak height, Peak area), and Other (MSC, Derivatives, UmPyFilters.AsLS, UmPyFilters.AsLS). A 'Filter Settings' dialog box is open over the 'Derivatives' filter, showing settings for 'Derivative order' (First), 'Polynomial order' (Quadratic), 'Points in each sub-model' (15), and 'Distance between each point' (1). The dialog includes a 'Preview' section with a list of observation IDs (c001 to c016) and a plot of 'Intensity' vs 'Raman shift' (3600 to 200). The plot shows a peak at 2958 cm⁻¹ with a fitted curve and a residual plot below it. The 'Original' peak is at 7.9852. The wizard interface also shows a list of selected filters at the bottom: '1. Savitzky-Golay' and '2. AsLS correction'.

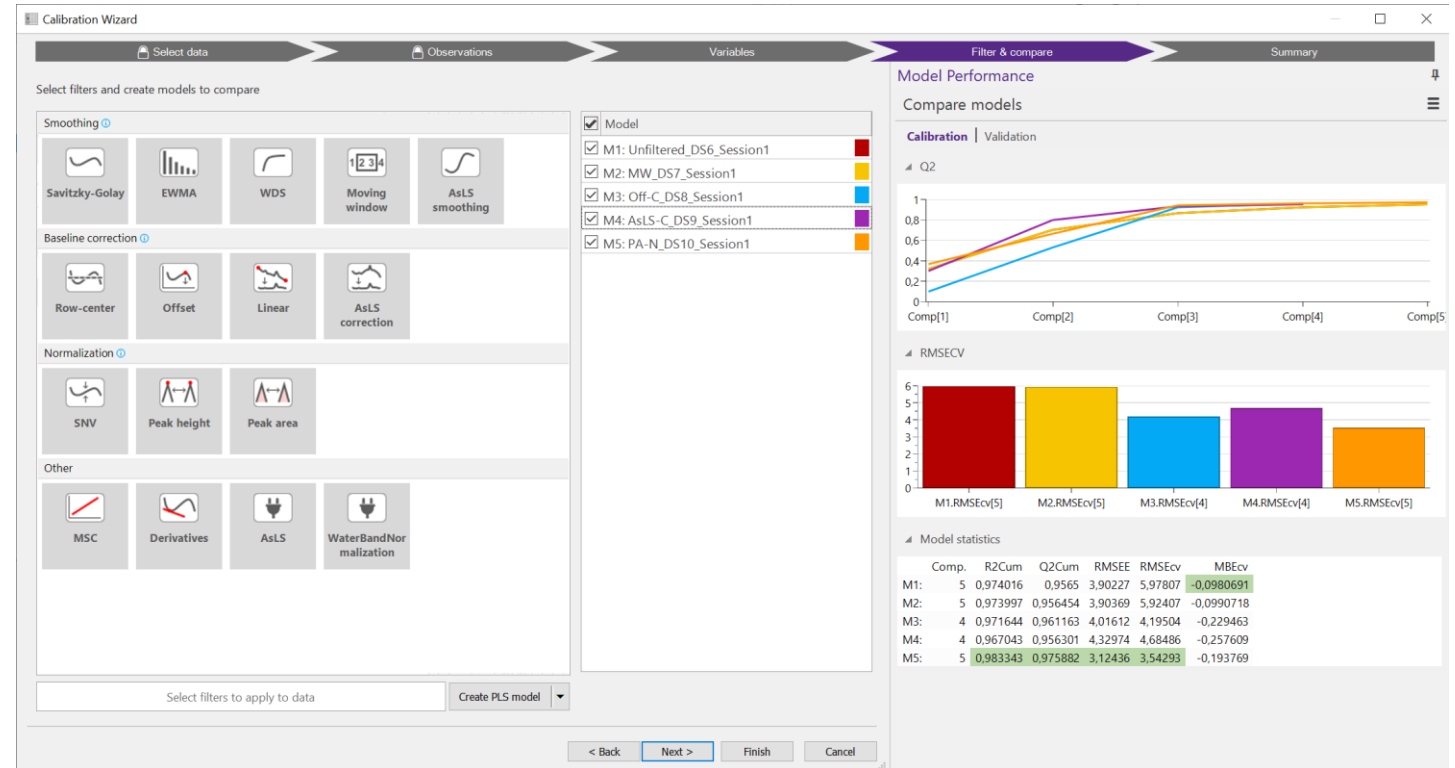
# Calibration Wizard

- Wizard for creating and comparing multivariate calibration models with 1 Y
  - Automatically available for spectroscopy projects
    - Spectral ID required
  - Creates a calibration session of models
    - Similar to Class models and Phase models
    - Workset observations consistent over session
- Existing calibration sessions can be opened and new data can be added
  - But only new prediction (validation) data
- Session report template available



# Calibration Wizard – Filter & Compare

- Same filter selection as in Preprocessing wizard
  - Select zero or many filters to apply
- Fit PLS or OPLS models
  - Access to cross validation group definition
- Model performance pane display model parameters and statistics
  - Calibration view (workset statistics)
  - Validation view (prediction set statistics)





Simplifying Progress

## SIMCA® 17 What's New – Batch Data and Database Import

**SARTORIUS**

# Batch Alignment – New Batch Alignment Option

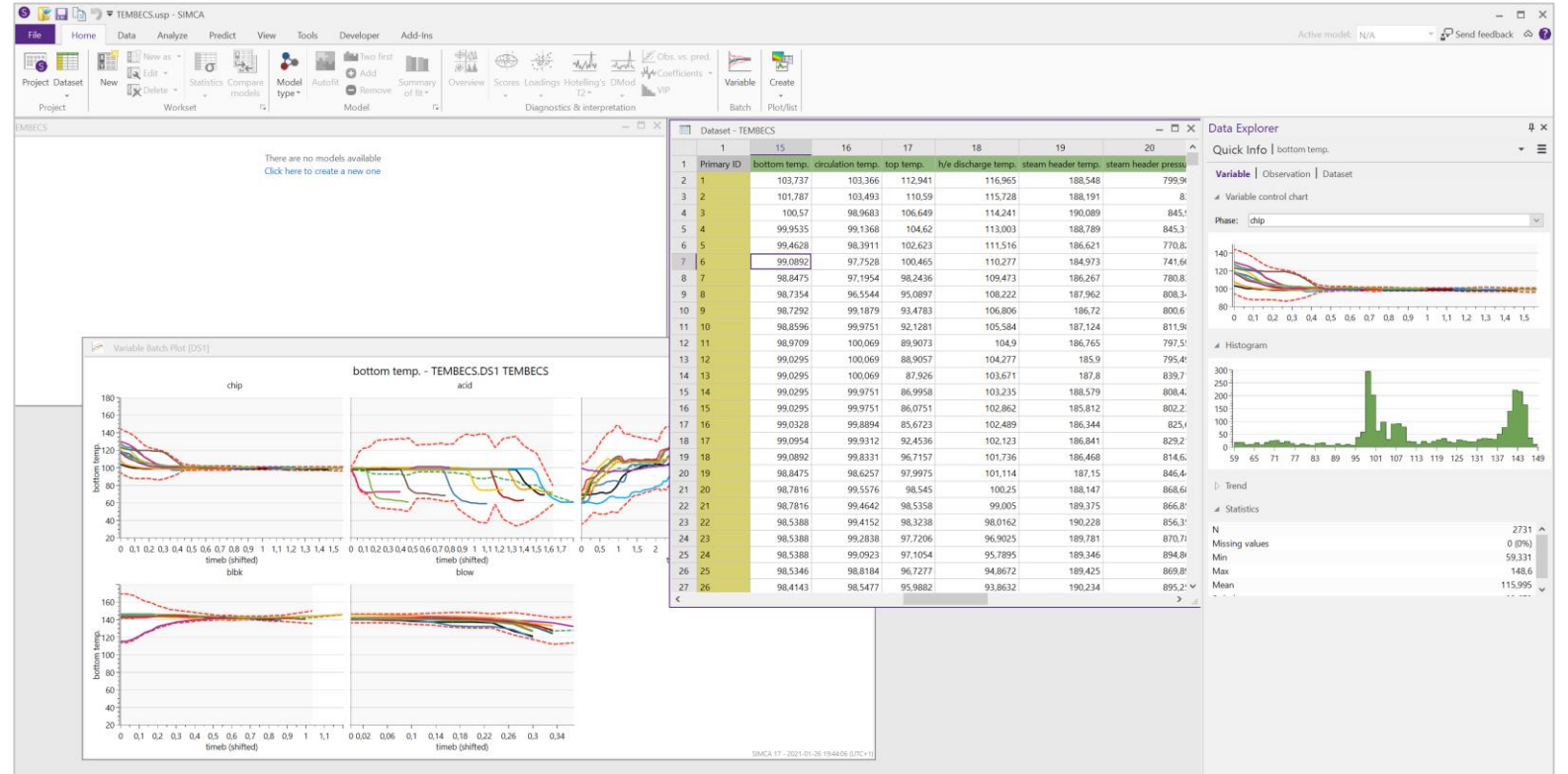
- New option to control the batch alignment vector
- The default SIMCA® alignment algorithm has been complemented with the possibility for user defined alignment vector
- For short batches (< 20 observations) there sometimes is a desire to control the alignment vector in detail
- In SIMCA® 17, it is possible to align Y according to an existing batch
  - Y alignment option in Workset dialog
  - Same batch for all phases
  - Short batches (<20 observations) have this as default

The screenshot displays the SIMCA 17 Workset dialog box, which is used for configuring data analysis parameters. The 'Variables' tab is active, showing a list of variables including 'Y Time (days)' which is highlighted. A 'Select Batch' dialog box is open, allowing the user to choose a representative batch for alignment. The 'Y-variable type' is set to 'Time or Maturity', and the 'Alignment' option is set to 'Alignment batch (Batch\_1)'. The 'Quick Info' panel on the right shows a histogram and trend plot for the selected variable.

Batch	Length	Included
SIMCA alignment	13	
Batch_2	13	✓
Batch_4	13	✓
Batch_5	13	✓
Batch_7	13	✓
Batch_8	13	✓
Batch_9	13	✓
Batch_10	13	✓
Batch_12	13	✓
Batch_13	13	✓
Batch_3	13	✓
Batch_11	13	✓
Batch_14	13	✓
Batch_1	13	✓
Batch_6	13	✓

# Batch Modelling Related Additions

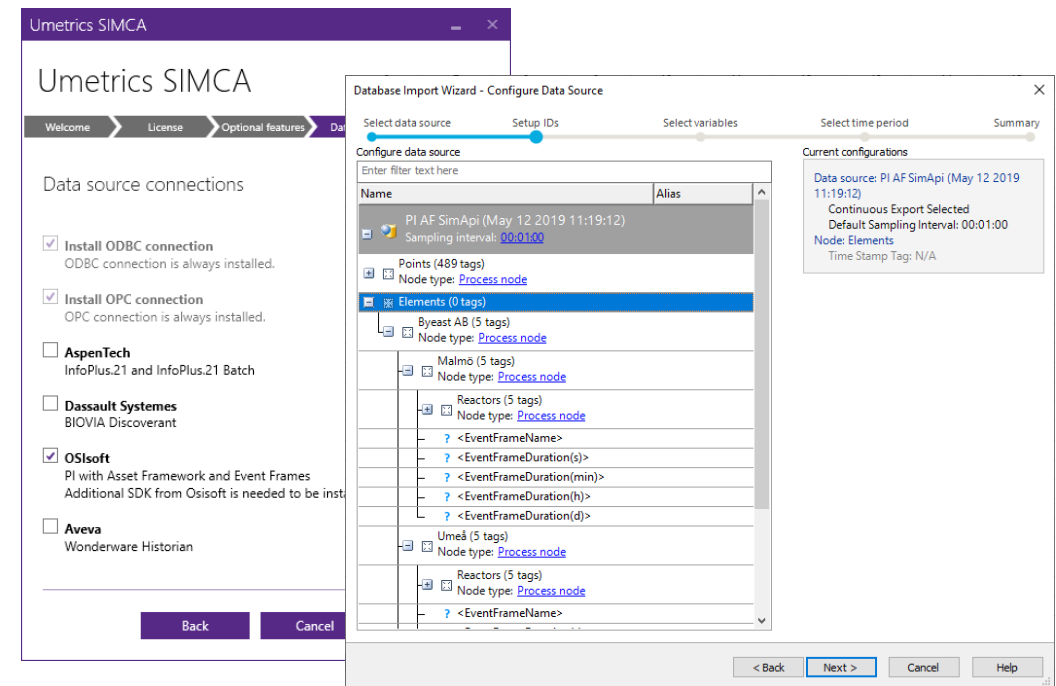
- Variable BCC from dataset
  - No need to first create a model
  - Also in Quick info
  - Select and Exclude batches possible





# SimApi:S and Database Interface Improvements

- To assist users with process data to extract data directly from the process database some common SimApi:s have been made available during installation
  - No need to download and install the SimApi after SIMCA® installation
  - More SimApi:s are available on Sartorius Data Analytics webpage (link in SIMCA® Help)
- Database interface improvements
  - Hierarchies and node trees visible in database import
  - Database import of batch data where the process node is missing a Batch ID tag is now possible





Simplifying Progress

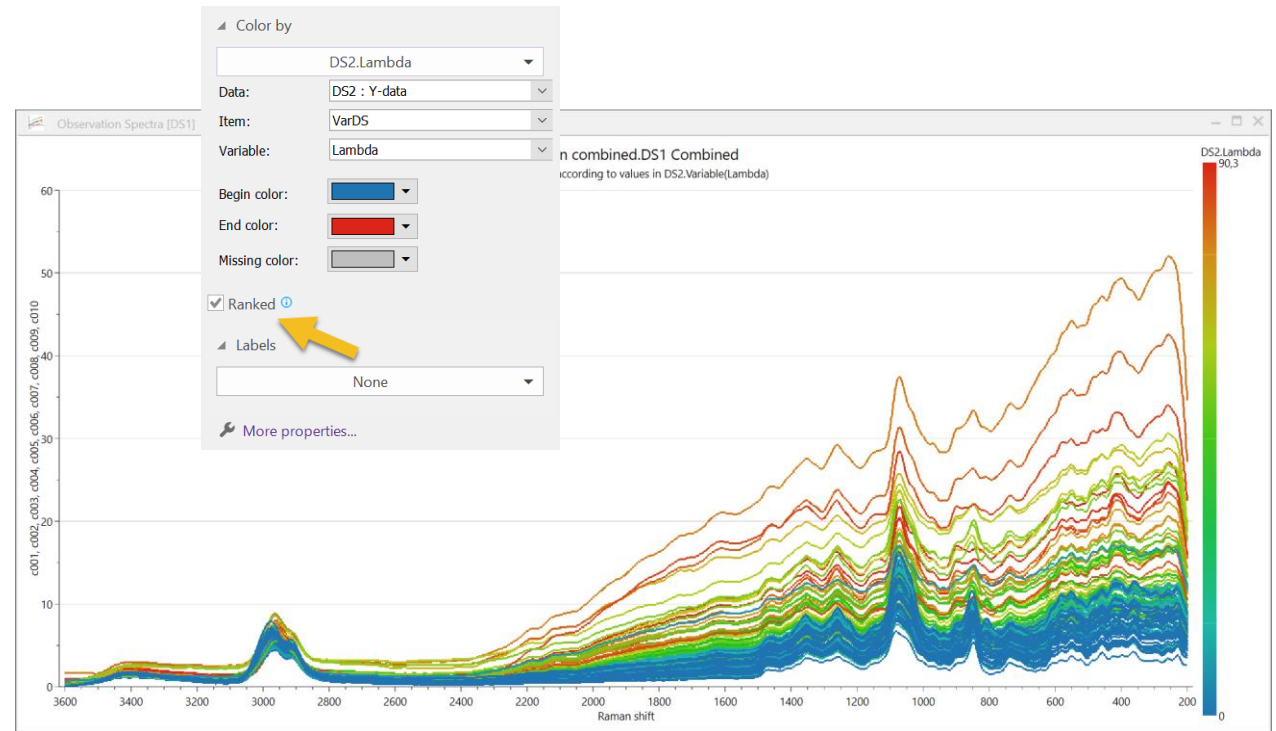


SIMCA® 17 What's New – Plot Interactivity  
and Other Improvements

SARTORIUS

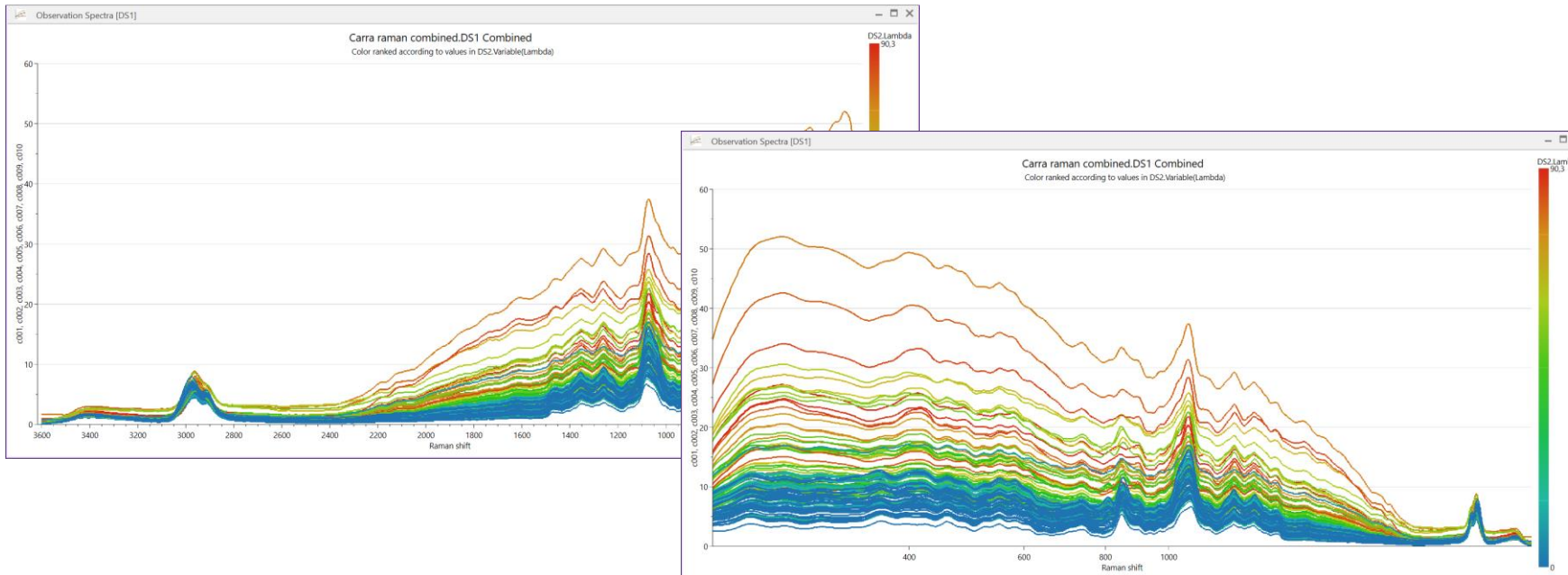
# Plot Interactivity - Color by Rank

- Color by rank added as an alternative for Continuous coloring
- Ranked color is useful when the coloring vector values have a skewed distribution
  - Ranked color distributes the colors linearly over the full vector range
  - Available in the properties pane



# Plot Interactivity – Axis Options

- Reverse and Log axis available in the properties pane
  - Not only for spectra plot



Data Explorer

Appearance

Axes | Legend | Styles | Titles

General

Axis font: [font icons]

Title font: [font icons]

Axis X

Min: 183

Max: 3617

Rotation: No rotation

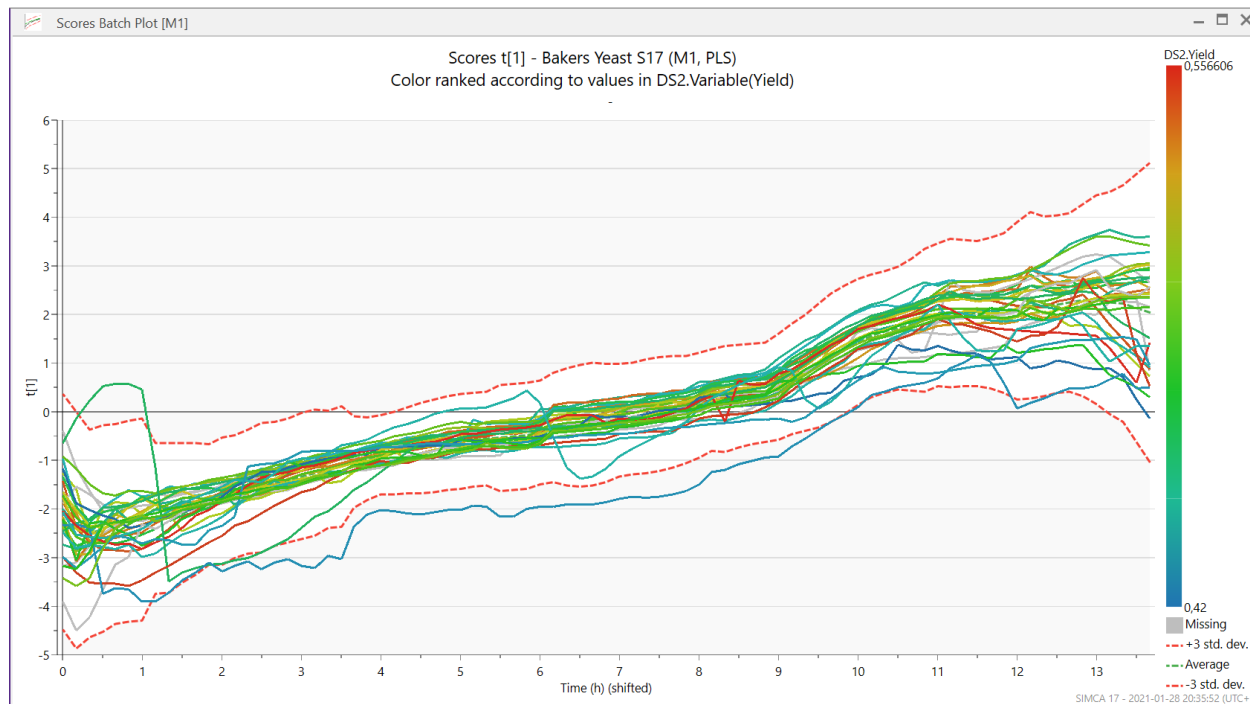
Title: Raman shift

Reverse axis

Log transform

# Plot Interactivity – BCC Coloring

- BCC can be colored by batch conditions



Data Explorer

Scores Batch Plot

Component: 1

Batch: [ All batches ]

Control limits: 3 std. dev.; Average batc...

Average: do not remove the average

Aligned

Color by: DS2.Yield

Data: DS2 : BC1\_Batch conditions

Item: VarDS

Variable: Yield

Begin color: [ Blue ]

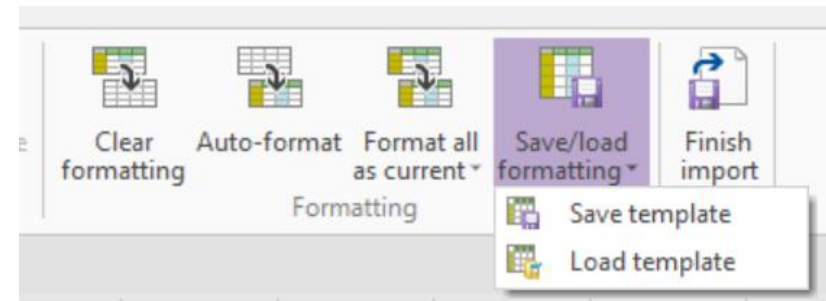
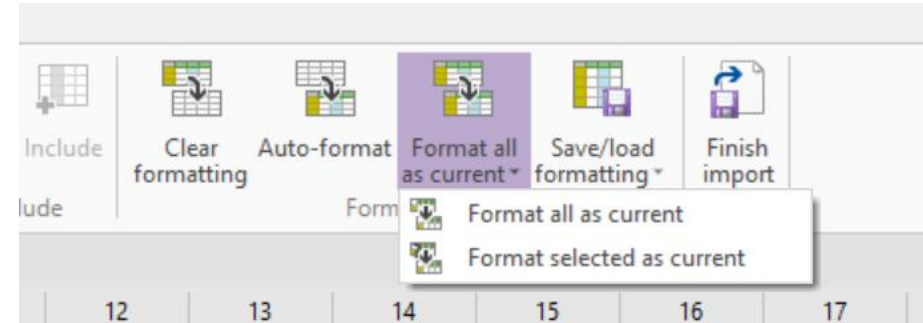
End color: [ Red ]

Missing color: [ Grey ]

Ranked

# File Import – Templates and Multiple File Import

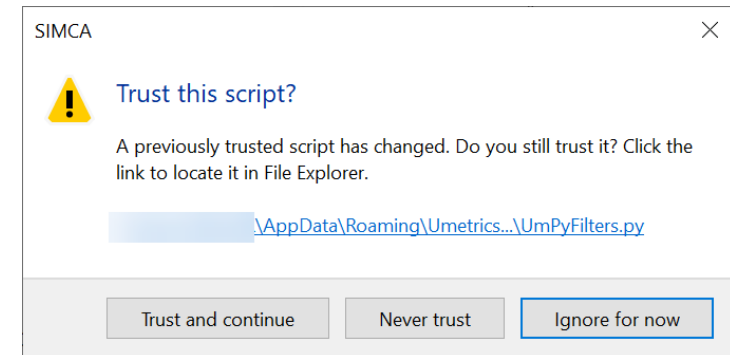
- Import multiple CSV, TXT and DIF files
  - SIMCA® supports multiple file import of text style file formats
- Template functionality during import
  - Template configures columns and rows
  - Copy current tab configuration to selected tabs
  - Save and load templates



# Python News and Additions

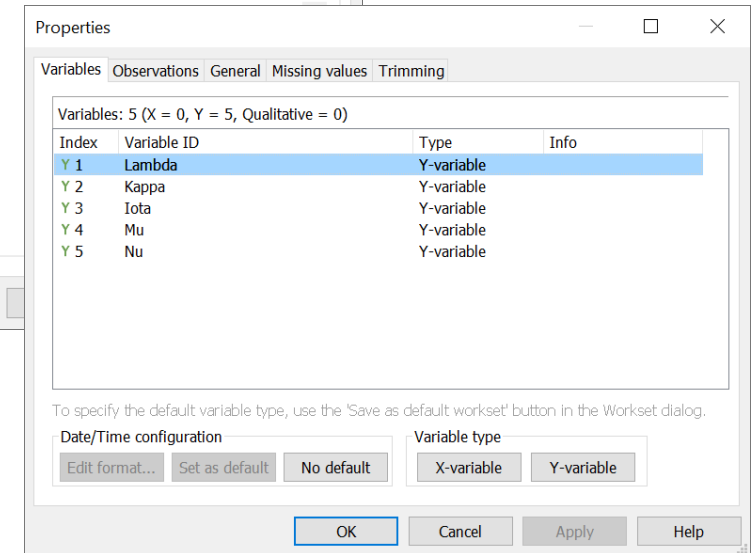
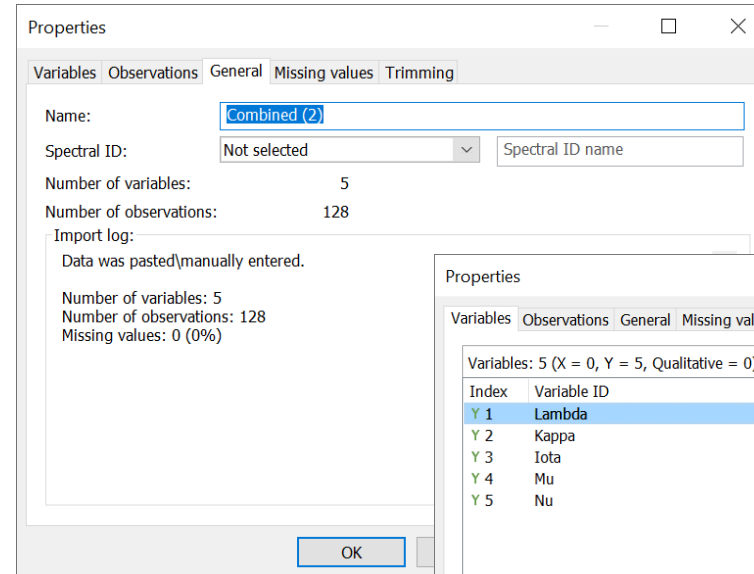
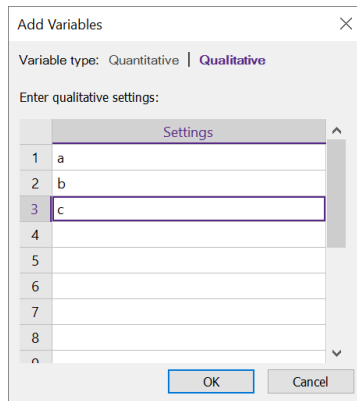
- Python updated to version 3.7.9
- NumPy, SciPy, and Pandas are pre-installed in the default virtual environment
  - Versions aligned with SIMCA<sup>®</sup>-Q and SIMCA<sup>®</sup>-online
- Python security improvements
  - User to trust new or updated Python scripts and Plugins
- New Python functionalities
  - Create, copy, and save generic plots
  - Plot properties in BCC
  - Labels and colors in plots
  - All new SIMCA<sup>®</sup> 17 added functionality

```
>>> import umypkg
>>> sys.version
'3.7.9 (tags/v3.7.9:13c94747c7, Aug 17 2020, 18:58:18) [MSC v.1900 64 bit (AMD64)]'
>>> umypkg.list()
Package          Version
-----
numpy            1.19.5
pandas           1.2.0
pip              20.2.3
python-dateutil 2.8.1
pytz             2020.5
scipy            1.6.0
setuptools       50.3.0
six              1.15.0
wheel            0.35.1
```



# Dataset Properties

- In dataset properties you can
  - Change the name of the dataset
  - Changing X and Y assignment
  - Set, or rename, Spectral ID
- Add qualitative variables to an existing dataset





# Dataset Merge

- In SIMCA® 17 un-matched content is retained in a new dataset

Dataset - New At-line data unmatched							
	1	2	3	4	5	6	7
1	Primary ID	\$BatchID	Sample time	OD	[LAC YSI] (mM)	[LAC] (mM)	[GLU] (mM)
2	942	B1	2007-05-31 09:22:00	0,995	2,67		
3	945	B2	2007-06-01 08:37:30	0,5	6,1	7,1	7,1
4	950	B2	2007-06-01 12:45:00	1,12	2,85	2,98117	4,21314

Merge Datasets

Merge datasets  
Merge two datasets by matching observations (side-by-side) or variables (top-bottom).

Side-by-side

Top-bottom

Left dataset (master): Process  
Right dataset: New At-line data

\$BatchID \$BatchID Exact

Timestamp Sample time Forward

Non matching rows: Use last good | **Insert missing**

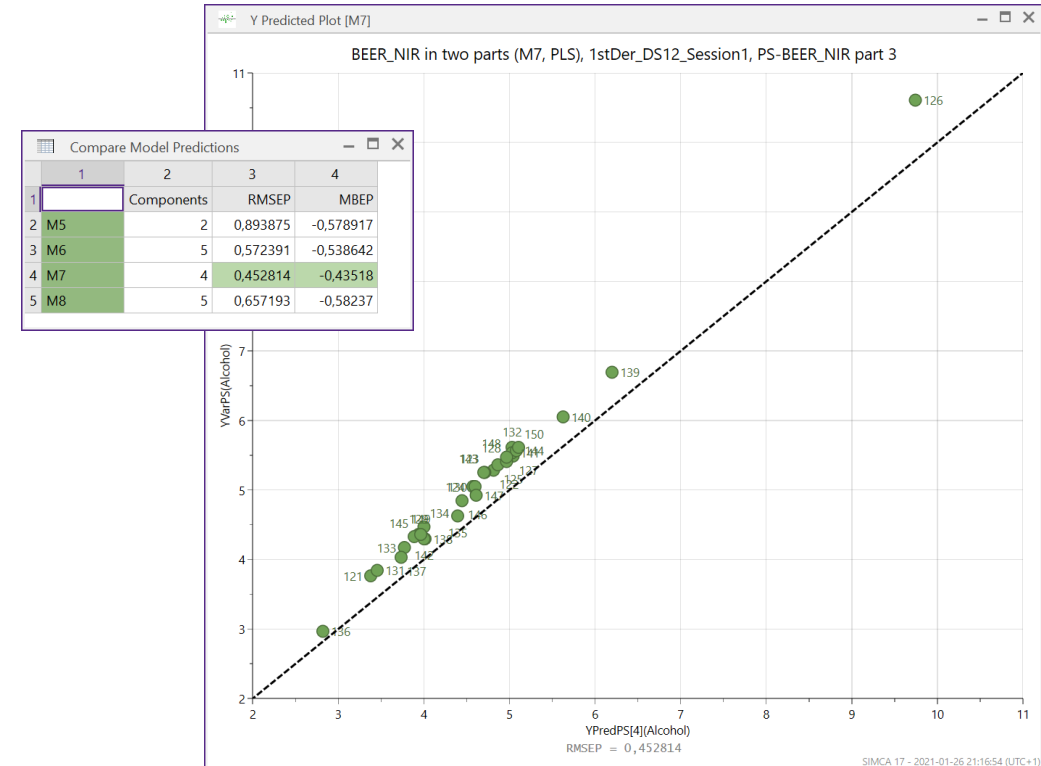
- Last good crossing over phases
- Keep matching values in a separate column
- Keep non matching data in a separate dataset

Left match 1,3 %, right match 80 %.

OK Cancel

# New Vectors – Mean Bias Error

- Mean Bias Error (MBE)
  - Average prediction error in original unit and with sign
  - Indicates if model has a bias in the predictions
    - i.e. over or under predicts
  - The new Identity line (1:1) helps to visualize bias
- Available bias vectors
  - MBEE
    - Estimated from workset observations
  - MBE<sub>cv</sub>
    - Calculated from the CV-round predictions of Workset observations
  - MBEP
    - Calculated from (true) prediction set observations
- Compare models list found in Home and Predict tabs
  - Prediction performance for models with same Y
  - Best value in each column in green



# SIMCA<sup>®</sup> Compatibility

- SIMCA<sup>®</sup> 17 is compatible with
  - SIMCA<sup>®</sup> -Q 17 and SIMCA<sup>®</sup> -online 17
- Save as SIMCA<sup>®</sup> 16 introduced
  - But not available for preprocessing functionality added in SIMCA<sup>®</sup> 17
- SIMCA<sup>®</sup> Qpe format
  - A special format for embedded predictions. Need special implementation to be used
    - Only for OEM customers
- Filters in SIMCA 17 *not compatible* with SIMCA 16
  - Smoothing
    - Savitzky-Golay Quartic, Quintic
    - EWMA
    - Moving window
    - AsLS smoothing
  - Baseline correction
    - Offset
    - Linear
    - AsLS correction
  - Normalization
    - Peak height
    - Peak area
  - Other
    - Derivatives 1st Quartic, Quintic
    - Derivatives 2nd Quartic, Quintic
    - Derivatives 3rd Quartic, Quintic
    - Derivatives 4th Quintic

# Performance – Some Examples

- Many aspects of SIMCA<sup>®</sup> performance and responsiveness have been addressed in SIMCA<sup>®</sup> 17
- Plot creation and interactivity
  - e.g. line plots and coloring of line plots
- Opening of projects is approximately 67% faster
- Drop-down lists get populated with values in a fraction (<1%) of the time
- And many more things to improve the SIMCA<sup>®</sup> experience

Thank You for Your Interest in  
SIMCA<sup>®</sup> 17

Don't forget to check out the instructional videos

**SARTORIUS**