



Simplifying Progress

SIMCA® 16 What's new

SARTORIUS

SIMCA® 16 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 adopted for both single and multi-block analysis.
- SIMCA® 16 delivers improvements in three focus areas for the current, experienced, user as well as the beginner
- Usability improvements
 - By providing new users with a guided introduction to SIMCA and existing users with smoother plot interactivity, quick raw data visualization, and relevant workflows we want to give all users a rewarding data analytic experience
- Model use and understanding
 - Giving users new and extended tools for interrogating and experimenting with generated models, we help turn model knowledge into real world interpretations
- New technology
 - Aiming at driving innovation within the multivariate analytics field we provide our users access to new and advanced analytical technology

SIMCA[®] 16 – Some highlights

- Usability improvements
 - A “Start screen” when no project is open
 - A guided “getting started” tour for beginners
 - Flexible merging of datasets
 - Context based ribbon structure
 - Workset wizard with integrated quick info pane
 - Data explorer side pane with new and updated information panes
 - Plot properties are in a context based side pane
 - An updated report generator
 - Improved legend handling for BCC multiplot and others
- Model use and understanding
 - Score space exploration
 - Multivariate solver
- New technology
 - MOCA – Multiblock Orthogonal Component Analysis
- Python additions and improvements
 - Install external packages inside SIMCA
 - Python plugin creation for File import formats and Data preprocessing (filters)

Learning what's new in SIMCA 16

- In the following slides an overview is given to the changes and additions made in SIMCA 16
- For more descriptions on how to use the features, please watch the videos that can be reached from the start page of SIMCA 16
- You can also search for Sartorius Data Analytics on YouTube

SARTORIUS

Simplifying Progress



Usability improvements

SIMCA[®] start screen and feeds

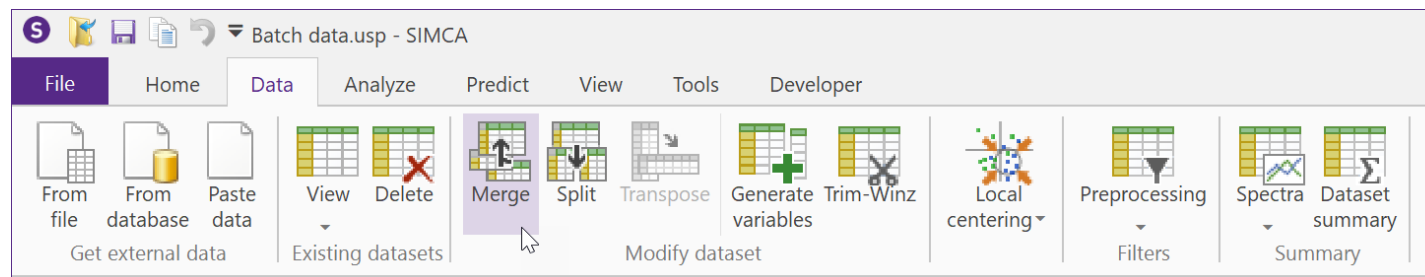
- When starting SIMCA without having any active projects, the start screen let's you chose what to do

The screenshot shows the SIMCA start screen with several callouts:

- Guided tour for beginners**: Points to the 'Guided tours' section in the left sidebar.
- What's new videos**: Points to the 'What's new' section in the left sidebar.
- Select type of project to create**: Points to the 'New' section with options for Regular project, Batch project, Omics project, and Spectroscopy project.
- List of recent projects**: Points to the 'Last opened project' and 'Older' sections.
- Links to recent blogs and videos**: Points to the 'Feeds' section on the right.

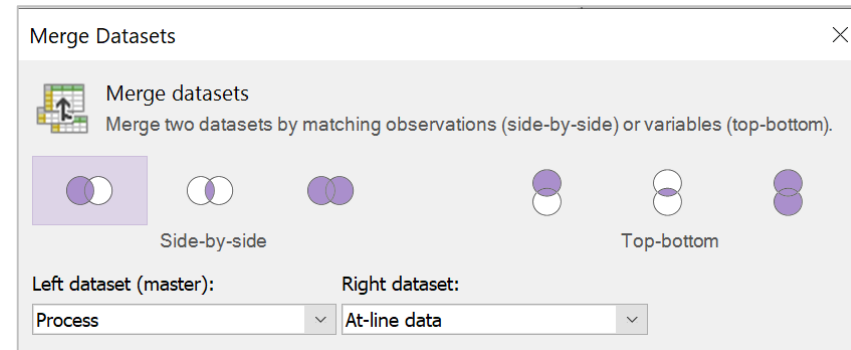
Flexible merging of datasets

- Many users struggle to align and merge data before they get the data into SIMCA
 - In earlier versions you could only merge datasets based on primary ID inside SIMCA
- In SIMCA 16 we introduce a flexible merging of dataset
 - Merge data Side-by-Side or Top-Bottom
 - Merge by identifiers and variables
- Merge is found on the Data tab



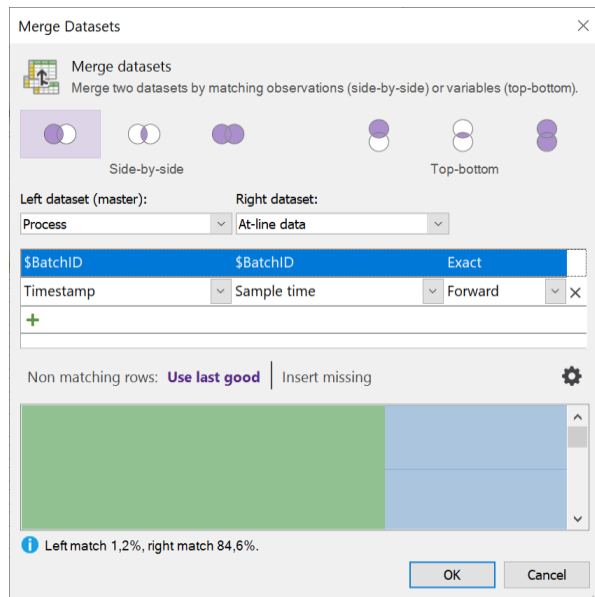
Flexible merging of datasets

- Types of merge
 - Master/slave, intersection or full merge
- Options
 - Exact, nearest and intrapolate
 - Fill with last good value or empty cells
- Typical application for Side-by-Side merging
 - At-line data sampled at long and irregular intervals with process data sampled at high frequency
- Typical application for Top-Bottom merging
 - Spectral data with different resolution or not identical wavelengths

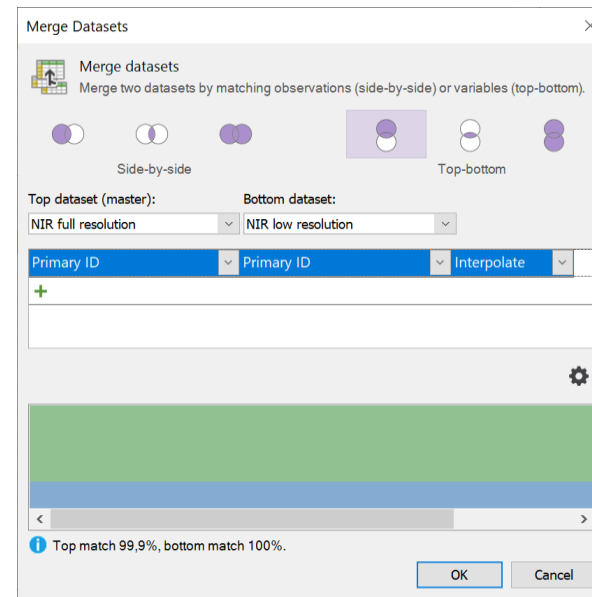


Flexible merging of datasets

- Side-by-Side example
 - Process data and At-line data
 - Left merge (master/slave)
 - Match by Batch ID and timestamps
 - Fill with last good value

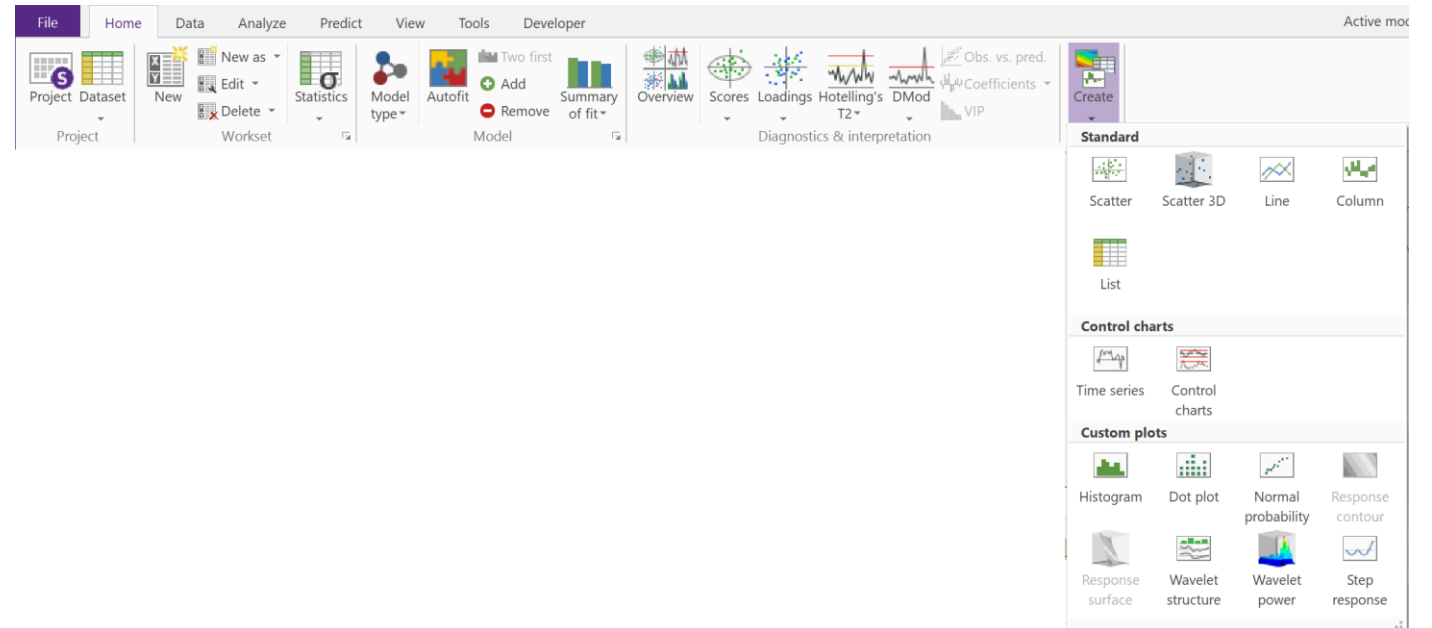
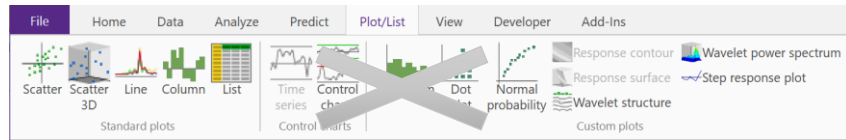


- Top-Bottom
 - Spectral data of different resolutions
 - Top merge (master/slave)
 - Match by Primary ID (Wavelength)
 - Intrapolate



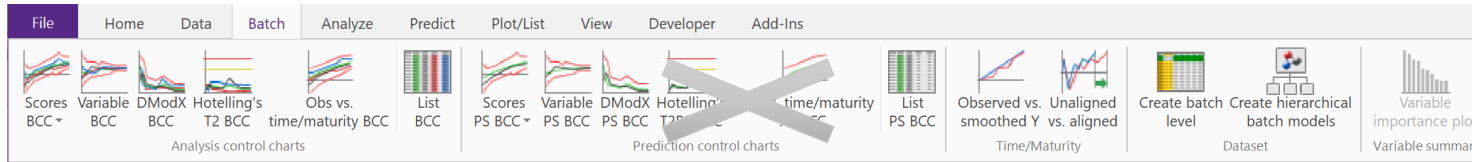
Cleaned ribbon structure

- Plot/List tab removed
 - All generic plots are available in the Create gallery on the Home tab



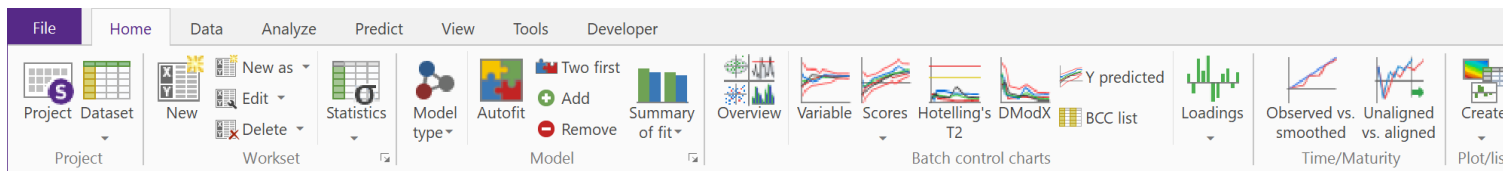
Context based ribbon structure

- Batch tab removed for batch projects

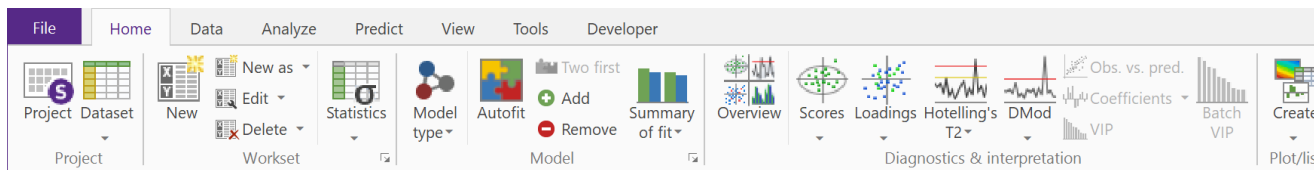


- The Home tab provides relevant plots for the selected model type (BEM or BLM)
 - The Batch prediction plots are now in the Predict tab

BEM



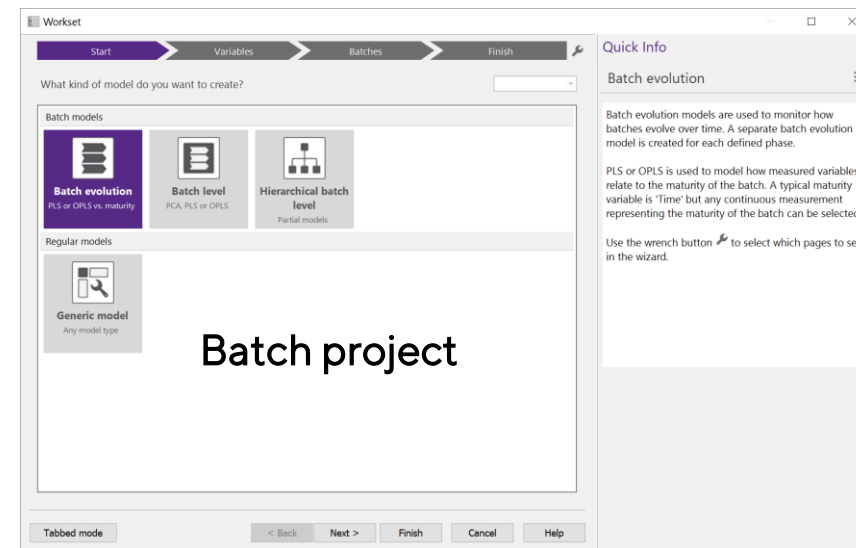
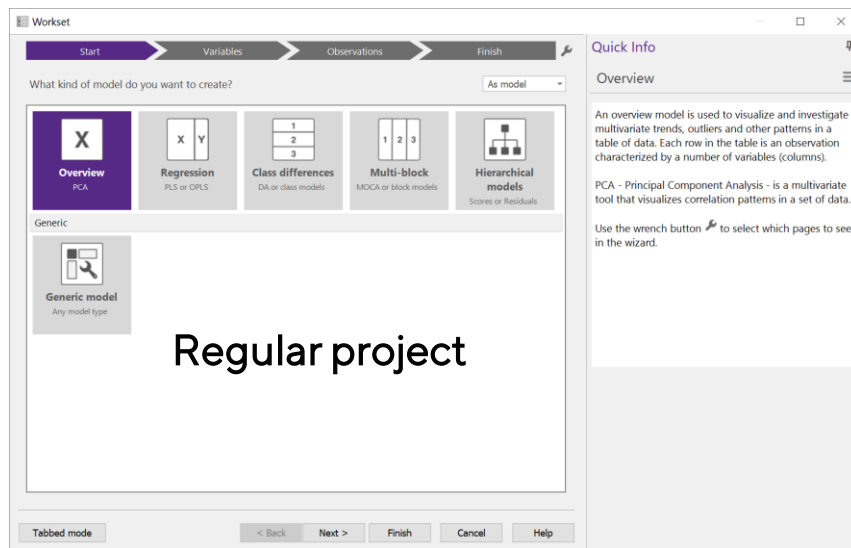
BLM



- Batch level dataset creation is now handled by the Workset wizard.

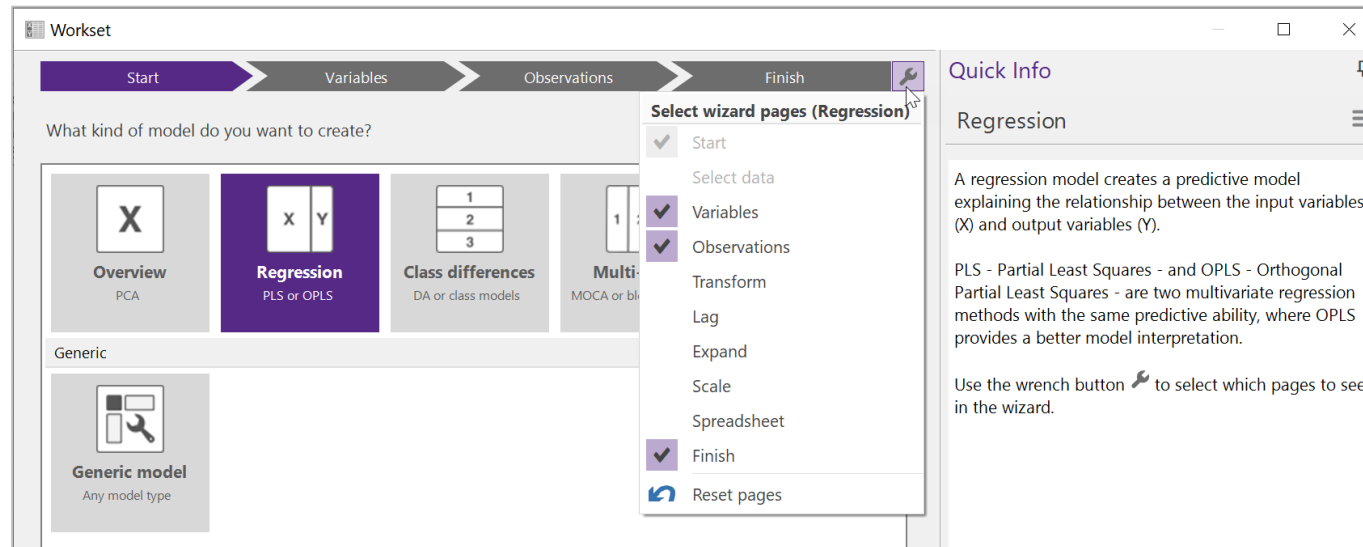
Workset wizard with integrated Quick Info

- The workset dialog has been converted to a workset wizard
- The wizard starts with an Start page where you define your objective
 - Different objectives for regular and batch projects
 - The selected objective controls what options will appear later in the wizard



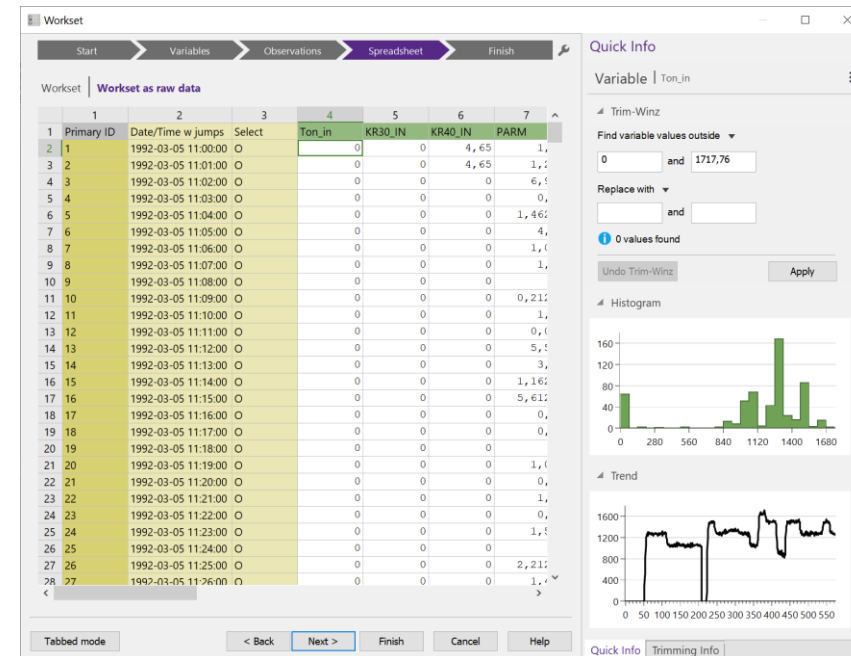
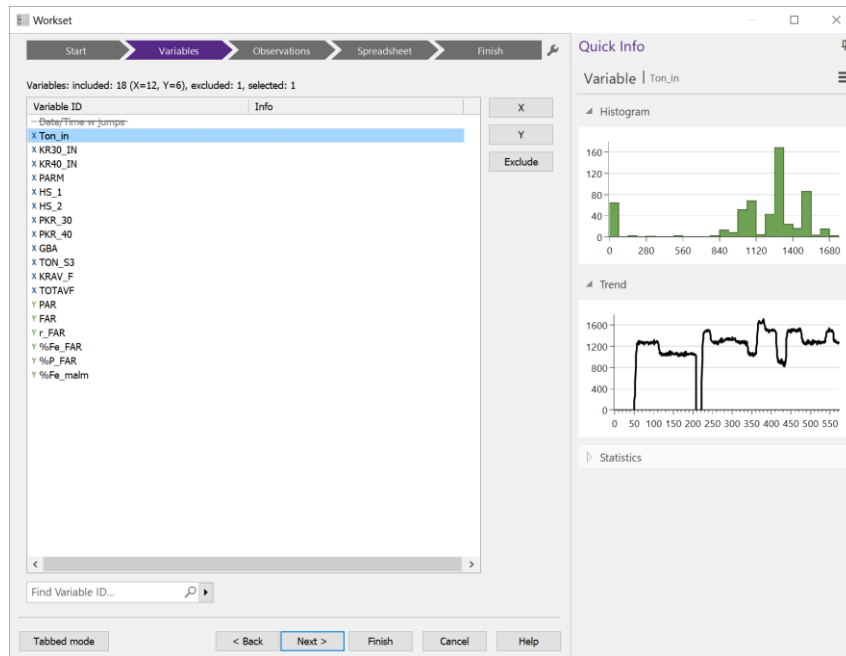
Workset wizard with integrated Quick Info

- A progress bar indicates where you are in the workset definition
 - Clicking Next will lead you through the workset creation
 - You can also click in the progress bar to go directly to a selected step
- The steps in the progress bar can be modified to include other steps
 - The modifications will be remembered by objective



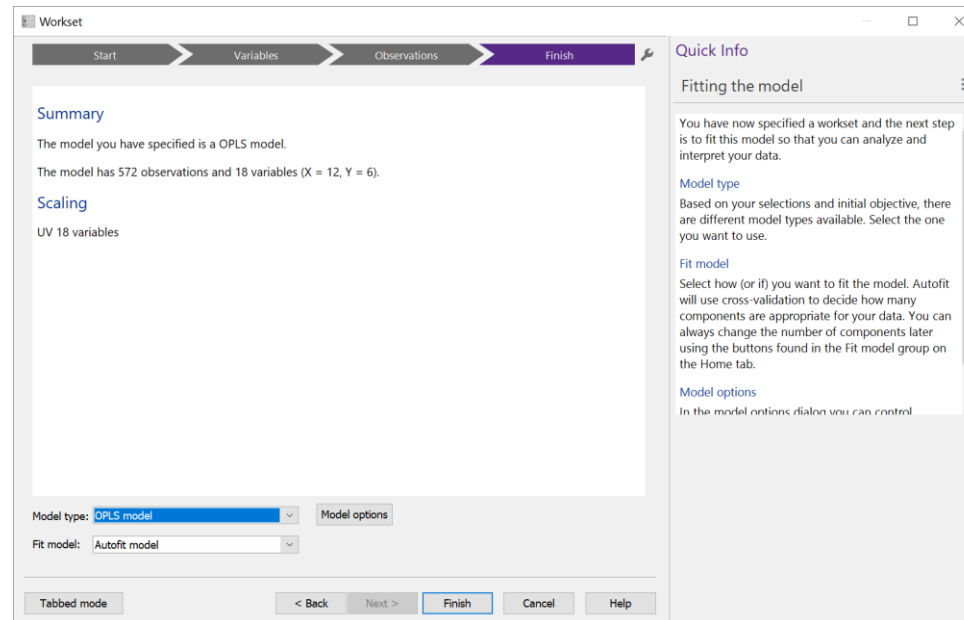
Workset wizard with integrated Quick Info

- The Updated Quick Info is now integrated on all steps in the workset wizard
- On the Spreadsheet step (needs to be added) you have access to Trim-Winz
 - Trimming done on the workset is only affecting the individual model and can be brought to SIMCA-online



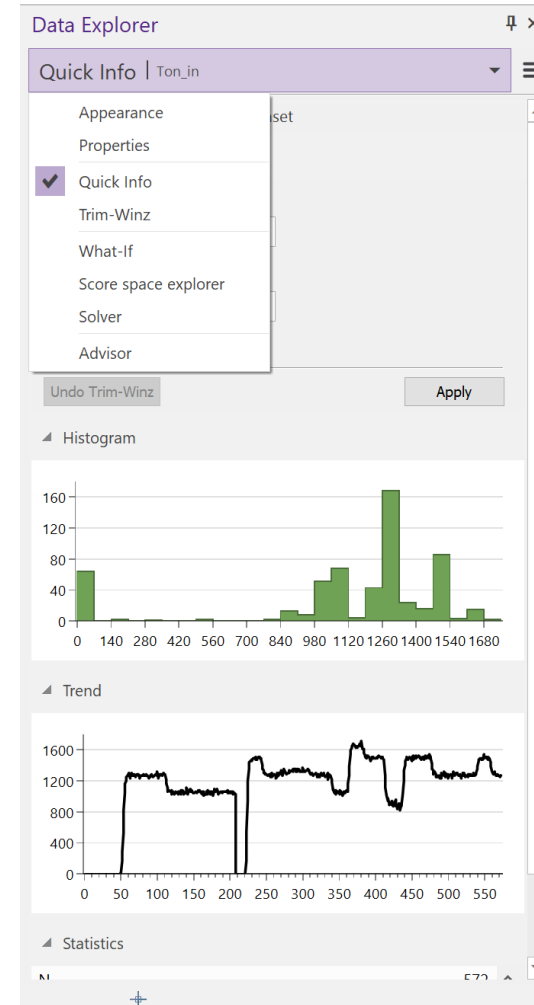
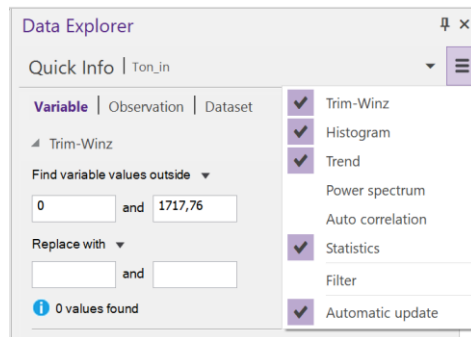
Workset wizard with integrated Quick Info

- The last step in the wizard let's you chose model type and if you want to fit the model automatically
 - The selection is remembered
- Model options (e.g. CV-groups and significance) can be accessed on the last page of the workset wizard



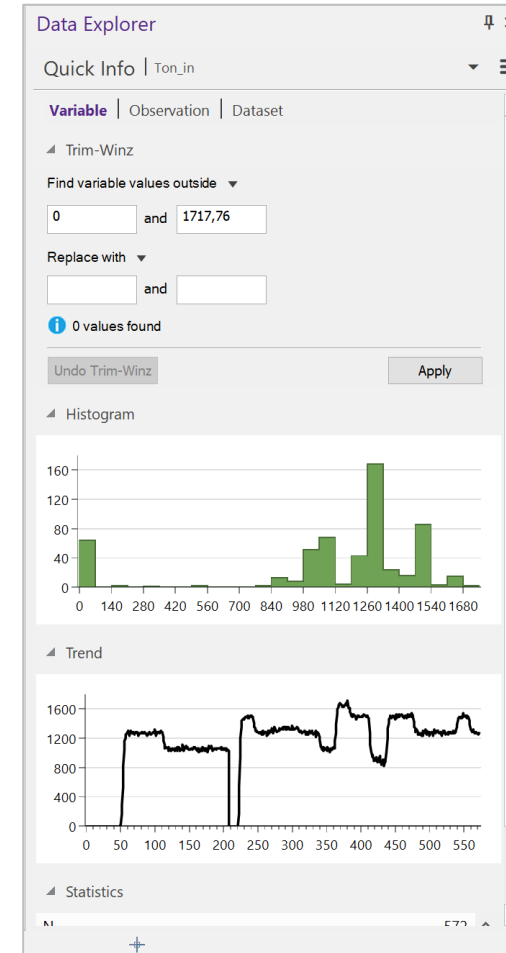
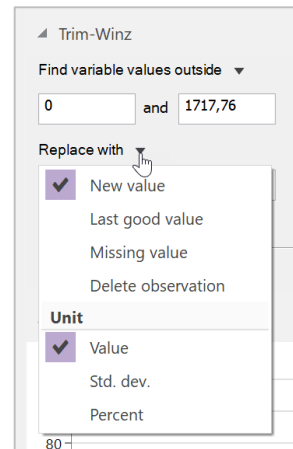
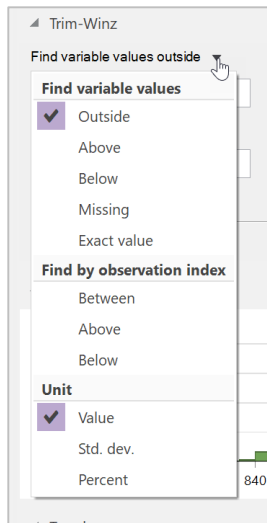
Data Explorer side pane

- The Data Explorer pane holds a collection of context based informative panes
 - Some have been modified and extended and some completely new side panes
- The Data Explorer pane will show information relevant to the active window in the SIMCA work area
- An options menu for customization is available for each individual pane



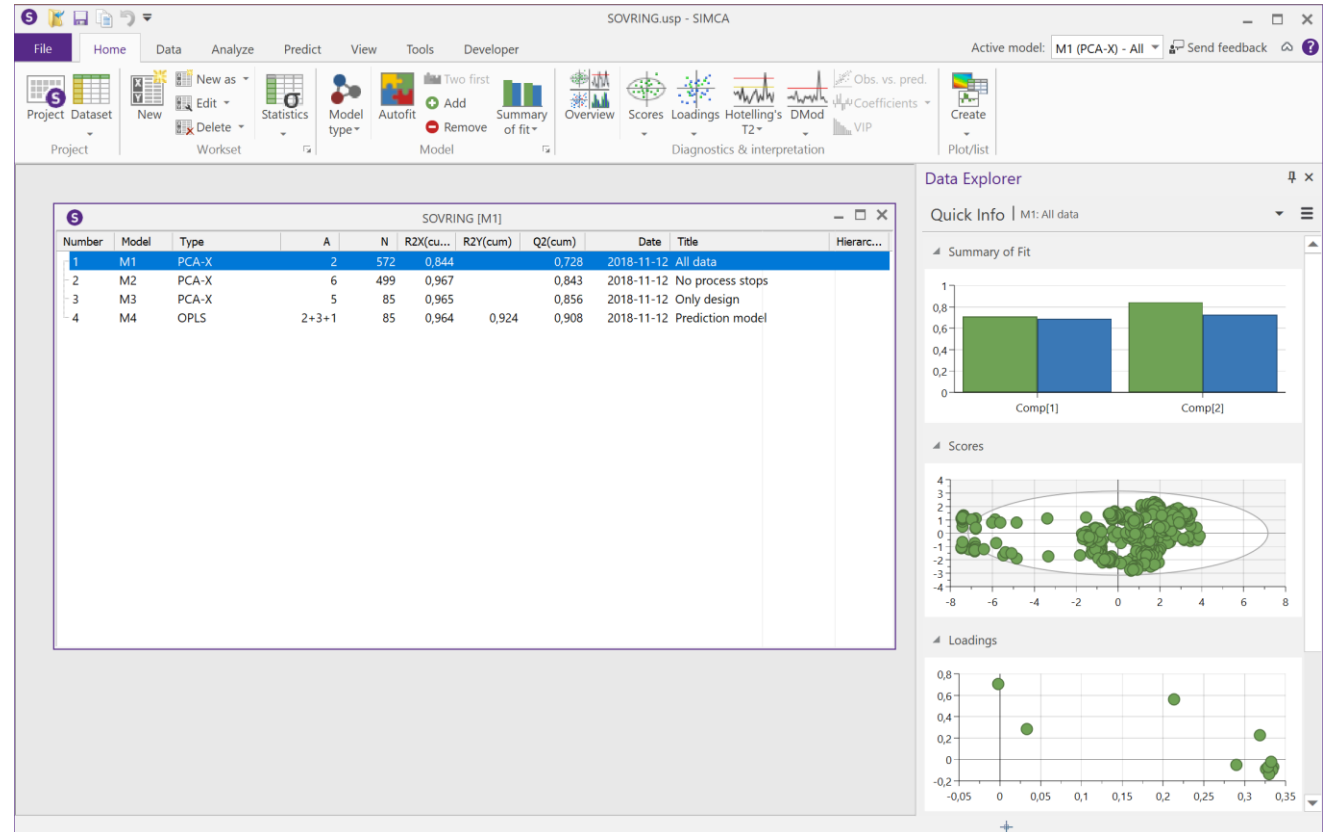
Data Explorer side pane – Quick Info with Trim-Winz

- An extended Quick Info is activated when clicking in lists, e.g. Datasets
- Trim-Winz has it's own section but is also included in the Quick Info for datasets
 - Trim-Winz has been extended with more search and replace functions including find "Missing" and "Exact"



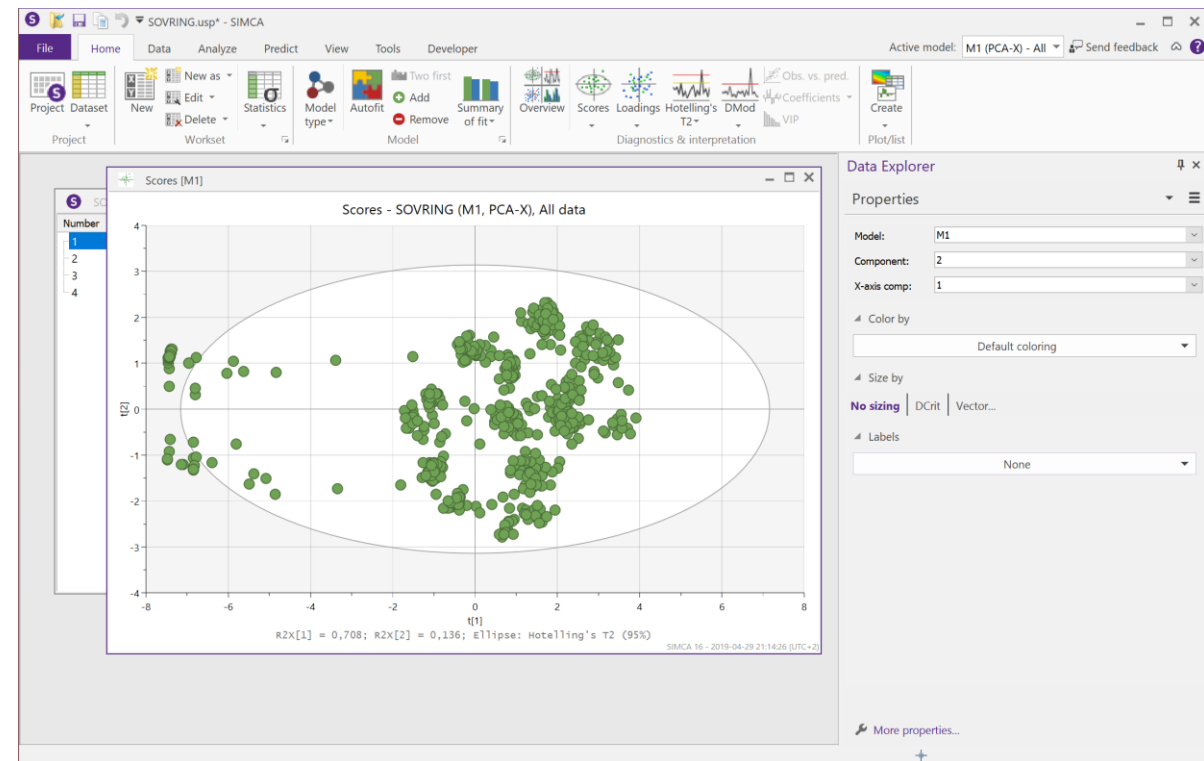
Data Explorer side pane – Quick Info on models

- When selecting a model in the project window, a selection of model plots appear
 - You can add or remove plots using the options menu next to the Quick Info header
 - Double-clicking one of these Quick info plots pops it out as a full size plot



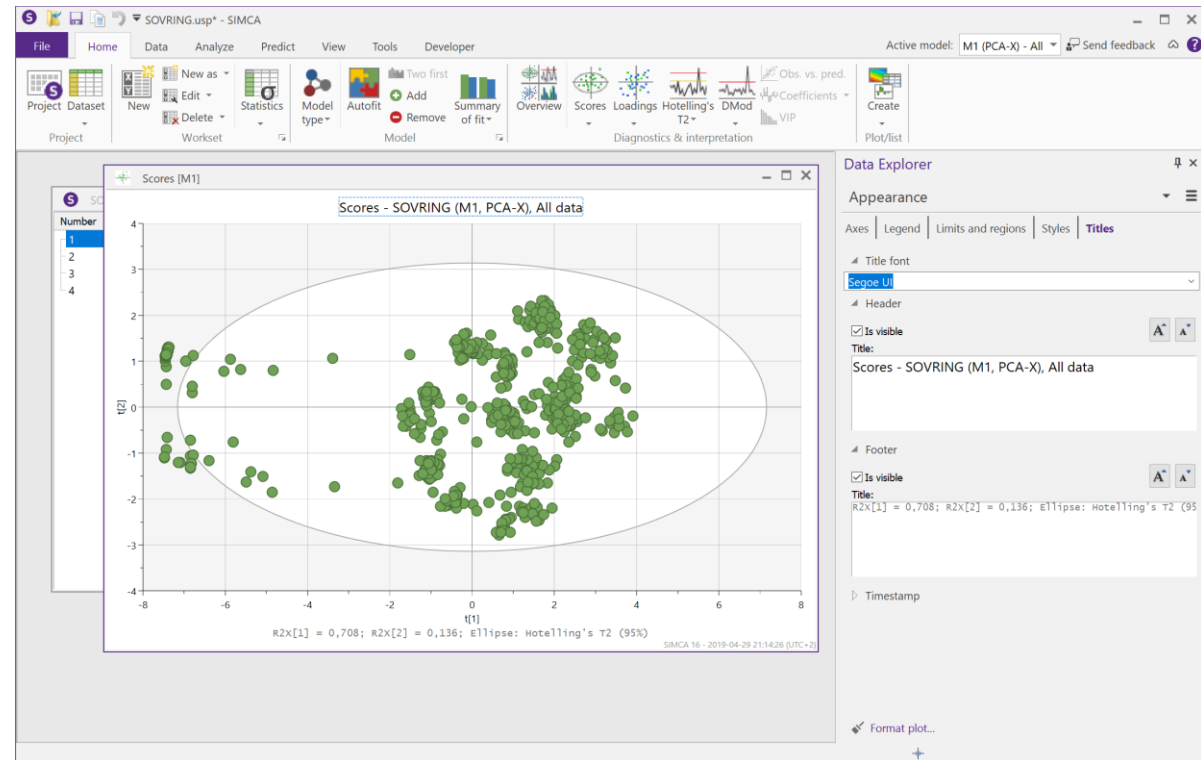
Data Explorer side pane – Plot properties

- The Properties pane opens up when a plot becomes active
 - Control content, colors and labels



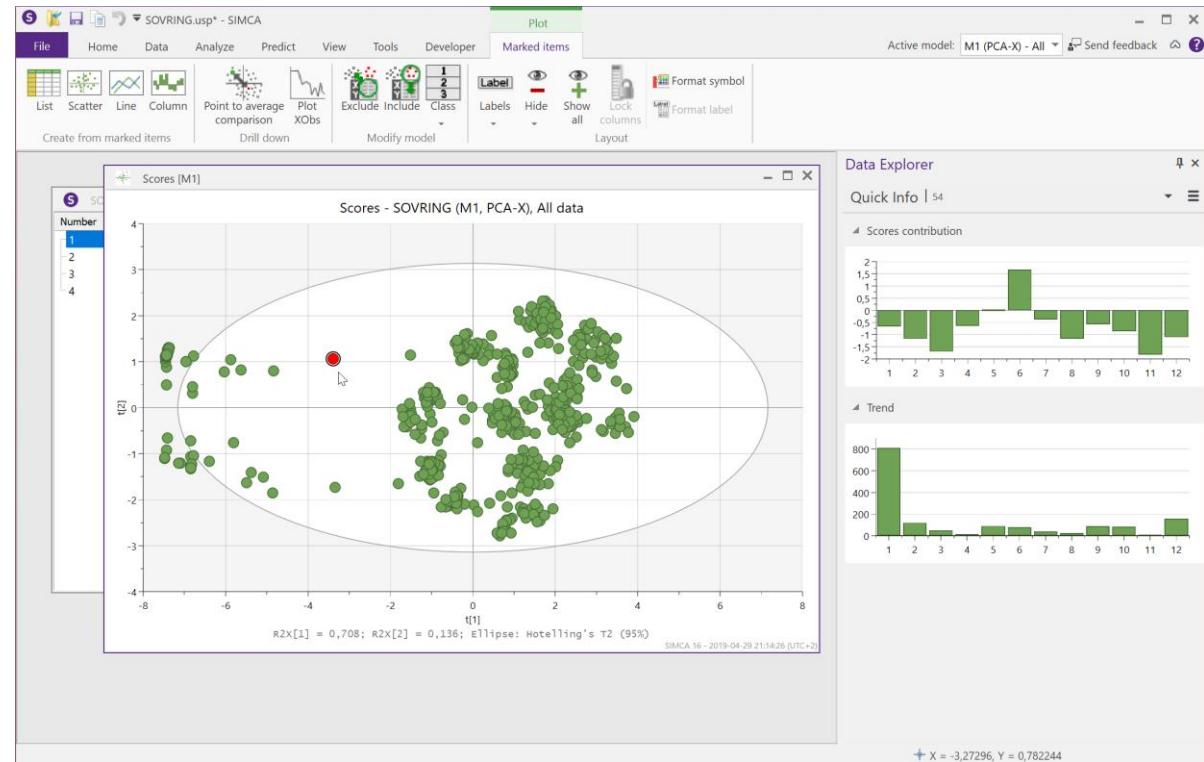
Data Explorer side pane – Plot appearance

- When clicking on a title or an axis, the new Appearance pane becomes active letting you modify the appearance of the active plot



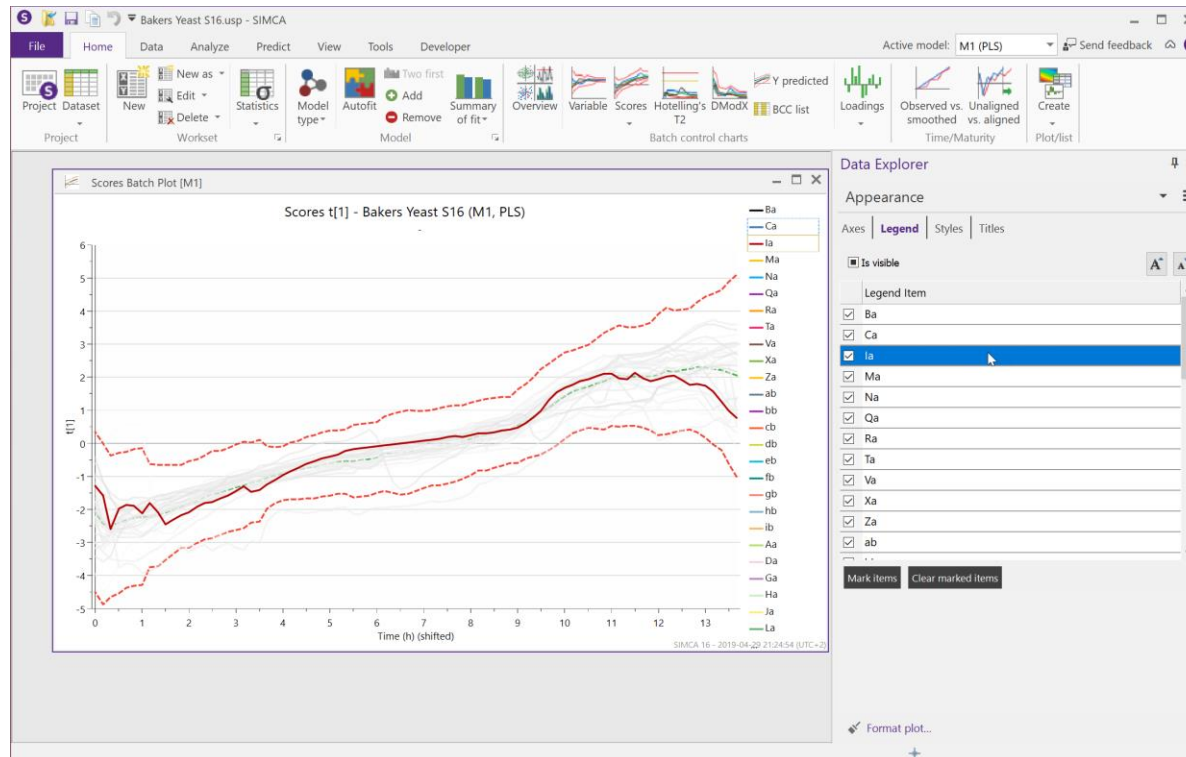
Data Explorer side pane – Observation contribution

- Selecting an individual point or a group of points in an observation plot will open up the Quick Info with a contribution plot



Data Explorer side pane – Legend interaction

- Clicking on the legend opens the new, scrollable, legend pane
 - The legend pane lets you highlight objects and select which objects to display



An updated report generator - Formats

- The report generator has been updated and now allows for saving and exchanging content in a variety of formats
 - The report always starts as HTML
 - The report is by default saved in the SIMCA project file and can also be exported to an external file

The screenshot displays the SIMCA report generator interface. The main window, titled "Bakers Yeast S16* - Report", shows a report preview with the following content:

Insert here the contribution for batches with time points outside 2 times the Dcrit

Batch Control Charts for the Workset batches

In the following plots you can display the batches aligned and non-aligned.

Scores Control Chart

Scores (t1) - Bakers Yeast S16 (M1, PLS)

The score control chart displays the selected score value (t[1] by default) over time for one or several selected batches. The chart also shows the average batch (green) and the 3 sigma limits (red). A behaved batch stays inside the red control limits during all its evolution, while a batch going outside indicates process problems.

Insert here contribution plots for region of batches outside limits

DModX control chart

DModX - Bakers Yeast S16 (M1, PLS)

The DModX control chart displays the evolution of DModX values for various batches over time, with a legend on the right side.

Report | HTML source

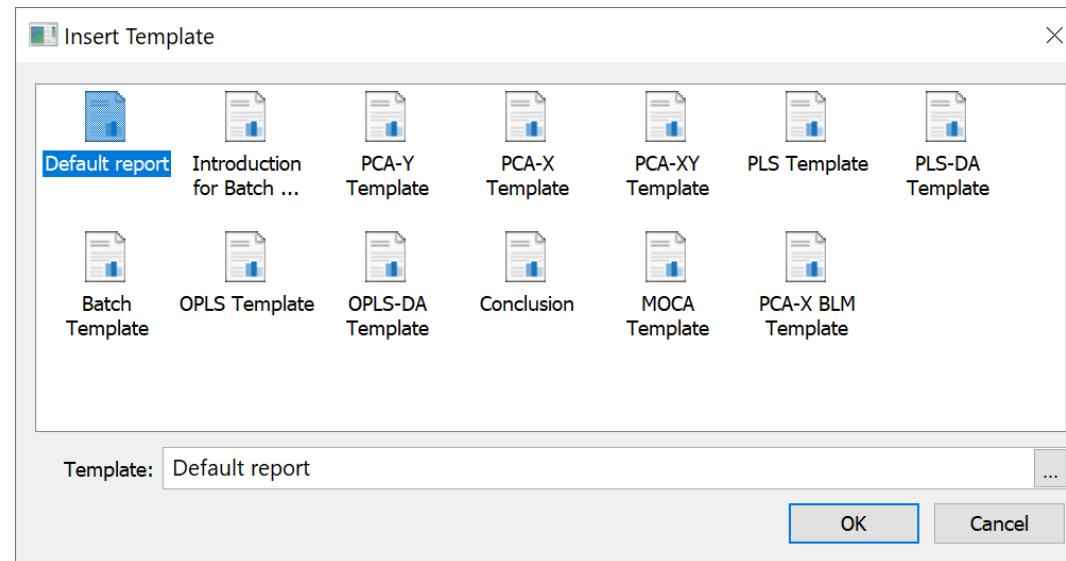
An "Export to PDF" button is visible in the top right corner of the report preview. A context menu is open over the "Export to PDF" button, listing the following options:

- Microsoft Excel
- Microsoft Publisher
- Microsoft Word
- Notepad
- Launch Windows App
- Google Chrome
- Internet Explorer

A second window titled "Bakers Yeast S16.pdf - Adobe Acrobat Reader DC" is open to the right, displaying the rendered PDF version of the report, which includes the same charts and text as the preview window.

An updated report generator - Templates

- Each model type has its own default report template
 - The included templates contain descriptive texts and some standard plots
 - The default templates can be modified to include other plots and descriptions to suit your needs
 - The templates can be shared with other SIMCA users



Multiplots for Summary of Fit and Batch Control Charts

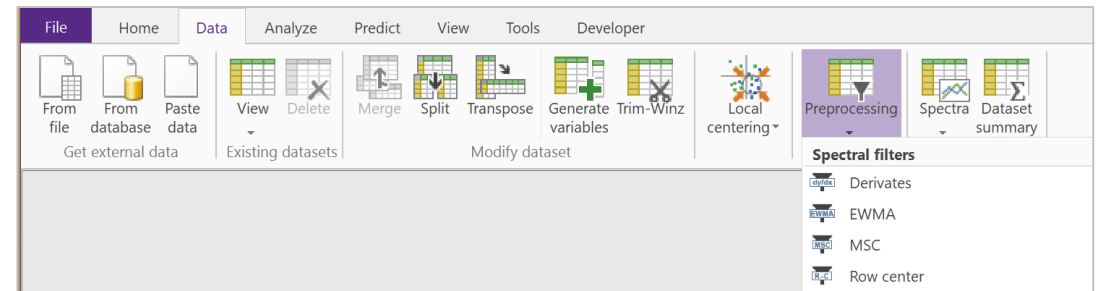
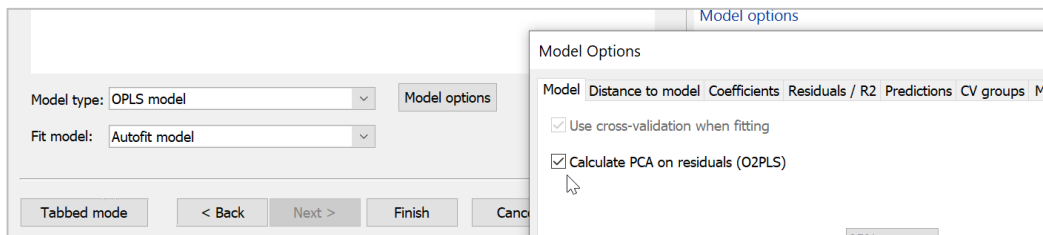
- Multiplots have been implemented for some plots
 - Summary of fit for class or phase models
 - Batch control charts (BCC:s) for multiple phases



SIMCA® 16 – Other additions and modifications

- OPLS has been extended to include all O2PLS calculations
 - O2PLS has therefore been removed
 - Select Regression objective in Workset wizard and on the Finish step, select OPLS
 - Use model options to include the additional O2PLS components

- Spectral filters have changed name to Preprocessing (on the Data tab)
 - The new Python preprocessing plugin has more capabilities than only spectral filtering



SARTORIUS

Simplifying Progress



Model use and understanding

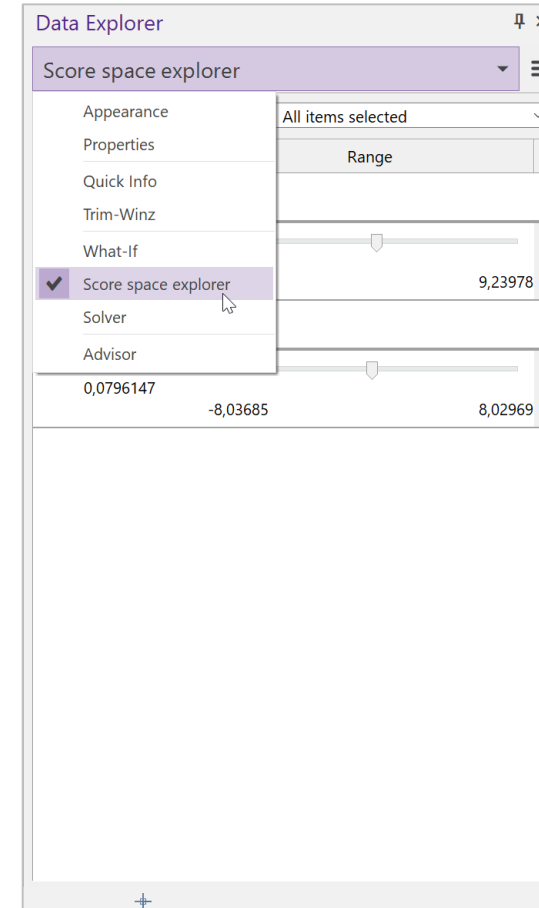
Model interrogation innovations – Score space explorer

- Works on regular projects
- Score space explorer
 - Several customers have asked for a tool to easier investigate and convert score points to original variables
- Score space explorer scenario
 - You have a PC model of historical batches of raw materials and see an area in the score plot that is empty and you want to know what type of material is missing



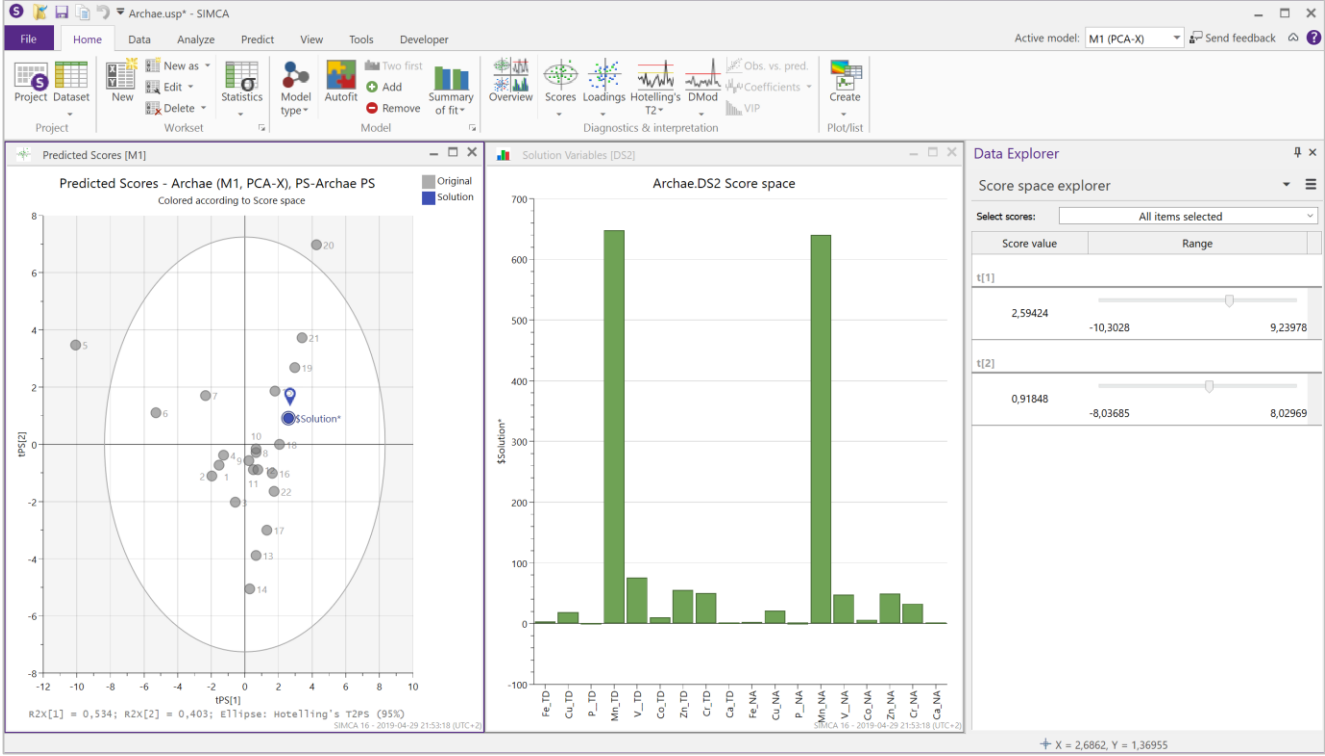
Score space explorer

- Finding the variable profile for a specific location in a score scatter plot
 - The multivariate correlation is retained
 - Found in the Data explorer pane
- The interface consist of two plots and a slider section in the side pane
 - Predicted score scatter plot
 - An observation variable column chart for the selected point
 - A set of score vector sliders (in the side pane)



Score space explorer

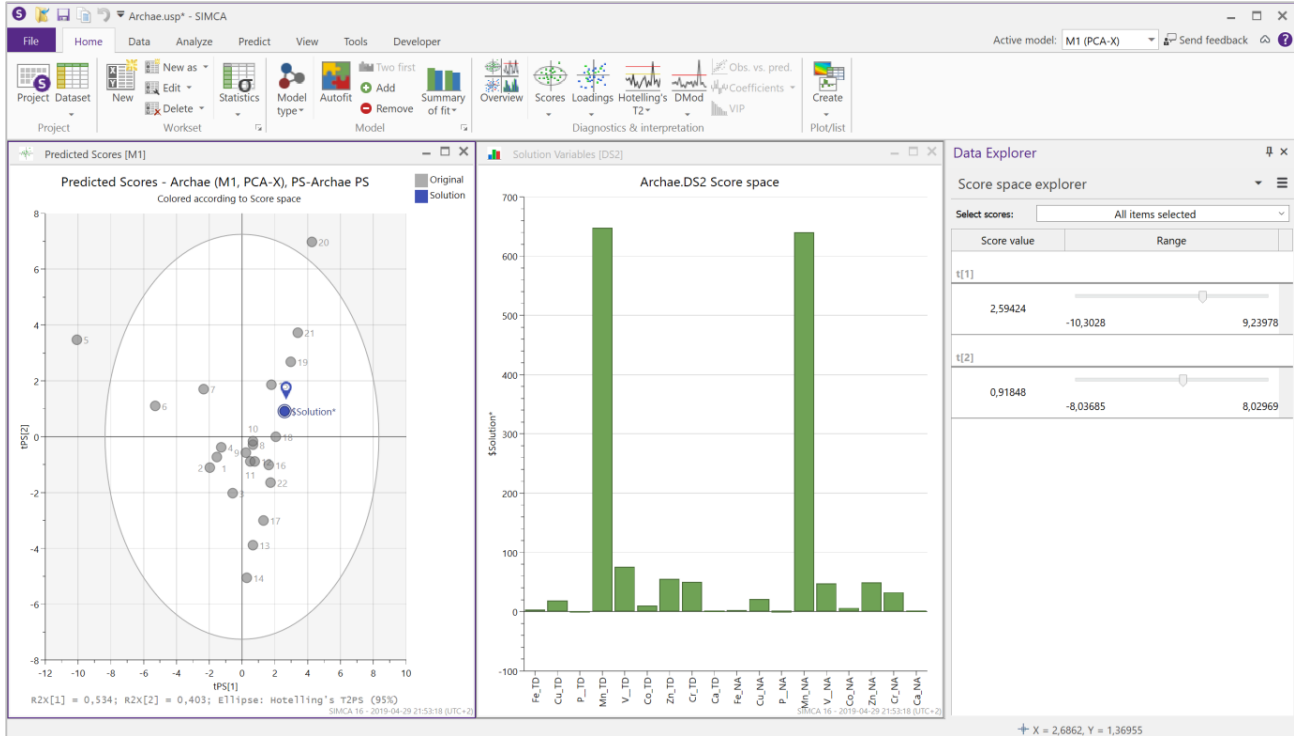
- The sliders and the score scatter plot indicate the selected point
- The variable plot provides the profile corresponding to the selected score point



Score space explorer

- You can interact with the Score space explorer in two ways
 - Use the sliders to move the selected point in the score plot or
 - Click in the scatter plot

The variable plot instantly displays the corresponding variable settings



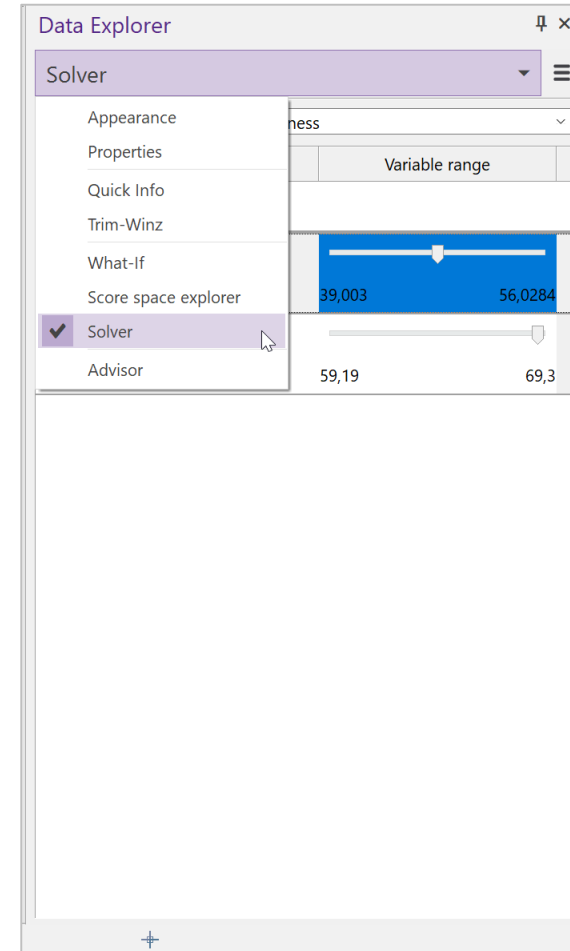
Model interrogation innovations – Multivariate solver

- Works on regular project OPLS models
- Multivariate solver
 - Similar to an optimizer but with fixed target values (i.e. not optimize or minimize)
 - Thanks to OPLS, the solver gives instant answer without iterations
- Solver scenario
 - You have a predictive OPLS model and some of the inputs have to change (e.g. a new batch of raw material) and you need to know how to compensate to reach the same output
 - Some X-variables are fixed and target Y is known



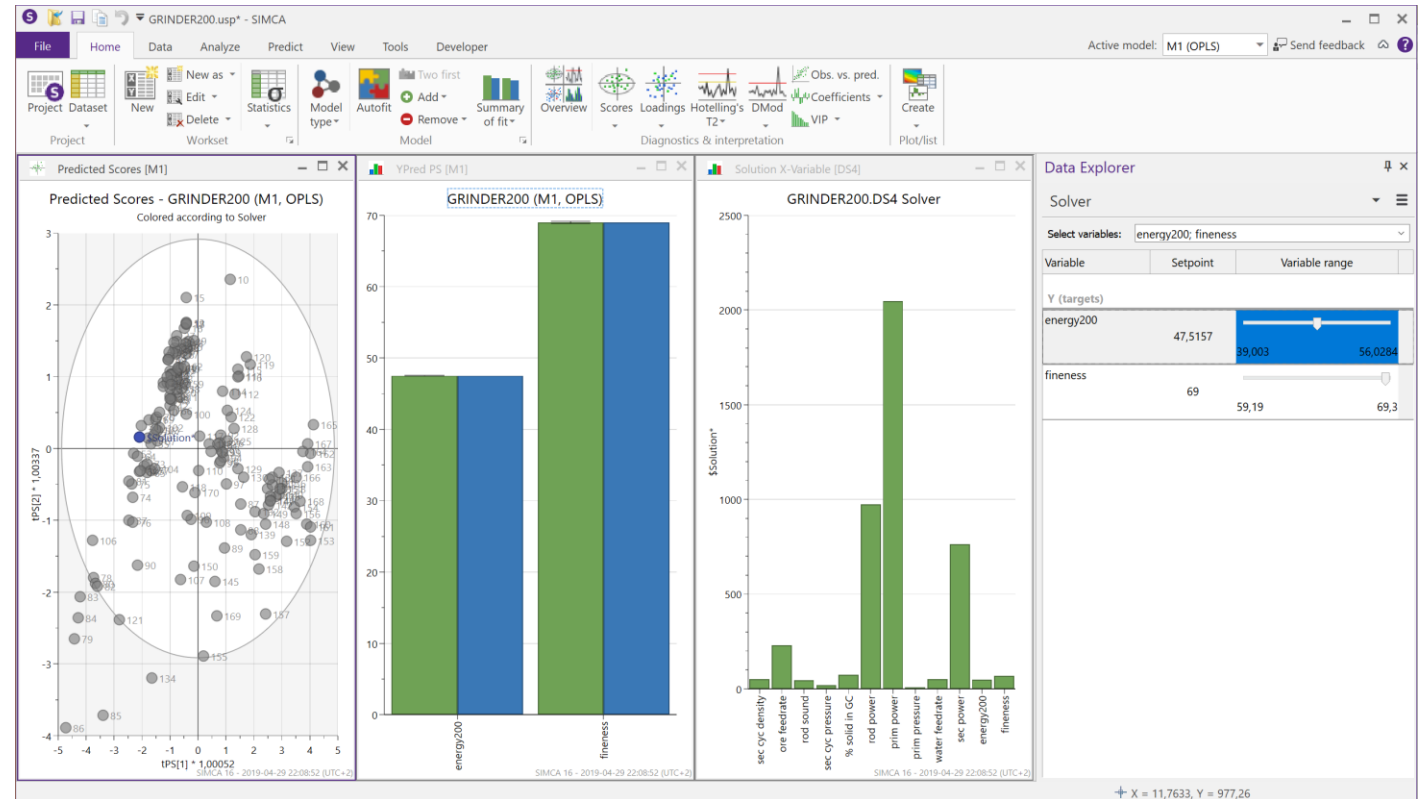
Multivariate solver

- The solver is accessed through the Data Explorer pane
- The solver works for single or multiple Y:s
 - It is recommended that the Y-variables are uncorrelated or only weakly correlated
- The solver works in two steps
 - The score values, corresponding to the desired target Y, are calculated
 - The variable settings for the score point are calculated
 - Only solutions where the multivariate pattern is retained are considered when no X-constraints are used
- When using constraints on X-variables the correlation pattern is broken
 - When the suggested solution no longer belongs to the model (PModX+) a warning will be given



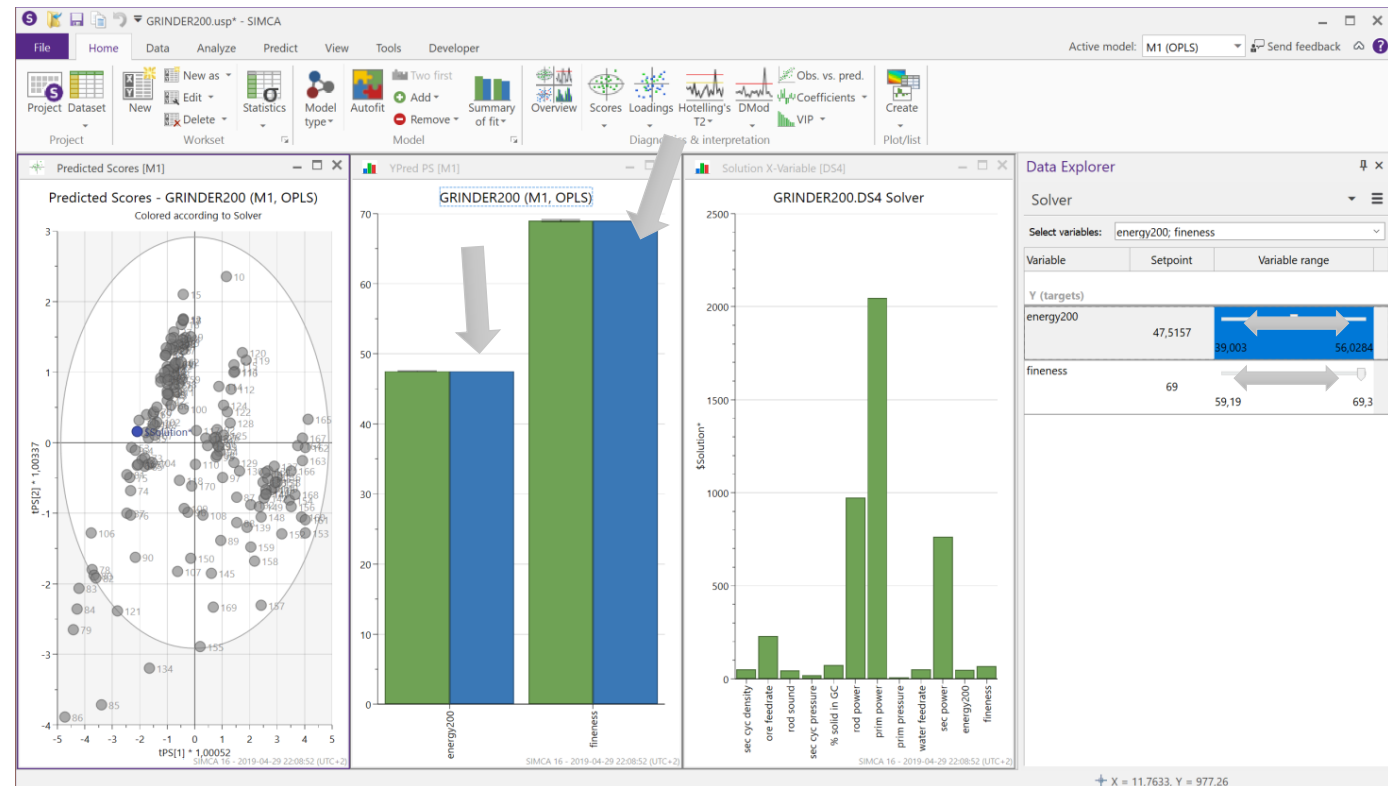
Multivariate solver

- The interface consist of three plots and a slider section in the side pane
 - Predicted score scatter plot
 - A column plot showing desired and predicted Y-values
 - An observation variable column chart for the suggested solution



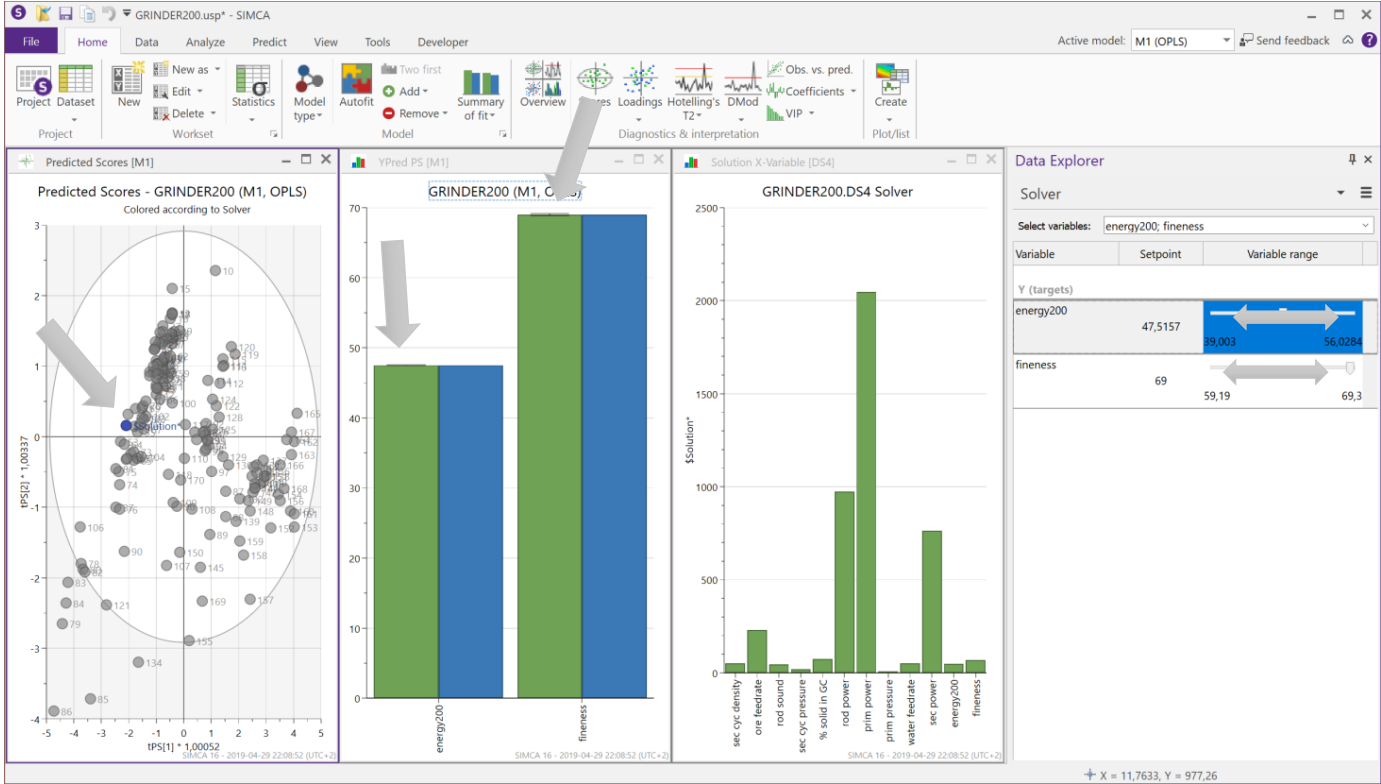
Multivariate solver

- The sliders are used to set the target values for Y
 - You can also type in the Setpoint field to the left of the sliders
- The desired Y target values are displayed as blue bars in the middle plot



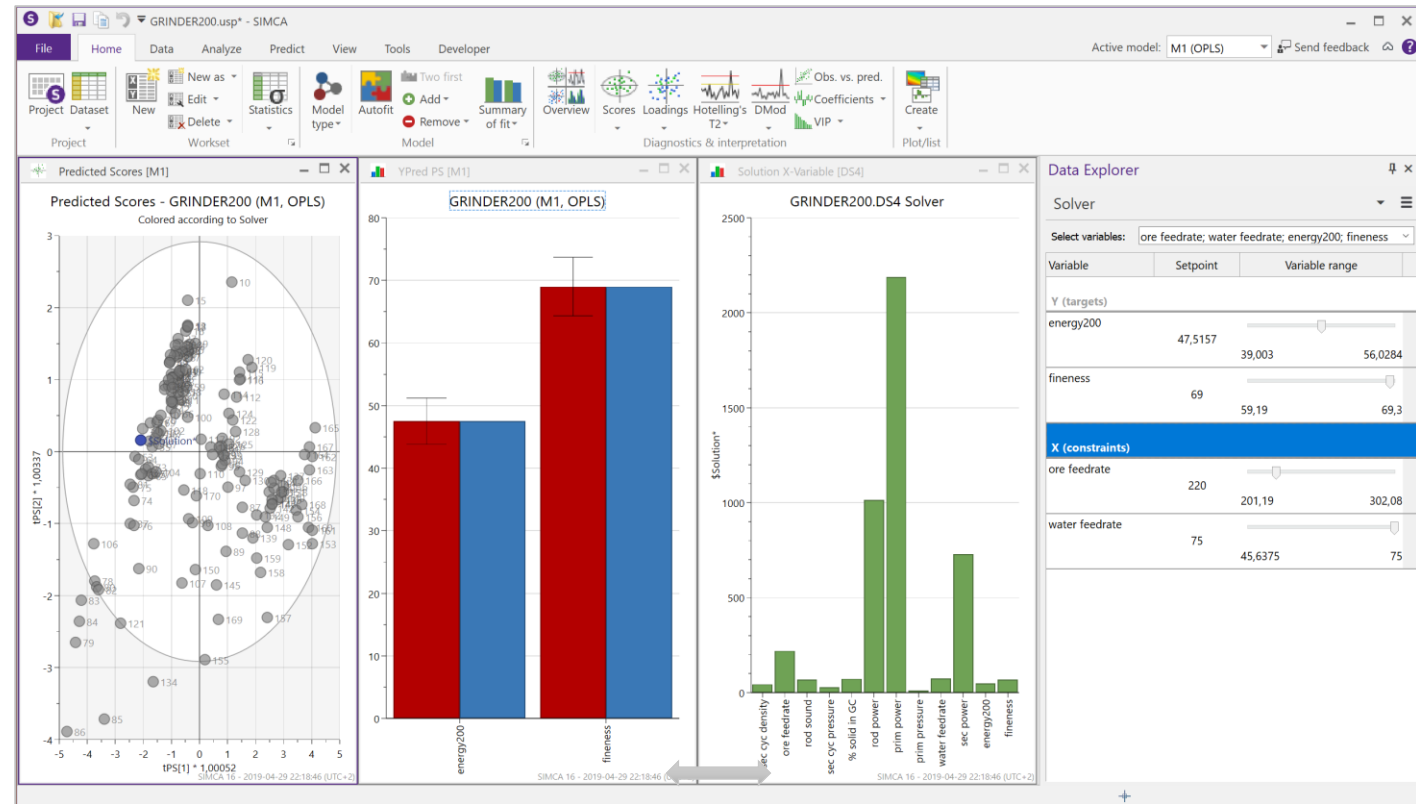
Multivariate solver

- The score scatter plot indicate the suggested solution
 - And the green bars in the middle plot show the predicted Y-values
- The variable plot to the right shows the variable profile for the suggested solution



Multivariate solver

- When using constraints on X-variables the correlation pattern is broken
 - Some level of X-constraint is OK but not too much
 - Select X-variables in Select variables menu and set the constraints using the sliders
 - When suggested solution no longer belongs to the model (PModX+) the prediction plot turns red



SARTORIUS

Simplifying Progress

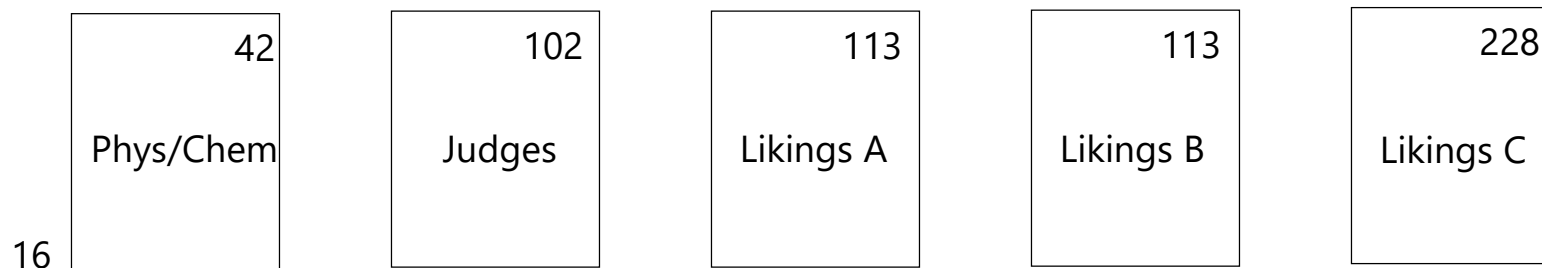


New technology

Multiblock Orthogonal Component Analysis - MOCA

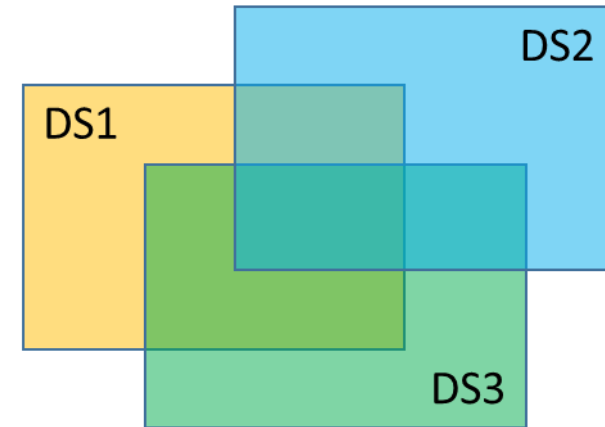
- A new and novel method for analyzing multiblock data is introduced
 - All blocks have the same observations but the variables differ
- Some data examples are:
 - Systems biology where samples are analyzed using e.g. metabolomics, proteomics, lipidomics
 - Manufacturing data where process signals are complemented with spectroscopy
 - Sensory data where expert panel data are compared to chemical analysis and consumer preference

Sensory data example – 5 blocks of data



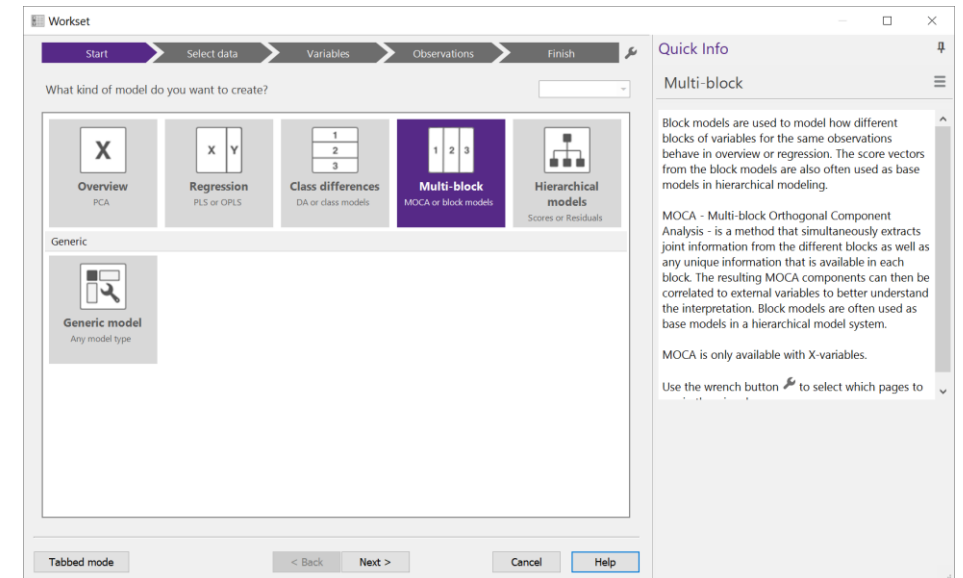
Multiblock Orthogonal Component Analysis - MOCA

- MOCA provides an overview of the information in all blocks simultaneously, similar to what PCA provides for one block and O2PLS provides for two blocks
- The method answers the questions
 - How much of the information in the blocks is common/joint among the blocks?
 - How much of the variation is unique for each specific block?
- MOCA is based on the OnPLS-algorithm which is using OPLS and PCA in an innovative way to find the joint and unique block information in a single analysis



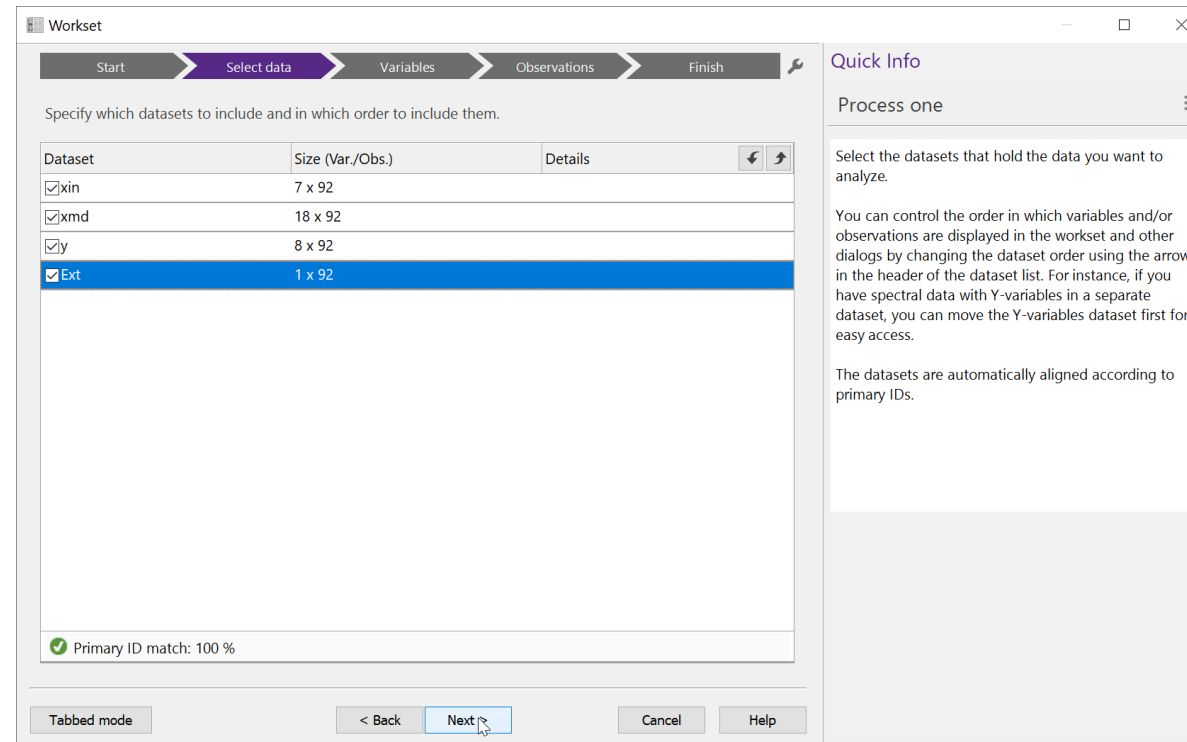
MOCA Analysis – Setting up the model

- A process dataset with three blocks of data in three separate datasets
 - 92 observations
 - x_{in} : 7 variables (input related variables)
 - x_{md} : 18 variables (intermediate process variables)
 - y : 8 variables (output related variables)
 - In addition there is one external data signal : Ext1
- Start the Workset wizard and select the Multiblock objective



MOCA Analysis – Setting up the model

- Select all datasets from the select data list
 - The blocks do not have to be in separate datasets but it is recommended



MOCA Analysis – Setting up the model

- In the variable step, set blocks according to the datasets

Workset

Start > Select data > Variables > Observations > Finish

Variables: included: 34 (X=34, Y=0), excluded: 0, selected: 1

Variable ID	Info	Dataset
x1in		xin
x2in		xin
x3in		xin
x4in		xin
x5in		xin
x6in		xin
x7in		xin
x8md		xmd
x9md		xmd
xamd		xmd
xbmd		xmd
xcmd		xmd
xdmd		xmd
xemd		xmd
xfmd		xmd
xgnx		xmd
xhnx		xmd
xinx		xmd
xjnx		xmd
xknx		xmd
xlnx		xmd
xmen		xmd
xnen		xmd
xoen		xmd
xvnen		xmd

Find Variable ID... 🔍

Blocks: Set block

Quick Info

Variable | x8md

Histogram

Trend

Statistics

MOCA Analysis – Setting up the model

- Variables not part of any block can still be kept as X-variables (Optional step)
 - E.g. sample treatments, run order, groups, external disturbances,...
 - Reset potential external variables so that they are defined as X but not assigned to any block

Workset

Start > Select data > Variables > Observations > Finish

Variables: included: 34 (X=34, Y=0), excluded: 0, selected: 1

Variable ID	Block	Info	Dataset
X xbmd	xmd		xmd
X xcmd	xmd		xmd
X xcmd	xmd		xmd
X xemd	xmd		xmd
X xfmd	xmd		xmd
X xgnx	xmd		xmd
X xhnx	xmd		xmd
X xinx	xmd		xmd
X xjnx	xmd		xmd
X xknx	xmd		xmd
X xlnx	xmd		xmd
X xmen	xmd		xmd
X xnen	xmd		xmd
X xoem	xmd		xmd
X xpen	xmd		xmd
X y1	y		y
X y2	y		y
X y3	y		y
X y4	y		y
X y5	y		y
X y6	y		y
X y7	y		y
X y8	y		y
X Ext1	Ext		Ext

Find Variable ID...

Buttons: X, Y, Exclude, Blocks: Set block, Tabbed mode, < Back, Next >, Finish, Cancel, Help

Quick Info

Variable | Ext1

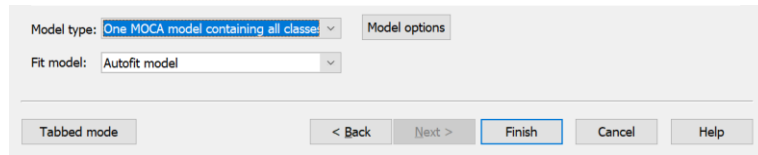
Histogram

Trend

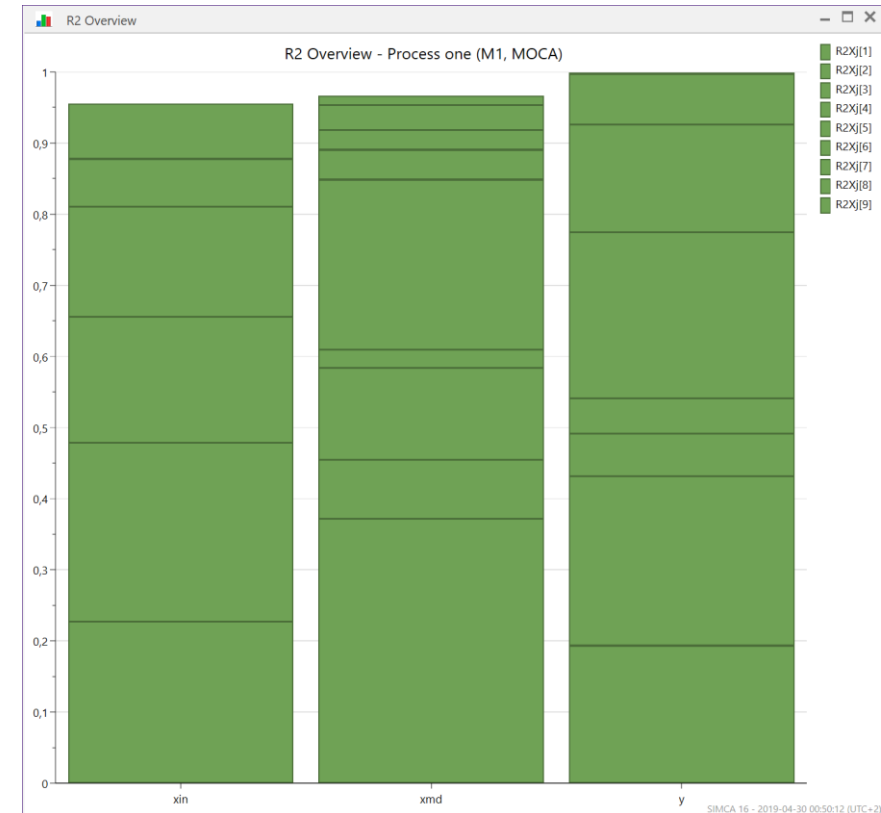
Statistics

MOCA Analysis – Fitting the model

- When reaching the Finish step in the Workset wizard, select MOCA analysis and Autofit model (only option for MOCA)



- The MOCA model R2 overview has one column for each block and summarizes the explained variation for all blocks



MOCA Analysis – Model window and R2 Overview

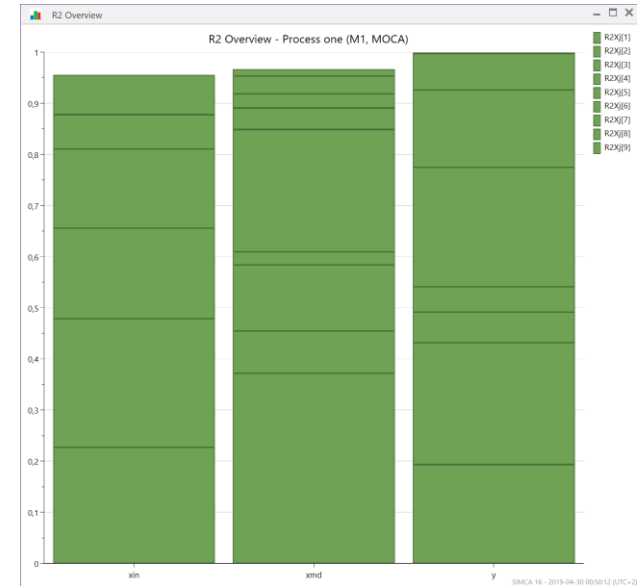
- The model window and the R2 overview show the same information
 - The explained variation by component and cumulative
- In this example we only have joint components
 - Joint components are green in the cumulative R2 overview (unique are blue)
 - The first four are globally joint (components including all blocks)
 - The next three (components 5-7) are locally joint between xmd and y
 - The last two (components 8 and 9) are locally joint between xin and xmd

Model M1

Model type: MOCA, observations (N): 92, variables (K): 34 (xin=7, xmd=18, y=8)

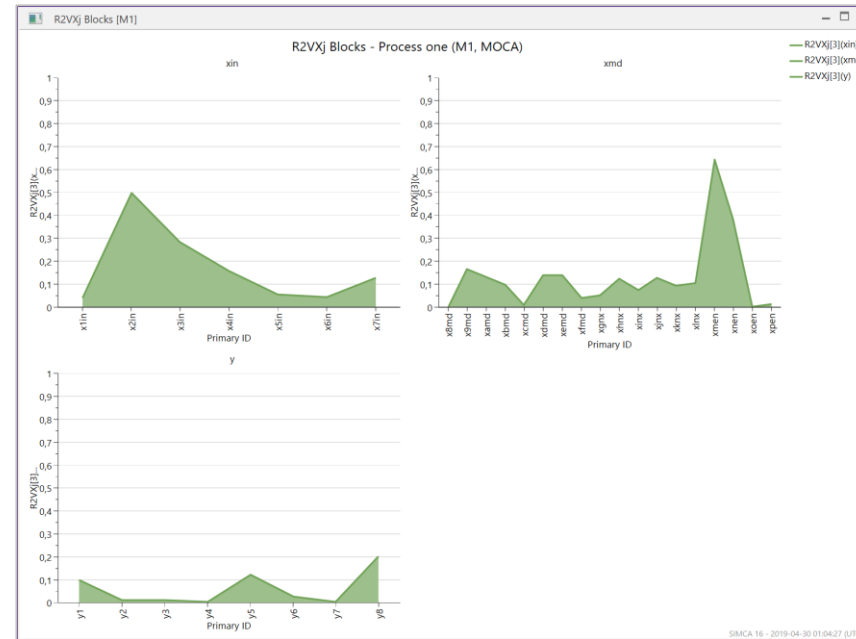
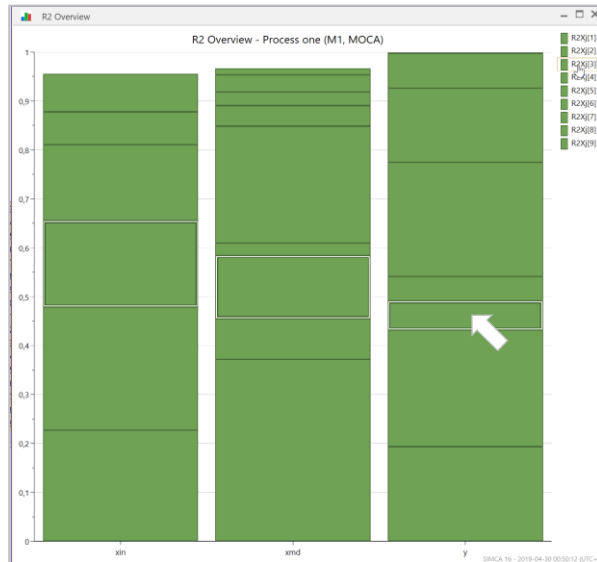
Title: Enter model title Workset Options

Component	R2X	R2X(cum)	R2X	R2X(cum)	R2X	R2X(cum)
Model		0,955		0,966		0,997
Blocks	xin		xmd		y	
Joint components		0,955		0,966		0,997
1	0,227	0,227	0,372	0,372	0,193	0,193
2	0,251	0,479	0,0831	0,455	0,238	0,431
3	0,177	0,656	0,129	0,584	0,06	0,491
4	0,155	0,81	0,0257	0,61	0,0495	0,541
5	--	--	0,239	0,849	0,234	0,774
6	--	--	0,0417	0,89	0,151	0,926
7	--	--	0,0277	0,918	0,0715	0,997
8	0,0674	0,878	0,0353	0,953	--	--
9	0,0775	0,955	0,0131	0,966	--	--
Unique compone...		0		0		0



MOCA Analysis – R2 overview drill down

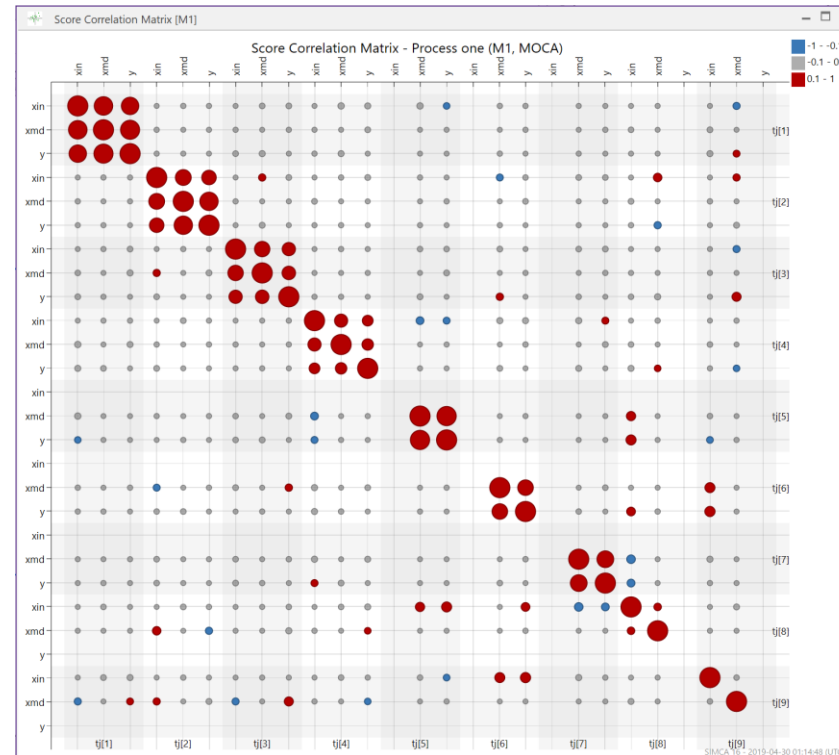
- Double clicking on one of the sections in the R2 overview brings up the explained variation of that component per variable
 - Displayed as a multiplot, one plot per block



MOCA Analysis – Scores correlation matrix

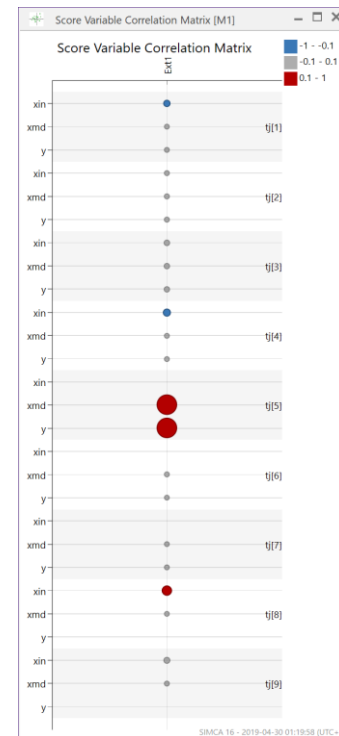
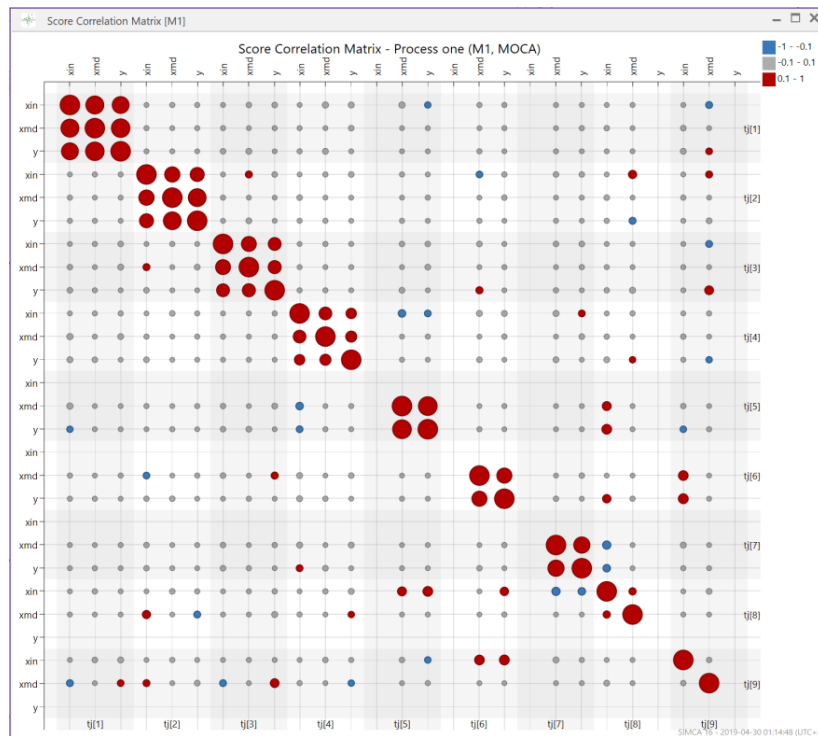
- The score correlation matrix is used to check the quality of the model
- With three blocks and 9 components the score correlation matrix is built up by $9 \times 3 = 27$ rows and columns divided into 3×3 squares (one for each component)
- The component squares on the diagonal are important. These are the only squares where we want to see large points.
 - The size of the points indicate the correlation strength

- This looks like a good model
 - Big points only in the 3×3 squares on the diagonal



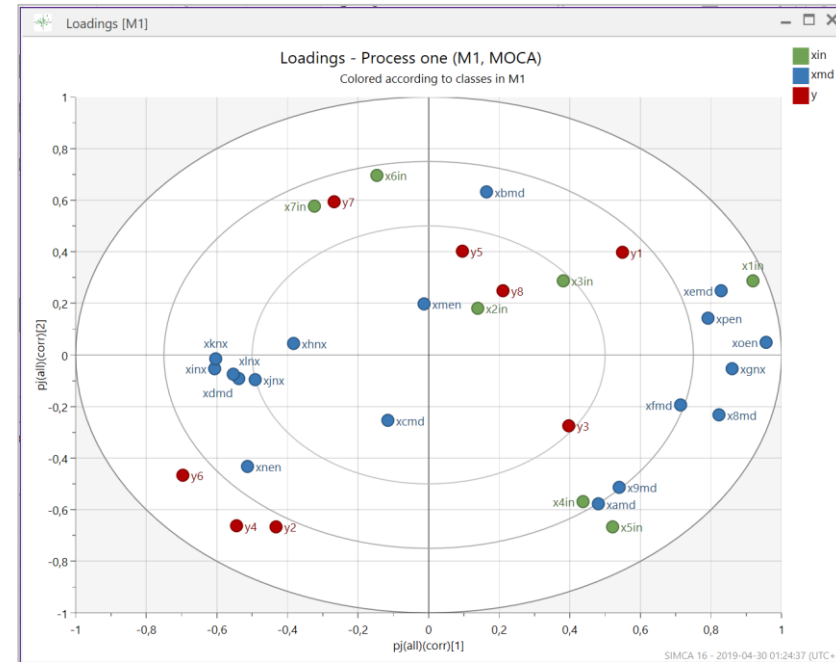
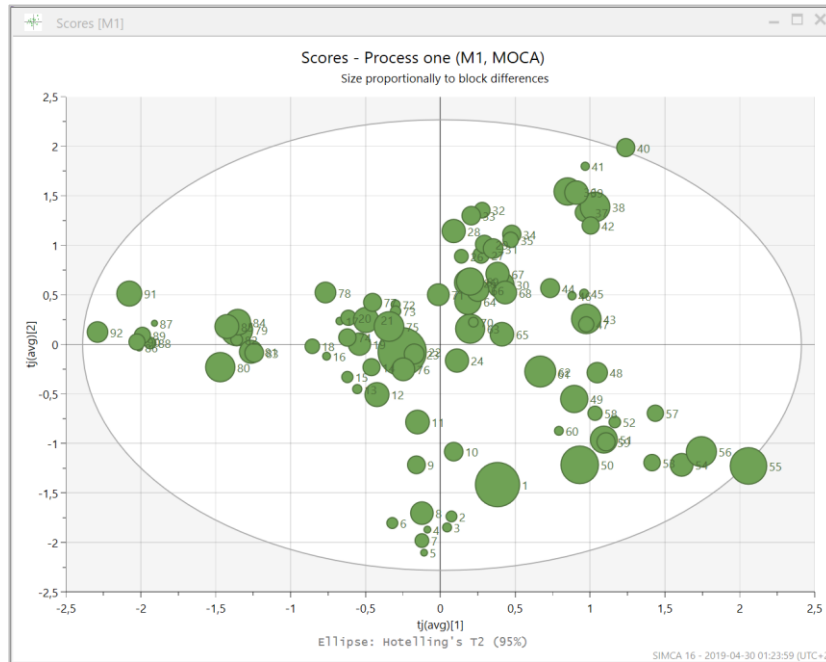
MOCA Analysis – Score variable correlation matrix (optional)

- If external vectors (not part of any block) were included as X-variables we can investigate if they are correlated to any of the MOCA score vectors
 - The external vector in this example seems to be correlated to the 5:th joint component



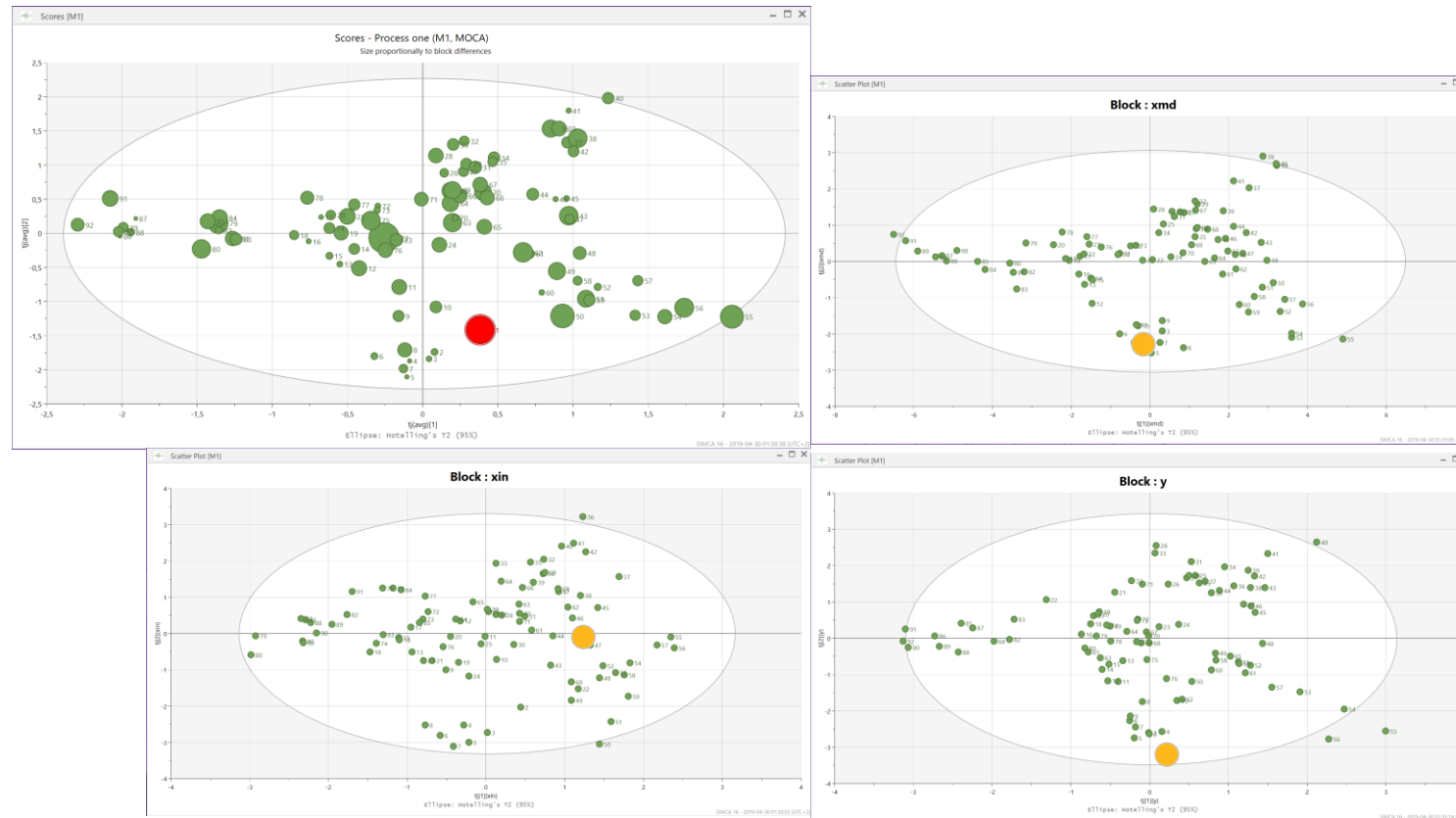
MOCA Analysis – Scores and loading scatter plots

- The score and loadings plots from the MOCA model are interpreted in the same way as normal PCA scatter plots
 - The point location in the MOCA score scatter plot is the average score value for all blocks, and the size reflects how different the location of the point is in the different blocks



MOCA Analysis – Score scatter plots

- Comparing the score plot from the MOCA model with the individual block scores
 - The location is almost the same in **xmd** and **y** for point “1” but very different in **xin**. This deviation in should be investigated further



SARTORIUS

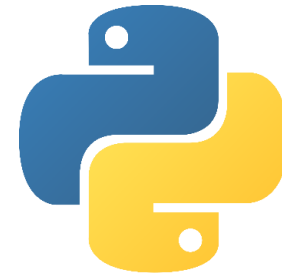
Simplifying Progress



Python additions and improvements

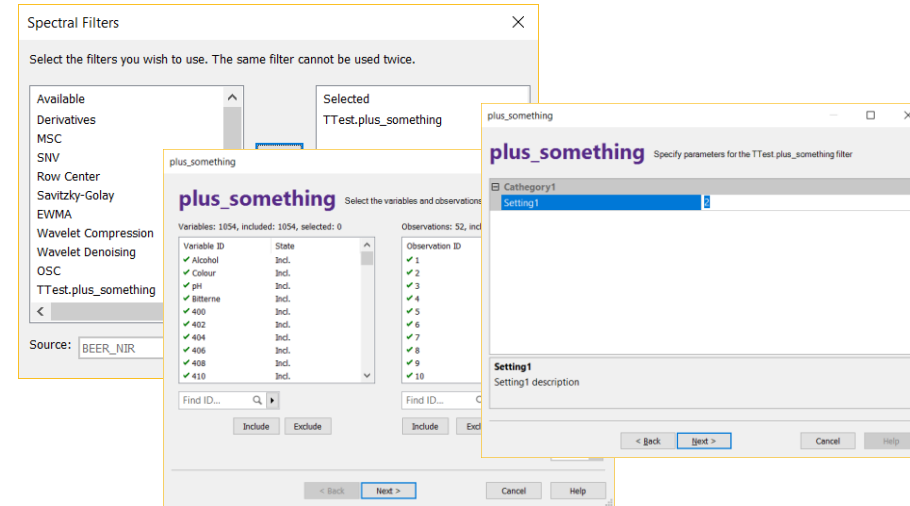
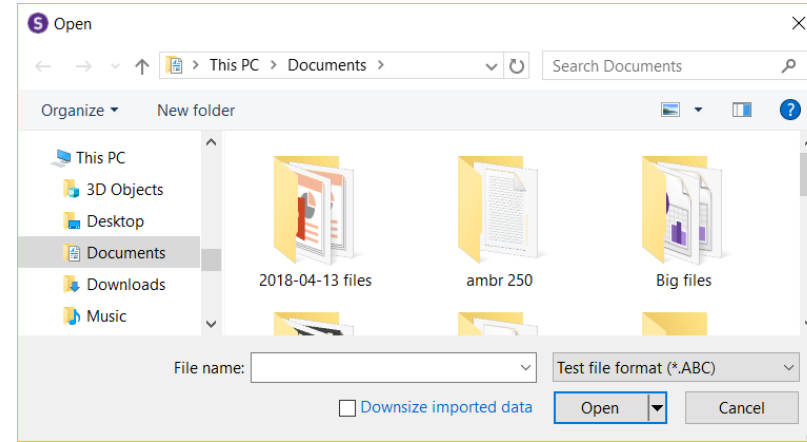
Python implementation news

- Customer feedback
 - Why do I have to install Python on the side to be able to access other packages (e.g. NumPy)?
 - How can I debug my scripts?
 - I have a spectral preprocessing method that works well in Python. When can that method be implemented in SIMCA?
- With SIMCA 16 we address all of these requests
- Python (3.7) is now fully installed with SIMCA
 - No need to install Python on the side
- Users can add additional Python packages directly in the console using PIP
 - Using virtual environments the user can also work with different combination of packages
- Python scripts can be debugged with VS Code



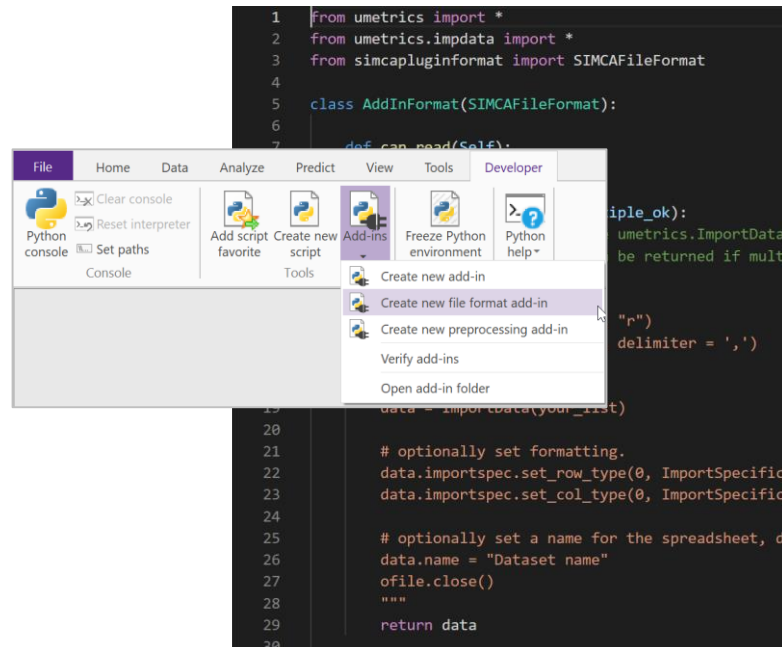
SIMCA Python plugins

- SIMCA supports two types of plugins
 - Users or third party companies can create custom plugins for both file reading and data preprocessing
 - Plugins can be shared with other users
- File reader plugins
 - Adds a new file type to the normal SIMCA file open dialog
- Data preprocessing plugin
 - Adds a new preprocessing choice to the list of implemented preprocessing techniques
 - Works as all the other filters (i.e. also for predictions)
- Plugins make SIMCA very flexible since users can create their own file import and data preprocessing
 - No need to wait for SSDA to implement functionality



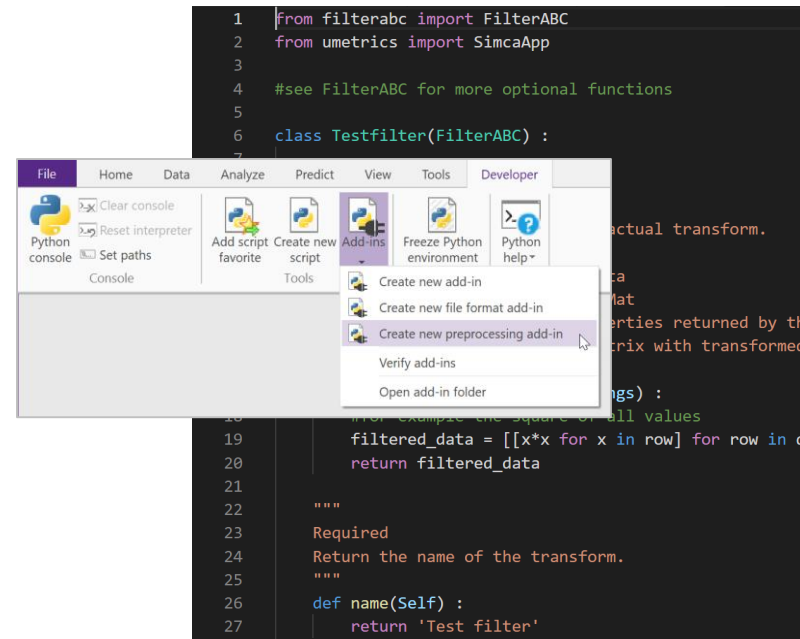
SIMCA Python file and spectral filter plugins

- Initiate the plugin creation in SIMCA and a boilerplate script is prepared for the user to add functional code to



The screenshot shows the SIMCA Python console interface. The 'Add-ins' menu is open, displaying options: 'Create new add-in', 'Create new file format add-in', 'Create new preprocessing add-in', 'Verify add-ins', and 'Open add-in folder'. The 'Create new file format add-in' option is selected. In the background, a Python script is visible with the following code:

```
1 from umetrics import *
2 from umetrics.impdata import *
3 from simcapluginformat import SIMCAFileFormat
4
5 class AddInFormat(SIMCAFileFormat):
6
7     def can_read(Self):
8
9         # optional set formatting.
10        data.importspec.set_row_type(0, ImportSpecific
11        data.importspec.set_col_type(0, ImportSpecific
12
13        # optionally set a name for the spreadsheet, d
14        data.name = "Dataset name"
15        ofile.close()
16        """
17        return data
```



The screenshot shows the SIMCA Python console interface. The 'Add-ins' menu is open, displaying options: 'Create new add-in', 'Create new file format add-in', 'Create new preprocessing add-in', 'Verify add-ins', and 'Open add-in folder'. The 'Create new preprocessing add-in' option is selected. In the background, a Python script is visible with the following code:

```
1 from filterabc import FilterABC
2 from umetrics import SimcaApp
3
4 #see FilterABC for more optional functions
5
6 class Testfilter(FilterABC) :
7
8     """
9     Required
10    Return the name of the transform.
11    """
12
13    def name(Self) :
14        return 'Test filter'
```

Thank you for your interest in
SIMCA[®] 16

Don't forget to check out instructional videos
Search for Sartorius Data Analytics on YouTube

SARTORIUS