

SIMCA

Multivariate Data Analysis Solution

Omics skin for SIMCA 17 Quick guide

9 February 2021

Introduction

This guide describes the interface and functionality unique to the Omics skin, complementing the SIMCA 17 Help.

The Omics skin was developed to offer a simplified and customized interface designed for a specific application area. Compared to the regular SIMCA interface, a few new features, some customized default settings and plot options, as well as shortcuts for typical routine operations were introduced to suit omics data and normal modeling procedures within this community. The plot library, model types and calculations are identical to those in the standard SIMCA.

Access and activate

The Omics skin is installed as part of the standard SIMCA 17 installation. If you wish to use them and they are not shown as options under **File | New**, enable them in **File | Options, SIMCA options, Skins** section. Set **Yes** on the skins you wish to enable.

Activate the skin by opening SIMCA and on the View tab clicking Omics. The skin is also activated by clicking **File | New | Omics project**. It is possible to switch skins at any time on the View tab.

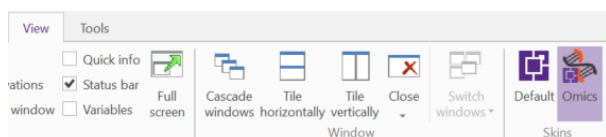


Figure 1: Set active skin on the View tab.

Omics skin features

Upon activation of the skin the appearance of SIMCA changes to the Omics skin layout, holding functions found in the default SIMCA skin as well as features unique to the Omics skin.



Figure 2: Omics skin Home tab.

The Home tab spans over features from dataset import to create report. The omics unique features are:

- Analysis wizard in the Workset group.
- Scaling and Transformation of variables in the active model in the Workset group.
- Descriptive statistics in the Validation group.

The Analysis wizard

The Analysis wizard offers a step-by-step guided analysis to identify discriminating parameters between two groups of samples/observations. It contains the following steps:

- Workset definition
- Raw data analysis (optional)
- Observation overview
- Discriminant analysis summary
- Identify discriminating parameters
- Create report

After completing the analysis wizard the user knows if there are discriminating parameters that separate the selected groups and also if the analysis is trustworthy.

To operate the Analysis wizard smoothly it is important that a Class ID or secondary observation ID has been specified at import.

Workset definition

1) Select type of data. Four types of commonly used data types are predefined. To each an appropriate default scaling is set. Scaling can be changed by clicking the blue "Scaling" text.

- MS – Pareto scaling
- NMR – Pareto scaling
- Chromatographic – Pareto scaling
- Identified metabolites – Unit variance scaling
- General – Unit variance scaling

- 2) Select dataset. Here it is possible to combine datasets if data is distributed over several datasets.
- 3) Select objective. Raw data analysis and 2 group comparison are optional. Data overview is always performed.
- 4) Define groups. Group ID must be a Class ID or secondary observation ID defined at import. Several items can be assigned as Reference or Group 1 but cannot belong to both Reference and Group 1.

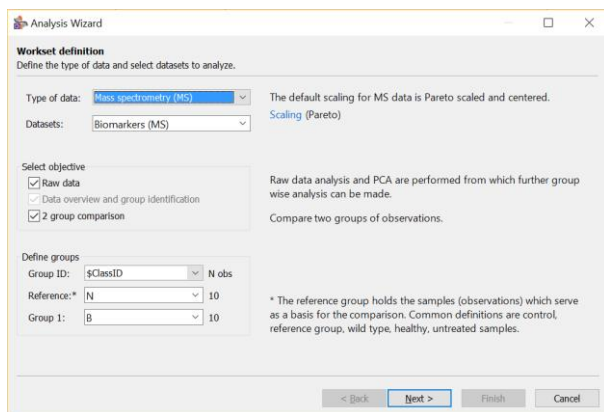


Figure 3: Workset definition page.

Raw data analysis

The raw data plot allows identification and exclusion of observations or variables which are obviously erroneous.

Mark a variable in the raw data plot to see the values for each observation in the right-hand variable plot.



Figure 4: Raw data analysis page.

Observation overview

Prior to discriminant analysis a Principal component analysis (PCA) model is calculated. The Observation overview plot (score plot) and the Distance to model plot highlight outliers, observations which do not fit the model.

On this page note that the toolbar allows:

- Assigning or reassigning group belonging – mark observations and click Assign.
- Excluding observations – mark an observation and click Exclude
- Creating a contribution plot showing how the marked observation differs from average – double-click a point, or mark it and click Group to average.

After introducing model changes, click Update to apply the changes and recalculate the PCA model.

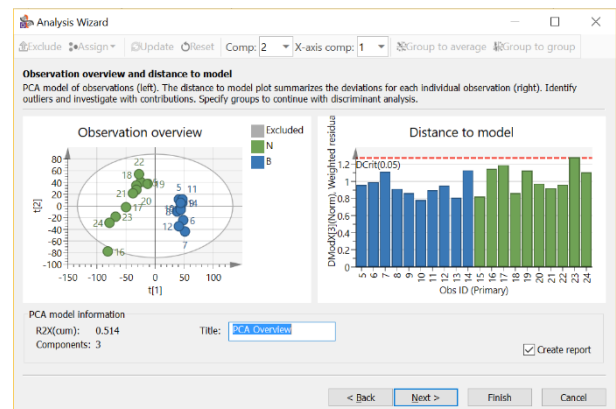
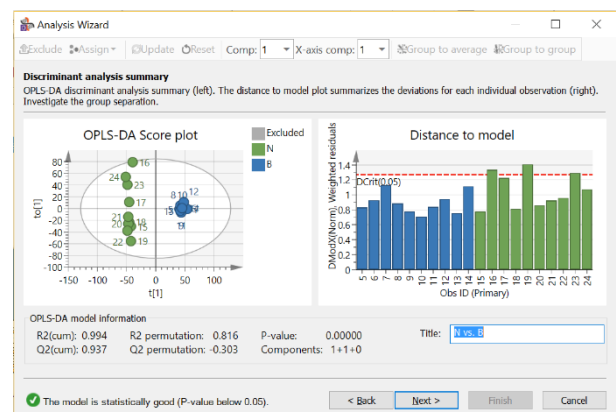


Figure 5: Observation overview page.

On this page it is possible to exit the Analysis wizard (Finish) and optionally create a report based on the PCA analysis.

Discriminant analysis summary

On the Discriminant analysis summary page the results from the calculated OPLS-DA model are visualized with OPLS-DA score and Distance to model plots.



model plots. A note at the bottom of the page indicates if the model is trustworthy or not.

Figure 6: Discriminant analysis summary page.

The OPLS-DA model created here is used in the next step to identify discriminating parameters.

Identify discriminating parameters

The S-plot or S-line plot – depending on data type selection at Workset definition stage – is used to identify parameters which strongly contribute to group separation. ROC plot shows the level of group separation, here with Area under curve = 1 the separation is perfect.

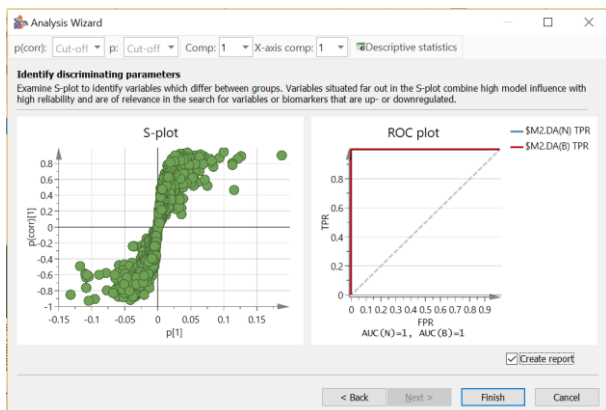


Figure 7: Identify discriminating parameters page with ROC plot.

Click a variable to see the raw data represented in the dot plot.

Define $p(\text{corr})$ and p cut-off to highlight discriminating parameters by typing a value for cut-off and pressing Enter to confirm.

Click Descriptive statistics to display statistics for the marked variables.

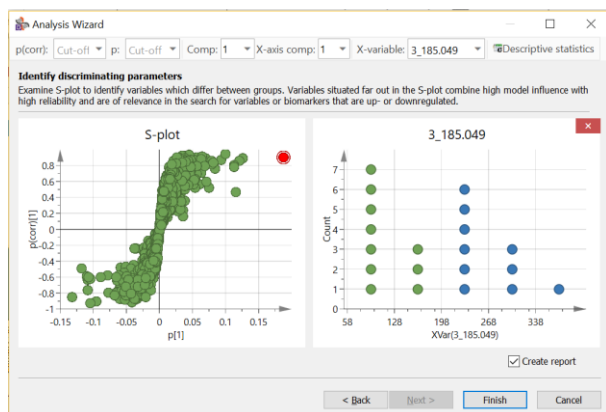


Figure 8: Identify discriminating parameters page with defined $p(\text{corr})$ and p cut-off values.

Finish and Create report

With the Create report check box selected, clicking Finish creates a report for the OPLS-DA analysis. The report feature is the same as used in standard SIMCA but with an omics template. See SIMCA help for information about how to customize reports.

Upon Finish four overview plots summarizing the analysis open up.

Omics Workset features

Changing transformation or scaling of variables using Transform or Scaling in the Workset group, immediately introduces the change in the active model, regardless of whether it is fitted or not. The new transformation or scaling is applied to all variables. After the change the model is instantly recalculated with the new settings.

If the wish is to investigate impact of different transformations and scaling, create a copy of the model first using New as or by marking the model in the Project window and pressing Ctrl+C, Ctrl+V.

Transform

Clicking Transform in the Workset group on the Home tab opens the following dialog:

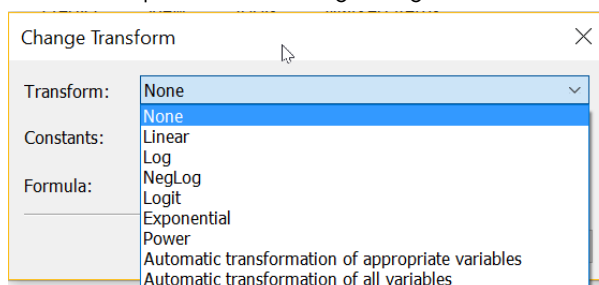


Figure 9: Change Transform dialog.

When selecting a transformation SIMCA applies the change to all variables in the currently active model.

Scaling

With the scaling feature in the Workset group on the Home tab it is possible to instantly change scaling for all variables in the active model.

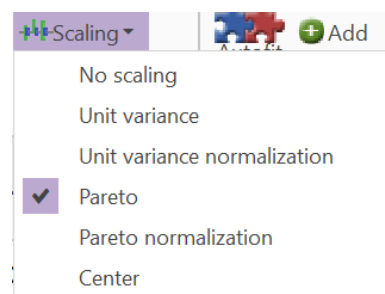


Figure 10: Scaling options.