Control Advisor Models for SIMCA-online

Introduction

SIMCA-online is a real-time prediction system for the process industry. From different types of data sources, such as MES, LIMS and process data historians, SIMCA-online automatically collects data, performs multivariate calculations and generates intuitive control charts that summarize the state of the process. Diagnosis of deviations and alarms are provided through interactive drill down features.

Control Advisor in SIMCA-online extends the real-time multivariate monitoring capabilities to include a Forecast mode for predictive monitoring and an Advised future mode for process optimization with model predictive control (MPC).

Control Advisor includes powerful predictive capabilities utilizing a combination of imputation and regression methods to forecast future trajectories of batch processes. Predictions of final qualities and yields are made available early in the process and deviations may be detected before they happen, allowing for proactive corrective action. The Advised future provides suggested process adjustments that may be implemented manually or automatically providing closed loop control.

Objective of the tutorial

The objective for this tutorial is to teach the reader how to create a model in SIMCA that can be used with the Control Advisor in SIMCA-online.

Theory and definitions

To make an informed decision how to steer a batch process during production a good estimate of the future, a forecast or imputation, based on historical data and the knowledge on how the system have behaved historically is needed. There are different methods to estimate the future values of a batch, but Imputation by Regression (IBR) has proven to be one of the most accurate and fast ones[1].

When the imputation of the future missing dependent variables has been made, optimization algorithms are deployed to find a more optimal future evolution for the batch. We call this the Advised future. The optimal setpoints found by SIMCA-online can then be applied to the process as "mid-batch" corrections performed by the user.

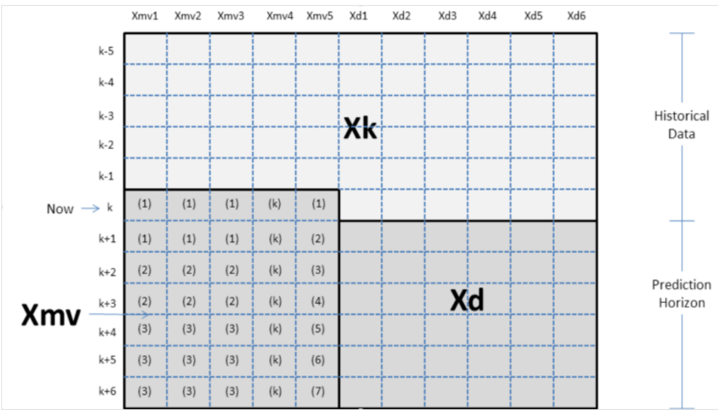


Figure 1: Manipulated variables (Xmv) are future variables that are controllable, and their future settings are known, such as setpoints. All other future X variables are called dependent variables (Xd). All values are known (Xk) up until the current time point.

With good forecasting capabilities, the future setpoints of the manipulated variables for batch processes can be optimized on a few discrete maturities called control maturities for one or more targets, such as:

- Optimizing Y, such as yield or quality (in Batch Level Models, BLM)

[1] S. Garcia-Munoz, T Kourti, J. MacGregor, Model Predictive Monitoring for Batch Processes, Ind. Eng. Chem. Res. (2004)

- Staying inside the model (in Batch Evolution Models, BEM or BLM)
- Staying close to the average batch from the model (in BEM)
- Keeping a dependent variable close to a trajectory (in BEM)
- Staying as close to the original manipulated variables as possible (in BEM)

For continuous projects, the targets could be to get as small variation as possible.

Observe that for the Advised future to work in the Control Advisor, there has to be enough variation in the manipulated variables in the historical data to span the region where the optimizer is allowed to work. To ensure this, it is recommended to perform a design of experiments with the manipulated variables at each control maturity.

Imputation

An imputation is a way to estimate what a missing value in a dataset most likely would be.

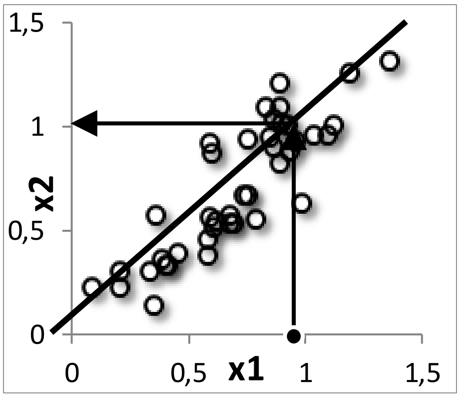


Figure 2: Basic principle of model-based imputation; Estimate the values for x2 based on a model and the value of x1. This is the case for PMP, SCP and CMR.

In the case for the Control Advisor in SIMCA-online, this means estimating the future dependent variables.

Examples of different imputation methods[2]

1. TRI: Trimmed Score Method
2. II: Iterative Imputation
3. PMP: Projection to the Model Plane
4. SCP: Single Component Projection
5. CMR: Conditional Mean Replacement
6. TSR: Trimmed Score Regression
7. IBR: Imputation by Regression

The imputation method used by SIMCA-online is IBR. All missing dependent values are predicted in a single pass algorithm.

The IBR method for batch processes use a batch level PLS prediction with all variables in the X block and the dependent variables as Y block. The batch level predictions of the Y block are then reorganized back to batch evolution dependent variables.

The IBR method for continuous processes uses a combination of lags and leads where the lagged variables is set to the X block and the dependent variables, the leads, are set to the Y block.

Data

This tutorial only includes one dataset for batch processes. Continuous projects are discussed at the end of this document.

[2] Arteaga and Ferrer (2002) Dealing with missing data in MSPC: several methods, different interpretations, some examples

The dataset **Baker Yeast Imputation.xlsb** contains a batch ID column, 11 variables and 33 batches. The first 7 variables (x1-x7) are dependent variables, **Xd,** variable x8 (**Temp_SP**) and x9 (**Air_SP**) are manipulated variables, **Xmv**. Variables x10 and x11 will be excluded from the model.

$BatchID	Ethanol	Temp	Molasses	NH3	Air	Level	pH	Temp_SP	Air_SP	Time (Days)	Time (Hrs)
Ba	0,03748	35,7145	478,196	23,7236	1842,99	41,738	4,74916	30	6715	0	0
Ba	0,0589	33,6153	1089,93	79,4786	2154,35	41,8353	4,72612	30	6715	1	0,041667
Ba	0,22471	31,4477	1110,98	81,7717	2342,81	41,9245	5,09084	30	6715	2	0,083333
Ba	0,38293	30,3056	1109,91	81,0109	2518,66	42,0182	5,04351	30	6715	3	0,125
Ba	0,52197	30,1153	1107,82	80,1929	2704,02	42,0932	5,11628	30	6715	4	0,166667
Ba	0,59908	30,3526	1125,21	80,9073	2878,89	42,194	5,01217	30	6715	5	0,208333
Ba	0,62089	30,5014	1172,53	86,2506	3059,62	42,2778	5,05936	30	6715	6	0,25
Ba	0,62648	30,492	1223,96	89,2116	3250,84	42,3894	5,11921	30	6715	7	0,291667

Table 1: Part of the Bakers Yeast Imputation dataset

Imputation model creation for batch models

Here is how to create an imputation model in SIMCA (each of the steps are described in detail below):

1. Import a dataset
2. Fit the BEM.
3. Expand the dataset to the batch level
4. Create the batch level dataset
5. Create the PLS imputation model

Import a dataset

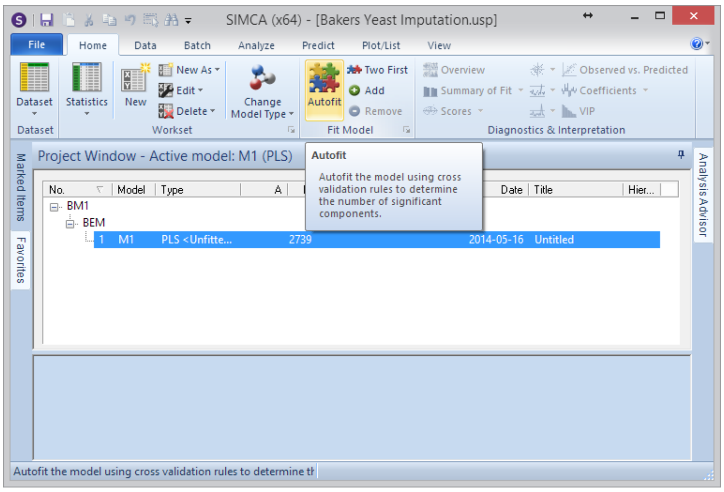
1. Open SIMCA and click File | New batch project.
2. Open the file called Bakers Yeast Imputation.xlsb.
3. Set all variables as Quantitative variables except the Batch ID column that should be set to Batch ID.

	Batch I	2	3	4	5	6	7	8	9	10	11	12
Primar	$BatchID	Ethanol	Temp	Molasses	NH3	Air	Level	pH	Temp_SP	Air_SP	Time (Day	Time (Hrs)
2	Ba	0,03748	35,7145	478,196	23,7236	1842,99	41,738	4,74916	30	6715	0	0
3	Ba	0,0589	33,6153	1089,93	79,4786	2154,35	41,8353	4,72612	30	6715	1	0,0416667
4	Ba	0,22471	31,4477	1110,98	81,7717	2342,81	41,9245	5,09084	30	6715	2	0,0833333
5	Ba	0,38293	30,3056	1109,91	81,0109	2518,66	42,0182	5,04351	30	6715	3	0,125
6	Ba	0,52197	30,1153	1107,82	80,1929	2704,02	42,0932	5,11628	30	6715	4	0,166667
7	Ba	0,59908	30,3526	1125,21	80,9073	2878,89	42,194	5,01217	30	6715	5	0,208333
8	Ba	0,62089	30,5014	1172,53	86,2506	3059,62	42,2778	5,05936	30	6715	6	0,25
9	Ba	0,62648	30,492	1223,96	89,2116	3250,84	42,3894	5,11921	30	6715	7	0,291667
10	Ba	0,6288	30,3847	1280,65	94,4121	3361,74	42,4982	5,06234	30	6715	8	0,333333
11	Ba	0,64116	30,3445	1327,71	98,291	3609,83	42,5999	5,0456	30	6715	9	0,375

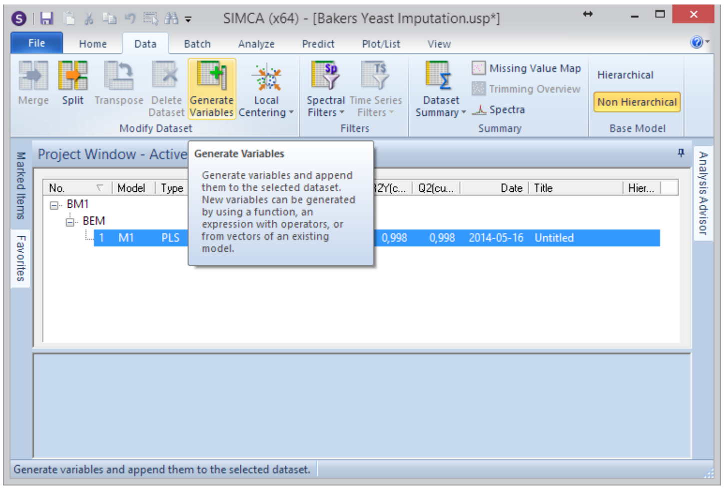
4. Click Finish on the Home tab to finalize the import.
5. Give a name to the Sartorius Stedim Data Analytics Solutions SIMCA Project file (USP-file) or keep the default name.

Fit the BEM and expand the dataset

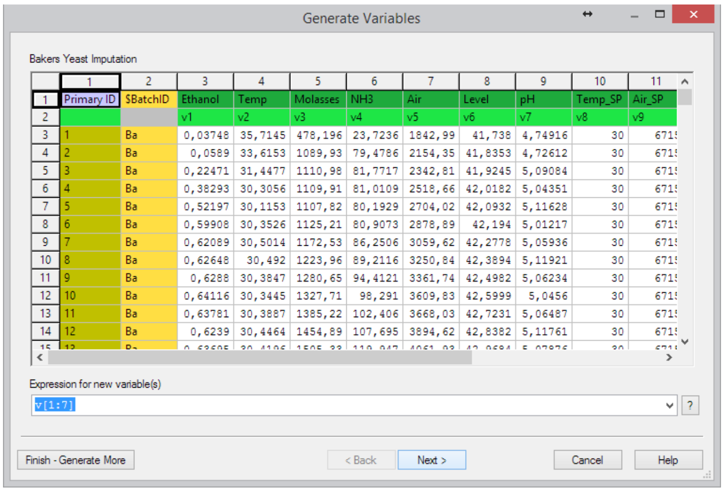
1. Fit the BEM.



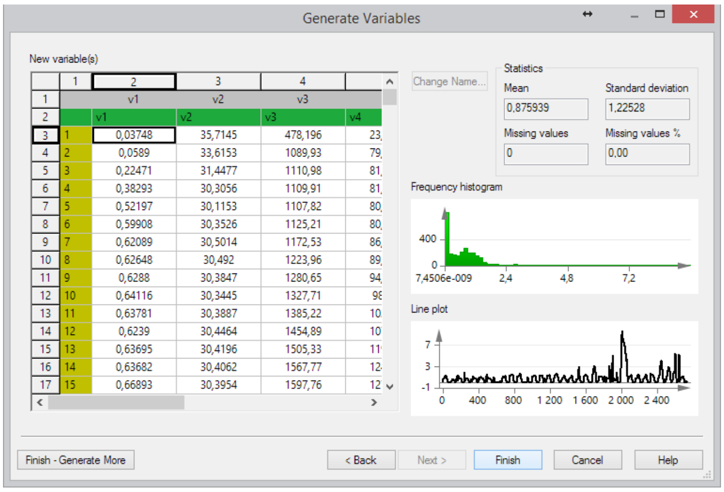
2. Make a copy of the Xd variables in the batch evolution dataset using Generate Variables
3. Open the Generate Variables dialog.



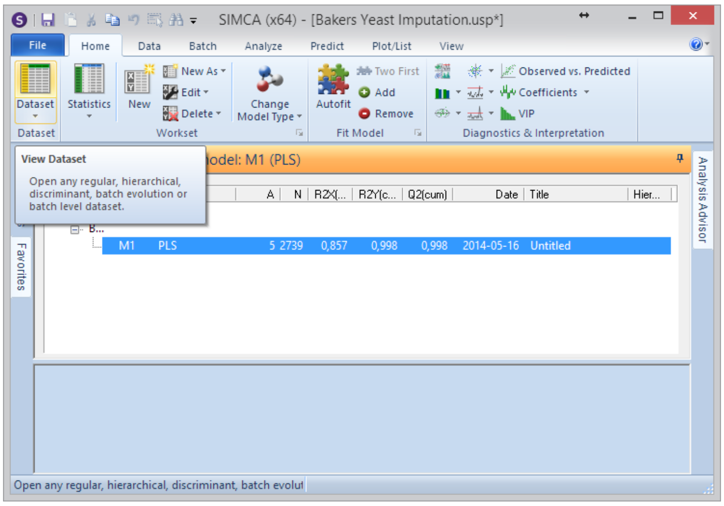
4. Generate a copy of variables 1 to 7 by typing v[1:7] in the Expression for new variables field. Click Next.



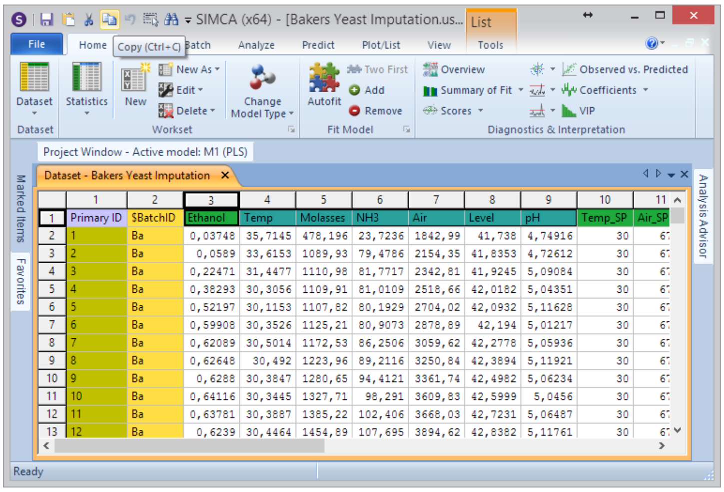
5. Click Finish.



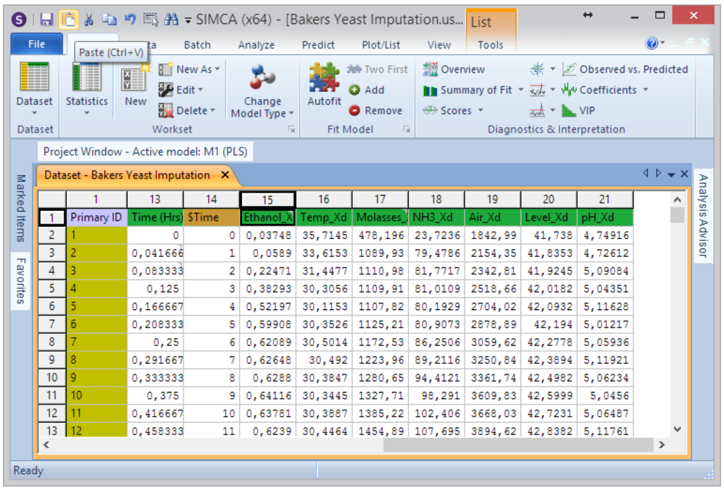
6. Rename the new generated variables in the dataset
7. Open the dataset.



8. Copy the original variable names into an editing tool, such as Notepad or Excel.



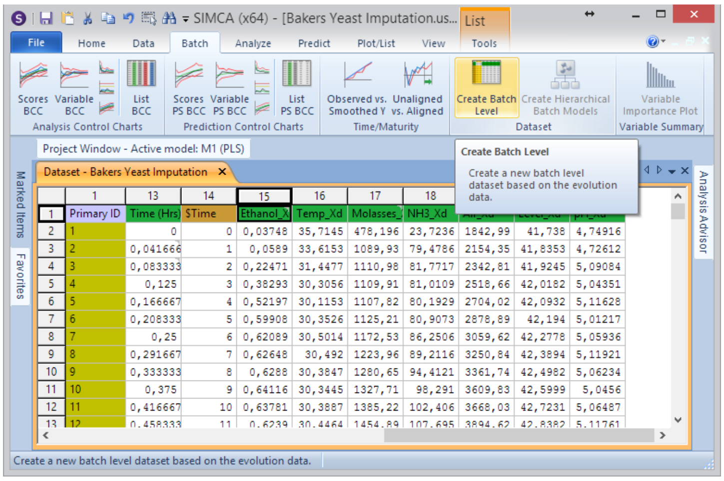
9. Add the suffix "_Xd" to each variable name and paste it in for the 7 new generated variables.



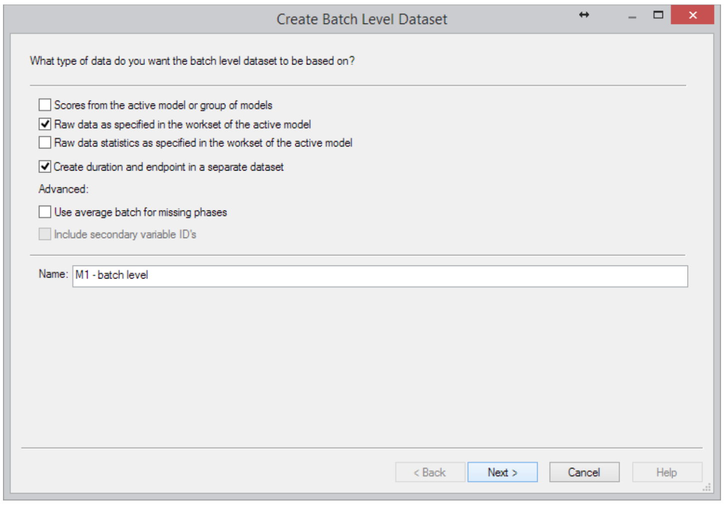
SARTORIUS

Create the batch level dataset

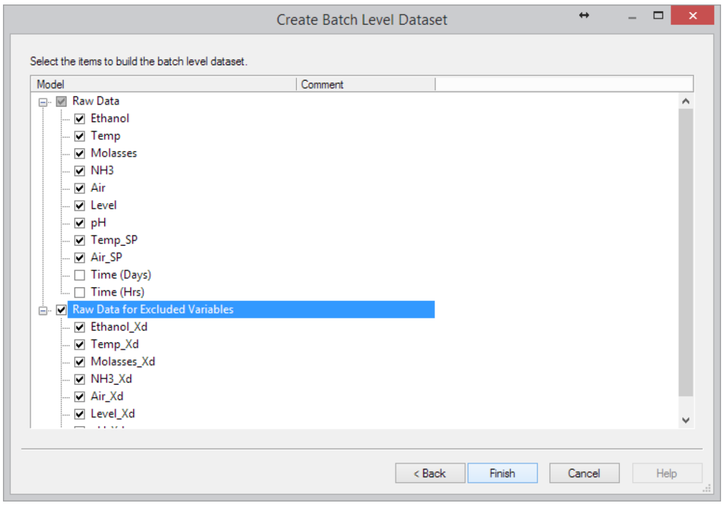
1. On the Batch tab, click **Create Batch Level** to open the **Create Batch Level Dataset** dialog.



2. Select Raw data as specified in the workset of the active model and click Next.

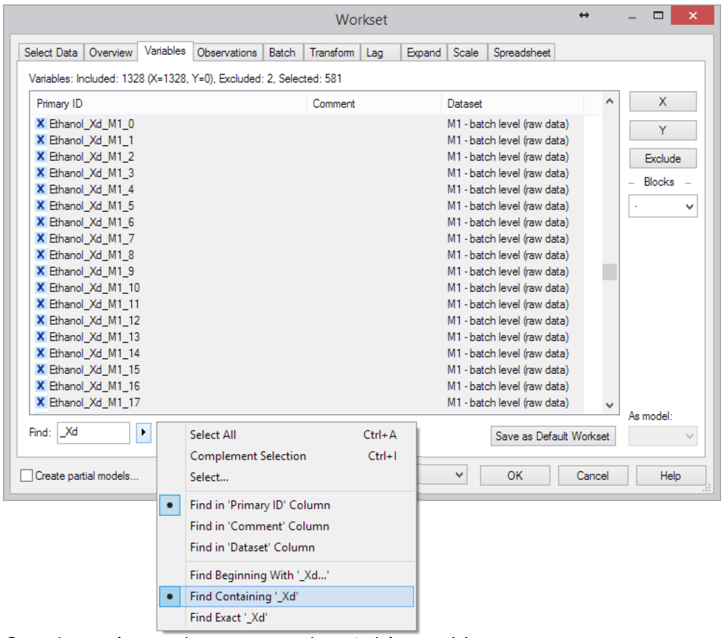


3. Select the generated variables ending with '_Xd' **under Raw Data for Excluded Variables** and click **Finish**.

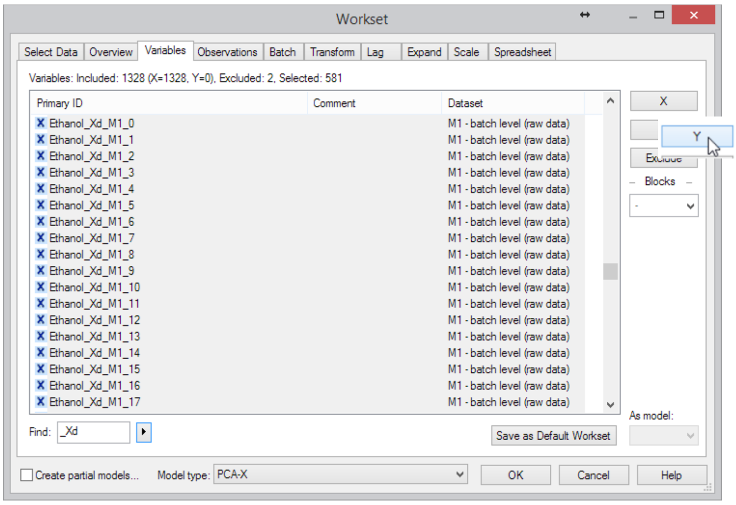


Create imputation model

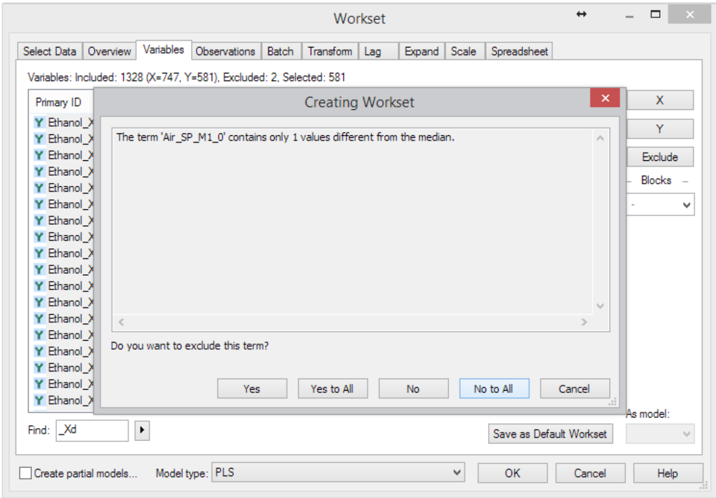
1. In the Workset dialog make sure that the **Create a batch level model** check box is selected. On the **Variables** tab select all generated variables for all time points by searching for the _Xd part of the variable ID.



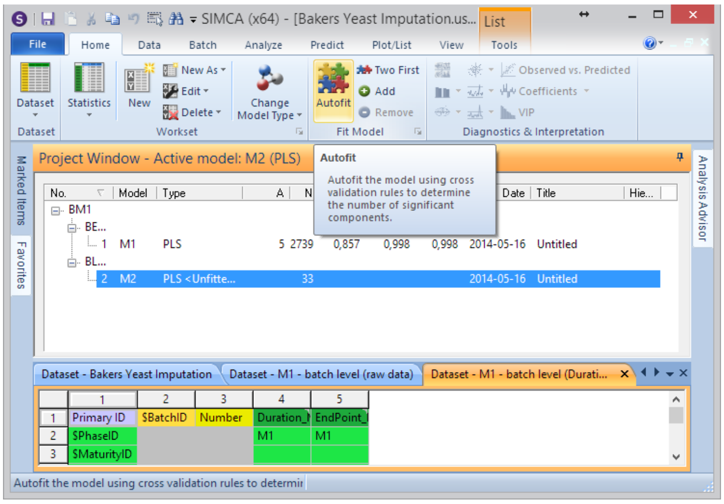
2. Set the selected generated variables as Y.



3. Do ***not*** exclude any variables or observations.



4. Fit the model.

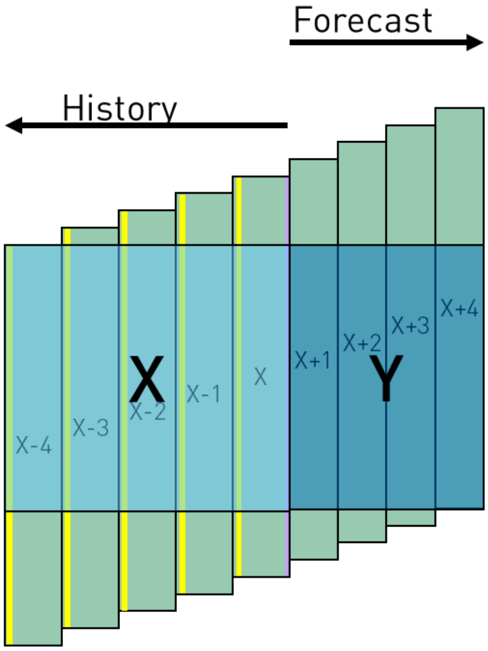


5. Rename the model to, for instance, ***Forecast model*** to easily find it in the SIMCA-online configuration.

Imputation model creation for continuous models

Here is how to create an imputation model in SIMCA (each of the steps are described in detail below):

1. Import a dataset
2. Create leads in the generate variables dialog
3. Create a new PLS model with lags



Import a dataset

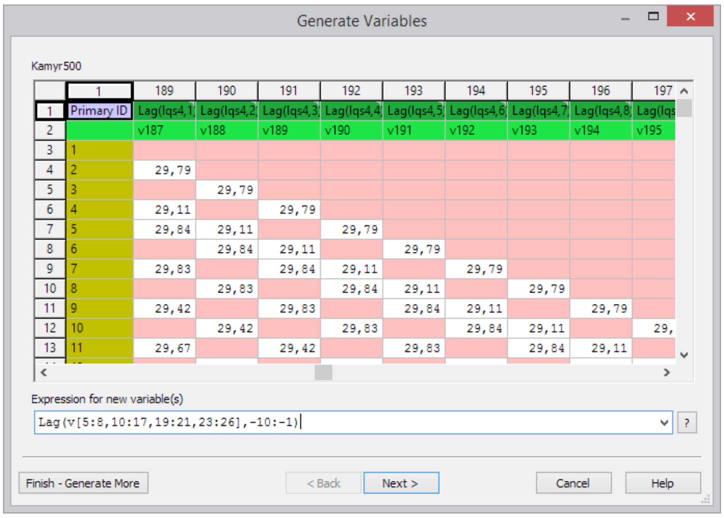
1. Open SIMCA and click File | New regular project.
2. Open the data file.

	Batch I	2	3	4	5	6	7	8	9	10	11	12
Primar	$BatchID	Ethanol	Temp	Molasses	NH3	Air	Level	pH	Temp_SP	Air_SP	Time (Day	Time (Hrs)
2	Ba	0,03748	35,7145	478,196	23,7236	1842,99	41,738	4,74916	30	6715	0	0
3	Ba	0,0589	33,6153	1089,93	79,4786	2154,35	41,8353	4,72612	30	6715	1	0,0416667
4	Ba	0,22471	31,4477	1110,98	81,7717	2342,81	41,9245	5,09084	30	6715	2	0,0833333
5	Ba	0,38293	30,3056	1109,91	81,0109	2518,66	42,0182	5,04351	30	6715	3	0,125
6	Ba	0,52197	30,1153	1107,82	80,1929	2704,02	42,0932	5,11628	30	6715	4	0,166667
7	Ba	0,59908	30,3526	1125,21	80,9073	2878,89	42,194	5,01217	30	6715	5	0,208333
8	Ba	0,62089	30,5014	1172,53	86,2506	3059,62	42,2778	5,05936	30	6715	6	0,25
9	Ba	0,62648	30,492	1223,96	89,2116	3250,84	42,3894	5,11921	30	6715	7	0,291667
10	Ba	0,6288	30,3847	1280,65	94,4121	3361,74	42,4982	5,06234	30	6715	8	0,333333
11	Ba	0,64116	30,3445	1327,71	98,291	3609,83	42,5999	5,0456	30	6715	9	0,375

3. Click Finish on the Home tab to finalize the import.
4. Give a name to the Sartorius Stedim Data Analytics Solutions SIMCA Project file (USP-file) or keep the default name.

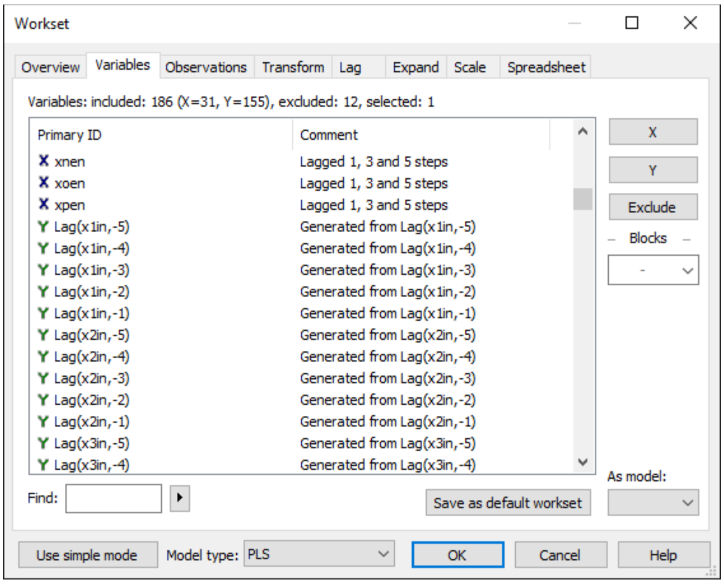
Create leads in the generate variables dialog

1. Create leads on the Xd data (only) to form the Y variables in the PLS imputation model also in the Generate variables dialog. Typical syntax **Lag(v[3:35], -10:-1)** means that variables 3 to 35 will be leaded 10 steps. Each variable will then be lagged -10, -9, -8,..., -1 steps.

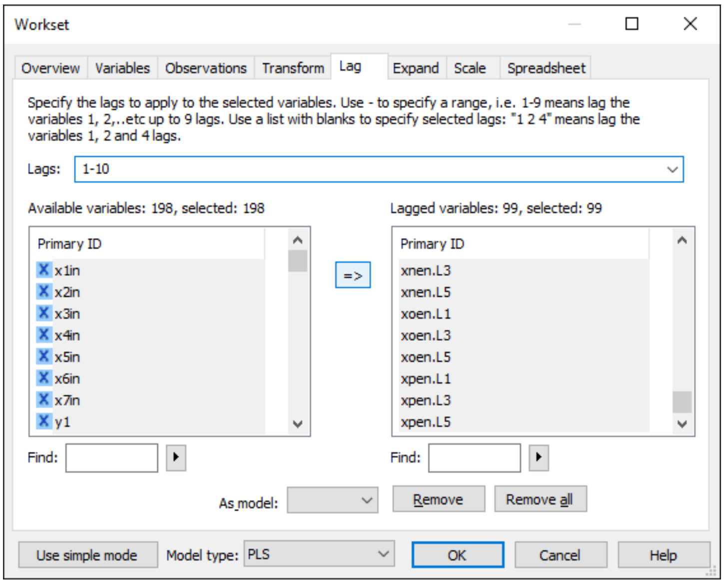


Create a new PLS model

1. Set all the leaded variables from above to Y and all of the Xd variables as X.



2. In the lag tab, select all the variables and lag them at least the same amount of observations as the leads.



3. Fit the model.
4. Rename the model to, for instance, Imputation model.