



Soft Sensor Toolbox V2.1 Quick Start Guide

Note: The current version will expire on December 30, 2021. New versions will be released.

Soft Sensor Analytics



Data preparation

Make sure the data are prepared following the guidelines below:

- Prepare the dataset using Excel
- the first column of a spreadsheet always contains timestamps
- the first row of a spreadsheet always contains the name of variables
- The data for each variable are recorded in each corresponding column
- For the created spreadsheet, select the entire timestamp column, right-click on this column and go to **“format cells”**, then go to **“Time”** or **“Custom”**, change the format of time information to a format similar to: **“mm/dd/yyyy h:mm”** or **“yyyy-mm-dd h:mm:ss”**.
- Save the dataset in **.xlsx** format

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Chemical	FeedT	FeedA	FeedB	FeedC	FeedD	Underflow	Underflow	Underflow	Underflow	Lab	DenN	DenFr
2	1/1/2016 0:00	0.022112	0.006653	-0.06091	-0.06707	0.127711	0.097012	-0.0143	-0.0069	0.016138	0.049496	-0.03981	0.202639	0.051406
3	1/1/2016 0:10	0.017209	0.006962	-0.06285	-0.06516	0.127958	0.097763	-0.01553	-0.00889	0.017511	0.048498	-0.03981	0.207361	0.053169
4	1/1/2016 0:20	0.015876	0.007888	-0.06473	-0.06411	0.130218	0.099531	-0.01709	-0.0086	0.017219	0.050171	-0.03981	0.210861	0.056008
5	1/1/2016 0:30	0.009537	0.005079	-0.06557	-0.06413	0.123251	0.096869	-0.01455	-0.00757	0.018022	0.048976	-0.03981	0.208895	0.053042
6	1/1/2016 0:40	0.001811	0.008856	-0.06253	-0.06579	0.129069	0.103874	-0.0168	-0.00939	0.017386	0.049496	-0.03981	0.208984	0.052436
7	1/1/2016 0:50	-0.00412	0.005274	-0.06094	-0.06531	0.122974	0.09481	-0.0169	-0.00742	0.018212	0.050445	-0.03981	0.21229	0.052739
8	1/1/2016 1:00	0.005994	0.007147	-0.06551	-0.06471	0.126308	0.101966	-0.01606	-0.00716	0.017317	0.047987	-0.03981	0.212114	0.053654
9	1/1/2016 1:10	-0.00593	0.003337	-0.0579	-0.06656	0.118972	0.090138	-0.01593	-0.00708	0.016339	0.047333	-0.03981	0.204069	0.051467

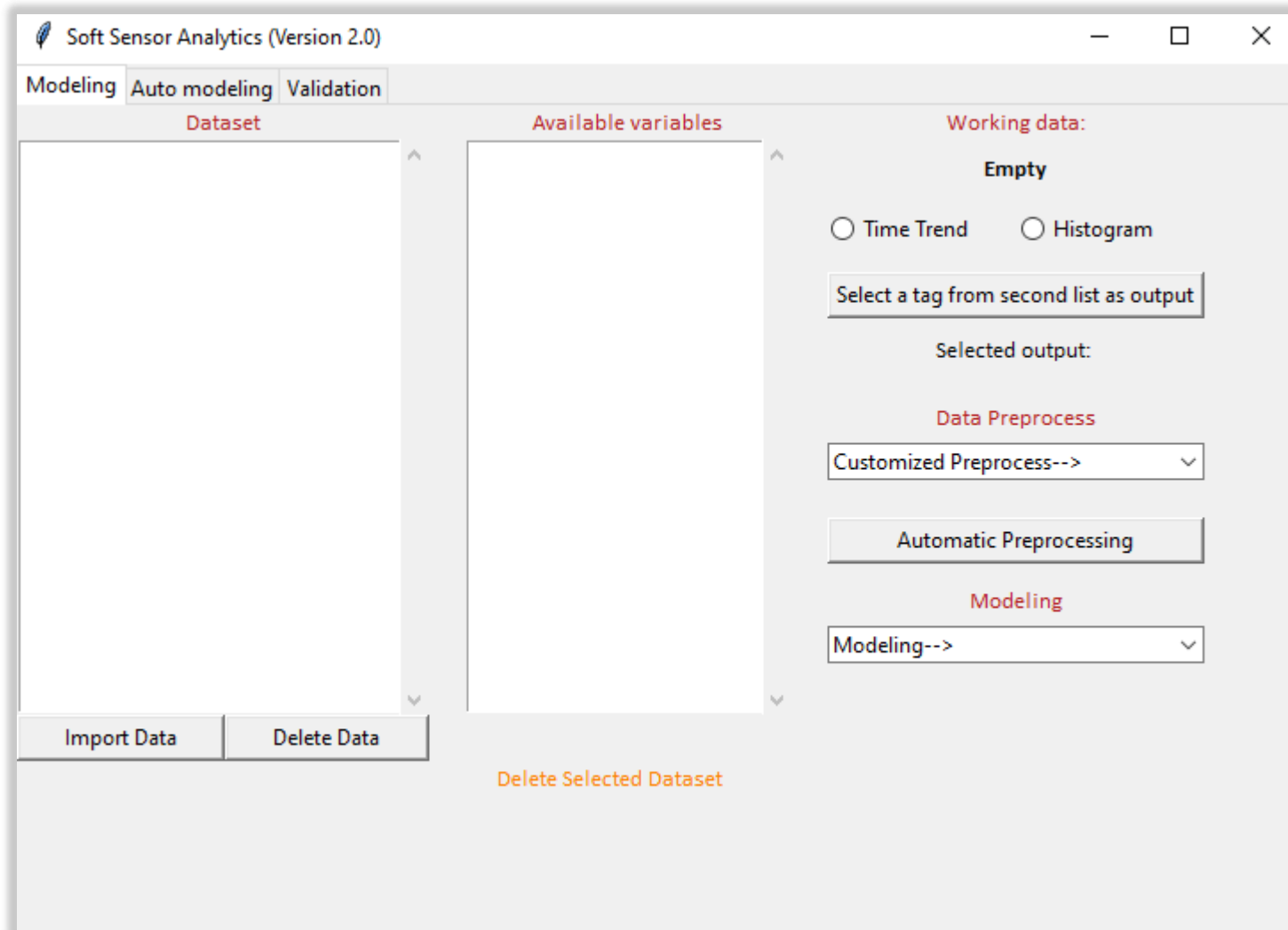
Soft Sensor Analytics



To start

Make sure the data sets to be used are in the same folder

Click **“SoftSensorAnalytics_v2.0.exe”** to open the program



Soft Sensor Analytics



An overview of the toolbox

We create this guide to help you minimize the learning curve.

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' application window. The interface is divided into several sections:

- Modeling**: Contains tabs for 'Modeling', 'Auto modeling', and 'Validation'. A callout box 'One-click modeling' points to the 'Auto modeling' tab, stating: 'Click to choose PLS or Gaussian Process Regression to build a model automatically'.
- Dataset**: A list of datasets with 'Data2' selected. Callout boxes 'Data management' (Import and delete data.) and 'Model management' (Compare the performance of different models) point to the 'Import Data' and 'Delete Data' buttons at the bottom of this section.
- Available variables**: A list of variables including 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'.
- Working data:** Shows 'Data2' with options for 'Time Trend' and 'Histogram' (labeled 'Data visualization'). Below this is a dropdown for 'Select a tag from second list as output' and a 'Selected output:' field.
- Data Preprocess**: A dropdown menu set to 'Customized Preprocess-->' (labeled 'Preprocessing methods') and an 'Automatic Preprocessing' button.
- Modeling**: A dropdown menu set to 'Modeling-->' (labeled 'Modeling algorithms').

Additional callout boxes include 'Data modeling' pointing to the 'Modeling' tab and 'Delete Selected Dataset' pointing to a button at the bottom of the interface.



Part I: Start Guide

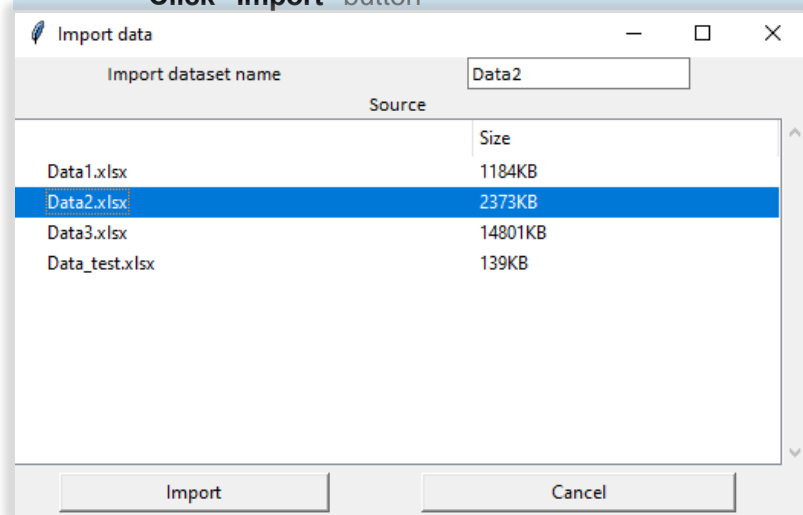
Soft Sensor Analytics



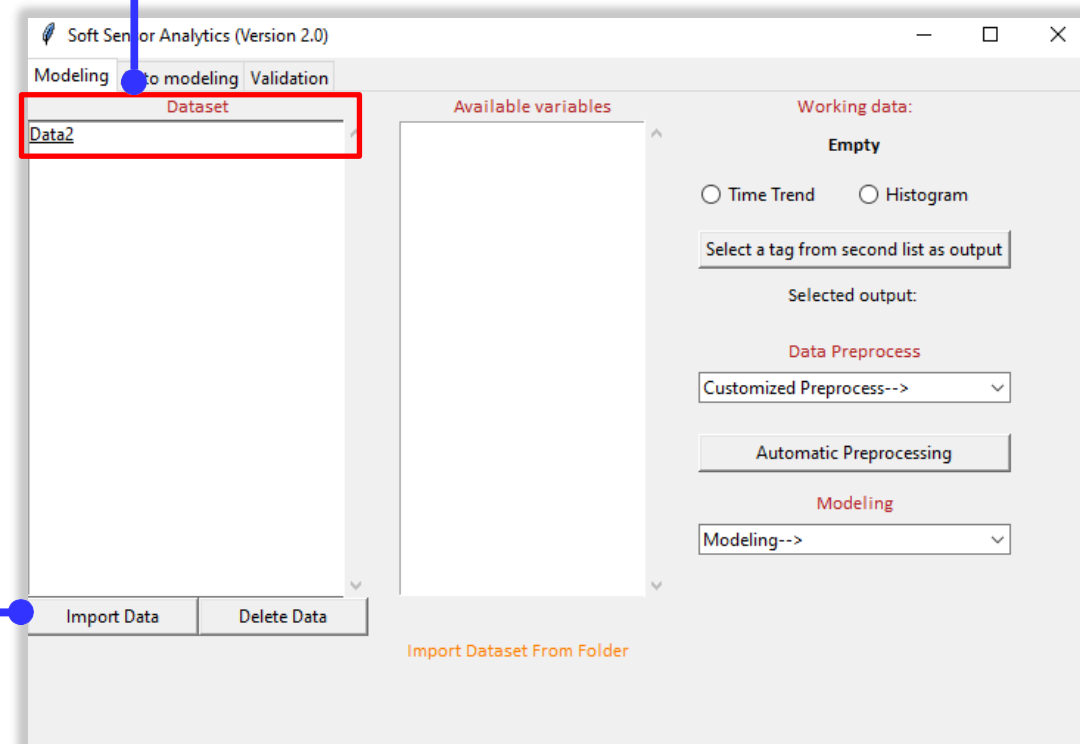
Quick Start Guide – Import data sets

Note: Please save data sets in the “.xlsx” format

- 1
- Click “Import Data” button
 - Select data file “Data2.xlsx”
 - Click “Import” button



“Data2” has been imported

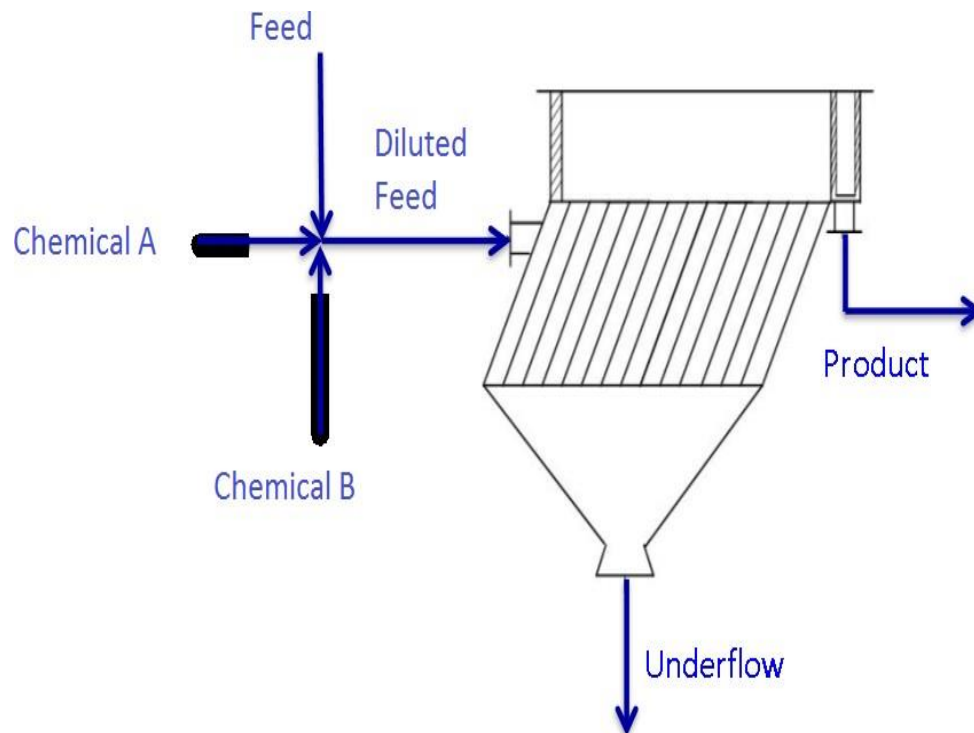


Soft Sensor Analytics



Quick Start Guide – A Case Study on Chemical Process (Data2)

The case study is based on normalized data for a chemical process. Suppose our objective is to estimate “Chemical A” using soft sensors. The data is saved as "Data2".



Symbol	Description
Chemical A	Chemical A flowrate
FeedA	Unit A Feed
FeedB	Unit B Feed
FeedC	Unit C Feed
FeedD	Unit D Feed
UnderflowA	Unit A underflow
UnderflowB	Unit B underflow
UnderflowC	Unit C underflow
UnderflowD	Unit D underflow
Lab	Chemical ratio value
DenN	Chemical A Density
DenFr	Feed Density

Soft Sensor Analytics



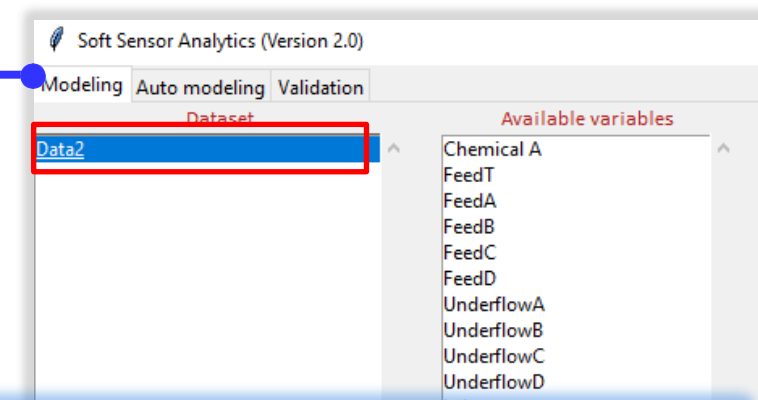
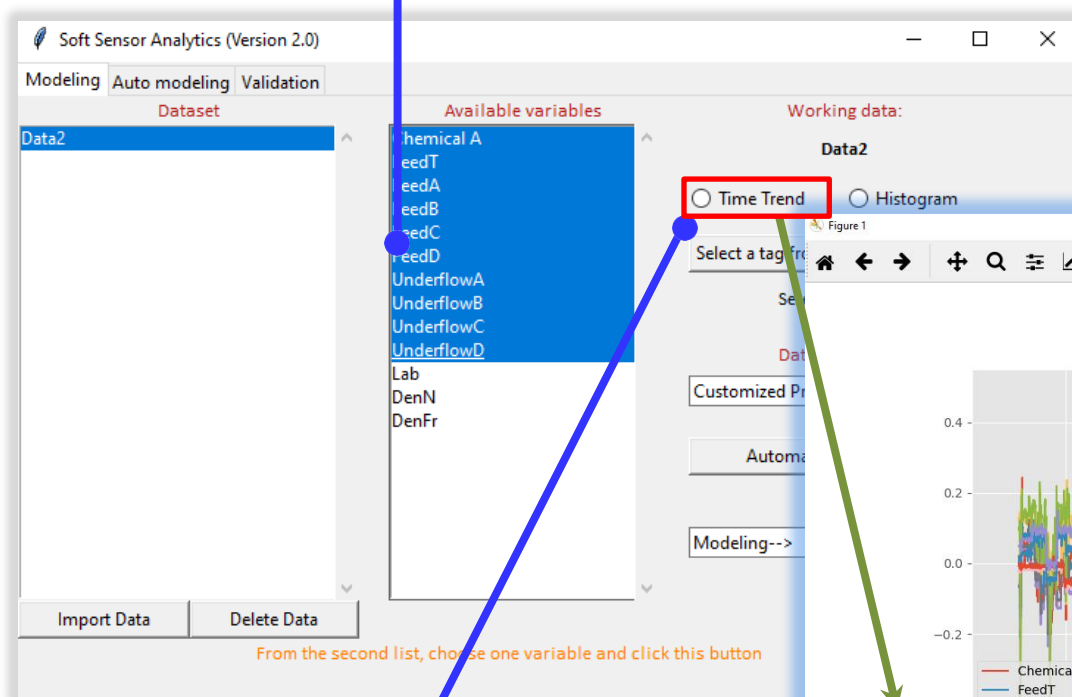
Visualize the selected data

3

Drag-select to select variables for visualization
Hold “CTRL” to multi-select if they are not consecutive

2

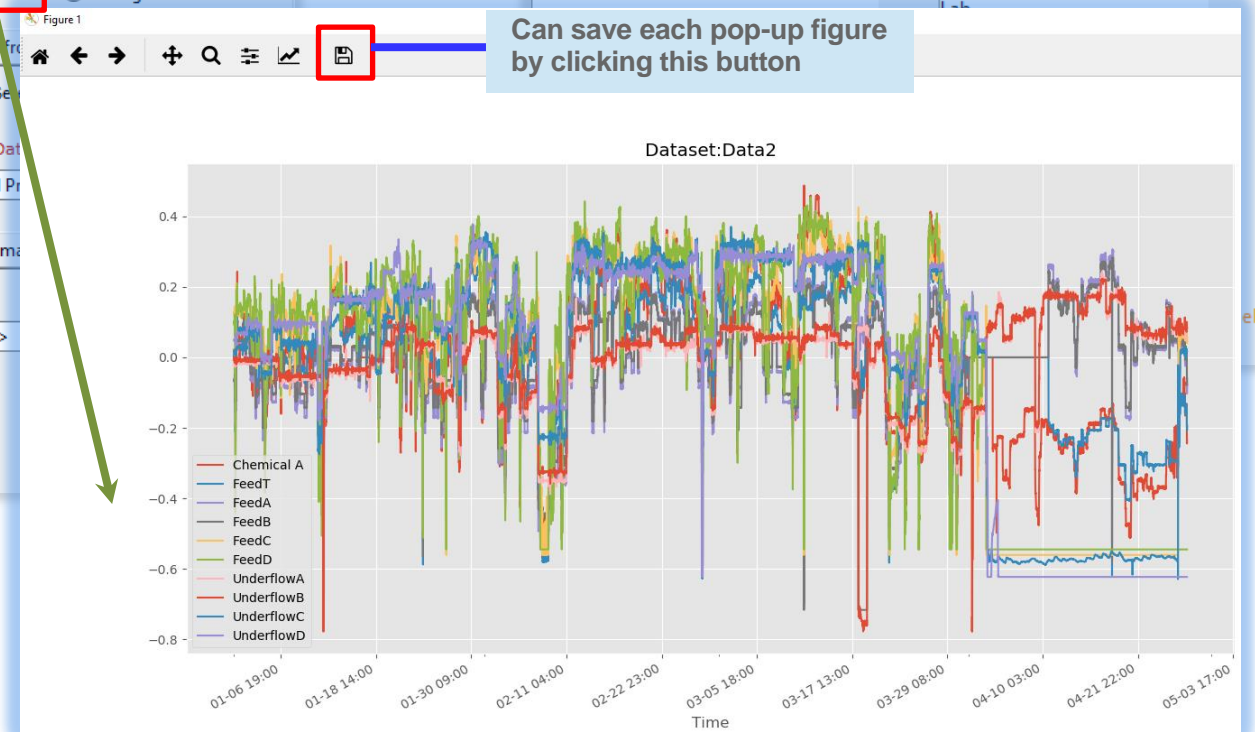
- Click on the **“Modeling”** tab.
- Click on **“Data2”** from the left workspace (i.e., “Dataset” column) to select it



Can save each pop-up figure by clicking this button

4

Click **“Time Trend”** to visualize the selected variables



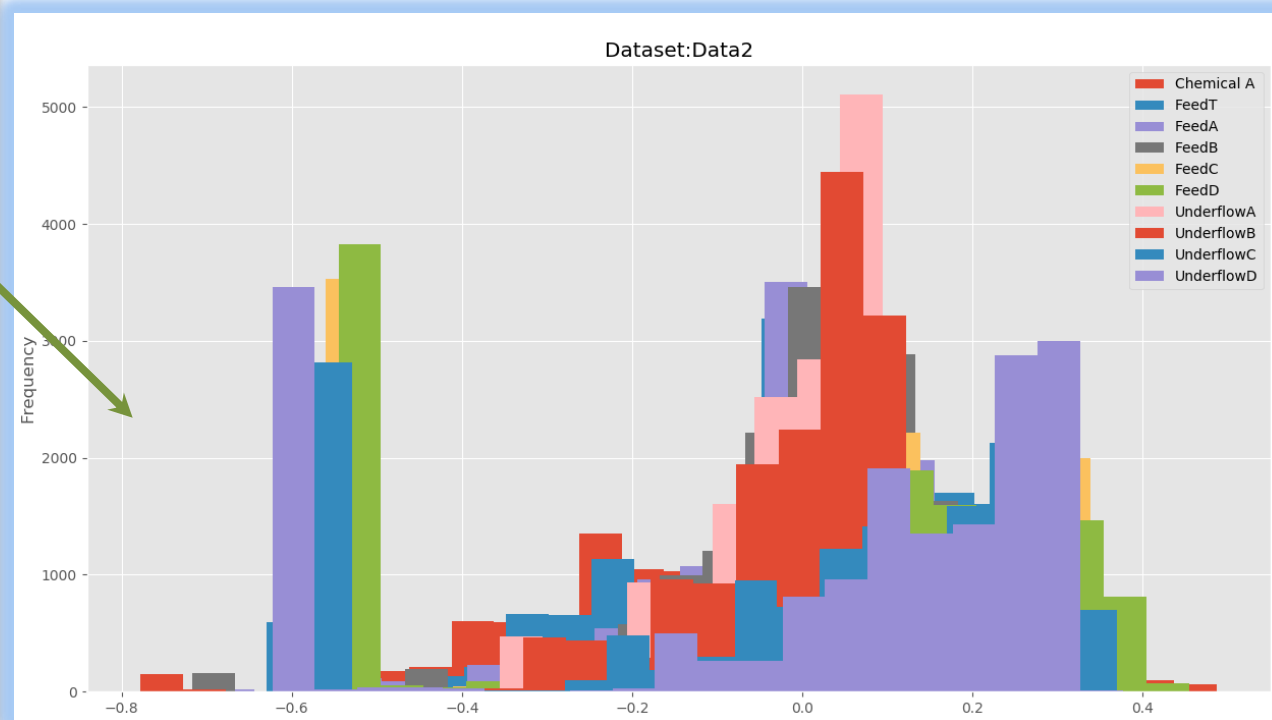
Note: Please close the current popup figures before taking next action

Soft Sensor Analytics



Visualize the selected data

The screenshot shows a software interface with a list of available variables on the left and data processing options on the right. The variables list includes Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, and DenFr. The right panel has sections for "Working data:" (with "Data2" selected and "Histogram" radio button highlighted), "Data Preprocess:" (with "Customized Preprocess-->" selected), and "Modeling:" (with "Modeling-->" selected). A blue arrow points from the "Histogram" button to the histogram chart on the right.



6

Click "Histogram" to generate histogram for selected variables

Soft Sensor Analytics



Select an output for modeling

7 Select output from second list

- Choose a variable from the second list
- Click "Select a tag from second list as output" button to select this variable as the output for modeling

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' application window. The 'Modeling' tab is active. The 'Dataset' list on the left contains 'Data2'. The 'Available variables' list in the center contains 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'. The 'Chemical A' variable is highlighted with a red box. The 'Working data:' section on the right shows 'Data2' selected. Below this, there are radio buttons for 'Time Trend' and 'Histogram'. A button labeled 'Select a tag from second list as output' is highlighted with a blue circle, and a red box highlights the 'Selected Output: <<Chemical A>>' text field. Below this are sections for 'Data Preprocess' (with a dropdown menu set to 'Customized Preprocess-->') and 'Modeling' (with a dropdown menu set to 'Modeling-->'). At the bottom, there are 'Import Data' and 'Delete Data' buttons, and a 'Delete Selected Dataset' link.

Soft Sensor Analytics



Data preprocessing – Remove missing data and outliers

8

This step is to remove missing data from the dataset

- Select "Data2" from Dataset column
- Click "Customized Preprocess -->" on the right
- Click "Remove missing data"; new data is "Data2_RM"

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Dataset' column on the left contains 'Data2' and 'Data2_RM', with 'Data2' selected. The 'Available variables' column lists various sensor tags like 'Chemical A', 'FeedT', 'FeedA', etc. The 'Working data:' section shows 'Data2' selected. The 'Data Preprocess' section is open, showing a list of preprocessing algorithms. The 'Remove missing data' option is highlighted with a red box. A green arrow points from the 'Remove missing data' option in the 'Data Preprocess' list to the 'Data2' entry in the 'Dataset' column. At the bottom, there are 'Import Data' and 'Delete Data' buttons.

This provides different data preprocessing algorithms

Soft Sensor Analytics



Data preprocessing – Remove missing data and outliers

9

This step is to remove outliers in the dataset

- Select "Data2_RM" from Dataset column
- Click "Customized Preprocess -->" on the right
- Click "Remove missing data"; new data is "Data2_RM_RO"

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Dataset' column on the left contains 'Data2', 'Data2_RM', and 'Data2_RM_RO'. The 'Available variables' column in the center lists various process variables like 'Chemical A', 'FeedT', 'FeedA', etc. The 'Working data:' section on the right shows 'Data2_RM' selected. Below it, the 'Data Preprocess' dropdown menu is open, with 'Remove outlier' selected. A red box highlights 'Data2_RM_RO' in the dataset list, and another red box highlights 'Remove outlier' in the dropdown menu. A green arrow points from the 'Remove outlier' option to the 'Data2_RM_RO' entry. At the bottom, there are 'Import Data' and 'Delete Data' buttons. A caption at the bottom reads: 'This provides different data preprocessing algorithms'.

Soft Sensor Analytics



Data preprocessing – Resample and estimate output delay

10

This step is to resample data (if two sampling rates exist) and estimate time delay between inputs and output

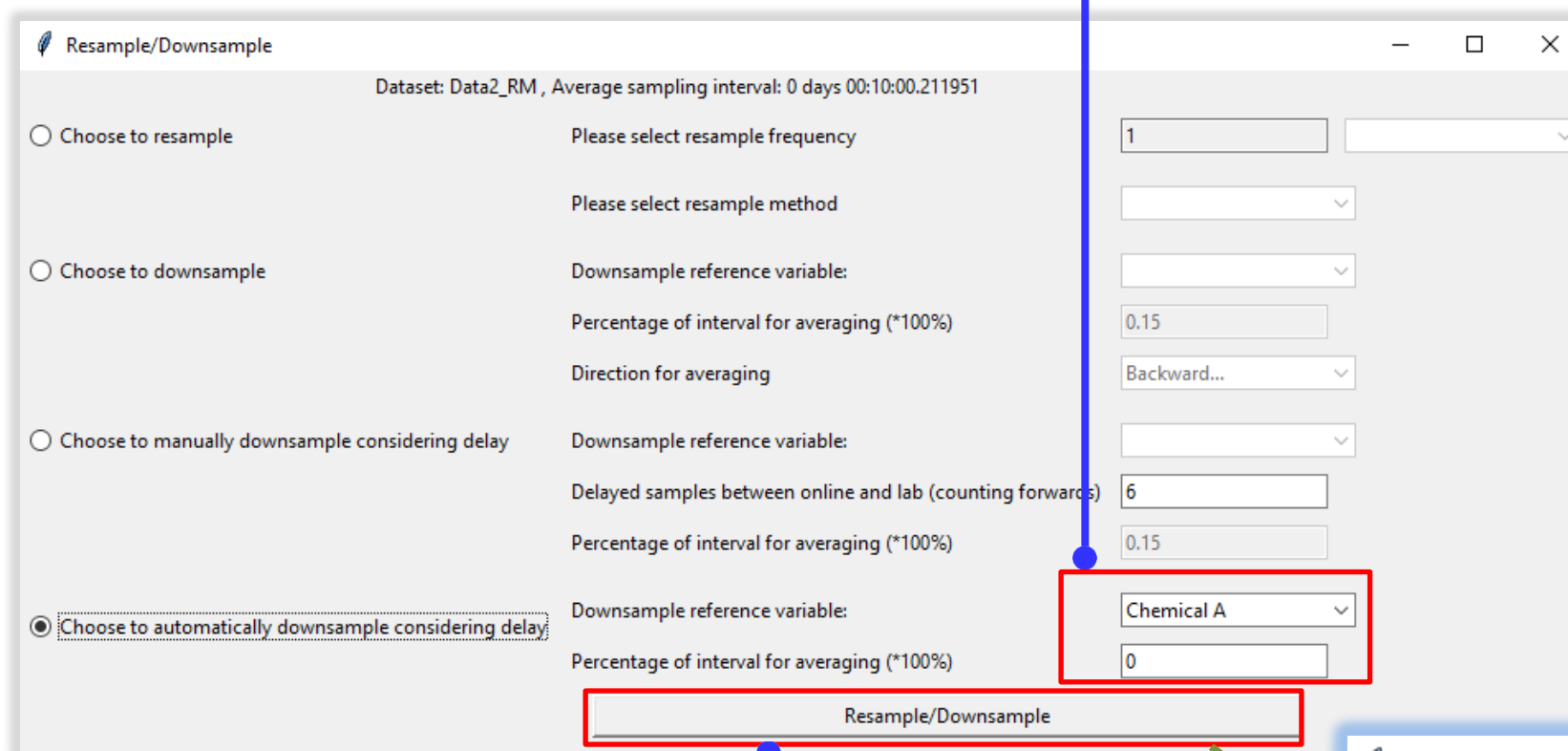
- Select “Data2_RM_RO” from Dataset column
- Click “Customized Preprocess -->” on the right
- Click “Resample/Downsample”; a new window will pop up

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Dataset' column on the left has 'Data2_RM_RO' selected and highlighted with a red box. The 'Available variables' column in the center lists various variables like 'Chemical A', 'FeedT', 'FeedA', etc. The 'Working data' section on the right shows 'Data2_RM_RO' selected. Below this, the 'Data preprocess' dropdown menu is open, with 'Resample/Downsample' selected and highlighted with a red box. A blue line connects the selected dataset to the preprocessing menu. At the bottom, a note states: 'This provides different data preprocessing algorithms'.

Data preprocessing – Resample and estimate output delay

11

- Select “**Chemical A**” as the reference (because it is the output)
- Enter “**0**” as the percentage of interval for averaging



The screenshot shows the 'Resample/Downsample' dialog box for the dataset 'Data2_RM'. The 'Choose to automatically downsample considering delay' option is selected. The 'Downsample reference variable' is set to 'Chemical A' and the 'Percentage of interval for averaging (*100%)' is set to '0'. The 'Resample/Downsample' button is highlighted with a red box.

Dataset: Data2_RM , Average sampling interval: 0 days 00:10:00.211951

Choose to resample

Please select resample frequency: 1

Please select resample method: [dropdown]

Choose to downsample

Downsample reference variable: [dropdown]

Percentage of interval for averaging (*100%): 0.15

Direction for averaging: Backward...

Choose to manually downsample considering delay

Downsample reference variable: [dropdown]

Delayed samples between online and lab (counting forwards): 6

Percentage of interval for averaging (*100%): 0.15

Choose to automatically downsample considering delay

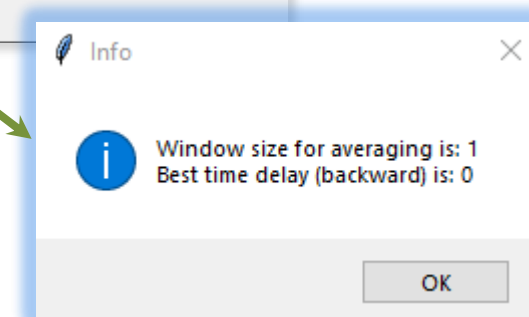
Downsample reference variable: Chemical A

Percentage of interval for averaging (*100%): 0

Resample/Downsample

12

- Click “Resample/Downsample” button



Soft Sensor Analytics



Data preprocessing – Resample and estimate output delay

A new dataset “Data2_RM_RO_RS” is generated for modeling

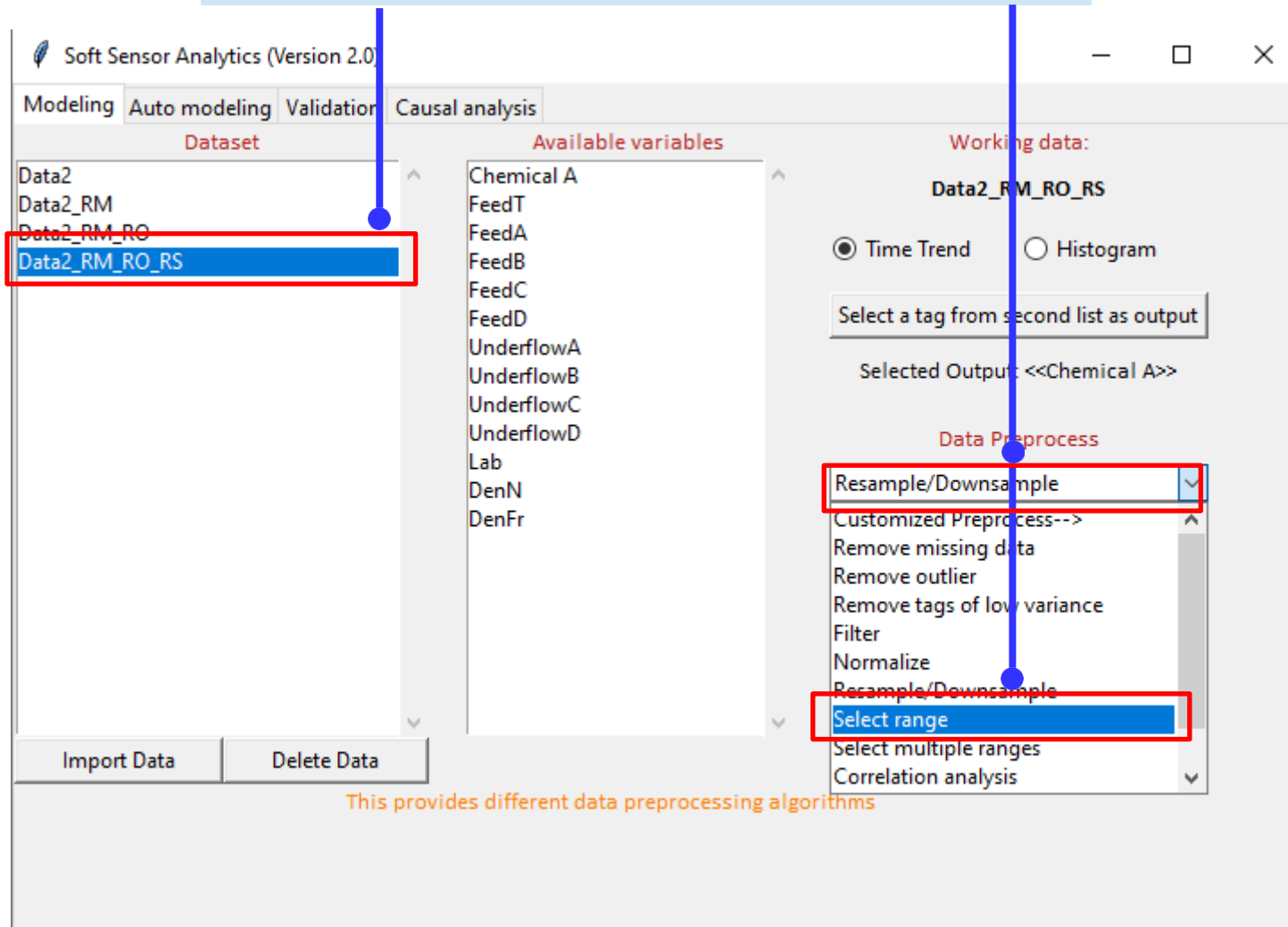
The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Modeling' tab is active. The 'Dataset' list on the left contains 'Data2', 'Data2_RM', 'Data2_RM_RO', and 'Data2_RM_RO_RS', with the last one highlighted in blue and a red box around it. A blue arrow points from the text above to this dataset. The 'Available variables' list on the right includes 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'. The 'Working data' section shows 'Data2_RM_RO_RS' selected, with 'Time Trend' selected over 'Histogram'. Below this, there is a button 'Select a tag from second list as output' and a text field 'Selected Output: <<Chemical A>>'. The 'Data Preprocess' section has a dropdown menu set to 'Resample/Downsample' and an 'Automatic Preprocessing' button. The 'Modeling' section has a dropdown menu set to 'Modeling-->'. At the bottom, there are 'Import Data' and 'Delete Data' buttons. An orange text instruction at the bottom reads: 'From the second list, choose one variable and click this button'.

Data preprocessing – Split data into two parts for training and validation

13

Get a smaller dataset for training

- Select “Data2_RM_RO_RS”
- Click “Customized Preprocess-->”, then Click “Select range”; a window will pop up, as shown in the next slide



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS

Available variables

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Working data:

Data2_RM_RO_RS

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Chemical A>>

Data Preprocess

Resample/Downsample
Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Select multiple ranges
Correlation analysis

Import Data Delete Data

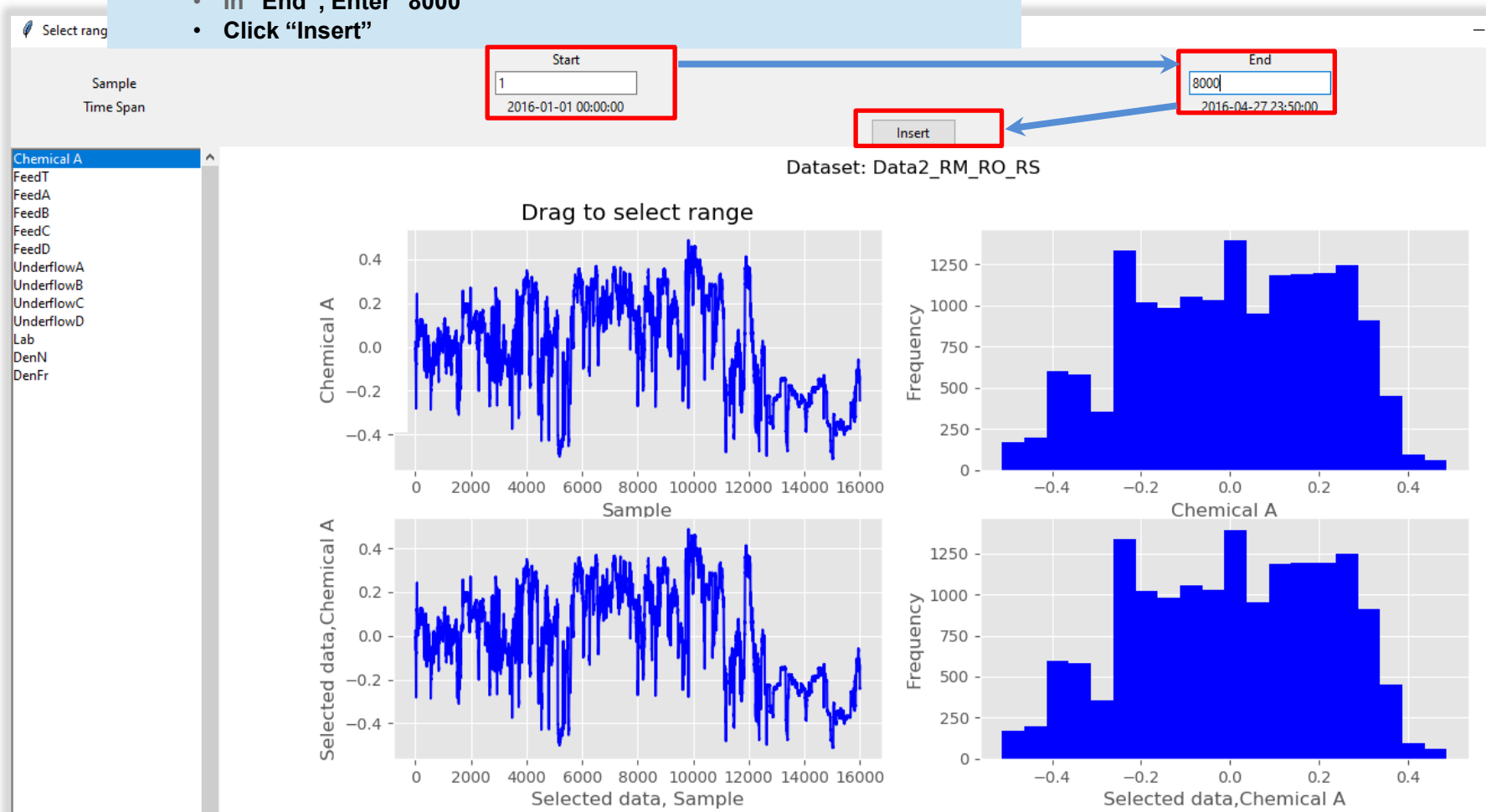
This provides different data preprocessing algorithms

Data preprocessing – Split data for training

13

To split data enter the range of dataset for training, do the following:

- Click “Chemical A” on the left column
- In “Start”, Enter “1”
- In “End”; Enter “8000”
- Click “Insert”

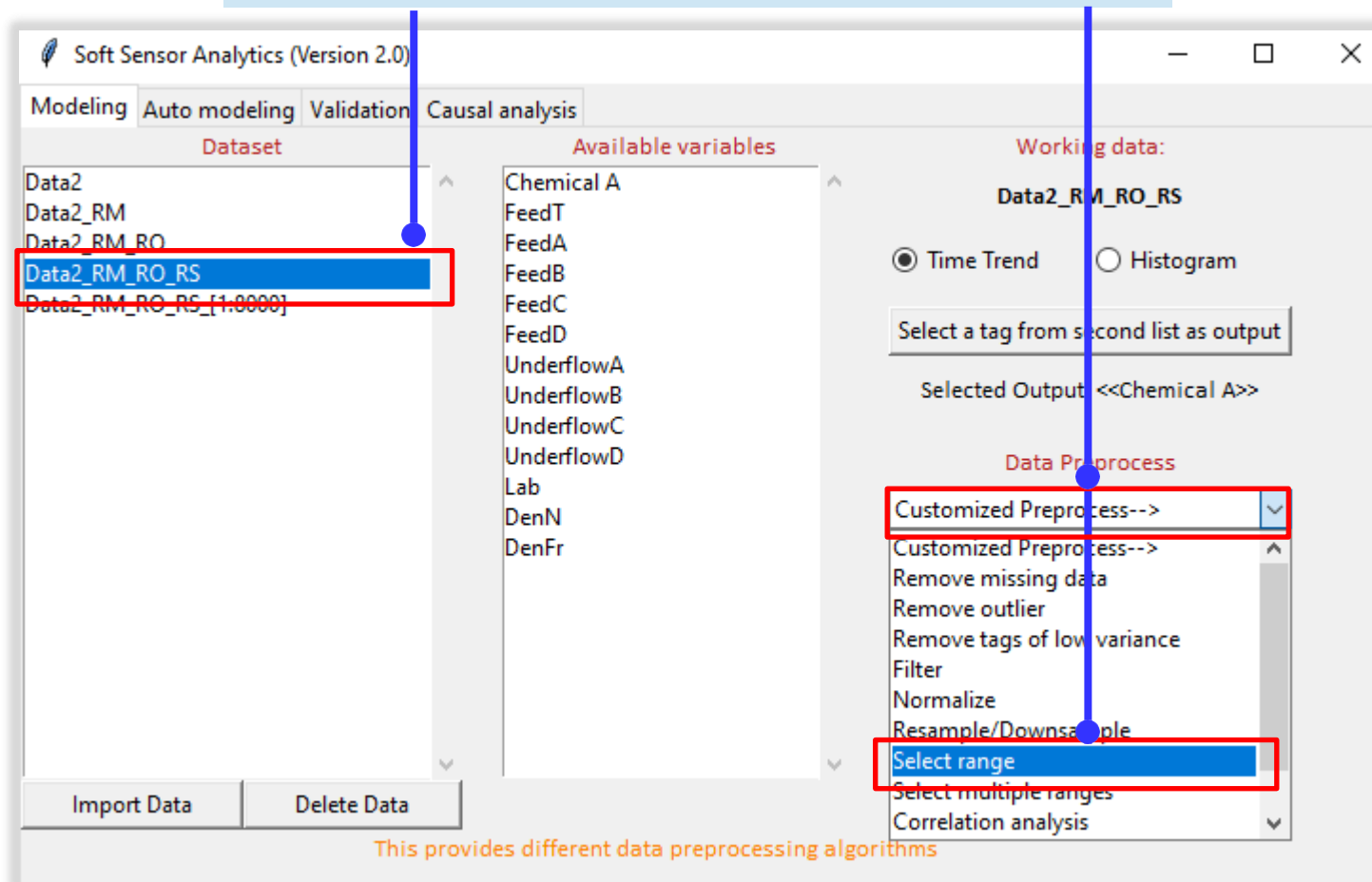


Data preprocessing – Split data into two parts for training and validation

14

Get a smaller dataset for validation

- Select “Data2_RM_RO_RS” again
- Click “Customized Preprocess-->”, then Click “Select range”; a window will pop up, as shown in the next slide



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

- Data2
- Data2_RM
- Data2_RM_RO
- Data2_RM_RO_RS**
- Data2_RM_RO_RS_{1.0000}

Available variables

- Chemical A
- FeedT
- FeedA
- FeedB
- FeedC
- FeedD
- UnderflowA
- UnderflowB
- UnderflowC
- UnderflowD
- Lab
- DenN
- DenFr

Working data: Data2_RM_RO_RS

Time Trend Histogram

Select a tag from second list as output

Selected Output <<Chemical A>>

Data Preprocess

- Customized Preprocess-->**
- Customized Preprocess-->
- Remove missing data
- Remove outlier
- Remove tags of low variance
- Filter
- Normalize
- Resample/Downsample
- Select range**
- Select multiple ranges
- Correlation analysis

Import Data Delete Data

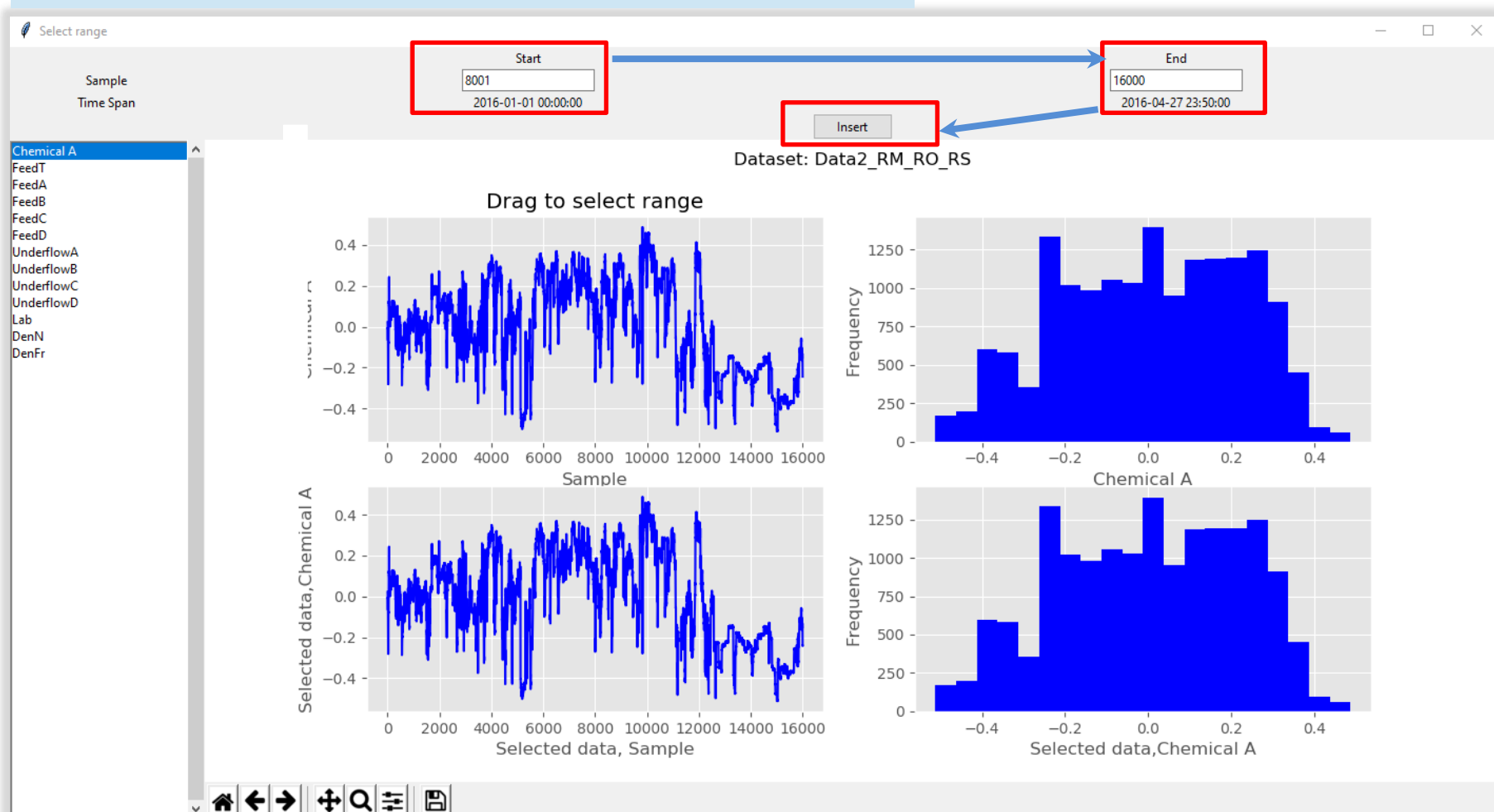
This provides different data preprocessing algorithms

Data preprocessing – Split data into smaller sets

13

To split data enter the range of dataset for validation, do the following:

- Click “Chemical A” on the left column
- In “Start”, Enter “8001”
- In “End”; Enter “16000”
- Click “Insert”



Note: Please close the current popup windows before taking next action

Soft Sensor Analytics



Generated data sets after pre-processing

Updated Dataset list

After preprocess two new datasets are generated
One can be used for training, other for validation

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Modeling' tab is active, and the 'Auto modeling' sub-tab is selected. The 'Dataset' list on the left includes 'Data2', 'Data2_RM', 'Data2_RM_RO', 'Data2_RM_RO_RS', 'Data2_RM_RO_RS_[1:8000]', and 'Data2_RM_RO_RS_[8001:16000]'. The 'Available variables' list in the center includes 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'. The 'Working data' section on the right shows 'Data2_RM_RO_RS' selected, with 'Time Trend' chosen over 'Histogram'. Below this, there is a dropdown menu for 'Select a tag from second list as output' with 'Selected Output: <<Chemical A>>' displayed. The 'Data Preprocess' section has a dropdown menu set to 'Customized Preprocess-->'. The 'Modeling' section has a dropdown menu set to 'Modeling-->'. At the bottom, there are 'Import Data' and 'Delete Data' buttons. A red box highlights the 'Data2_RM_RO_RS_[1:8000]' dataset, and an orange text box at the bottom right says 'From the second list, choose one variable and click this button'.

Soft Sensor Analytics



A more flexible way to generate datasets for training/validation

This feature enables users to select discontinuous windows and connect them together to form a new data set for training and/or validation

Get a smaller dataset for training

- Select "Data2_RM_RO_RS"
- Click "Customized Preprocess-->", then Click "Select multiple ranges"; a window will pop up, as shown in the next slide.

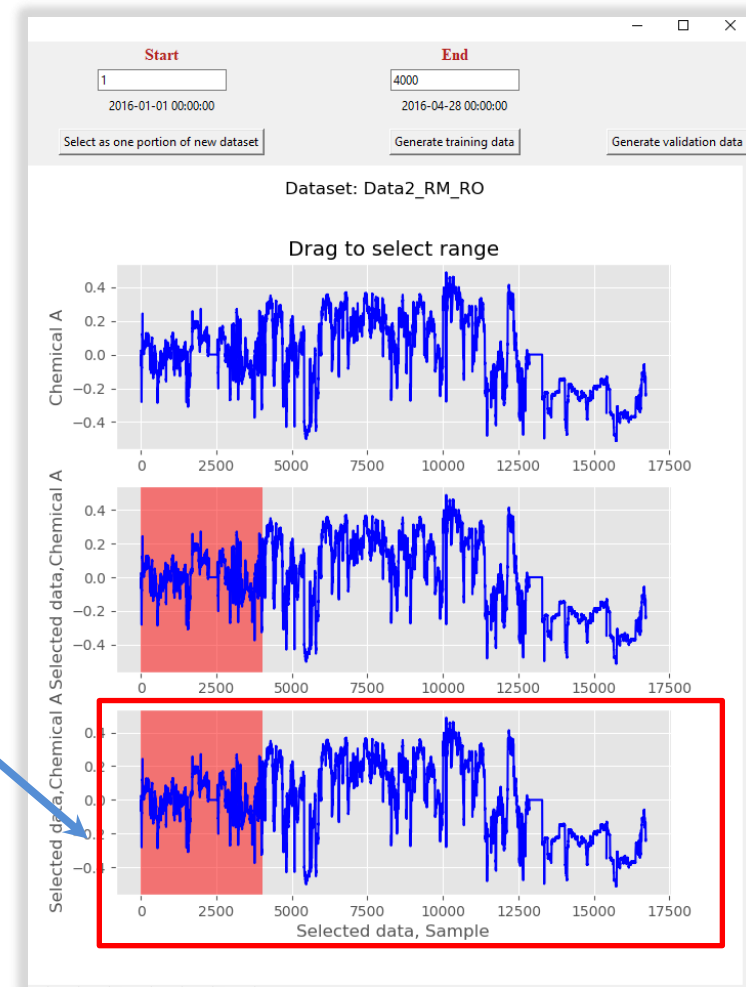
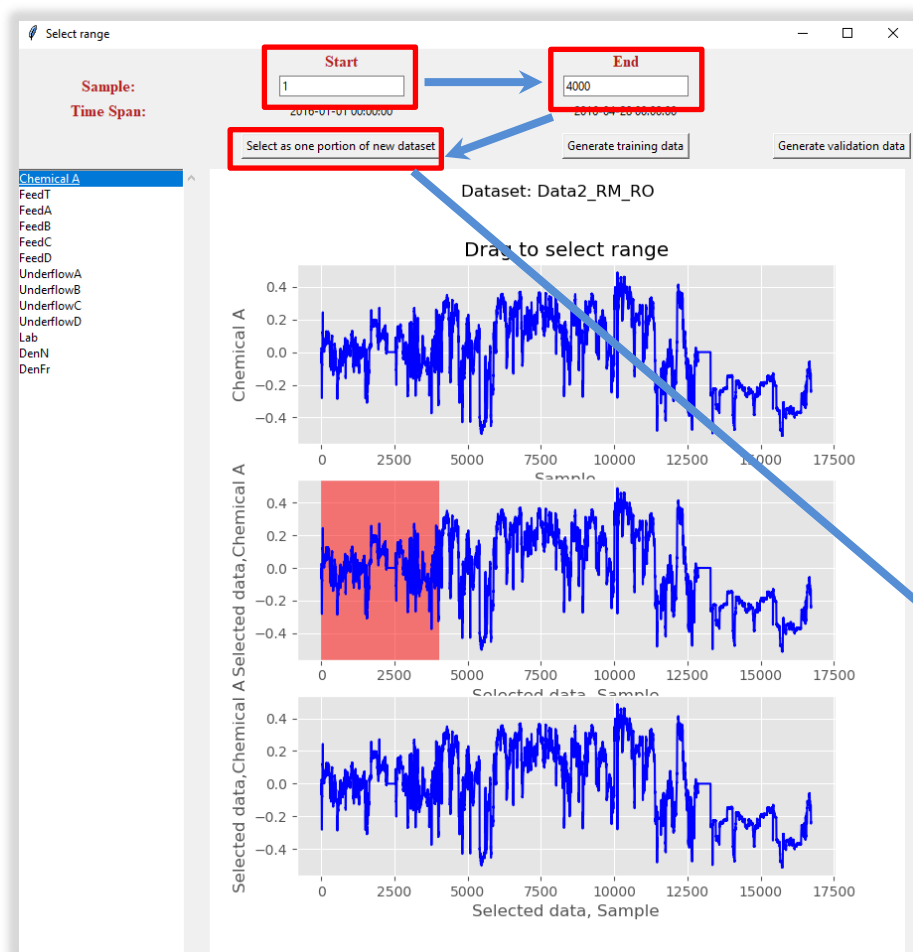
The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Dataset' list on the left includes 'Data2_RM_RO_RS', which is highlighted with a red box. The 'Available variables' list in the center includes 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'. The 'Working data' section on the right shows 'Data2_RM_RO_RS' selected, with 'Time Trend' and 'Histogram' options. The 'Data Preprocess' dropdown menu is open, showing 'Customized Preprocess-->' selected, with a red box around it. Below it, the 'Select multiple ranges' option is also highlighted with a red box. A blue line connects the 'Data2_RM_RO_RS' dataset to the 'Select multiple ranges' option in the dropdown menu. At the bottom, there are 'Import Data' and 'Delete Data' buttons.

This provides different data preprocessing algorithms

A more flexible way to generate datasets for training/validation

To select the first portion of the new data set, do the following:

- Click **“Chemical A”** on the left column
- In **“Start”**, Enter **“1”**
- In **“End”**; Enter **“4000”**
- Click **“Select as one portion of new dataset”**



Soft Sensor Analytics

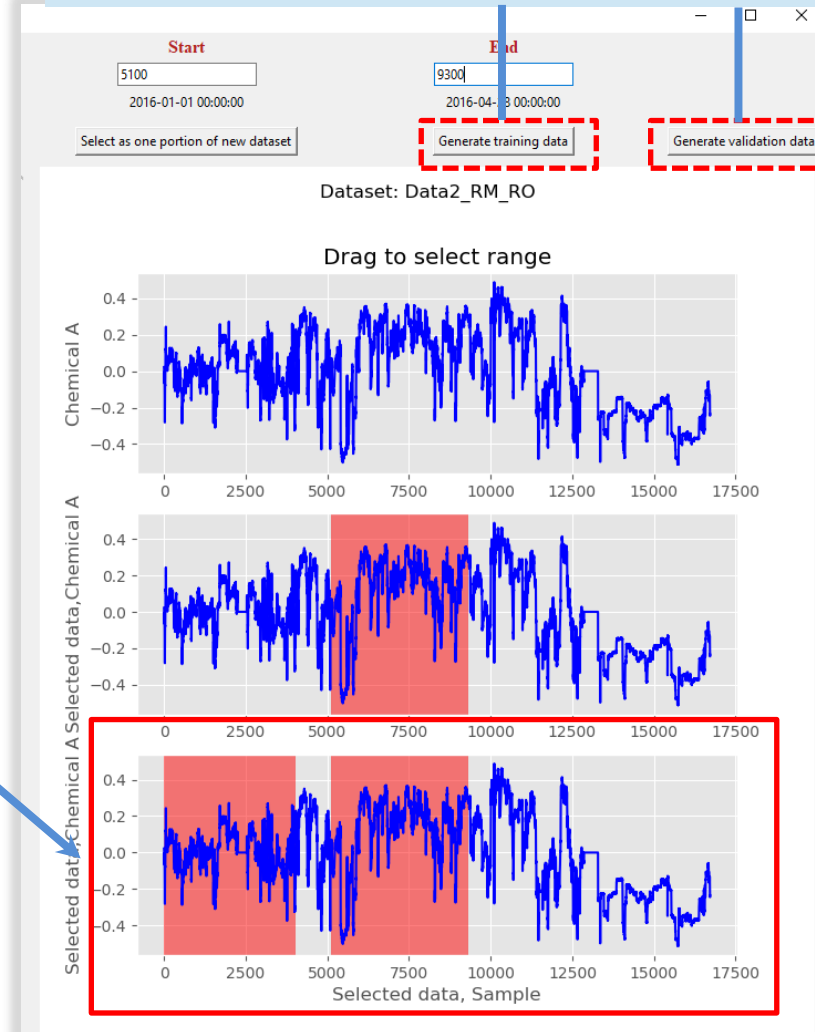
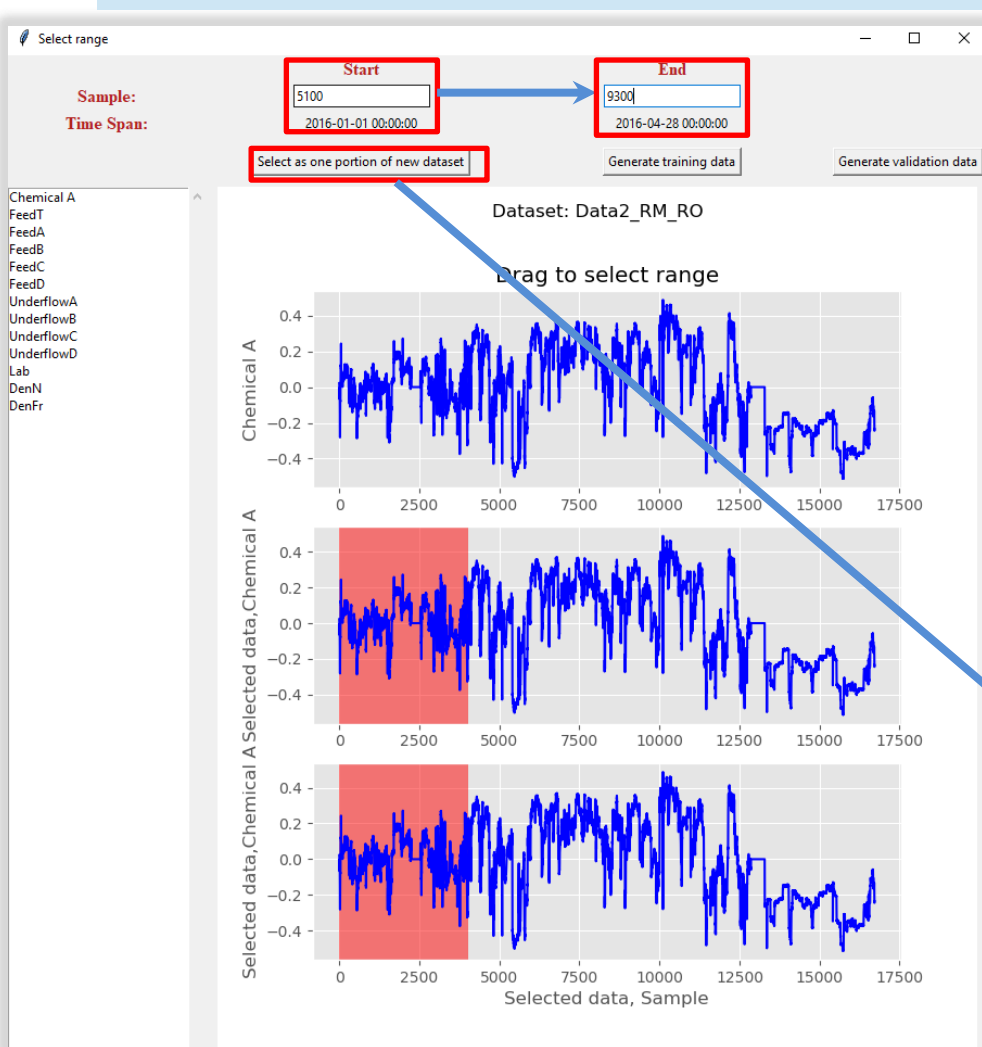


A more flexible way to generate datasets for training/validation

To select the second portion of the new data set, do the following:

- Click “**Chemical A**” on the left column
- In “**Start**”, Enter “**5100**”
- In “**End**”; Enter “**9300**”
- Click “**Select as one portion of new dataset**”

If the selected ranges seem to be good, then may click “**Generate training data**” or “**Generate validation data**” to generate a new data set that contains discontinuous windows of samples for training and/or validation.



Preprocessing – Correlation analysis

Conduct correlation analysis to check the influential variables as well as the co-linearity of the data-set.

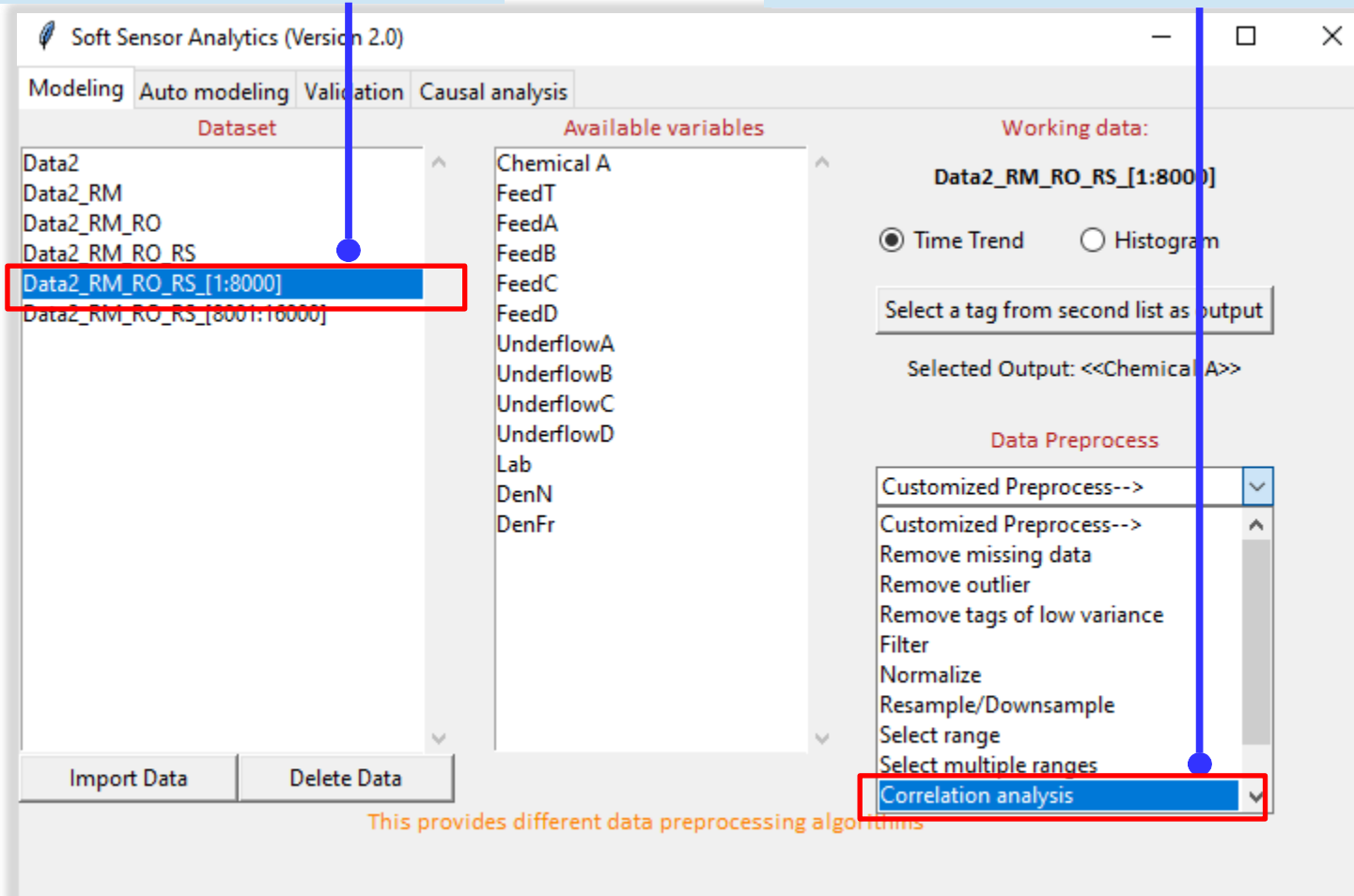
1 Select data set for correlation analysis

- Click “Modeling” tab
- Select “Data2_RM_RO_RS_[1:8000]”

2 Choose method

Choose method

- Click “Customized Preprocessing-->” on the right
- Click “Correlation Analysis”, a window will pop up



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset Available variables Working data:

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS
Data2_RM_RO_RS_[1:8000]
Data2_RM_RO_RS_[8001:16000]

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Chemical A>>

Data Preprocess

Customized Preprocess-->
Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Select multiple ranges
Correlation analysis

Import Data Delete Data

This provides different data preprocessing algorithms

Soft Sensor Analytics



Preprocessing - Correlation analysis

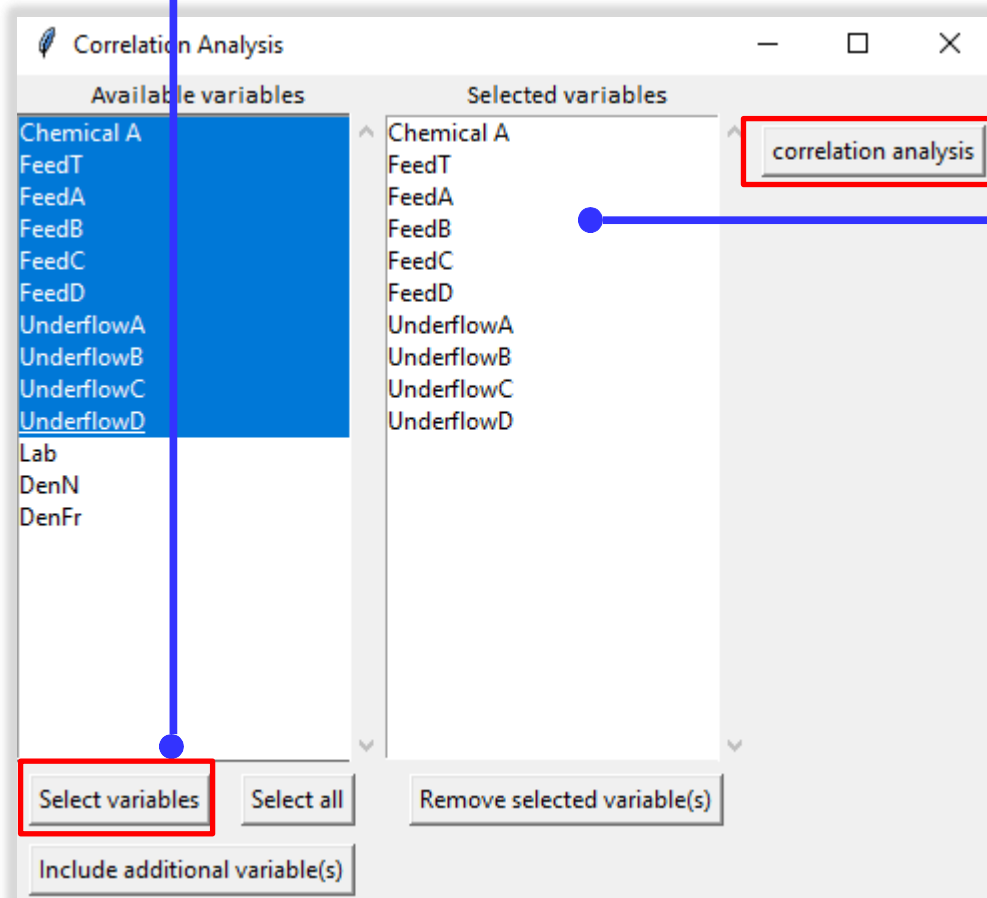
3

Choose variables to analyze

- **Drag-select** (can hold "CTRL") in the left list to select variables "**Chemical A; FeedT; FeedA;...; UnderflowD**"
- Click "**Select variables**"

Selected variables

The variables that are selected manually for correlation analysis are shown in the right column



4

Correlation analysis

Click "**correlation analysis**", a heatmap will be available, as given in the next slide

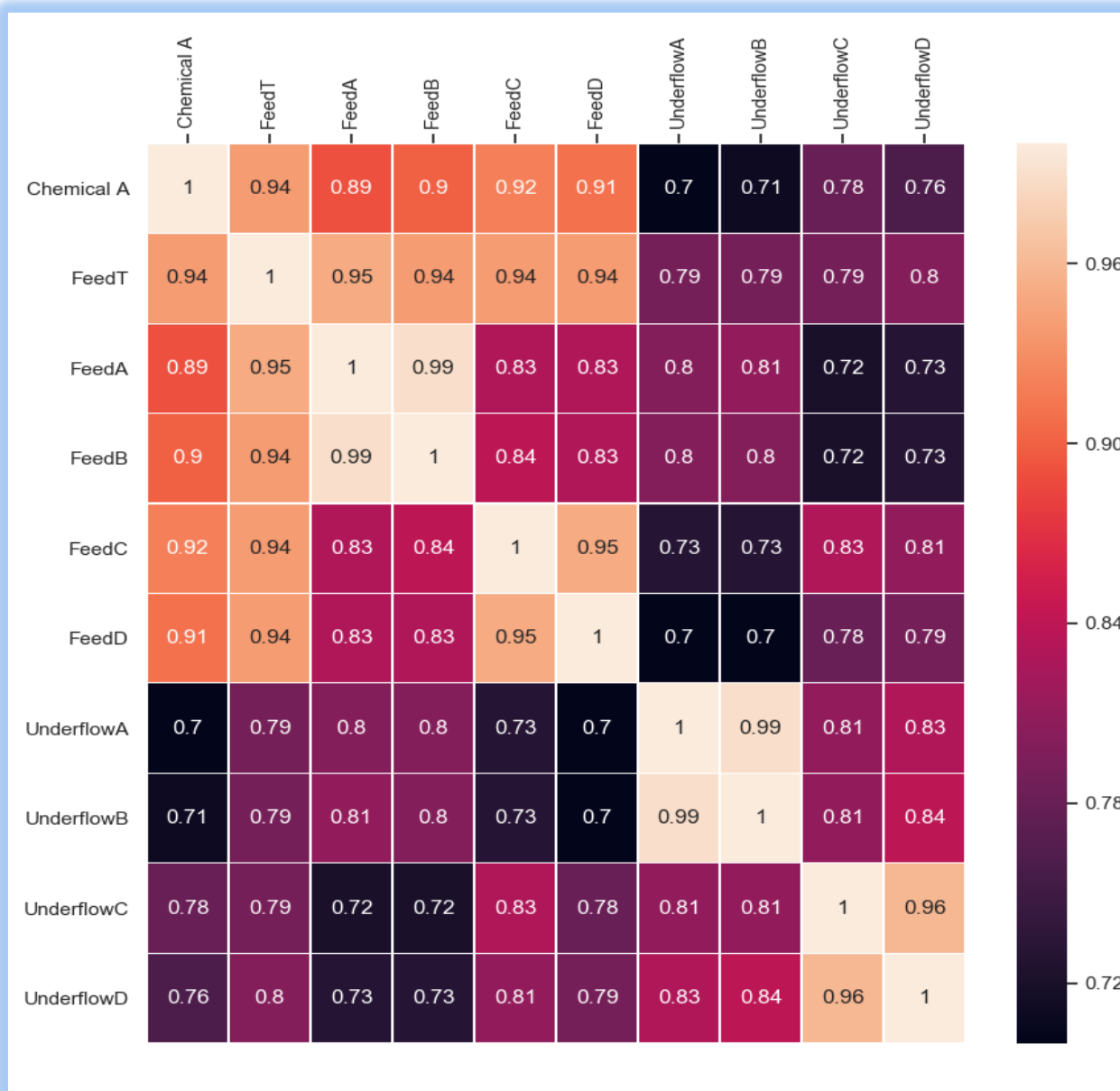


Correlation analysis – Result

Results

Variables are highly correlated to each other if the absolute value of the corresponding index is large.

One can select more variables and get more information from the correlation test results.



Note: Please close the current popup figures before taking next action



Correlation analysis

If one would like to choose all variables (the number of the variables should not be too large), then can click **“Select all”**

Selected variables

The variables that are selected manually for correlation analysis are shown in the right column

If one would like to deselect a few variables, then select in the right list, then click **“Remove selected variable(s)”** to remove the ones that you would like to deselect.

If one would like to add additional variables, then select in the right list, then click **“Included additional variable(s)”** to remove the ones that you would like to deselect.

Preprocessing– Mutual information analysis

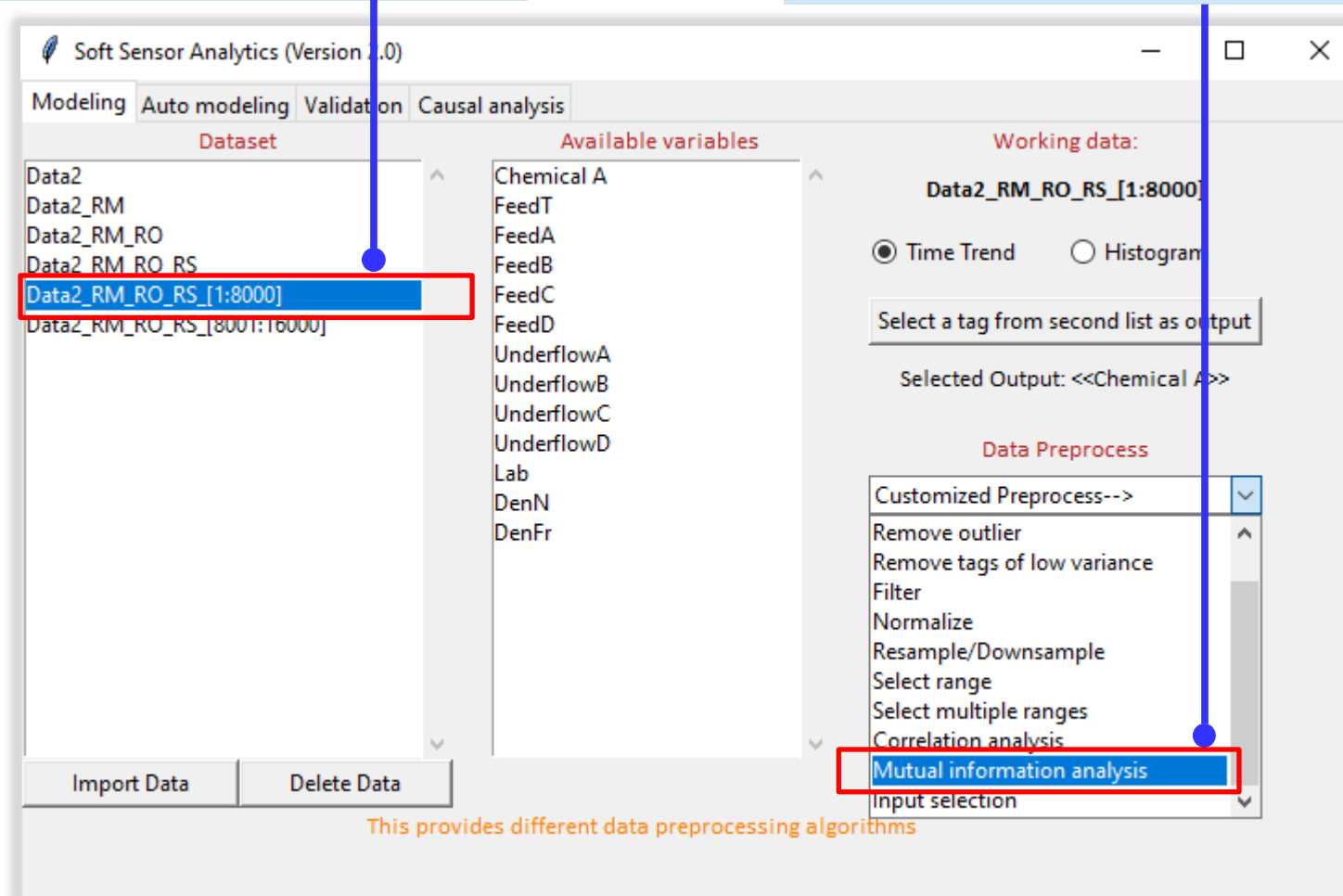
Data preprocess: mutual information is a measure of the mutual dependence between the two variables. The analysis procedure is like correlation analysis.

1 Select data set for Mutual information

- Click “Modeling” tab
- Select “Data2_RM_RO_RS_[1:8000]”

2 Choose method

- Click “Preprocessing-->” on the right
- Click “Mutual information analysis”, a window pops up



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS
Data2_RM_RO_RS_[1:8000]
Data2_RM_RO_RS_[8001:16000]

Available variables

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Working data:

Data2_RM_RO_RS_[1:8000]

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Chemical A>>

Data Preprocess

Customized Preprocess-->

Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Select multiple ranges
Correlation analysis
Mutual information analysis
Input selection

Import Data Delete Data

This provides different data preprocessing algorithms



Quick Start Guide – Mutual information analysis

Data preprocess: mutual information is a measure of the mutual dependence between the two variables. The analysis procedure is like correlation analysis.

3

Choose variables to analyze

- **Drag-select** (can hold **CTRL**) in the left list to select variables **“Chemical A; FeedT; FeedA;...; UnderflowD”**
- Click **“Select variables”**

Selected variables

The variables that are selected manually for mutual information analysis are shown in the right list

A screenshot of a software window titled "Mutual information analysis". The window is divided into two main sections: "Available variables" on the left and "Selected variables" on the right. In the "Available variables" list, several items are highlighted in blue: "Chemical A", "FeedT", "FeedA", "FeedB", "FeedC", "FeedD", "Underflow A", "Underflow B", "Underflow C", and "Underflow D". Below this list is a button labeled "Select variables", which is highlighted with a red box. Other buttons include "Select all", "Include additional variable(s)", and "Remove selected variable(s)". In the "Selected variables" list, the same set of variables is listed. A red box highlights the "Mutual information analysis" button located between the two lists. Blue lines with circular endpoints connect the "Select variables" button to the text in step 3, the "Mutual information analysis" button to the text in step 4, and the "Selected variables" list to the text in the "Selected variables" box.

4

Mutual information analysis

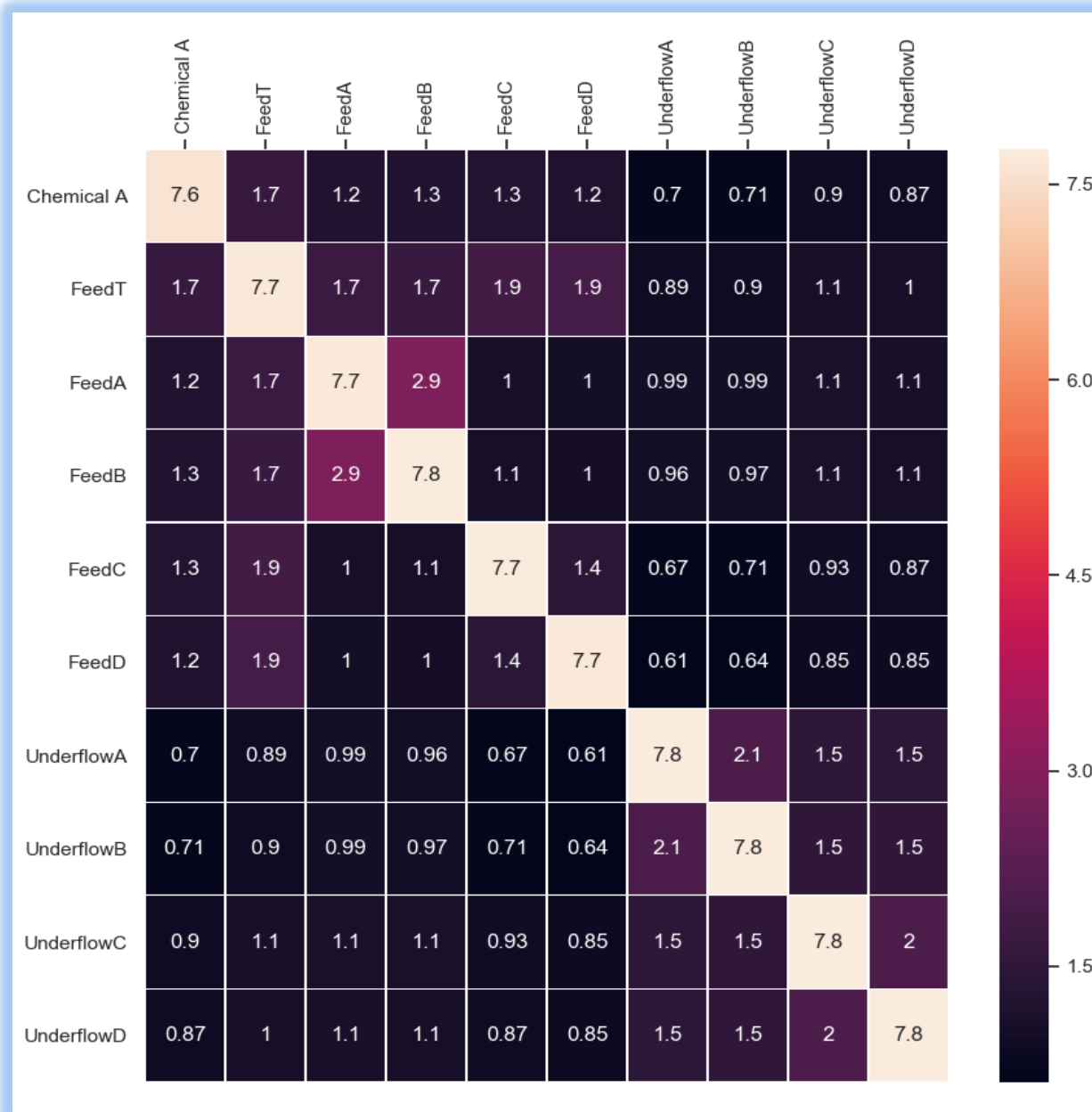
Then, Click **“Mutual information analysis”**



Mutual information analysis – Result

Results

Variables are highly dependent on each other tend to have larger indices based on the mutual information analysis.



Note: Please close the current popup figures before taking next action

Soft Sensor Analytics



Preprocessing - Mutual information analysis

If one would like to choose all variables (the number of the variables should not be too large), then can click **“Select all”**

Selected variables

The variables that are selected manually for correlation analysis are shown in the right column

If one would like to deselect a few variables, then select in the right list, then click **“Remove selected variable(s)”** to remove the ones that you would like to deselect.

If one would like to add additional variables, then select in the right list, then click **“Included additional variable(s)”** to remove the ones that you would like to deselect.

Soft Sensor Analytics



Selection of inputs

1

Select data and go to auto input selection

- Select "Data2_RM_RO_RS [1:8000]"
- Go to "Data Preprocess" and select "Auto input selection", a new window will pop up

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' application window. The 'Modeling' tab is active, and the 'Auto modeling' sub-tab is selected. The interface is divided into three main sections: 'Dataset', 'Available variables', and 'Working data'.
- The 'Dataset' list on the left contains several entries, with 'Data2_RM_RO_RS [1:8000]' highlighted in a red box.
- The 'Available variables' list in the center contains various process variables like 'Chemical A', 'FeedT', 'FeedA', etc.
- The 'Working data' section on the right shows 'Data2_RM_RO_RS [1:8000]' as the selected data source. Below this, there are radio buttons for 'Time Trend' (selected) and 'Histogram'. A text box prompts the user to 'Select a tag from second list as output', with 'Selected Output: <<Chemical A>>' displayed below it.
- The 'Data Preprocess' section at the bottom right features a dropdown menu currently set to 'Customized Preprocess-->'. A red box highlights the 'Input selection' option at the bottom of this list.
- At the bottom left, there are 'Import Data' and 'Delete Data' buttons.
- A blue line connects the selected dataset in the 'Dataset' list to the 'Input selection' option in the 'Data Preprocess' dropdown. Another blue line connects the 'Input selection' option to the 'Data Preprocess' section header.

This provides different data preprocessing algorithms

Automatic selection of inputs

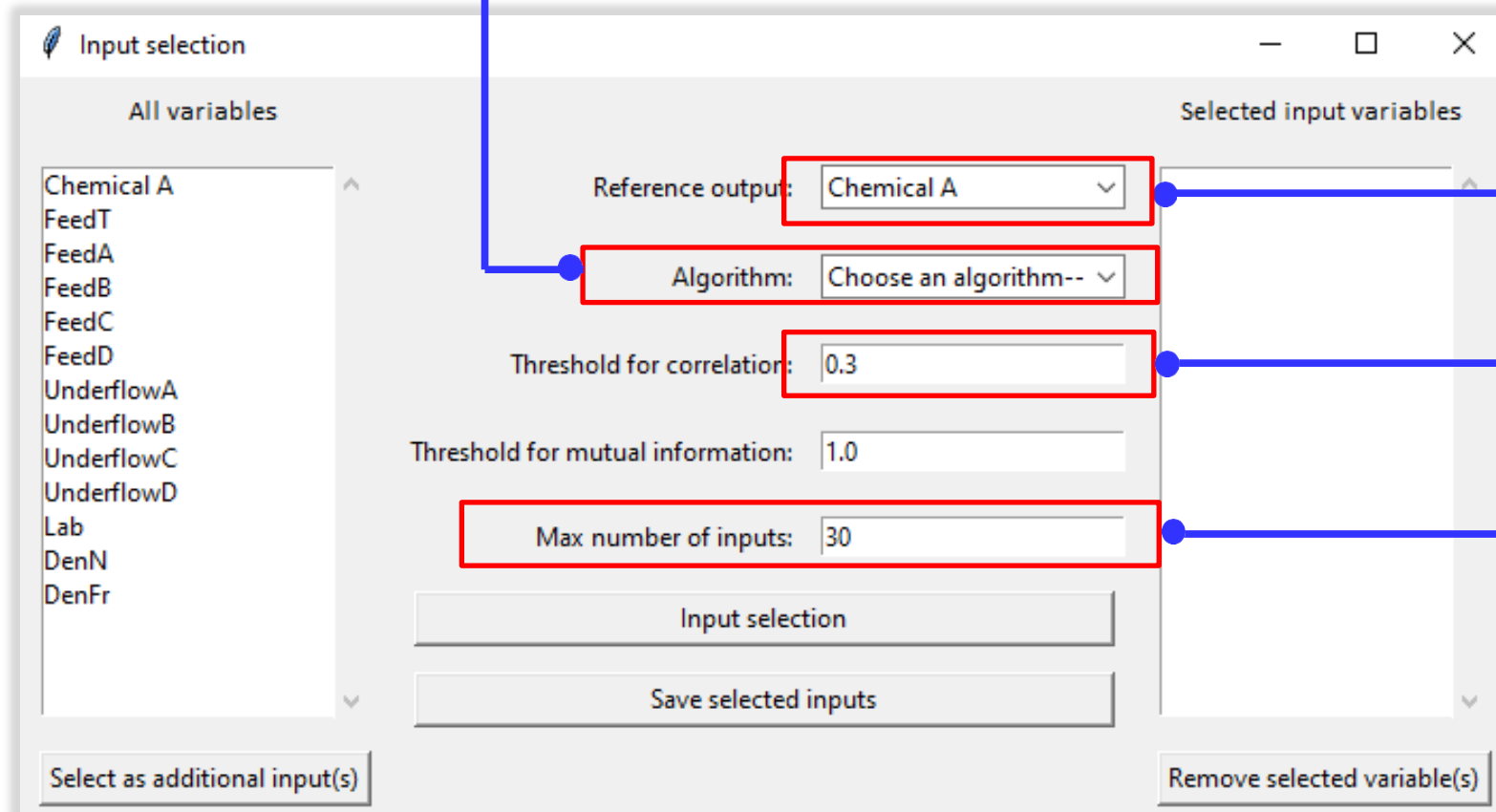
2

- Choose an algorithm for automatic input selection
- Go to “Algorithm”
 - Choose “Correlation analysis”

3

Adjust setting as needed

- “**Reference output**” is the output to be predicted
- “**Threshold for correlation**” is the lower bound for selecting inputs. This value should be non-negative, as absolute values of correlation coefficients are considered
- “**Max number of inputs**” is the upper bound on the number of inputs that can be used for modeling



Automatic selection of inputs

4

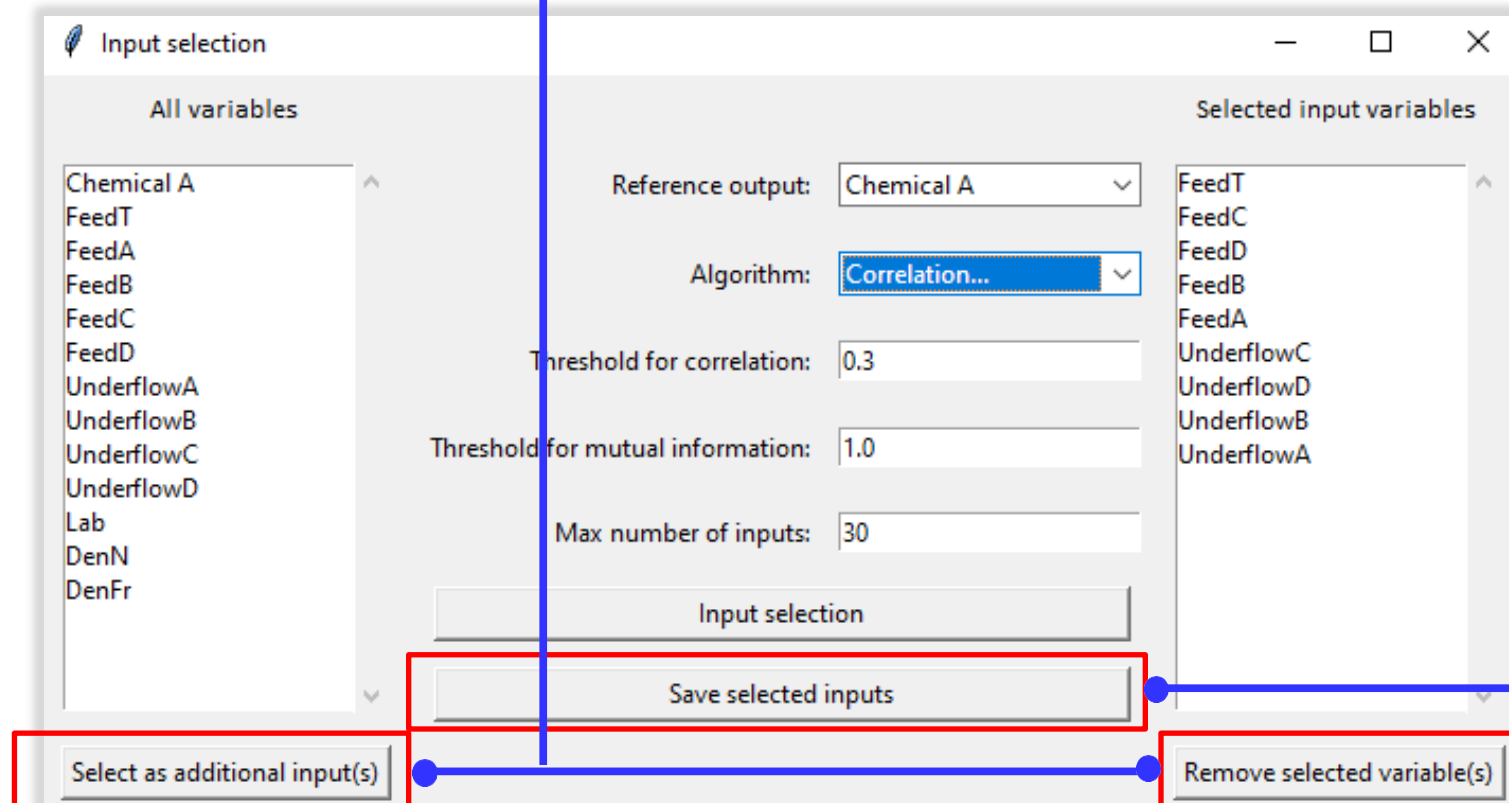
Make changes after automatic selection

- “Select as additional input(s)” can add more to the right list
- “Remove selected variable(s)” can remove variables from the right list

5

Save inputs selected automatically

- Click “Save selected inputs”
These inputs will be made available to modeling algorithms



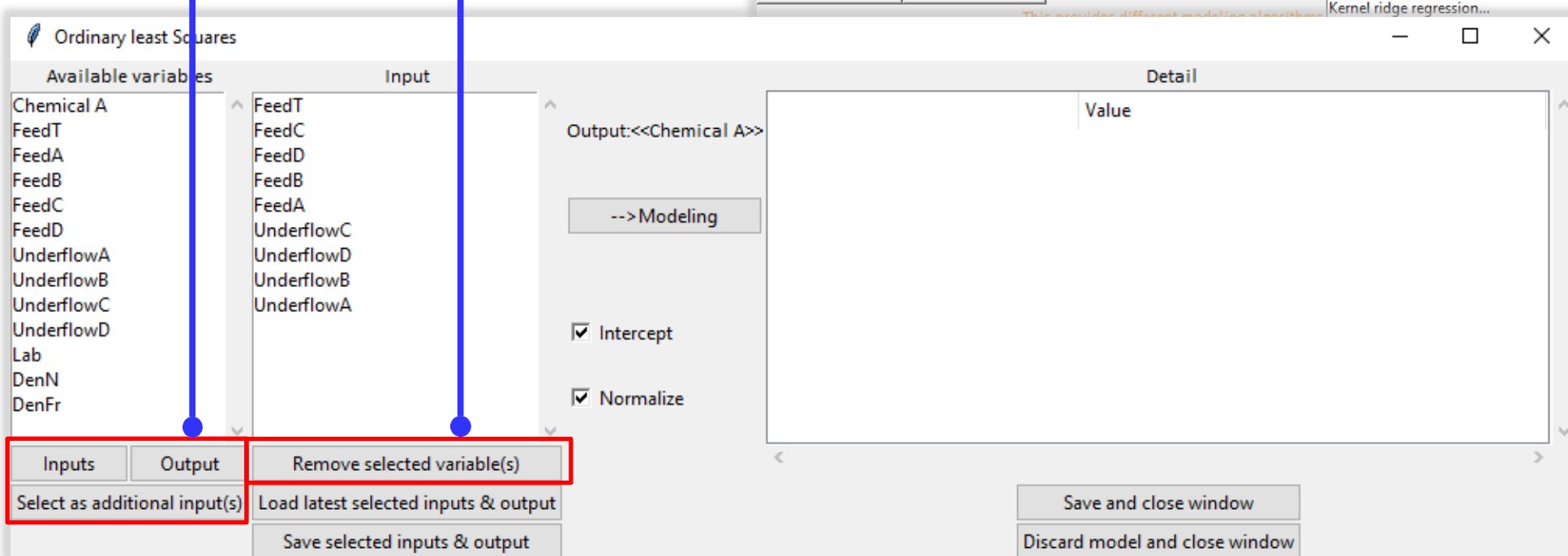
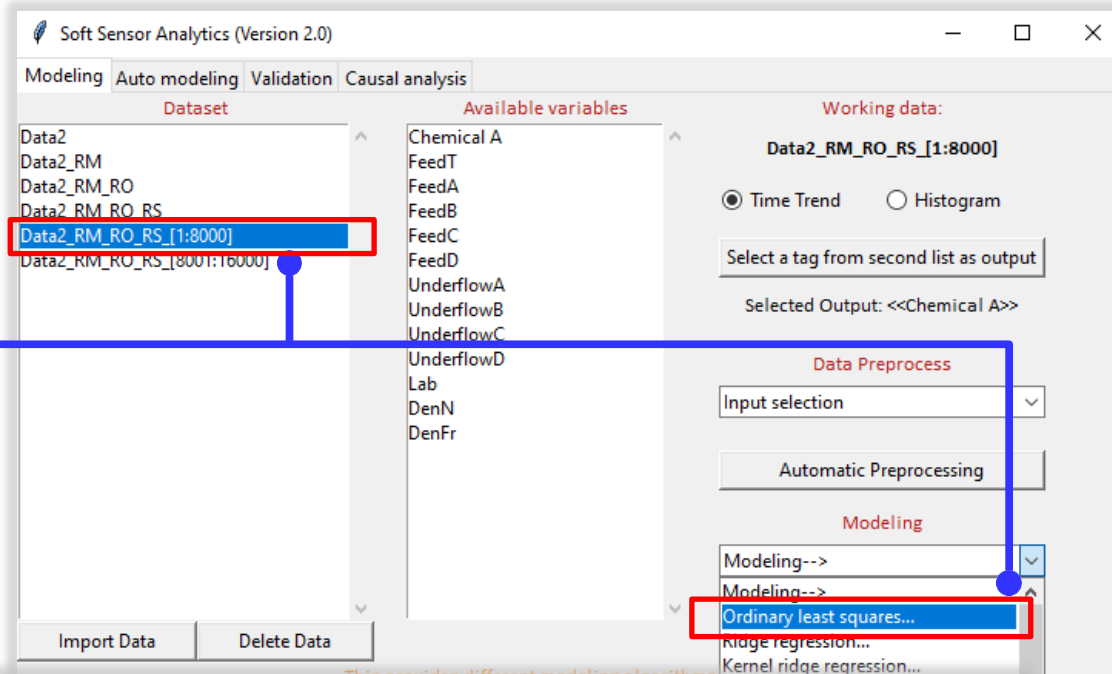
Soft Sensor Analytics



Modeling – ordinary least squares

- 1 Choose dataset and modeling method
- Click the “Modeling” tab
 - Click Data2_RM_RO_RS [1:8000] (training)
 - Go to “Modeling-->”, Click “Ordinary least squares...”, a window pops up as below

- 2 Selected inputs and output are loaded automatically
- If any change is necessary, the highlighted buttons can be used to add/remove variables



Soft Sensor Analytics



Modeling – ordinary least squares (select inputs manually)

4

Modeling
Click “-->Modeling” to perform regression

Ordinary least Squares

Available variables

Input

Output: <<Chemical A>>

-->Modeling

Intercept

Normalize

Value

Save and close window

Discard model and close window

3

Options
Tick “Intercept” or/and “Normalize”.

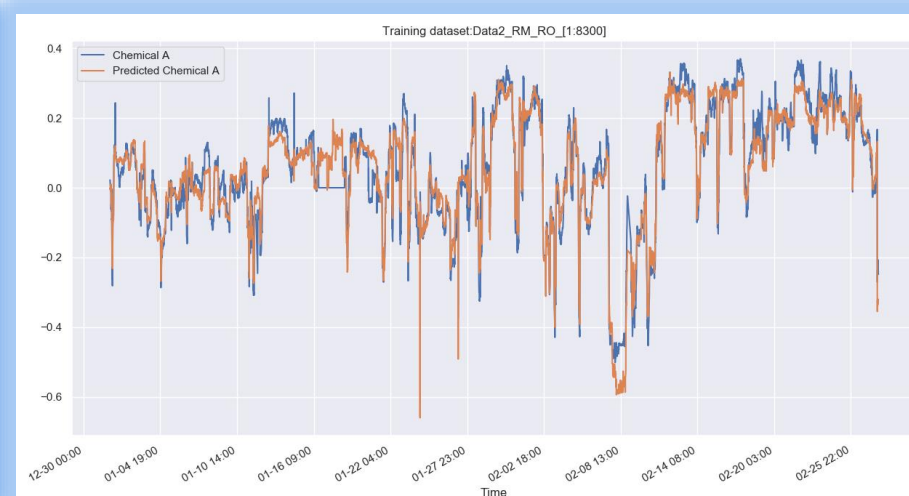
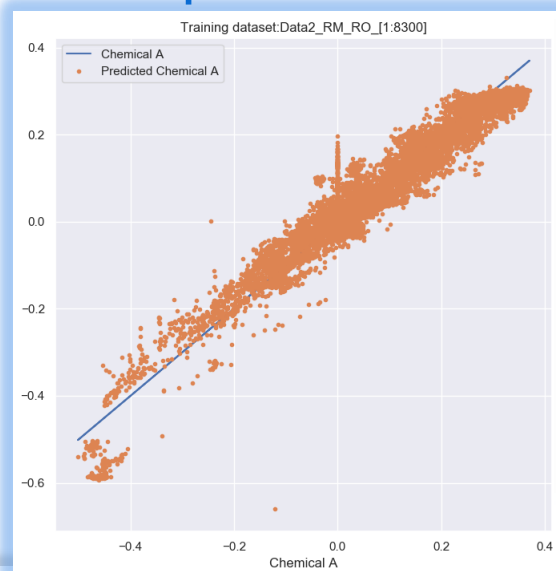
Soft Sensor Analytics



Results – ordinary least squares

Results

- Scatter plot and time trend are generated
- Model information is shown in “Detail” column



Ordinary least Squares

Available variables

- Chemical A
- FeedT
- FeedA
- FeedB
- FeedC
- FeedD
- UnderflowA
- UnderflowB
- UnderflowC
- UnderflowD
- Lab
- DenN
- DenFr

Input

- FeedT
- FeedC
- FeedD
- FeedB
- FeedA
- UnderflowC
- UnderflowD
- UnderflowB
- UnderflowA

Output: <<Chemical A>>

Method

Structure

- x(0)
- x(1)
- x(2)
- x(3)
- x(4)
- x(5)
- x(6)

Intercept

Normalize

Detail

Value

- Ordinary least Squares
- $y = ax + c$
- ...
- FeedT
- FeedC
- FeedD
- FeedB
- FeedA
- UnderflowC
- UnderflowD

5 Save model

- Close all figures
- Click “Save” to save this model for validation.

Save and close window

Discard model and close window

Soft Sensor Analytics



Modeling – Save current or Load previous set of inputs/output

7

If previous set of inputs/output is better, it can be recovered by clicking “**Load latest selected inputs & output**”

Then, modeling can be conducted again

Ordinary least Squares

Available variables

Input

Output: <<Chemical A>>

--> Modeling

Intercept

Normalize

Remove selected variable(s)

Load latest selected inputs & output

Save selected inputs & output

Save and close window

Discard model and close window

Method	Value
Structure	Ordinary least Squares
	$y = ax + c$
x(0)	FeedT
x(1)	FeedC
x(2)	FeedD
x(3)	FeedB
x(4)	FeedA
x(5)	UnderflowC
x(6)	UnderflowD

6

If any changes are made to the inputs or/and output, and if the performance is good, the set of the inputs/output variables can be saved by clicking “**Save selected inputs & output**”.

Soft Sensor Analytics



Modeling – partial least squares

- 1 Choose dataset and modeling method
- Click the “Modeling” tab
 - Click Data2_RM_RO_RS [1:8000] (training)
 - Go to “Modeling-->”, Click “Partial least squares...”, a window pops up as below

- 2 Selected inputs and output are loaded automatically
- If any change is necessary, the highlighted buttons can be used to add/remove variables

This provides different modeling algorithms

Soft Sensor Analytics



Modeling – Partial least squares (PLS)

4

Analysis

Click “-->Analysis” to find a good number of latent variables

A figure will pop up as show in next slide

Partial least squares regression

Available variables

Input

Output: <<Chemical A>>

Detail

Value

-->Analysis

-->Modeling

Normalize

Number of latent variables:

5

Inputs Output Remove selected variable(s)

Select as additional input(s) Load latest selected inputs & output

Save selected inputs & output

Save and close window

Discard model and close window

3

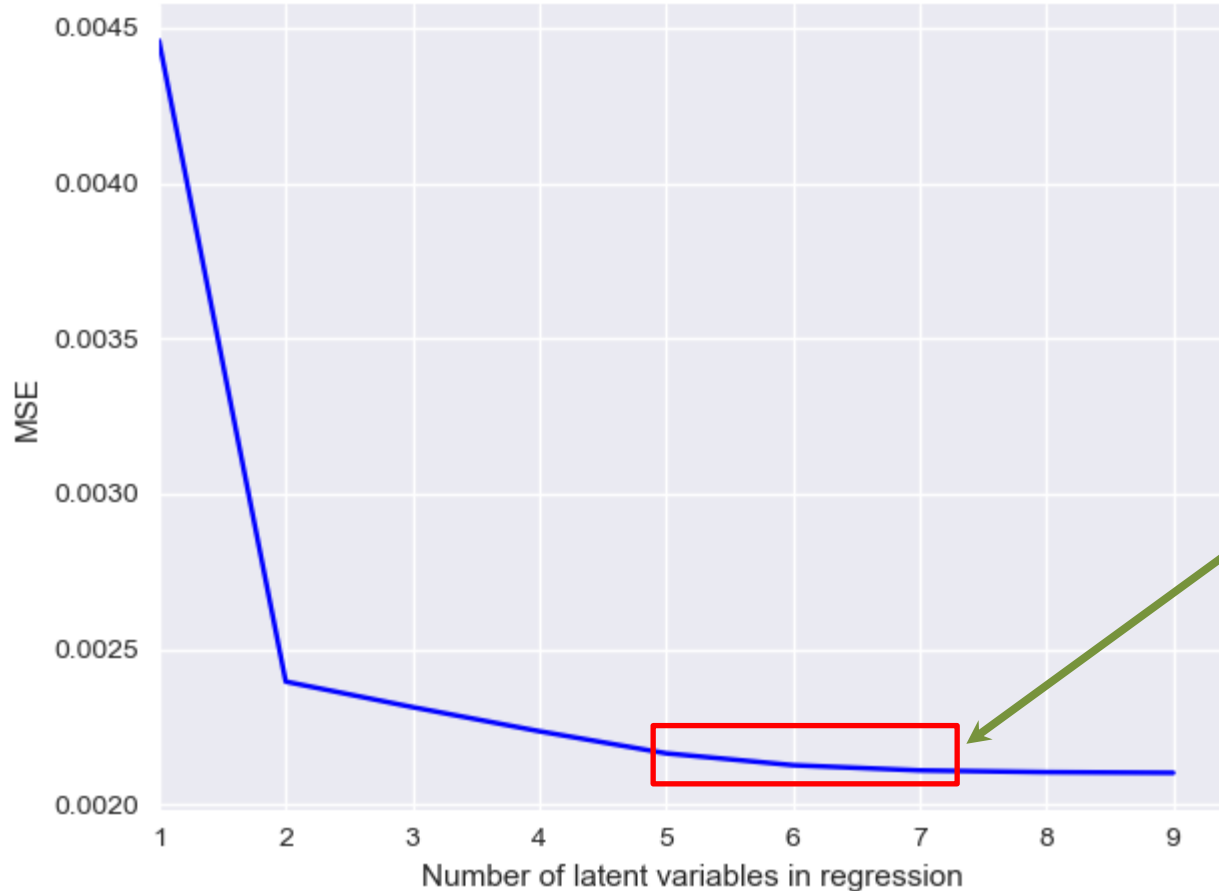
Options

Tick “Normalize” if this is needed.

Soft Sensor Analytics



Modeling – PLS (determine # of latent variables)



Choose number of latent variables

Number of latent variables is a key hyper-parameter of PLS. One optional analysis is provided to help you.

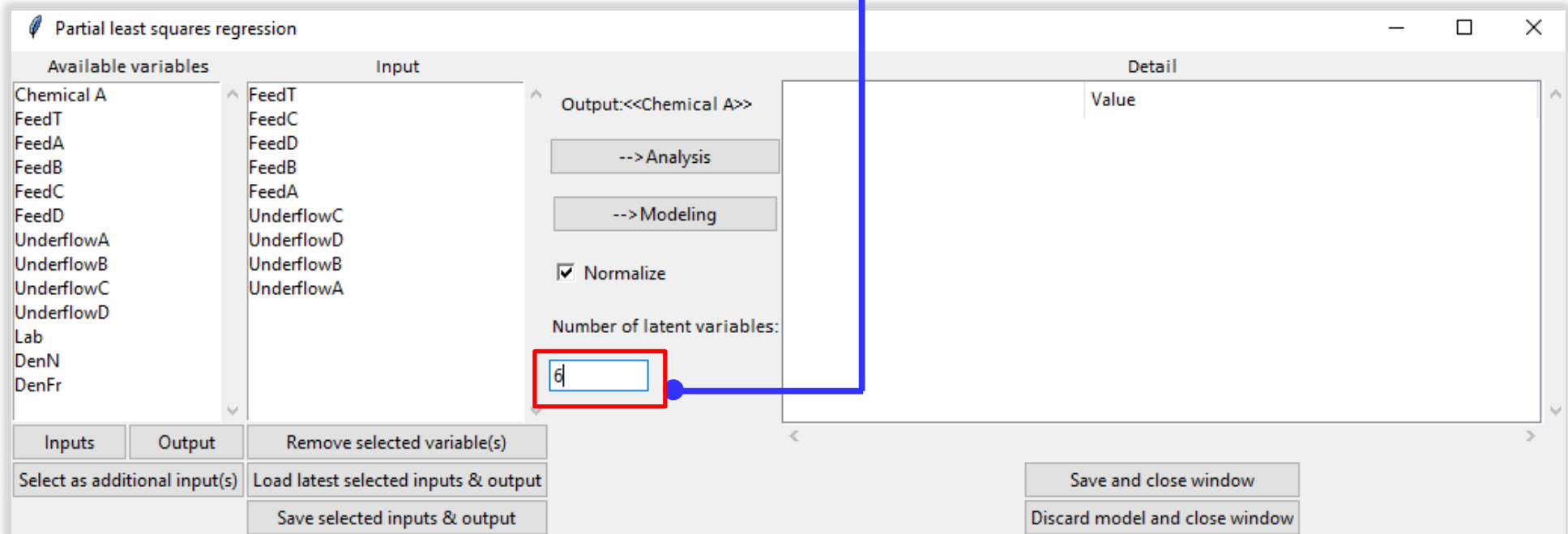
By clicking 'Analysis', the left figure pops up. Based on MSE values, 5, 6, or 7 is a good choice.

Modeling – PLS (determine # of latent variables)

5

Determine # of latent variables

- Close all figures
- Enter “6” in the “number of latent variables” box



Partial least squares regression

Available variables

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Input

FeedT
FeedC
FeedD
FeedB
FeedA
UnderflowC
UnderflowD
UnderflowB
UnderflowA

Output:<<Chemical A>>

--> Analysis

--> Modeling

Normalize

Number of latent variables:

6

Inputs Output Remove selected variable(s)

Select as additional input(s) Load latest selected inputs & output

Save selected inputs & output

Save and close window

Discard model and close window

Soft Sensor Analytics



7

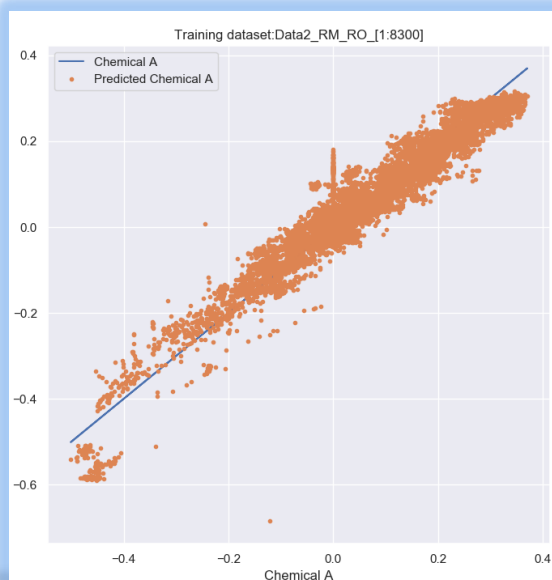
- Close figures
- Click “Save” to save this model

6

Regression

Click “-->Modeling”

- scatter plot and time trend are generated
- model information is shown in “Detail” column



Soft Sensor Analytics



Modeling – Gaussian regression

- 1 Choose dataset and modeling method
- Click the “Model” tab
 - Click **Data2_RM_RO_RS [1:8300]** (training data)
 - Go to “Modeling-->”, Click “GPR regression...”, a window pops up as below

- 2 Selected inputs and output are loaded automatically
- If any change is necessary, the highlighted buttons can be used to add/remove variables

This provides different modeling algorithms

Soft Sensor Analytics



Modeling – Gaussian process regression

The screenshot shows a software window titled "Gaussian process regression". It is divided into several sections: "Available variables" on the left, "Input" in the middle-left, "Output: <<Chemical A>>" in the middle-right, and "Detail" on the far right. The "Available variables" list includes Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, and DenFr. The "Input" section lists FeedT, FeedC, FeedD, FeedB, FeedA, UnderflowC, UnderflowD, UnderflowB, and UnderflowA. The "Output" section is set to "Chemical A" and has a "--> Modeling" button. The "Kernel" is set to "1.0*RBF(1.0)". The "Alpha" parameter is highlighted with a red box and has a value of "0.0001". A blue line connects this box to a callout box in the bottom right. The "Detail" section has a "Value" field. At the bottom, there are buttons for "Inputs", "Output", "Remove selected variable(s)", "Select as additional input(s)", "Load latest selected inputs & output", "Save selected inputs & output", "Save and close window", and "Discard model and close window".

3 Options
Enter different values if necessary, default is "0.0001"

Soft Sensor Analytics



Modeling – Gaussian process regression

The screenshot shows the "Gaussian process regression" software interface. The window is divided into several sections:

- Available variables:** A list of variables including Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, and DenFr.
- Input:** A list of selected input variables: FeedT, FeedC, FeedD, FeedB, FeedA, UnderflowC, UnderflowD, UnderflowB, and UnderflowA.
- Output:** The output is set to "Chemical A".
- Kernel:** A dropdown menu is highlighted with a red box, showing the selected kernel function: "1.0*RBF(1.0)".
- Alpha:** A text input field containing the value "0.0001".
- Detail:** A large empty area for displaying model details.
- Buttons:** A "Modeling" button is located below the output field. At the bottom, there are buttons for "Save and close window" and "Discard model and close window".

4

Kernel function selection
"1.0*RBF(1.0)" is selected as
the kernel function

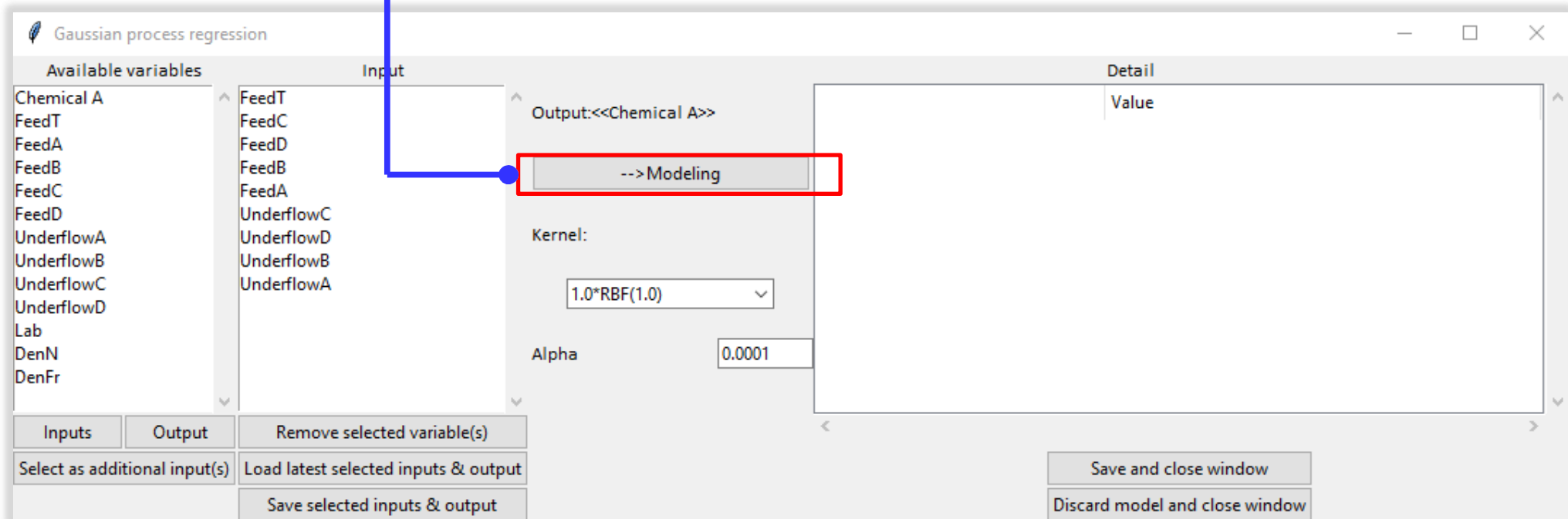
Modeling – Gaussian process regression

5

Regression

Click "-->Modeling"

- scatter plot and time trend are generated as shown in the next slides



Gaussian process regression

Available variables

Chemical A
FeedT
FeedC
FeedA
FeedB
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Input

FeedT
FeedC
FeedD
FeedB
FeedA
UnderflowC
UnderflowD
UnderflowB
UnderflowA

Output: <<Chemical A>>

-->Modeling

Kernel:
1.0*RBF(1.0)

Alpha
0.0001

Detail

Value

Inputs Output Remove selected variable(s)

Select as additional input(s) Load latest selected inputs & output

Save selected inputs & output

Save and close window

Discard model and close window

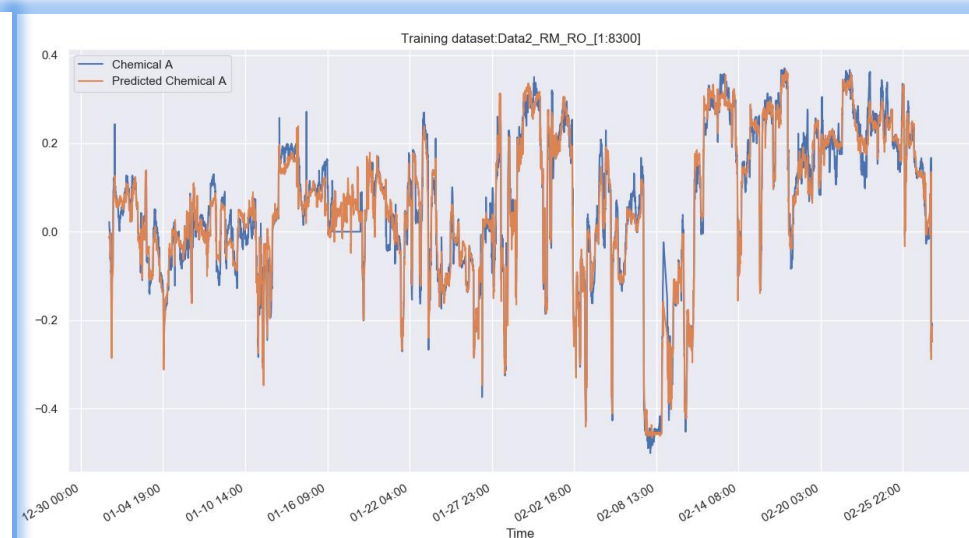
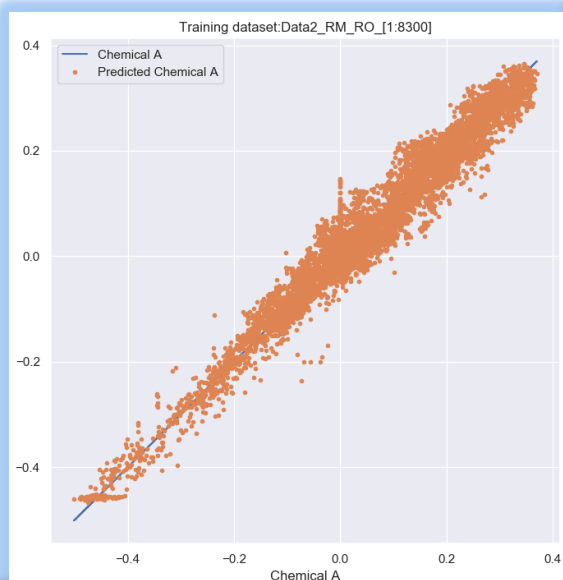
Soft Sensor Analytics



Results – Gaussian process regression

Results

- Scatter plot and time trend are generated
- Model information is shown in “Details” column



Available variables

Input

Output: <<Chemical A>>

Kernel: 1.0*RBF(1.0)

Alpha: 0.0001

Method: Stochastic gradient descent regression

Performance

Detail

Value: Chemical A, 0.0001

6 Save model

- Close all figures
- Click “Save” to save this model for validation

Save and close window

Discard model and close window

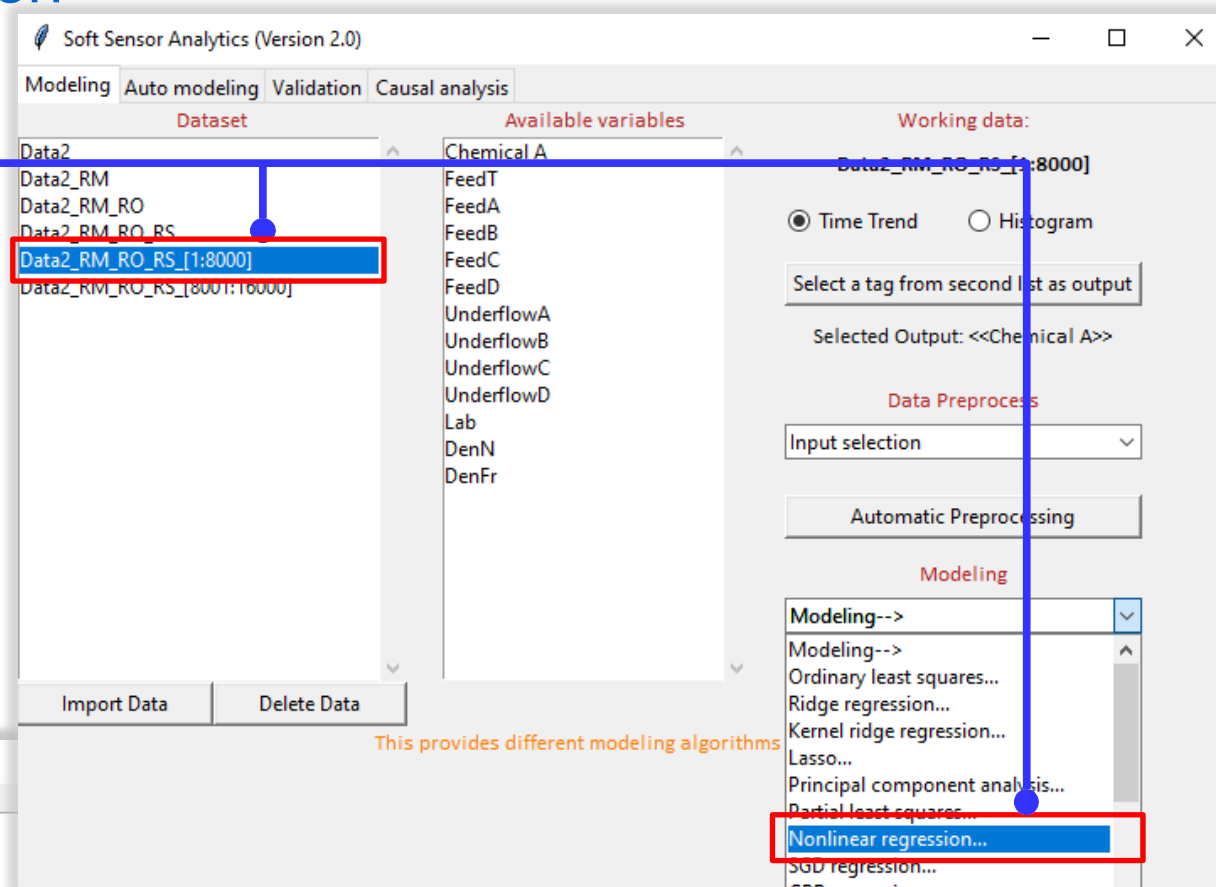
Soft Sensor Analytics



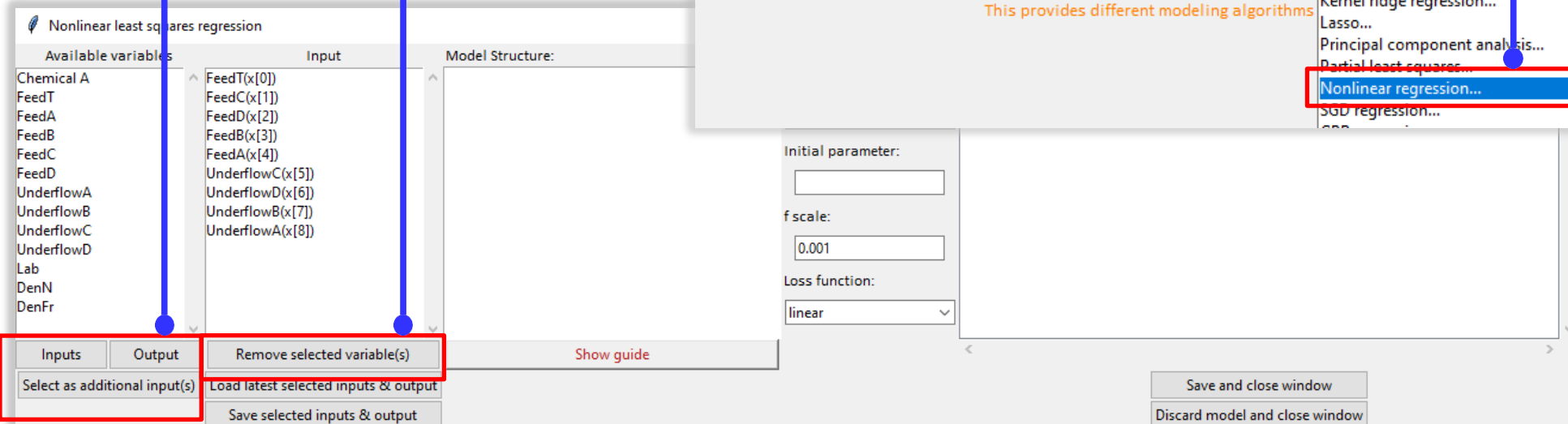
Modeling – nonlinear regression

- 1 Choose dataset and modeling method
- Click the “Model” tab
 - Click **Data2_RM_RO_RS [1:8300]** (training data)
 - Go to “Modeling-->”, Click “Nonlinear regression...”, a window pops up as below

2 Selected inputs and output are loaded automatically
If any change is necessary, the highlighted buttons can be used to add/remove variables



This provides different modeling algorithms



Soft Sensor Analytics



Modeling – nonlinear regression

3

Press the “**Show guide**” button to see how to enter a model structure and other information about nonlinear regression

The screenshot shows the "Nonlinear least squares regression" software interface. The window is divided into several sections:

- Available variables:** A list of variables including Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, and DenFr.
- Input:** A list of input variables including FeedT(x[0]), FeedC(x[1]), FeedD(x[2]), FeedB(x[3]), FeedA(x[4]), UnderflowC(x[5]), UnderflowD(x[6]), UnderflowB(x[7]), and UnderflowA(x[8]).
- Model Structure:** A large empty text area for defining the model structure.
- Output: <<Chemical A>>(y):** A section for defining the output variable, with a "Value" field.
- Initial parameter:** A text input field.
- f scale:** A text input field containing the value "0.001".
- Loss function:** A dropdown menu currently set to "linear".
- Buttons:** A "--> Modeling" button, a "Show guide" button (highlighted with a red box), and buttons for "Save and close window" and "Discard model and close window".
- Bottom Panel:** Buttons for "Inputs", "Output", "Remove selected variable(s)", "Select as additional input(s)", "Load latest selected inputs & output", and "Save selected inputs & output".

A blue line with a dot at the end points from the "Show guide" button to the instruction box above.

Soft Sensor Analytics



Modeling – nonlinear regression

4

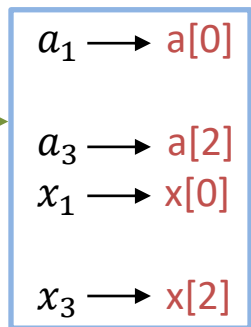
Set nonlinear model structure

- **Enter** model structure following the way in the red box
- Commonly used are given in the right table

Note: this model structure is used as a simple example

$$y = \frac{a_1 + a_2 x_1^2}{a_3 + a_4 x_2^2 + \sqrt{|x_3|}}$$

$$y = (a[0] + a[1]*x[0]**2) / (a[2] + a[3]*x[1]**2 + \text{sqrt}(\text{abs}(x[2])))$$



button	Math	Operation name
a+b	$a + b$	addition
a-b	$a - b$	subtraction
a*b	$a \times b$	multiplication
a/b	$a \div b$	division
a//b	$[a \div b]$	floor division (e.g. 5//2 = 2)
a%b	$a \text{ mod } b$	modulo
-a	$-a$	negation
abs(a)	$ a $	absolute value
a**b	a^b	exponent
sqrt(a)	\sqrt{a}	square root
log(a)	$\log(a)$	natural logarithm
exp(a)	e^a	natural exponent
sin(a)	$\sin a$	trigonometric sine
cos(a)	$\cos a$	trigonometric cosine
tan(a)	$\tan a$	trigonometric tangent
arcsin(a)	$\sin^{-1} a$	trigonometric inverse sine
arccos(a)	$\cos^{-1} a$	trigonometric inverse cosine
arctan(a)	$\tan^{-1} a$	trigonometric inverse tangent

Nonlinear least squares regression

Available variables: Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, DenFr

Input: UnderflowD(x[0]), UnderflowC(x[1]), FeedC(x[2])

Model Structure: $y = (a[0] + a[1]*x[0]**2) / (a[2] + a[3]*x[1]**2 + \text{sqrt}(\text{abs}(x[2])))$

Output: <<Lab>>(y)

Initial parameter: 0,0,0,0

f scale: 0.001

Loss function: linear

Method: Nonlinear least squares

Structure: $y = (a[0] + a[1]*x[0]**2) / (a[2] + a[3]*x[1]**2 + \text{sqrt}(\text{abs}(x[2])))$

Performance: linear

Buttons: Inputs, Output, Remove selected variable(s), Add additional input(s), Load latest selected inputs & outputs, Save selected inputs & outputs, Show guide, Save and close window, Discard model and close window

Four initial values are chosen as the model involves four parameters a[0] - a[3]



Modeling – nonlinear regression

4

Set nonlinear model structure

- **Enter** model structure following the way in the red box
- Commonly used operations are given in the right table

$$y = a[0]*x[0]**2 + a[1]*\text{sqrt}(\text{abs}(x[4])) + a[2]*x[3] + a[3]*\sin(x[1]) + a[4]*\log(\text{abs}(x[2])) + a[5]*\exp(x[8]) + a[6]/x[7] + a[7] + a[8]*x[5] + a[9]/x[6]$$

button	Math	Operation name
a+b	$a + b$	addition
a-b	$a - b$	subtraction
a*b	$a \times b$	multiplication
a/b	$a \div b$	division
a//b	$[a \div b]$	floor division (e.g. $5//2 = 2$)
a%b	$a \bmod b$	modulo
-a	$-a$	negation
abs(a)	$ a $	absolute value
a**b	a^b	exponent
sqrt(a)	\sqrt{a}	square root
log(a)	$\log(a)$	natural logarithm
exp(a)	e^a	natural exponent
sin(a)	$\sin a$	trigonometric sine
cos(a)	$\cos a$	trigonometric cosine
tan(a)	$\tan a$	trigonometric tangent
arcsin(a)	$\sin^{-1} a$	trigonometric inverse sine
arccos(a)	$\cos^{-1} a$	trigonometric inverse cosine
arctan(a)	$\tan^{-1} a$	trigonometric inverse tangent

Nonlinear least squares regression

Available variables: Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, DenFr

Input: FeedT(x[0]), FeedC(x[1]), FeedA(x[2]), FeedB(x[3]), FeedA(x[4]), UnderflowC(x[5]), UnderflowD(x[6]), UnderflowB(x[7]), UnderflowA(x[8])

Model Structure: $y = a[0]*x[0]**2 + a[1]*\text{sqrt}(\text{abs}(x[4])) + a[2]*x[3] + a[3]*\sin(x[1]) + a[4]*\log(\text{abs}(x[2])) + a[5]*\exp(x[8]) + a[6]/x[7] + a[7] + a[8]*x[5] + a[9]/x[6]$

Output: <<Chemical A>>(y)

--> Modeling

Initial parameter: 0,0,0,0,0,0,0,0,0,0

f scale: 0.001

Loss function: linear

Value

Save and close window

Discard model and close window

Ten initial values are chosen as the model involves ten parameters a[0] - a[9]

Modeling – nonlinear regression

5

Set initial value

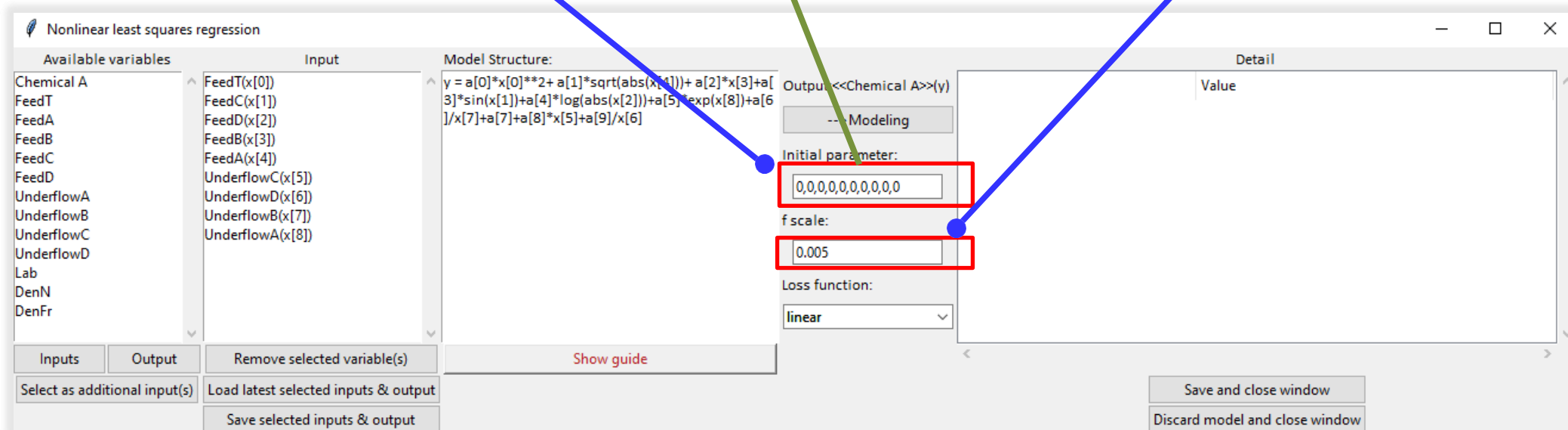
- Enter “0,0,0,0,0,0,0,0,0,0” in the “Initial parameter” box

6

Set f scale value

- Enter “0.001” in the “f scale” box

10 zeros are given as we need to estimate ten parameters



Nonlinear least squares regression

Available variables: Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, DenFr

Input: FeedT(x[0]), FeedC(x[1]), FeedD(x[2]), FeedB(x[3]), FeedA(x[4]), UnderflowC(x[5]), UnderflowD(x[6]), UnderflowB(x[7]), UnderflowA(x[8])

Model Structure: $y = a[0]*x[0]**2 + a[1]*\text{sqrt}(\text{abs}(x[1])) + a[2]*x[3] + a[3]*\sin(x[1]) + a[4]*\log(\text{abs}(x[2])) + a[5]*\exp(x[8]) + a[6]/x[7] + a[7] + a[8]*x[5] + a[9]/x[6]$

Output: <<Chemical A>>(y)

Initial parameter: 0,0,0,0,0,0,0,0,0,0

f scale: 0.005

Loss function: linear

Buttons: Modeling, Show guide, Save and close window, Discard model and close window



Modeling – nonlinear regression

7

Choose loss function

- Select “huber”

Nonlinear least squares regression

Available variables

Available variables	Input
Chemical A	FeedT(x[0])
FeedT	FeedC(x[1])
FeedA	FeedD(x[2])
FeedB	FeedB(x[3])
FeedC	FeedA(x[4])
FeedD	UnderflowC(x[5])
UnderflowA	UnderflowD(x[6])
UnderflowB	UnderflowB(x[7])
UnderflowC	UnderflowA(x[8])
UnderflowD	
Lab	
DenN	
DenFr	

Model Structure:

$$y = a[0]*x[0]**2 + a[1]*\sqrt{\text{abs}(x[4])} + a[2]*x[3] + a[3]*\sin(x[1]) + a[4]*\log(\text{abs}(x[2])) + a[5]*\exp(x[8]) + a[6]/x[7] + a[7] + a[8]*x[5] + a[9]/x[6]$$

Output: <<Chemical A>>(y)

--> Modeling

Initial parameter:

f scale:

Loss function:

huber

linear

soft

cauchy

arctan

Save and close window

Discard model and close window



Modeling – nonlinear regression

8

Regression

Click "-->Modeling"

- scatter plot and time trend are generated as shown in the next slides

Nonlinear least squares regression

Available variables: Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, DenFr

Input: FeedT(x[0]), FeedC(x[1]), FeedD(x[2]), FeedB(x[3]), FeedA(x[4]), UnderflowC(x[5]), UnderflowD(x[6]), UnderflowB(x[7]), UnderflowA(x[8])

Model Structure: $y = a[0]*x[0]**2 + a[1]*\text{sqrt}(\text{abs}(x[4])) + a[2]*x[3] + a[3]*\sin(x[1]) + a[4]*\log(\text{abs}(x[2])) + a[5]*\exp(x[8]) + a[6]*x[7] + a[7] + a[8]*x[5] + a[9]/x[6]$

Out: <<Chemical A>>(y)

--> Modeling

Initial parameter: 0,0,0,0,0,0,0,0,0

f scale: 0.005

Loss function: huber

Buttons: Inputs, Output, Remove selected variable(s), Show guide, Select as additional input(s), Load latest selected inputs & output, Save selected inputs & output, Save and close window, Discard model and close window

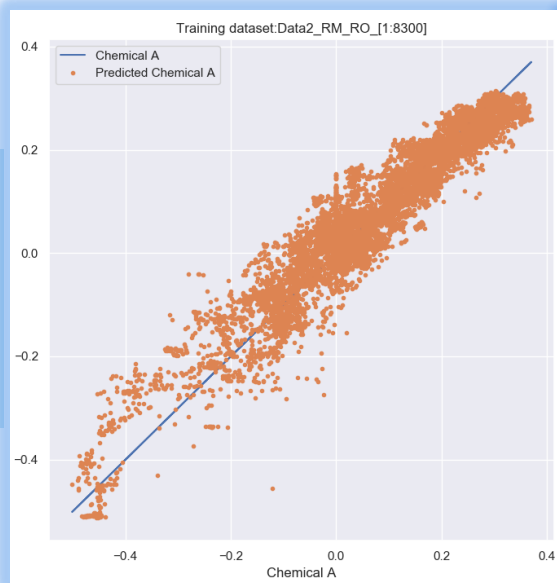
Soft Sensor Analytics



Results – nonlinear regression

Results

- Scatter plot and time trend are generated
- Model information is shown in “Detail” column



Nonlinear least squares regression

Available variables	Input
Chemical A	FeedT(x[0])
FeedT	FeedC(x[1])
FeedA	FeedD(x[2])
FeedB	FeedB(x[3])
FeedC	FeedA(x[4])
FeedD	UnderflowC(x[5])
UnderflowA	UnderflowD(x[6])
UnderflowB	UnderflowB(x[7])
UnderflowC	UnderflowA(x[8])
UnderflowD	
Lab	
DenN	
DenFr	

Model Structure:

$$y = a[0]*x[0]**2 + a[1]*\text{sqrt}(\text{abs}(x[4])) + a[2]*x[3] + a[3]*\sin(x[1]) + a[4]*\log(\text{abs}(x[2])) + a[5]*\exp(x[8]) + a[6]/x[7] + a[7] + a[8]*x[5] + a[9]/x[6]$$

Output: <<Chemical A>>(y)

--> Modeling

Initial parameter: 0,0,0,0,0,0,0,0,0

f scale: 0.005

Loss function: huber

Method Structure

- Method: Nonlinear least squares
- Structure: x, a
- loss function: linear
- Performance: ...

Detail

Value
Nonlinear least squares
y = a[0]*x[0]**2 + a[1]*sqrt(abs(x[4])) + a[2]*x[3] + a[3]*sin(x[1]) + a[4]*log(abs(x[2])) + a[5]*exp(x[8]) + a[6]/x[7] + a[7] + a[8]*x[5] + a[9]/x[6]
Chemical A
linear

Save model

Save and close window

Discard model and close window

8

- Close all figures
- Click “Save” to save model

Soft Sensor Analytics



Validation of the models

All the saved models based on four methods are shown under the "Validation"

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling **Validation** Causal analysis

Dataset Model Working data: Empty Detail

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS
Data2_RM_RO_RS_[1:8000]
Data2_RM_RO_RS_[8001:16000]

Model 1 (OLS)
Model 2 (PLS)
Model 3 (GPR)
Model 4 (Nonlinear)

Value

Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Ordinary least Squares
Structure	$y = ax + c$
x	...
y	Chemical A
c	-0.040440882213770846
a	...
Performance	...

Delete a model

Left Click To Select Dataset/Right Click To Manipulate

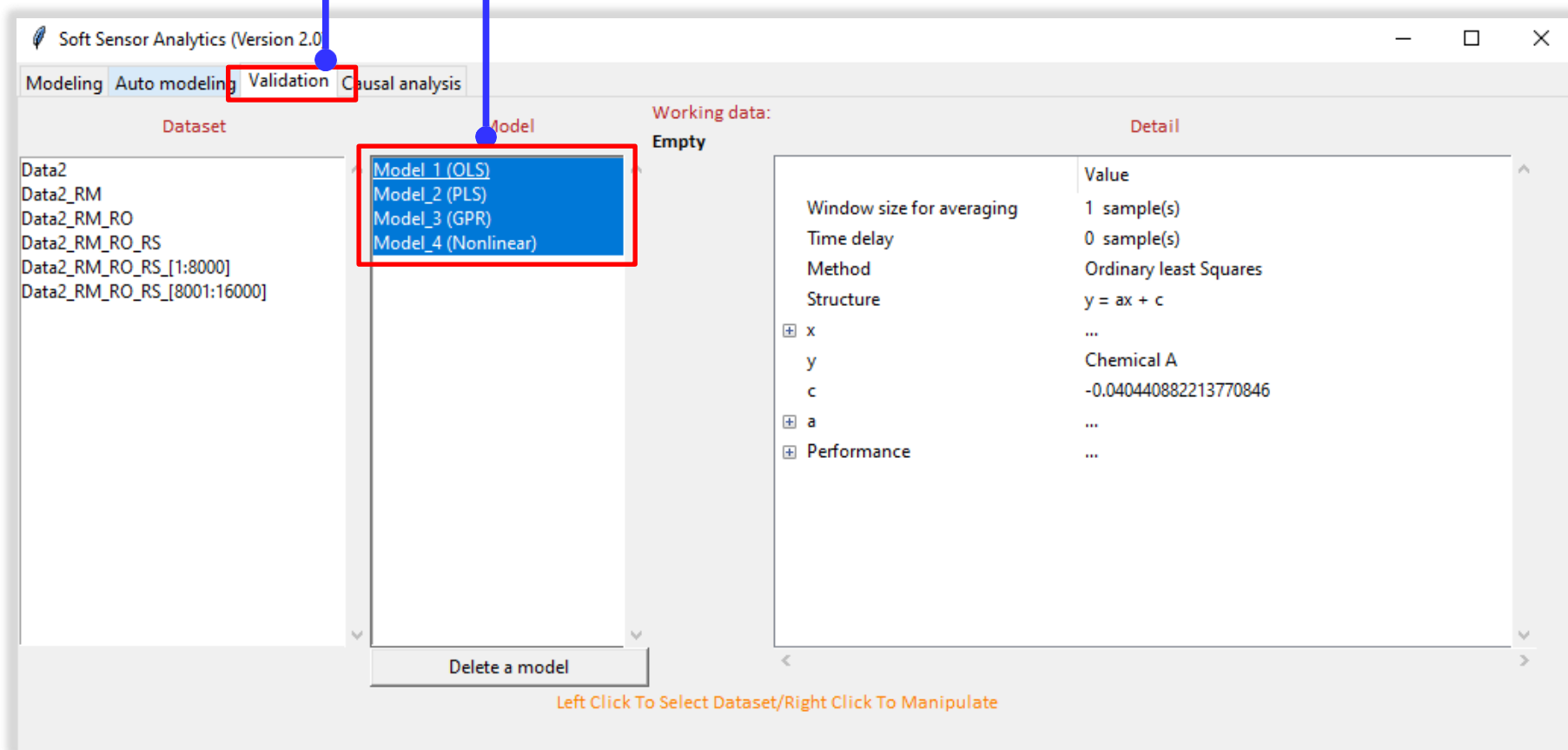
Validation of the models

Model Validation: the models based on different methods can be validated based on the second dataset.

1

Select models for validation

- Click “Validation” tab on the top
- Click sequentially to select the four models for validation and comparison



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

- Data2
- Data2_RM
- Data2_RM_RO
- Data2_RM_RO_RS
- Data2_RM_RO_RS_[1:8000]
- Data2_RM_RO_RS_[8001:16000]

Model

- Model_1 (OLS)
- Model_2 (PLS)
- Model_3 (GPR)
- Model_4 (Nonlinear)

Working data: Empty

Detail

	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Ordinary least Squares
Structure	$y = ax + c$
x	...
y	Chemical A
c	-0.040440882213770846
a	...
Performance	...

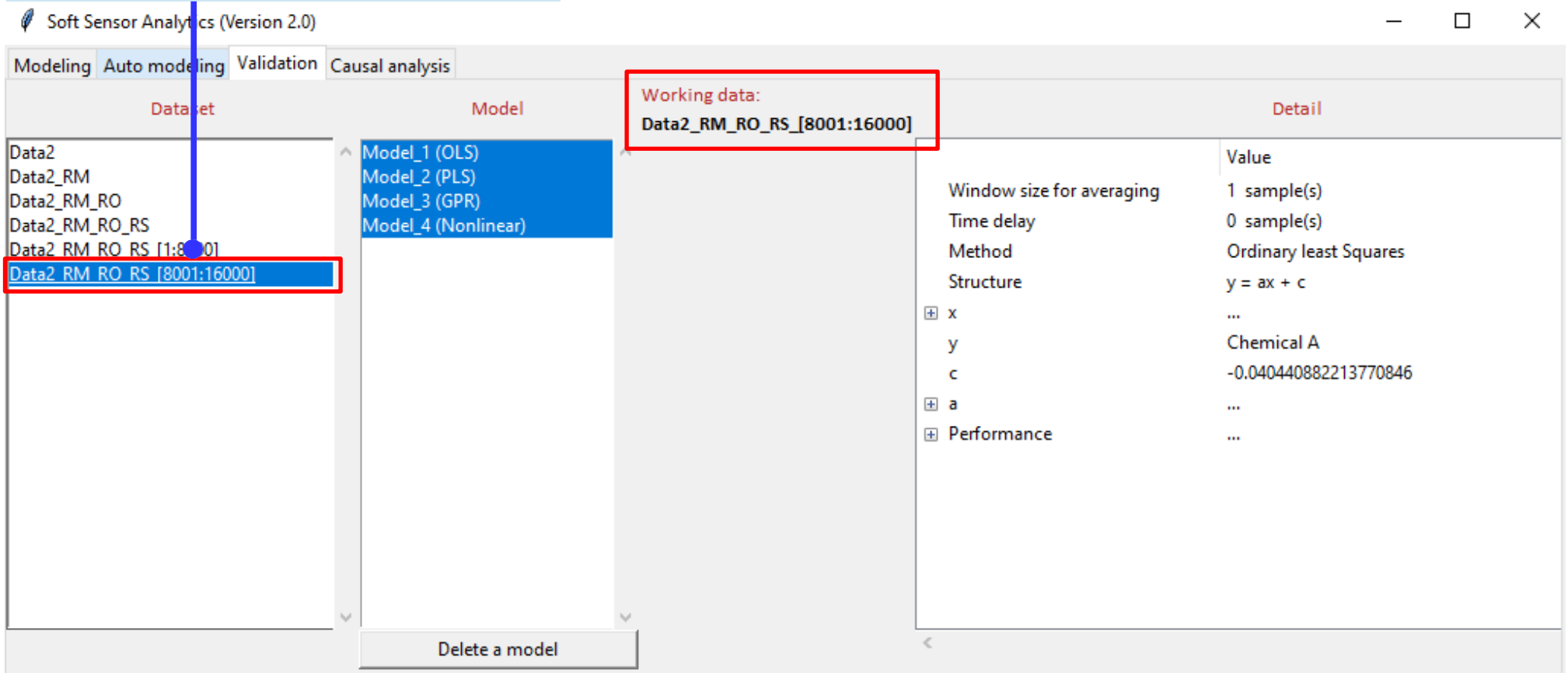
Delete a model

Left Click To Select Dataset/Right Click To Manipulate

Validation of the models

Model Validation: the models based on different methods can be validated based on the second dataset.

- 2
- Select data set for validation
 - Click “Data2_RM_RO_RS_[8001:16000]” to use it for validation



Parameter	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Ordinary least Squares
Structure	$y = ax + c$
x	...
y	Chemical A
c	-0.040440882213770846
a	...
Performance	...

Click To Select Model/Right Click To Manipulate

Soft Sensor Analytics

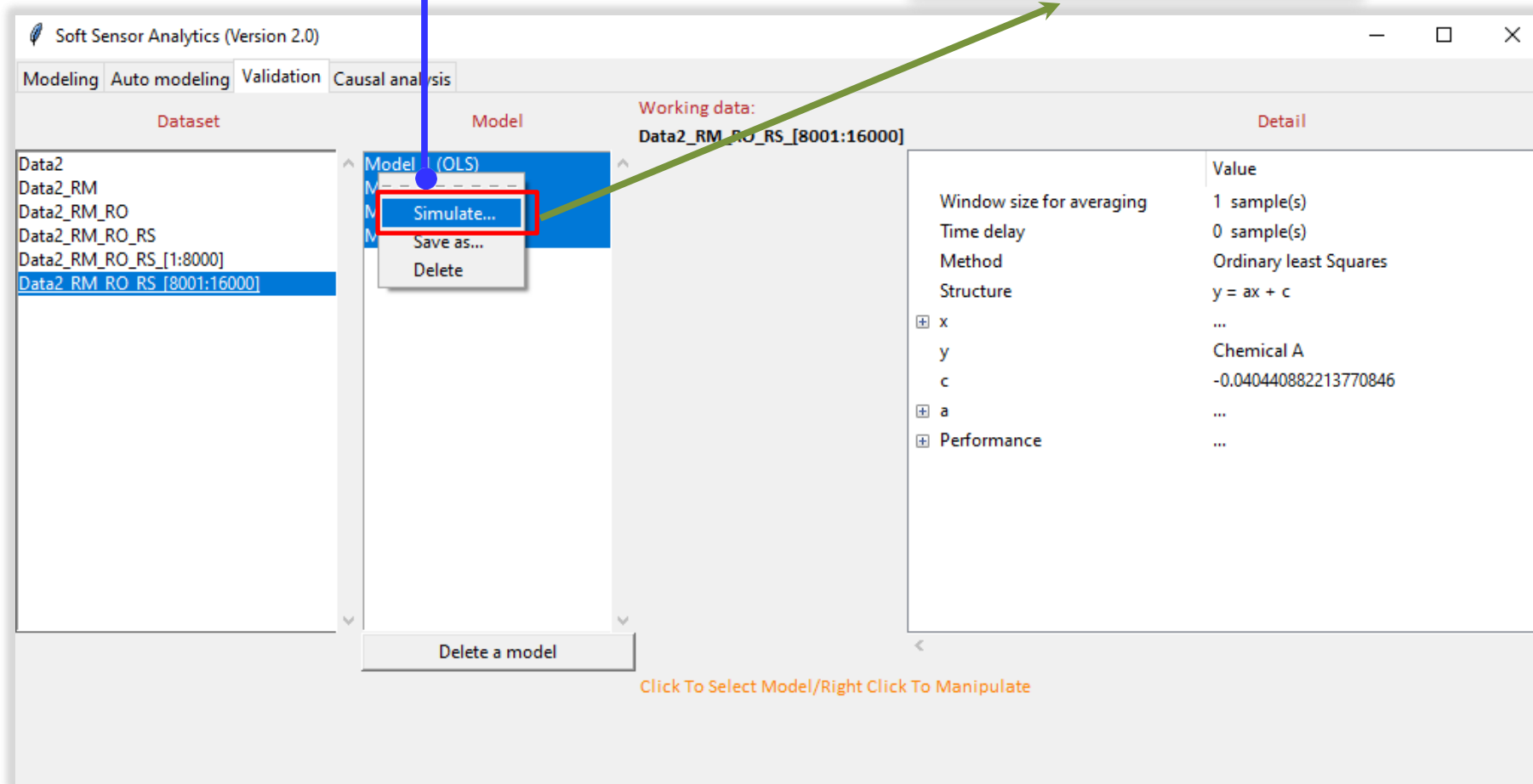
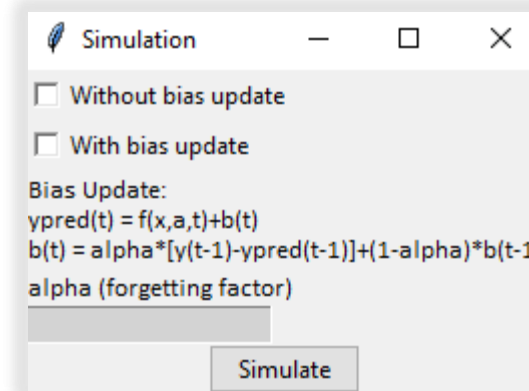


Validation of the models

3

Choose “Simulate” to continue

- **Right-Click** on the four models
- **Click “Simulate...”** button, a window will pop up



Soft Sensor Analytics

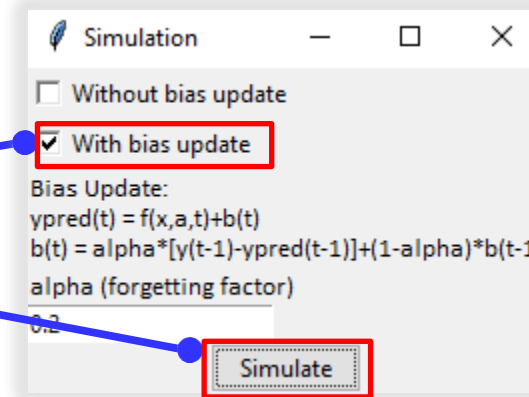


Validation of the models

4

Model validation

- Tick “With bias update” box
- Click “Simulate” button



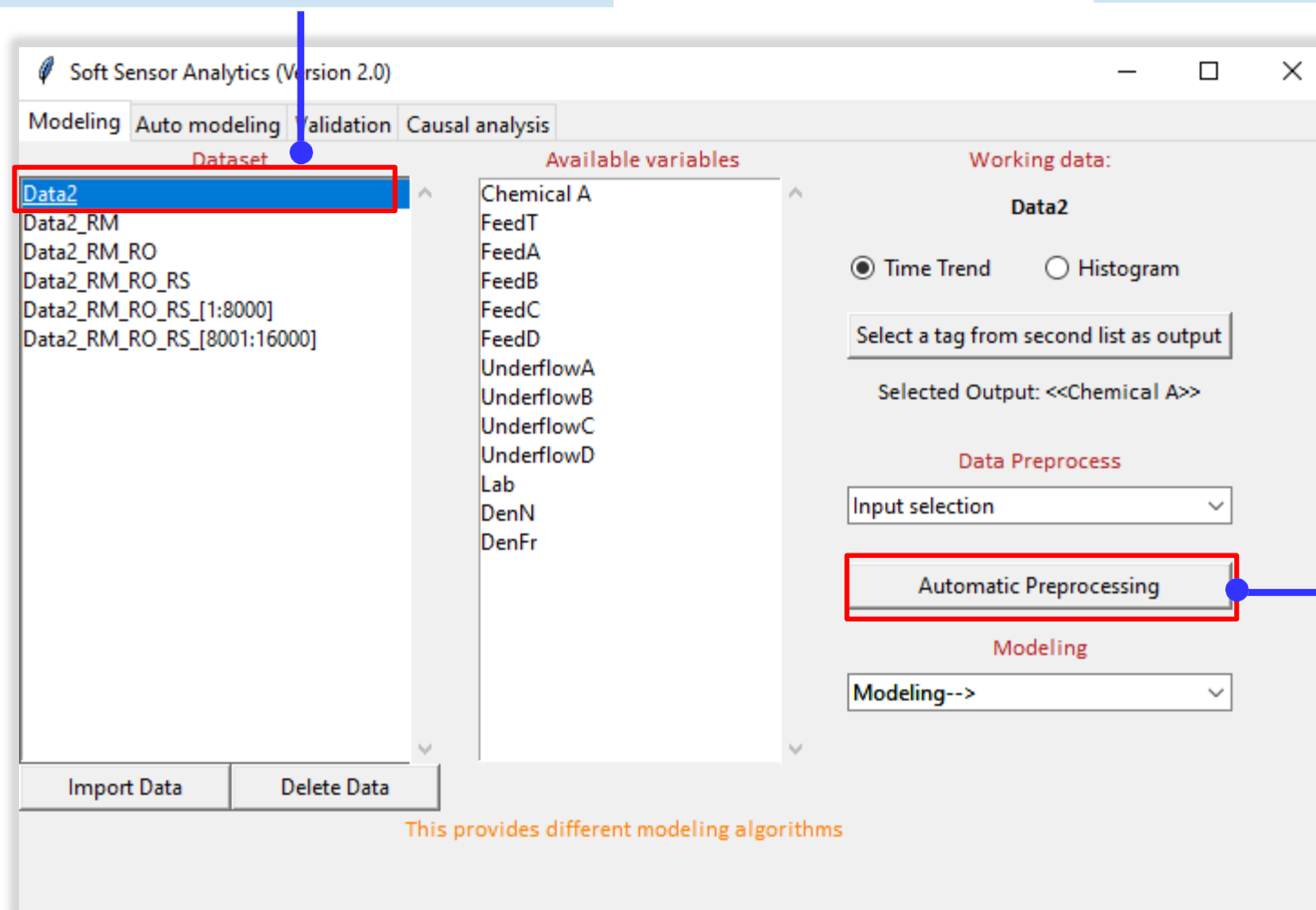
Simulation dataset:Data2_RM_RO_RS_[8001:16000]



Alternative way to process data quickly -- Automatic Preprocessing

- 1
- Under the “Modeling” tab, choose dataset to work on
- Choose original data “Data2” on the left list
The variables in this dataset will appear in the second list

- 2
- Automatic Preprocessing
- Click “Automatic Preprocessing” button



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

Data2

Data2_RM

Data2_RM_RO

Data2_RM_RO_RS

Data2_RM_RO_RS_[1:8000]

Data2_RM_RO_RS_[8001:16000]

Available variables

Chemical A

FeedT

FeedA

FeedB

FeedC

FeedD

UnderflowA

UnderflowB

UnderflowC

UnderflowD

Lab

DenN

DenFr

Working data:

Data2

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Chemical A>>

Data Preprocess

Input selection

Automatic Preprocessing

Modeling

Modeling-->

Import Data Delete Data

This provides different modeling algorithms

Automatic Preprocessing

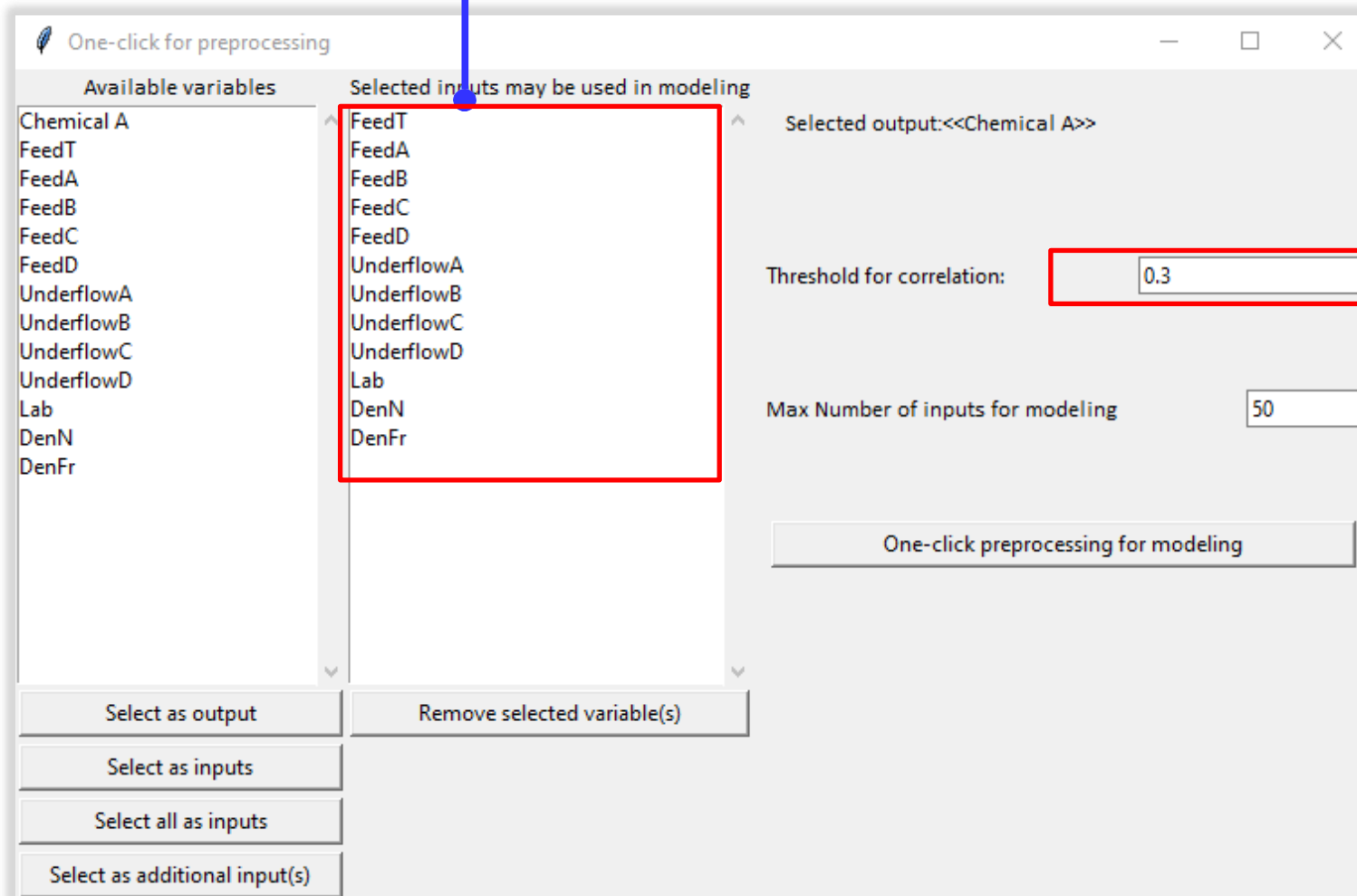
3

- Click **“Select all as inputs”** button
All the variables except the output will be shown in the second column as candidate inputs

4

Threshold selection

- Enter a value between 0 and 1 as the lower bound on the correlation coefficients for input recommendation, 0.3 is the default value



One-click for preprocessing

Available variables

Selected inputs may be used in modeling

Selected output: <<Chemical A>>

Threshold for correlation: 0.3

Max Number of inputs for modeling 50

One-click preprocessing for modeling



Automatic Preprocessing

5

Threshold selection

- Enter a positive integer as an upper bound on the number of inputs that will be included in modeling, 50 is the default value

The screenshot shows a software interface titled "One-click for preprocessing". It has two main columns: "Available variables" on the left and "Selected inputs may be used in modeling" on the right. The "Available variables" list includes Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, and DenFr. The "Selected inputs" list includes FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, and DenFr. Below the "Available variables" list are four buttons: "Select as output", "Select as inputs", "Select all as inputs", and "Select as additional input(s)". Below the "Selected inputs" list is a button: "Remove selected variable(s)". To the right of the "Selected inputs" list, there is a section for "Selected output: <<Chemical A>>". Below this, there is a "Threshold for correlation:" field with the value "0.3". Below that is a "Max Number of inputs for modeling" field with the value "9". At the bottom of the dialog is a large button labeled "One-click preprocessing for modeling".

6

Automatic Preprocessing

- Click "One-click preprocessing for modeling" button



Automatic Preprocessing

Results

- On the main window, two new datasets are generated. One is for modeling, the other one is for model validation.

Results – input selection

A set of inputs for modeling has also been generated. Next step when going to one of the regression algorithms, auto-selected inputs will be shown

Then, modeling can be conducted with minimum effort

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' application window. The 'Modeling' tab is active. The 'Dataset' list on the left includes 'Data2_for_training' and 'Data2_for_validation', which are highlighted with a red box. The 'Available variables' list in the center includes 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'. The 'Working data:' section on the right shows 'Data2' selected, with 'Time Trend' chosen over 'Histogram'. Below this, there is a dropdown menu for 'Input selection' and a button for 'Automatic Preprocessing'. At the bottom right, a dropdown menu for 'Modeling' is highlighted with a red box and contains the option 'Modeling-->'. A blue line connects the 'Data2_for_validation' dataset to the 'Modeling-->' option. At the bottom of the window, there are buttons for 'Import Data' and 'Delete Data', and a status bar with the text 'Select Variable and Plot Histogram'.

Another way to model quickly -- Automatic modeling

1

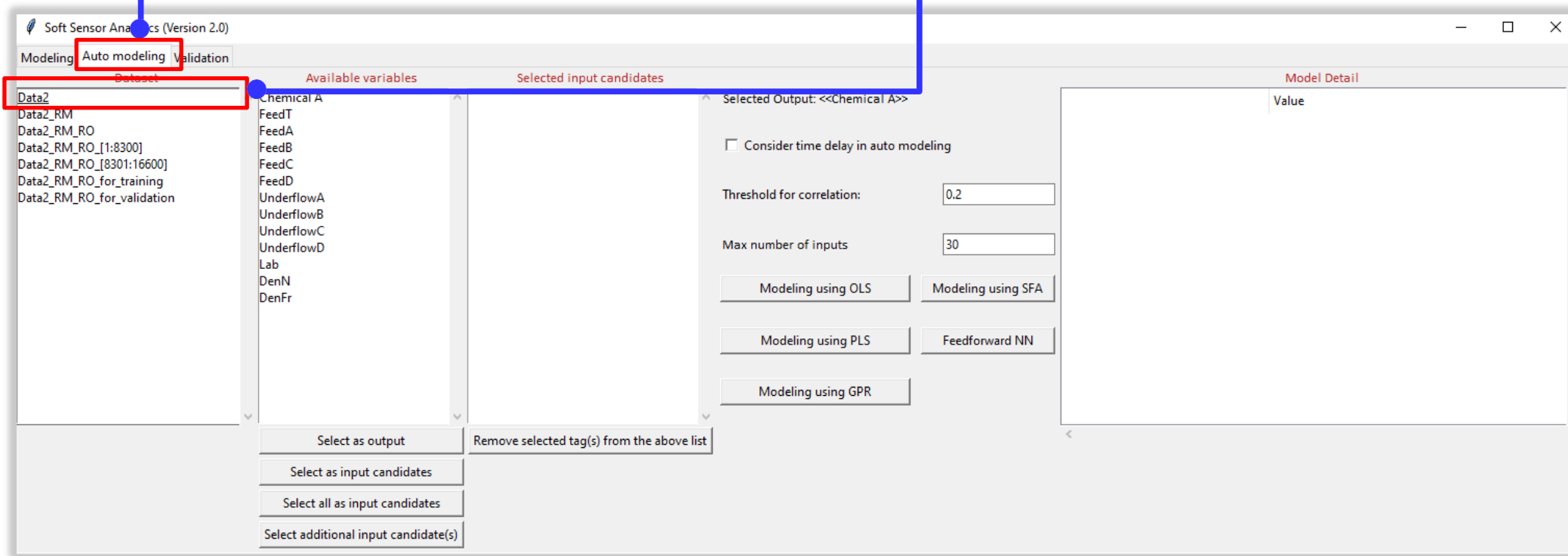
Automatic modeling

- Click the “Auto modeling” tab

2

Choose dataset to work on

- Choose original data “Data2” on the left list
The variables in this dataset will appear in the second list



Soft Sensor Analytics (Version 2.0)

Modeling **Auto modeling** Validation

Dataset

- Data2
- Data2_RM
- Data2_RM_RO
- Data2_RM_RO_[1:8300]
- Data2_RM_RO_[8301:16600]
- Data2_RM_RO_for_training
- Data2_RM_RO_for_validation

Available variables

- Chemical A
- FeedT
- FeedA
- FeedB
- FeedC
- FeedD
- UnderflowA
- UnderflowB
- UnderflowC
- UnderflowD
- Lab
- DenN
- DenFr

Selected input candidates

Selected Output: <<Chemical A>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

Select as input candidates

Select all as input candidates

Select additional input candidate(s)

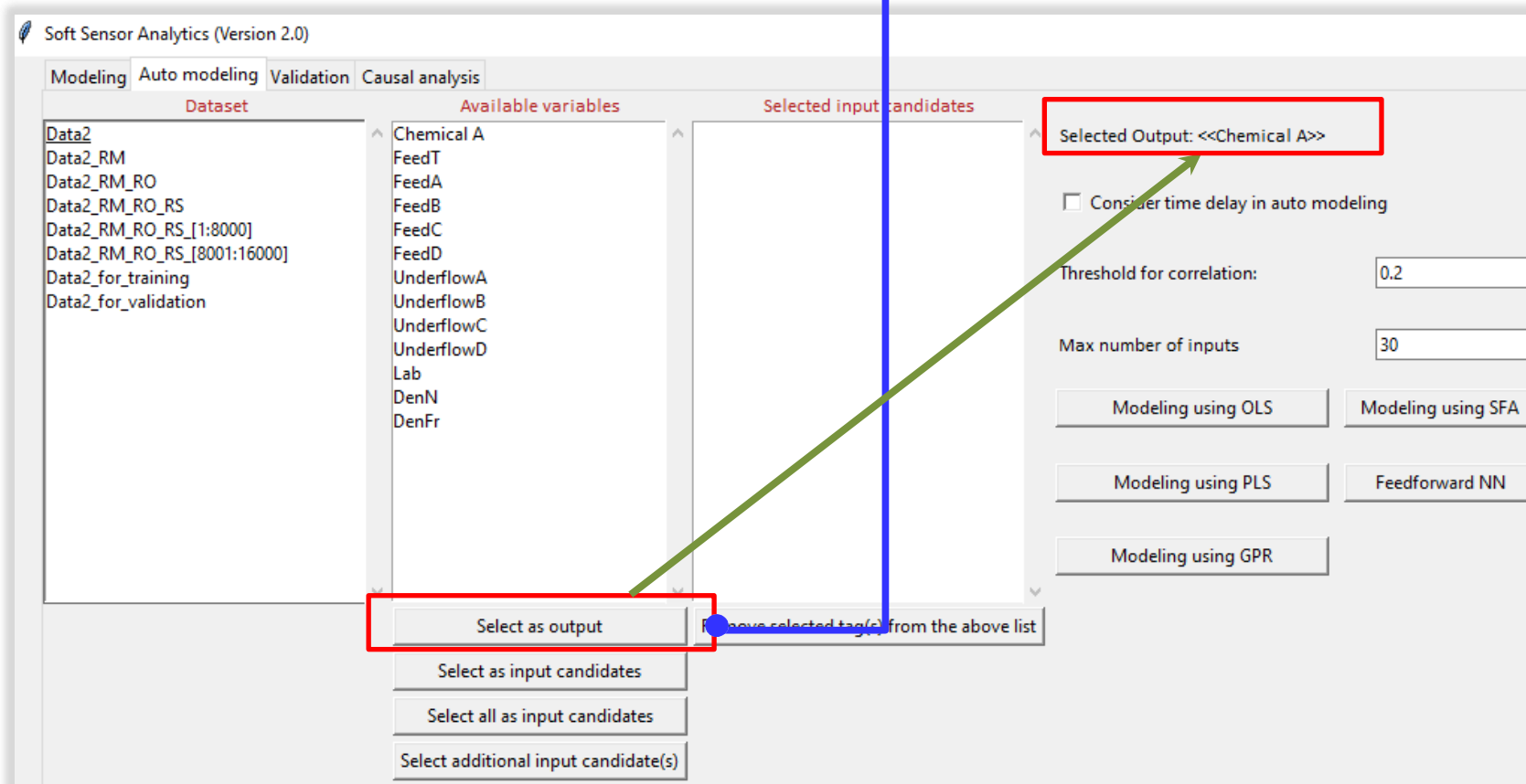
Model Detail

Value

Automatic modeling – inputs/output selection

3

- Previously selected output is shown on this window
- Select another output for modelling if that needs to be changed



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset Available variables Selected input candidates

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS
Data2_RM_RO_RS_[1:8000]
Data2_RM_RO_RS_[8001:16000]
Data2_for_training
Data2_for_validation

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Selected Output: <<Chemical A>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

Select as input candidates

Select all as input candidates

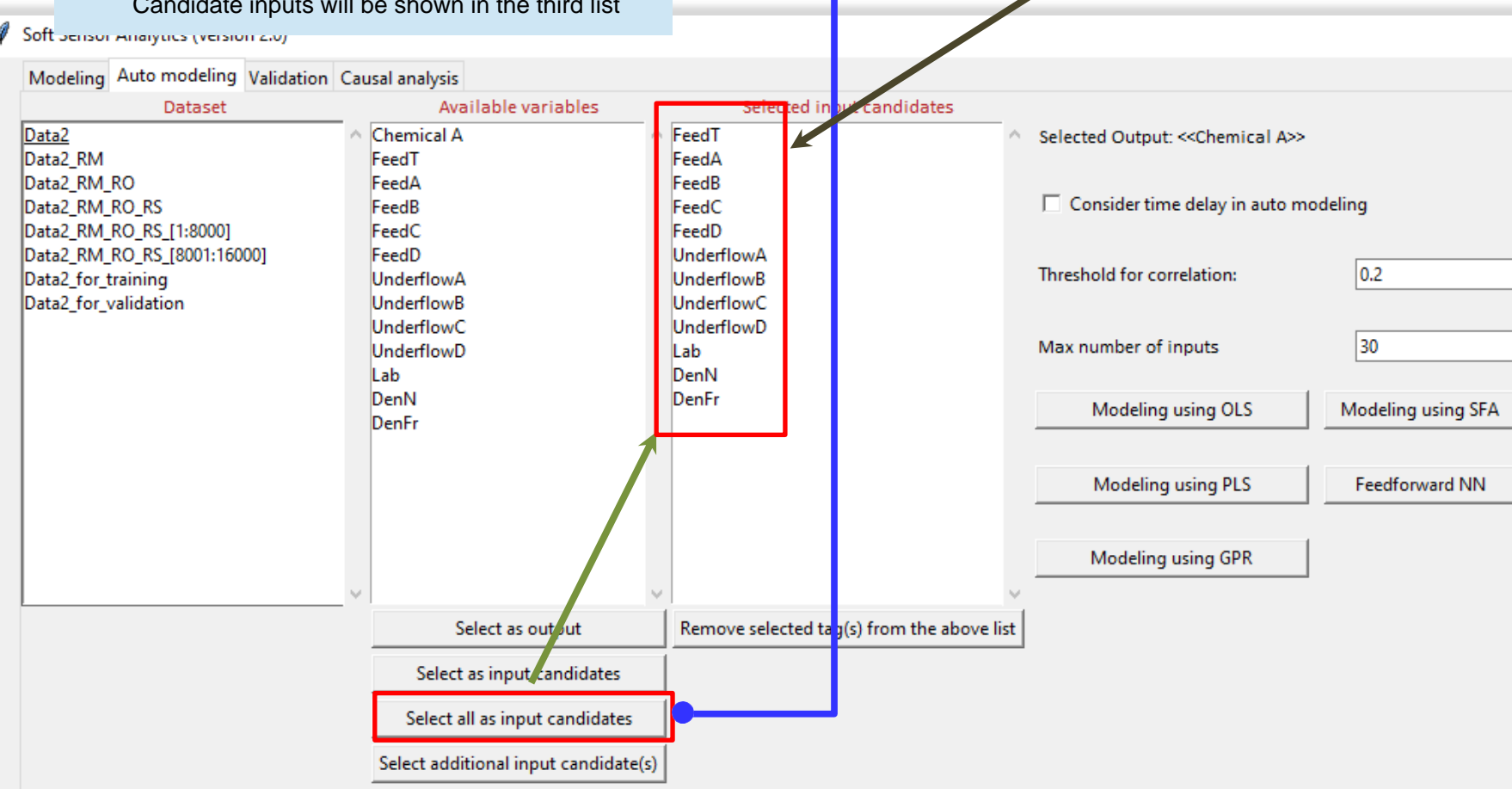
Select additional input candidate(s)

Automatic modeling – inputs/output selection

4

Select inputs that may be used for modelling
• Click “Select all as input candidates”
Candidate inputs will be shown in the third list

The already selected output will not appear on this list when selecting all as inputs

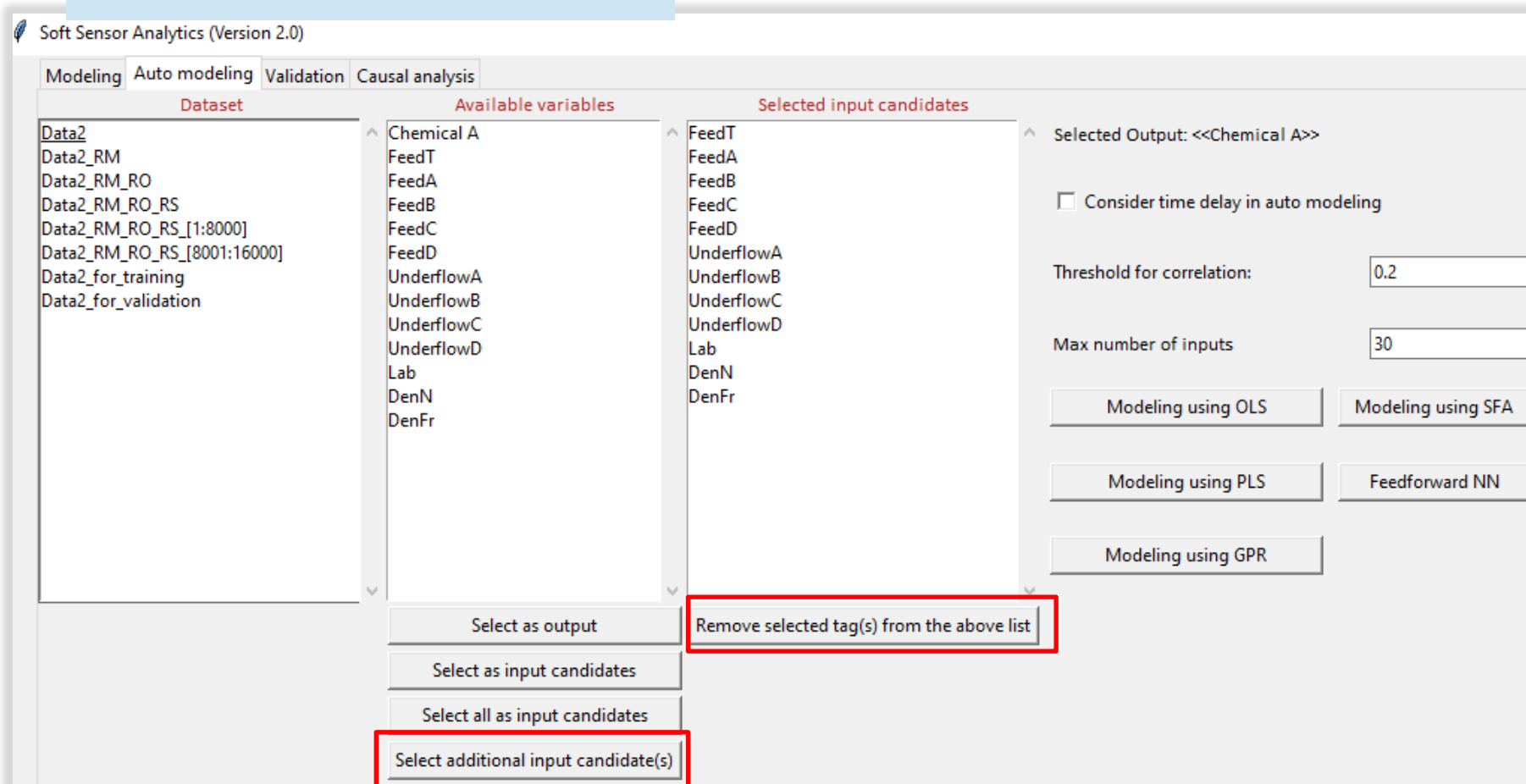


Note: We can also use “Select as input candidates” button to select potential inputs from all variables based on available information/knowledge

Automatic modeling – inputs/output selection

The selected inputs can be modified

- Click “**Remove selected tag(s) from the above list**” to remove unnecessary variables
- Click “**Select as additional input(s)**” to include more variables



The screenshot displays the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Auto modeling' tab is active, showing a workflow for selecting inputs and outputs for modeling. The interface is divided into several sections:

- Dataset:** A list of datasets including 'Data2', 'Data2_RM', 'Data2_RM_RO', 'Data2_RM_RO_RS', 'Data2_RM_RO_RS_[1:8000]', 'Data2_RM_RO_RS_[8001:16000]', 'Data2_for_training', and 'Data2_for_validation'.
- Available variables:** A list of variables including 'Chemical A', 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'.
- Selected input candidates:** A list of variables including 'FeedT', 'FeedA', 'FeedB', 'FeedC', 'FeedD', 'UnderflowA', 'UnderflowB', 'UnderflowC', 'UnderflowD', 'Lab', 'DenN', and 'DenFr'.
- Selected Output:** A dropdown menu showing '<<Chemical A>>'.
- Modeling options:** A set of buttons for 'Modeling using OLS', 'Modeling using SFA', 'Modeling using PLS', 'Feedforward NN', and 'Modeling using GPR'.
- Parameters:** A checkbox for 'Consider time delay in auto modeling', a 'Threshold for correlation' field set to '0.2', and a 'Max number of inputs' field set to '30'.

At the bottom of the interface, there are four buttons for modifying the selected inputs and output:

- 'Select as output'
- 'Remove selected tag(s) from the above list' (highlighted with a red box)
- 'Select as input candidates'
- 'Select all as input candidates'
- 'Select additional input candidate(s)' (highlighted with a red box)



Automatic modeling

5

Set threshold on correlation coefficient

- This is a lower bound on correlation coefficients for selecting input variables
- **Set "0.2"** as the threshold

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset Available variables Selected input candidates

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS
Data2_RM_RO_RS [1:8000]
Data2_RM_RO_RS [8001:16000]
Data2_for_training
Data2_for_validation

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Selected Output: <<Chemical A>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

Select as input candidates

Select all as input candidates

Select additional input candidate(s)



Automatic modeling

6

- Set max number of inputs
- Set "10" as the max number of inputs

7

- One-click modeling using OLS
- Click "Modeling using OLS" to build a model

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset Available variables Selected input candidates

Data2
Data2_RM
Data2_RM_RO
Data2_RM_RO_RS
Data2_RM_RO_RS_[1:8000]
Data2_RM_RO_RS_[8001:16000]
Data2_for_training
Data2_for_validation

Chemical A
FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

FeedT
FeedA
FeedB
FeedC
FeedD
UnderflowA
UnderflowB
UnderflowC
UnderflowD
Lab
DenN
DenFr

Selected Output: <<Chemical A>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 10

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

Select as input candidates

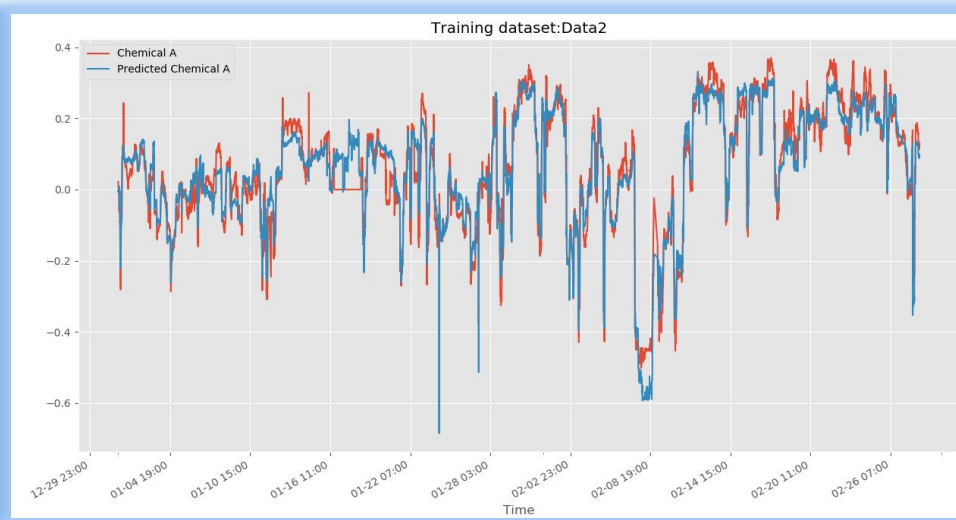
Select all as input candidates

Select additional input candidate(s)

Automatic modeling - Results

Results

- Scatter plot and time trend are generated
- Model information is shown in “**Details**” column



Soft Sensor Analytics (Version 2.0)

Modeling | Auto modeling | Validation | Causal analysis

Dataset: Data2

Available variables: Chemical A, FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, DenFr

Selected input candidates: FeedT, FeedA, FeedB, FeedC, FeedD, UnderflowA, UnderflowB, UnderflowC, UnderflowD, Lab, DenN, DenFr

Selected Output: <<Chemical A>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 10

Modeling using OLS | Modeling using SFA

Modeling using PLS | Feedforward NN

Modeling using GPR

Model Detail

Method	Value
Structure	Ordinary least squares
x	...
y	Chemical A
c	-0.0411000006805774
a	...
Performance	...



Part II: Exercise

Soft Sensor Analytics



The NIR data set (Data3.xlsx)

The other case study is based on NIR spectrum data (**Data3**) which is widely used to predict key specifications of product in Oil and Gas industry. In this case study the NIR spectrum is used to predict the cloud point (Target) of the diesel.

Symbol	Description	Availability
WL1	Spectrum at wavelength 801mm	Real-time
WL2	Spectrum at wavelength 802mm	Real-time
WL3	Spectrum at wavelength 803mm	Real-time
WL4	Spectrum at wavelength 804mm	Real-time
WL5	Spectrum at wavelength 805mm	Real-time
WL6	Spectrum at wavelength 806mm	Real-time
WL7	Spectrum at wavelength 807mm	Real-time
WL8	Spectrum at wavelength 808mm	Real-time
WL9	Spectrum at wavelength 809mm	Real-time
.....		
WL893	Spectrum at wavelength 1694mm	Real-time
Target	Cloud point of diesel	12 hours

Soft Sensor Analytics



Data preparation

Make sure the data are prepared following the guidelines below:

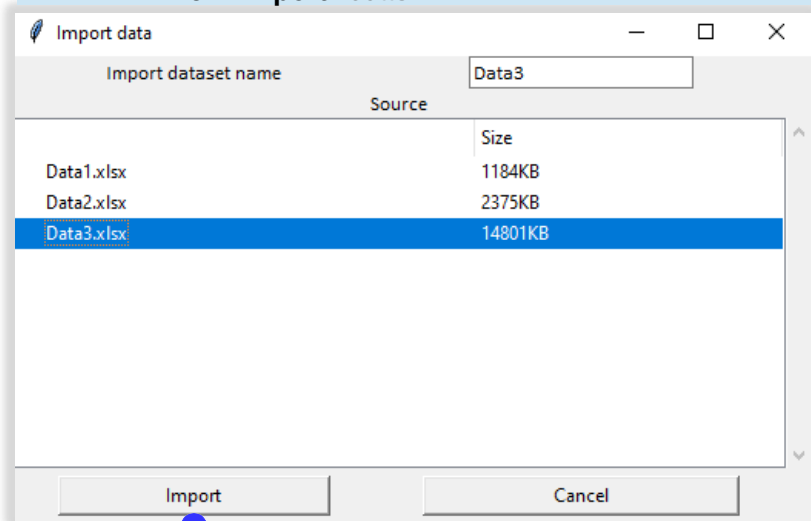
- Prepare the dataset using Excel
- the first column of a spreadsheet always contains timestamps
- the first row of a spreadsheet always contains the name of variables
- The data for each variable are recorded in each corresponding column
- For the created spreadsheet, select the entire timestamp column, right-click on this column and go to **“format cells”**, then go to **“Time”** or **“Custom”**, change the format of time information to a format similar to: **“mm/dd/yyyy h:mm”** or **“yyyy-mm-dd h:mm:ss”**.
- Save the dataset in **.xlsx** format

	A	B	C	D	E	F	G	H	I	J	K	L
1		Target	WL1	WL2	WL3	WL4	WL5	WL6	WL7	WL8	WL9	WL10
2	2015-01-01 0:00:00	-0.07822643	0.0256686	0.02878536	0.03384522	0.03534522	0.03242624	0.03159448	0.03125008	0.03282767	0.03543364	0.03744536
3	2015-01-01 6:00:00	-0.16201361	0.03198536	0.03043356	0.0295839	0.02636711	0.02308324	0.02121744	0.02114328	0.02129045	0.02325356	0.02608751
4	2015-01-02 0:00:00	-0.24328717	0.00586299	0.00561063	0.00086565	-0.00180007	-0.00425344	-0.00845396	-0.00970758	-0.00920677	-0.01026148	-0.01005259
5	2015-01-02 6:00:00	0.00388501	0.00377835	0.00544399	0.00629621	0.00521003	0.00288794	0.0012895	0.00161601	-0.00067395	-0.00431455	-0.00693641
6	2015-01-03 0:00:00	-0.00952094	0.00011715	0.00357057	0.00906133	0.01114327	0.00936595	0.00473976	0.00025206	-0.00558715	-0.00910827	-0.01476694
7	2015-01-03 6:00:00	-0.09582173	0.00211787	0.00559673	0.0054469	0.00274916	0.00254984	0.00082777	-0.00223123	-0.00610669	-0.01327314	-0.01457559
8	2015-01-04 0:00:00	-0.13729639	0.01230002	0.01227312	0.01089373	0.01036102	0.00988568	0.0066826	0.00118898	-0.00105129	-0.00395547	-0.00409901
9	2015-01-04 6:00:00	0.0340484	-0.00452169	-0.00491841	-0.00696038	-0.0089948	-0.00927903	-0.00454388	-0.00208087	0.00298439	0.01058669	0.01114148
10	2015-01-05 0:00:00	-0.11132236	-0.00122167	-0.00401728	-0.00514907	-0.00645921	-0.0066714	-0.00742462	-0.0083381	-0.00390465	0.00255266	0.00126136
11	2015-01-05 6:00:00	-0.12053895	0.00533969	0.00112026	-0.00333622	-0.00359396	-0.00187073	0.00194792	0.00031079	0.00299285	0.0085019	0.00653453

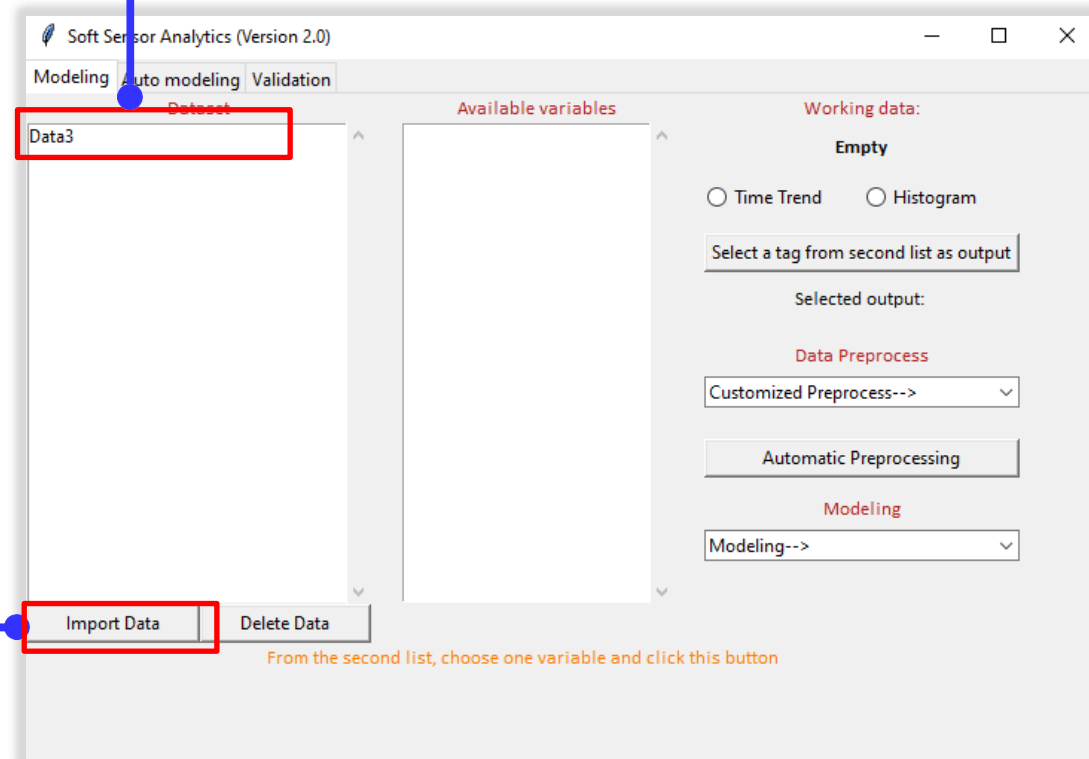
Import data sets

Note: Please save data sets in the “.xlsx” format

- 1
- Click “Import Data” button
 - Select data file “Data3.xlsx”
 - Click “Import” button



“Data3” has been imported



Soft Sensor Analytics



Current window with data3 imported

1

- Click on the “Data” tab.
- “Data3” has already been imported as highlighted in the right window“

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Available variables

Working data: Data3

Time Trend Histogram

Select a tag from second list as output

Selected output:

Data Preprocess

Customized Preprocess-->

Automatic Preprocessing

Modeling

Modeling-->

Import Data Delete Data

Import Dataset From Folder

Soft Sensor Analytics



Variables in Data3

2

Choose Data3 to work on it

- Click on "Data3" from left workspace (i.e., column "Dataset")

The variables in Data3 are shown in the right column.
Scroll down to view all the variables.

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' application window. The 'Modeling' tab is active. On the left, under the 'Dataset' column, 'Data3' is selected and highlighted with a red box. A blue arrow points from a text box to this selection. In the center, the 'Available variables' list is shown, containing 'Target' and 'WL1' through 'WL19'. A green arrow points from a text box to this list. On the right, the 'Working data' section shows 'Data3' selected, with options for 'Time Trend' and 'Histogram'. Below that, there are dropdown menus for 'Select a tag from second list as output', 'Data Preprocess' (set to 'Customized Preprocess-->'), and 'Modeling' (set to 'Modeling-->'). At the bottom, there are buttons for 'Import Data', 'Delete Data', and 'Import Dataset From Folder'.

Soft Sensor Analytics



Visualize several variables of Data3

3

Click to select variables for visualization

4

Click "Time Trend" to visualize the selected variables

The screenshot displays the 'Soft Sensor Analytics (Version 2.0)' interface. On the left, the 'Dataset' list includes 'Data3'. The 'Available variables' list contains WL44 through WL63, with 'WL51' selected. A blue line connects this selection to the 'Time Trend' radio button in the 'Working data:' section, which is also highlighted with a red box. A green arrow points from the 'Time Trend' button to a time-series plot titled 'Dataset: Data3'. The plot shows multiple data series (Target, WL1, WL8, WL12, WL13, WL17, WL51) over time from 01-29 06:00 to 08-24 04:00. The y-axis ranges from -0.6 to 0.6. A legend on the right identifies the series. Below the plot, the text 'Select Variable and Plot Histogram' is visible.

Soft Sensor Analytics



Visualize several variables of Data3

5

Click to select variables for visualization

6

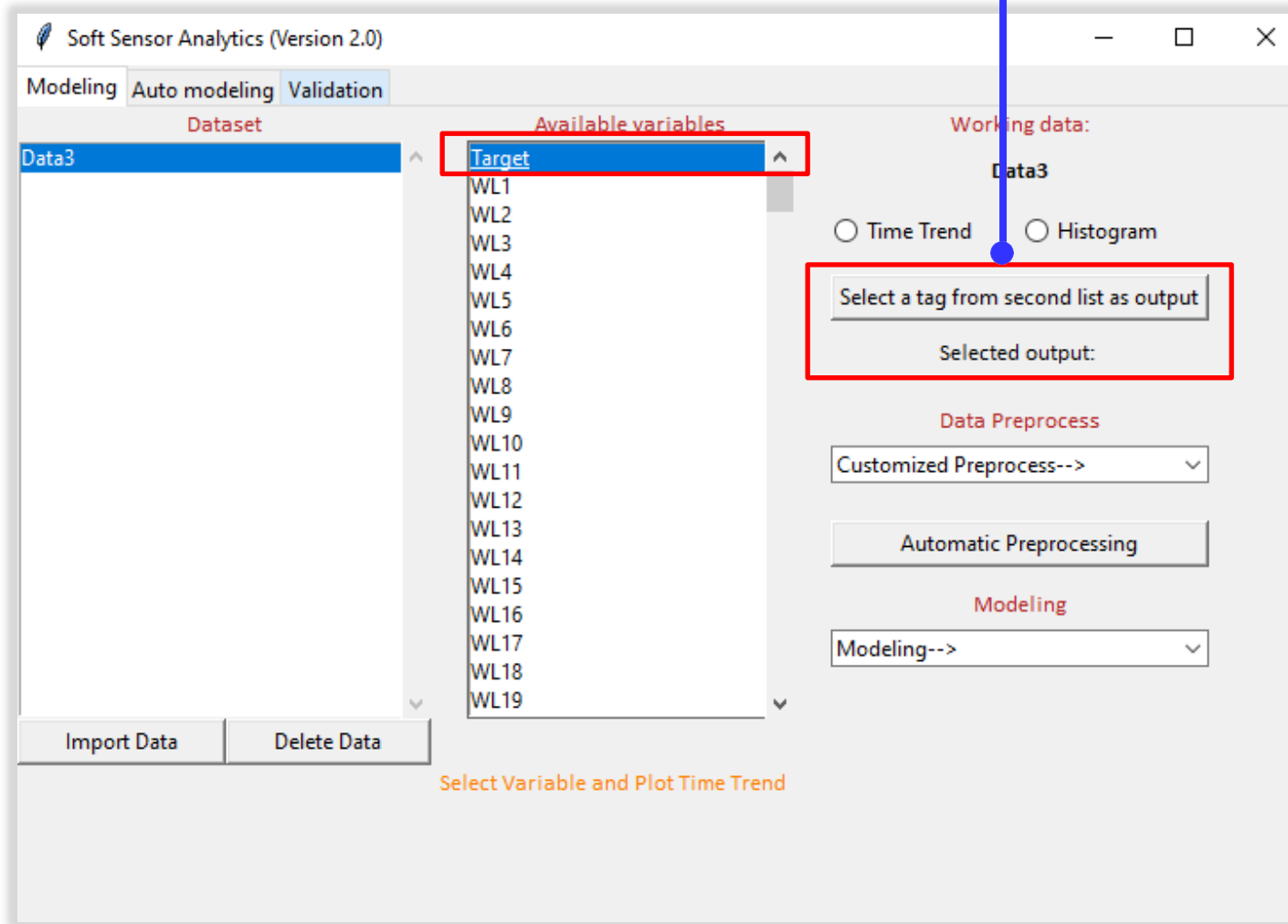
Click "Histogram" to visualize the selected variables

The screenshot displays the 'Soft Sensor Analytics (Version 2.0)' interface. On the left, the 'Dataset' list contains 'Data3'. The 'Available variables' list includes WL40 through WL59. A blue circle highlights the 'WL50' variable, with a blue arrow pointing to the 'Histogram' radio button in the 'Working data' section. The 'Histogram' radio button is also highlighted with a red box and a blue arrow. Below the 'Histogram' button is a text box 'Select a tag from second list as output' and a 'Selected output:' label. A green arrow points from the 'Histogram' button to a histogram plot titled 'Dataset: Data3'. The histogram shows the frequency distribution of the selected variables, with a legend on the right listing 'Target', 'WL2', 'WL3', 'WL12', 'WL14', 'WL17', 'WL50', 'WL56', and 'WL59'. The x-axis ranges from -0.6 to 0.6, and the y-axis (Frequency) ranges from 0 to 700. The histogram bars are colored according to the legend, with a prominent blue bar for WL50.

Select an output for modeling

7 Select output from second list

- Choose a variable “**Target**” from the second list
- Click “**Select a tag from second list as output**” button to select this variable as the output for modeling



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Data3

Available variables

Target

WL1

WL2

WL3

WL4

WL5

WL6

WL7

WL8

WL9

WL10

WL11

WL12

WL13

WL14

WL15

WL16

WL17

WL18

WL19

Working data:

Data3

Time Trend Histogram

Select a tag from second list as output

Selected output:

Data Preprocess

Customized Preprocess-->

Automatic Preprocessing

Modeling

Modeling-->

Import Data Delete Data

Select Variable and Plot Time Trend

Soft Sensor Analytics



Data preprocessing – Remove missing data and outliers

8

This step is to remove missing data from the dataset

- Select “Data3” from Dataset column
- Click “Customized Preprocess -->” on the right
- Click “Remove missing data”; new data is “Data3_RM”

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Data3
Data3_RM

Available variables

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19

Working data:

Data3

Time Trend Histogram

Select a tag from second list as output

Selected output:

Data Preprocess

Customized Preprocess-->
Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Correlation analysis
Mutual information analysis

Import Data Delete Data

This provides different data preprocessing algorithms

Soft Sensor Analytics



Data preprocessing – Remove missing data and outliers

9

This step is to remove outliers in the dataset

- Select "Data3_RM" from Dataset column
- Click "Customized Preprocess -->" on the right
- Click "Remove missing data"; new data is "Data3_RM_RO"

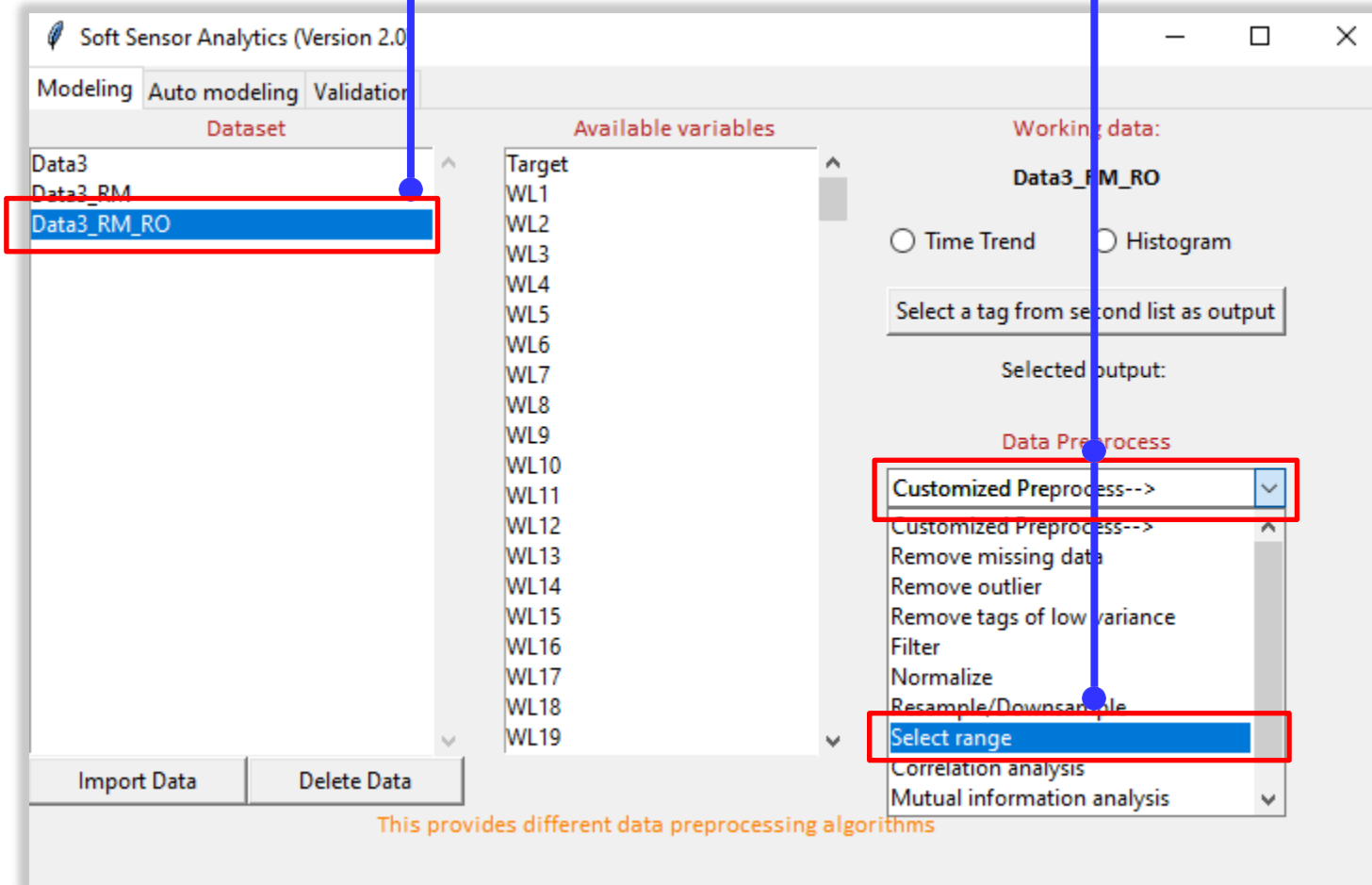
The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Dataset' column on the left contains 'Data3', 'Data3_RM', and 'Data3_RM_RO'. The 'Available variables' column lists 'Target' and 'WL1' through 'WL19'. The 'Working data' section shows 'Data3_RM_RO' and options for 'Time Trend' and 'Histogram'. The 'Data Preprocess' section is expanded, showing a list of preprocessing algorithms: 'Customized Preprocess-->', 'Remove missing data', 'Remove outlier', 'Remove tags of low variance', 'Filter', 'Normalize', 'Resample/Downsample', 'Select range', 'Correlation analysis', and 'Mutual information analysis'. A red box highlights 'Remove outlier' in the 'Data Preprocess' list. A blue box highlights 'Data3_RM' in the 'Dataset' column. A green arrow points from the 'Remove outlier' option to the 'Data3_RM' entry. A blue dot is placed on the 'Data3_RM' entry, with a blue line extending upwards. At the bottom, a caption reads: 'This provides different data preprocessing algorithms'.

Data preprocessing – Split data into two parts for training and validation

10

Get a smaller dataset for training

- Select “Data3_RM_RO”
- Click “Customized Preprocess-->”, then Click “Select range”; a window will pop up, as shown in the next slide



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Data3
Data3_RM
Data3_RM_RO

Available variables

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19

Working data:
Data3_RM_RO

Time Trend Histogram

Select a tag from second list as output

Selected output:

Data Preprocess

Customized Preprocess-->
Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Correlation analysis
Mutual information analysis

Import Data Delete Data

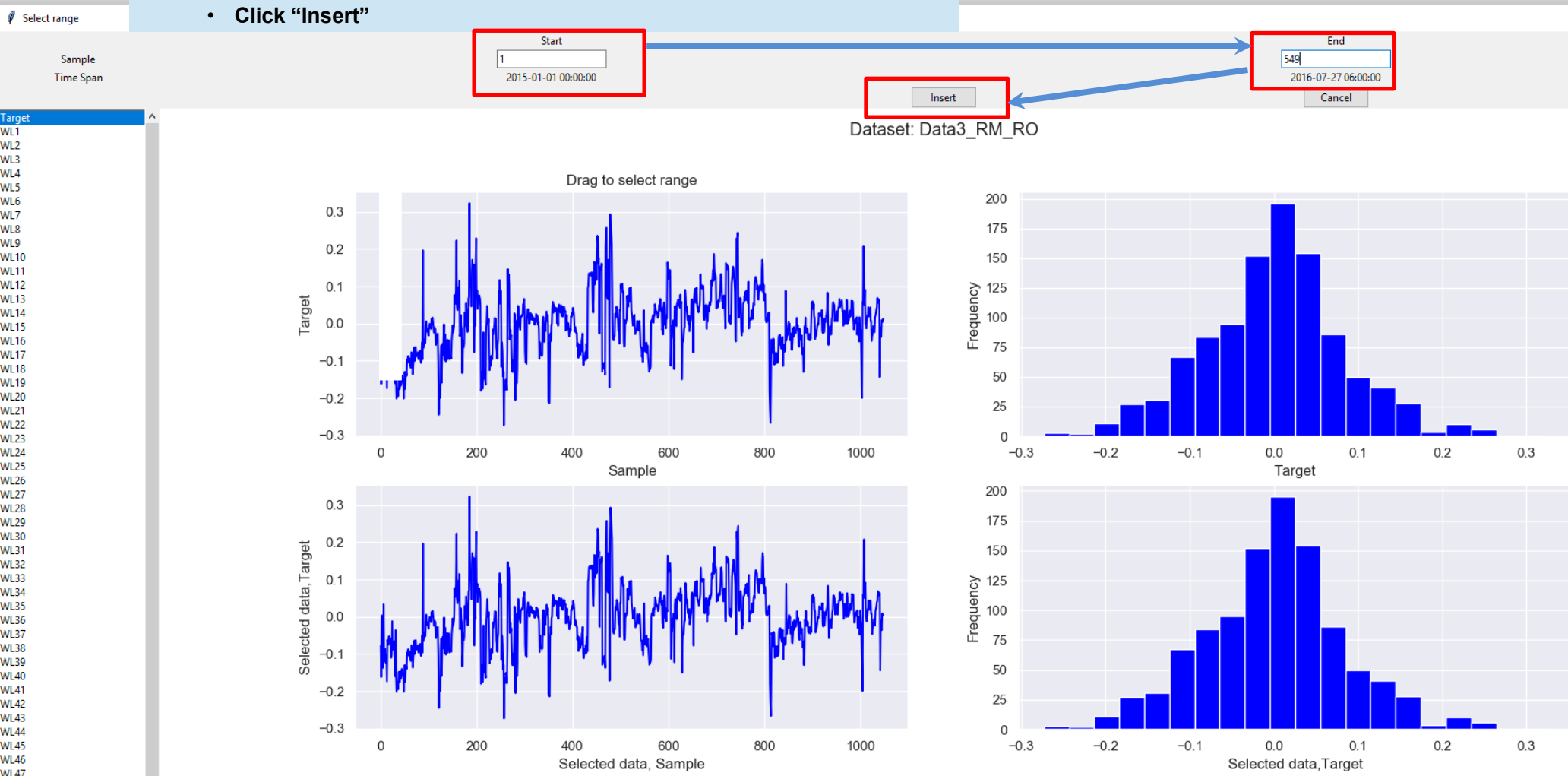
This provides different data preprocessing algorithms

Data preprocessing – Split data into smaller sets

11

To split data enter the range of dataset for training, do the following:

- Click “Target” on the left column
- In “Start”, Enter “1”
- In “End”; Enter “549”
- Click “Insert”



The screenshot shows the data preprocessing interface for the dataset "Data3_RM_RO". The "Start" field is set to "1" (2015-01-01 00:00:00) and the "End" field is set to "549" (2016-07-27 06:00:00). The "Insert" button is highlighted. Below the input fields, there are two line plots and two histograms. The top-left plot is titled "Drag to select range" and shows the "Target" variable over "Sample" (0 to 1000). The bottom-left plot is titled "Selected data, Target" and shows the selected data over "Selected data, Sample" (0 to 1000). The top-right plot is a histogram of the "Target" variable, and the bottom-right plot is a histogram of the "Selected data, Target" variable. Both histograms show a distribution centered around 0.0.

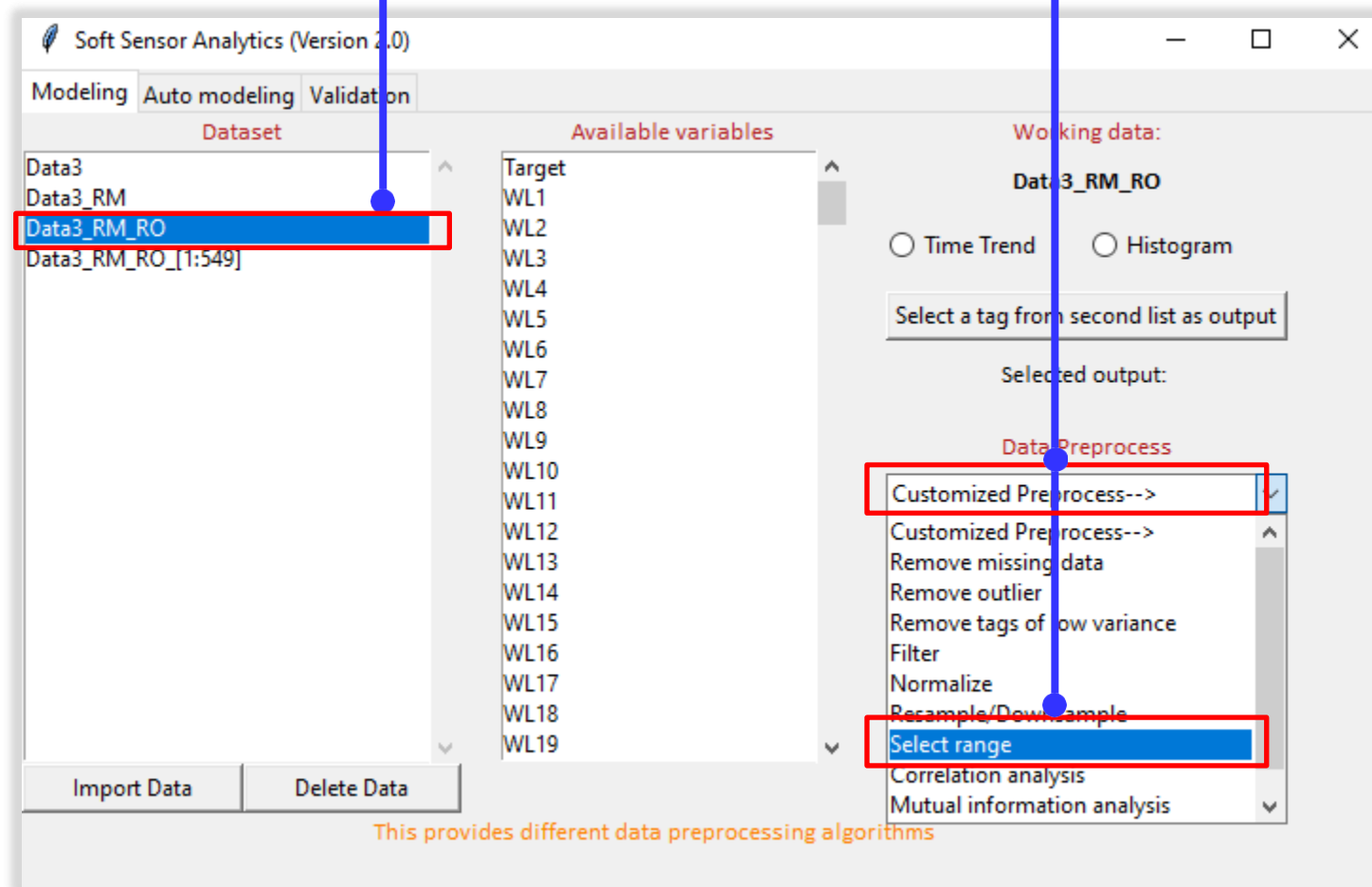
Note: Please close the current popup figures before taking next action

Data preprocessing – Divide data into two parts for training and validation

12

Get a smaller dataset for validation

- Select “Data3_RM_RO” again
- Click “Customized Preprocess-->”, then Click “Select range”; a window will pop up, as shown in the next slide



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Data3
Data3_RM
Data3_RM_RO
Data3_RM_RO_[1:549]

Available variables

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19

Working data:
Data3_RM_RO

Time Trend Histogram

Select a tag from second list as output

Selected output:

Data Preprocess

Customized Preprocess-->
Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Down sample
Select range
Correlation analysis
Mutual information analysis

Import Data Delete Data

This provides different data preprocessing algorithms

Soft Sensor Analytics

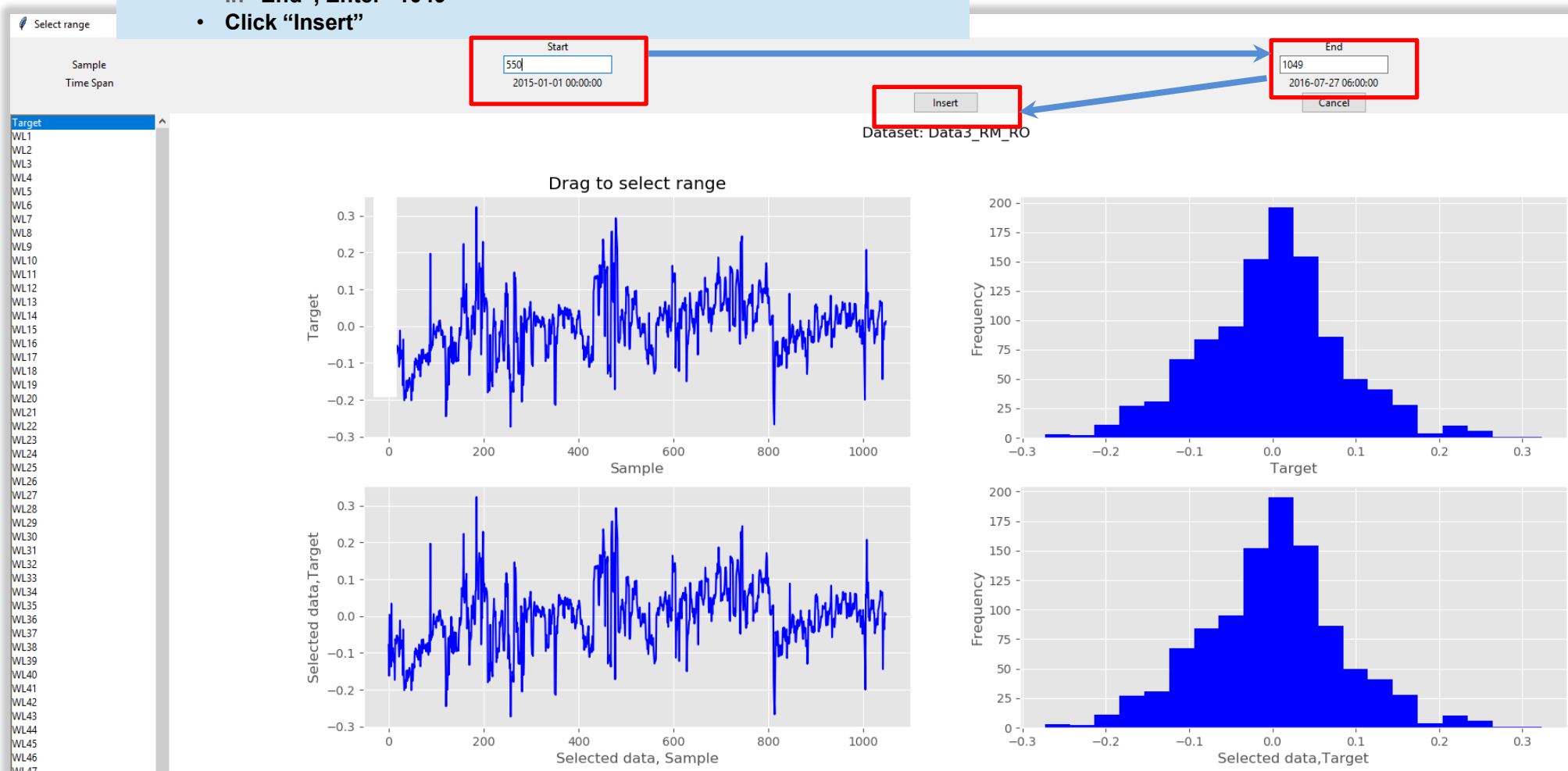


Data preprocessing – Divide data into two parts

13

To split data enter the range of dataset for validation, do the following:

- Click “Target” on the left column
- In “Start”, Enter “550”
- In “End”; Enter “1049”
- Click “Insert”



Note: Please close the current popup figures before taking next action

Soft Sensor Analytics



Generated data sets after pre-processing

Updated Dataset list

After preprocess two new datasets are generated
One can be used for training, other for validation

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Dataset' list on the left includes 'Data3', 'Data3_RM', 'Data3_RM_RO', 'Data3_RM_RO_[1:549]', and 'Data3_RM_RO_[550:1049]'. The 'Available variables' list in the center includes 'Target' and 'WL1' through 'WL19'. The 'Working data' section on the right is set to 'Data3_RM_RO' and includes options for 'Time Trend' and 'Histogram', a 'Select a tag from second list as output' field, and a 'Data Preprocess' dropdown menu set to 'Customized Preprocess-->'. There are also buttons for 'Automatic Preprocessing' and a 'Modeling' dropdown menu set to 'Modeling-->'. At the bottom, there are 'Import Data' and 'Delete Data' buttons, and a note that says 'Select Variable and Plot Histogram'.

Data preprocessing – Correlation analysis

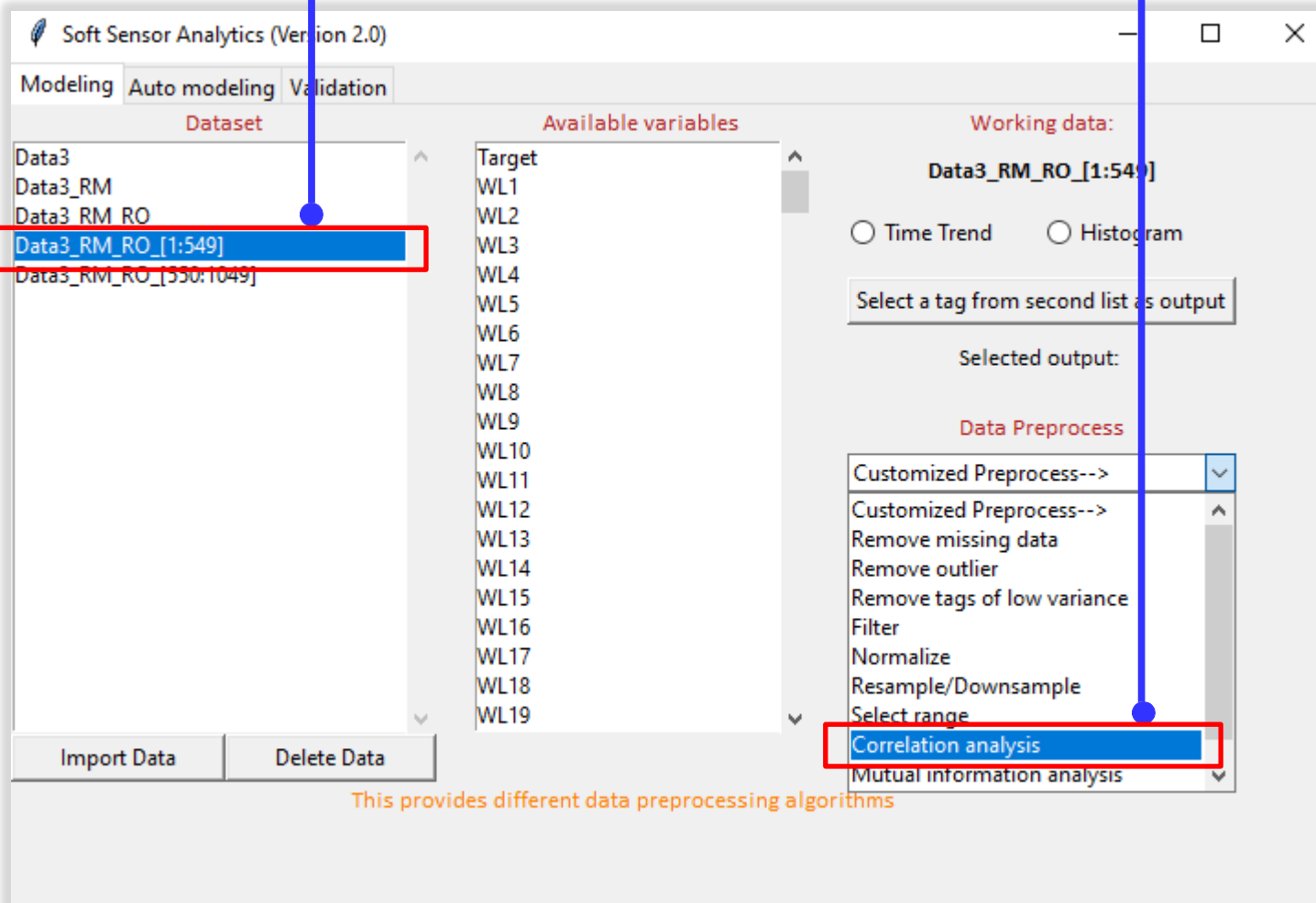
Conduct correlation analysis to check the influential variables as well as the co-linearity of the data-set.

1 Select data set for correlation analysis

- Click “Modeling” tab
- Select “Data3_RM_RO_[1:549]”

2 Choose method

- Click “Customized Preprocessing-->” on the right
- Click “Correlation Analysis”, a window will pop up



The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' software interface. The 'Modeling' tab is active. In the 'Dataset' list, 'Data3_RM_RO_[1:549]' is selected and highlighted with a red box. A blue line points from this selection to the 'Available variables' list, which contains 'Target' and 'WL1' through 'WL19'. Another blue line points from the 'Data3_RM_RO_[1:549]' selection to the 'Working data:' field, which also displays 'Data3_RM_RO_[1:549]'. Below this, there are radio buttons for 'Time Trend' and 'Histogram', and a text box for 'Select a tag from second list as output'. The 'Data Preprocess' section is expanded, showing a list of options: 'Customized Preprocess-->', 'Customized Preprocess-->', 'Remove missing data', 'Remove outlier', 'Remove tags of low variance', 'Filter', 'Normalize', 'Resample/Downsample', 'Select range', 'Correlation analysis', and 'Mutual information analysis'. The 'Correlation analysis' option is highlighted with a red box. At the bottom, there are 'Import Data' and 'Delete Data' buttons. A caption at the bottom reads: 'This provides different data preprocessing algorithms'.

Soft Sensor Analytics



Preprocessing - Correlation analysis

3

Choose variables to analyze

- **Drag-select** (can hold "CTRL") in the left list to select variables "Target, WL1, WL50, WL100, ..."
- **Click "Select variables"**

Selected variables

The variables that are selected manually for correlation analysis are shown in the right column

The screenshot shows a window titled "Correlation Analysis" with two columns: "Available variables" and "Selected variables". The "Available variables" column lists "Target" and "WL1" through "WL19". The "Selected variables" column lists "Target" and "WL1" through "WL850". A red box highlights the "Select variables" button at the bottom left. A blue line connects this button to the "3" callout. Another red box highlights the "correlation analysis" button at the bottom right of the "Selected variables" column. A blue line connects this button to the "4" callout. A third blue line connects the "Selected variables" callout to the "Selected variables" column.

4

Correlation analysis

Click "correlation analysis", a heatmap will be available, as given in the next slide

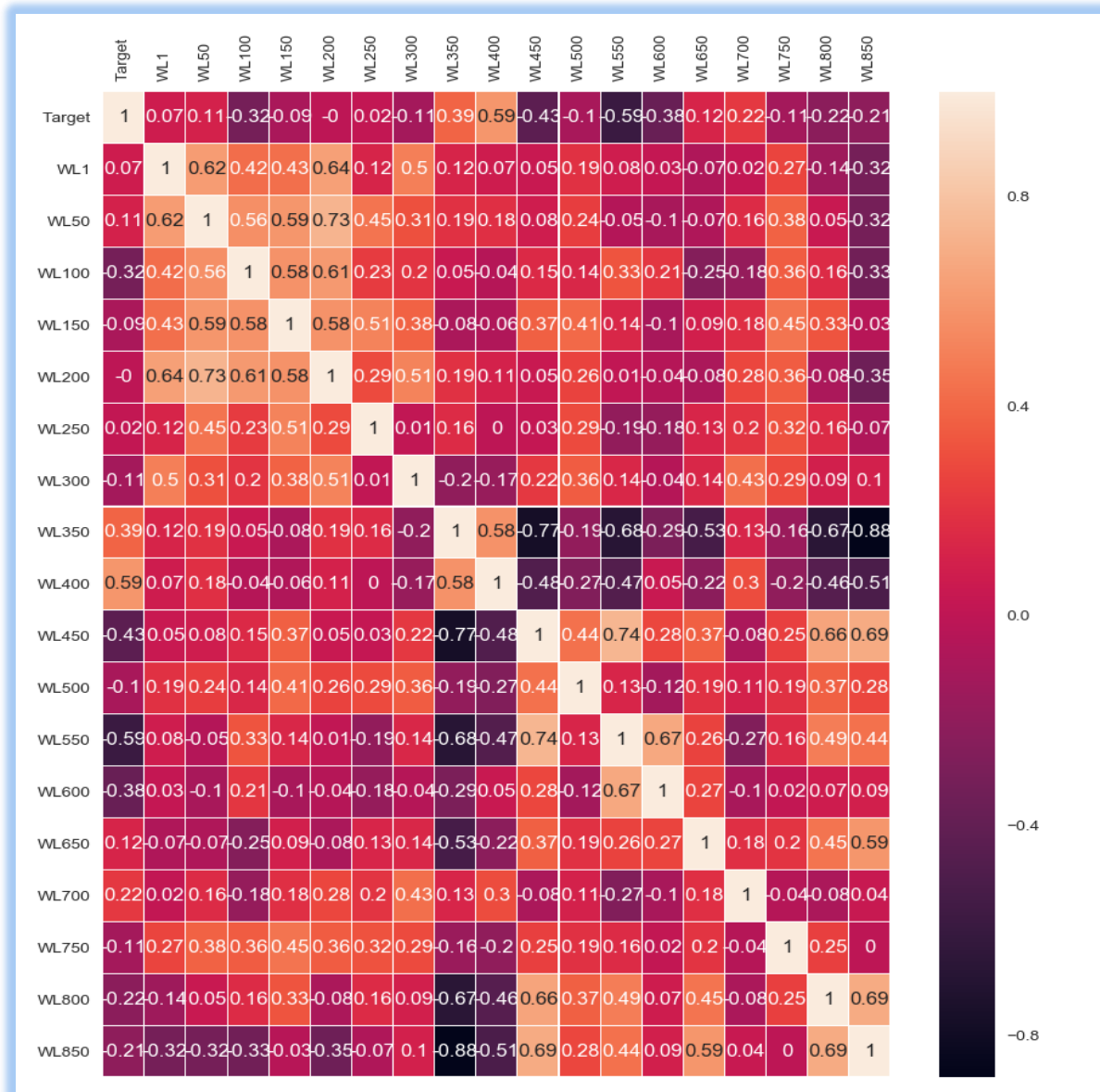


Correlation analysis – Result

Results

Variables are highly correlated to each other if the absolute value of the corresponding index is large.

One can select more variables and get more information from the correlation test results.



Soft Sensor Analytics



Automatic selection of inputs

1

Select data and go to auto input selection

- Select "Data3_RM_RO_[1:549]"
- Go to "Data Preprocess" and select "Auto input selection", a new window will pop up

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Data3' dataset is selected in the 'Dataset' list, with 'Data3_RM_RO_[1:549]' highlighted in a red box. The 'Available variables' list includes 'Target' and 'WL1' through 'WL19'. The 'Working data' section shows 'Data3_RM_RO_[1:549]' selected, with options for 'Time Trend' and 'Histogram'. The 'Data Preprocess' section is expanded, showing a list of preprocessing algorithms, with 'Input selection' highlighted in a red box. A blue line connects the highlighted data source to the 'Input selection' option. At the bottom, a red box highlights the 'Input selection' option, with a blue line connecting it to the 'Data Preprocess' section. Below the screenshot, the text 'This provides different data preprocessing algorithms' is visible.

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Data3
Data3_RM
Data3_RM_RO
Data3_RM_RO_[1:549]
Data3_RM_RO_[550:1049]

Available variables

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19

Working data:
Data3_RM_RO_[1:549]

Time Trend Histogram

Select a tag from second list as output

Selected output:

Data Preprocess

Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Correlation analysis
Mutual information analysis
Input selection

Import Data Delete Data

This provides different data preprocessing algorithms

Automatic selection of inputs

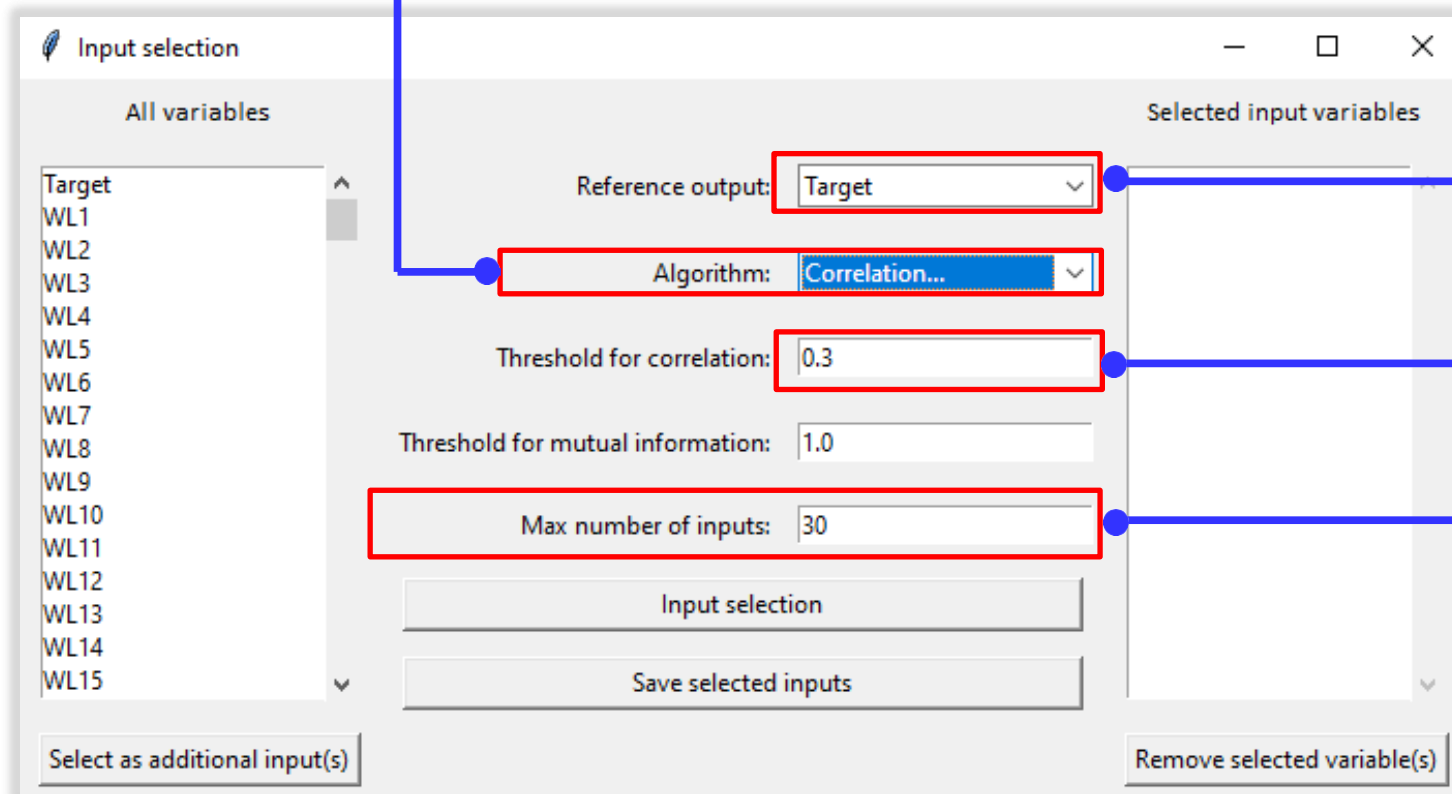
2

- Choose an algorithm for automatic input selection
- Go to “Algorithm”
 - Choose “Correlation analysis”

3

Adjust setting as needed

- “**Reference output**” is the output to be predicted
- “**Threshold for correlation**” is the lower bound for selecting inputs. This value should be non-negative, as absolute values of correlation coefficients are considered
- “**Max number of inputs**” is the upper bound on the number of inputs that can be used for modeling



Automatic selection of inputs

4

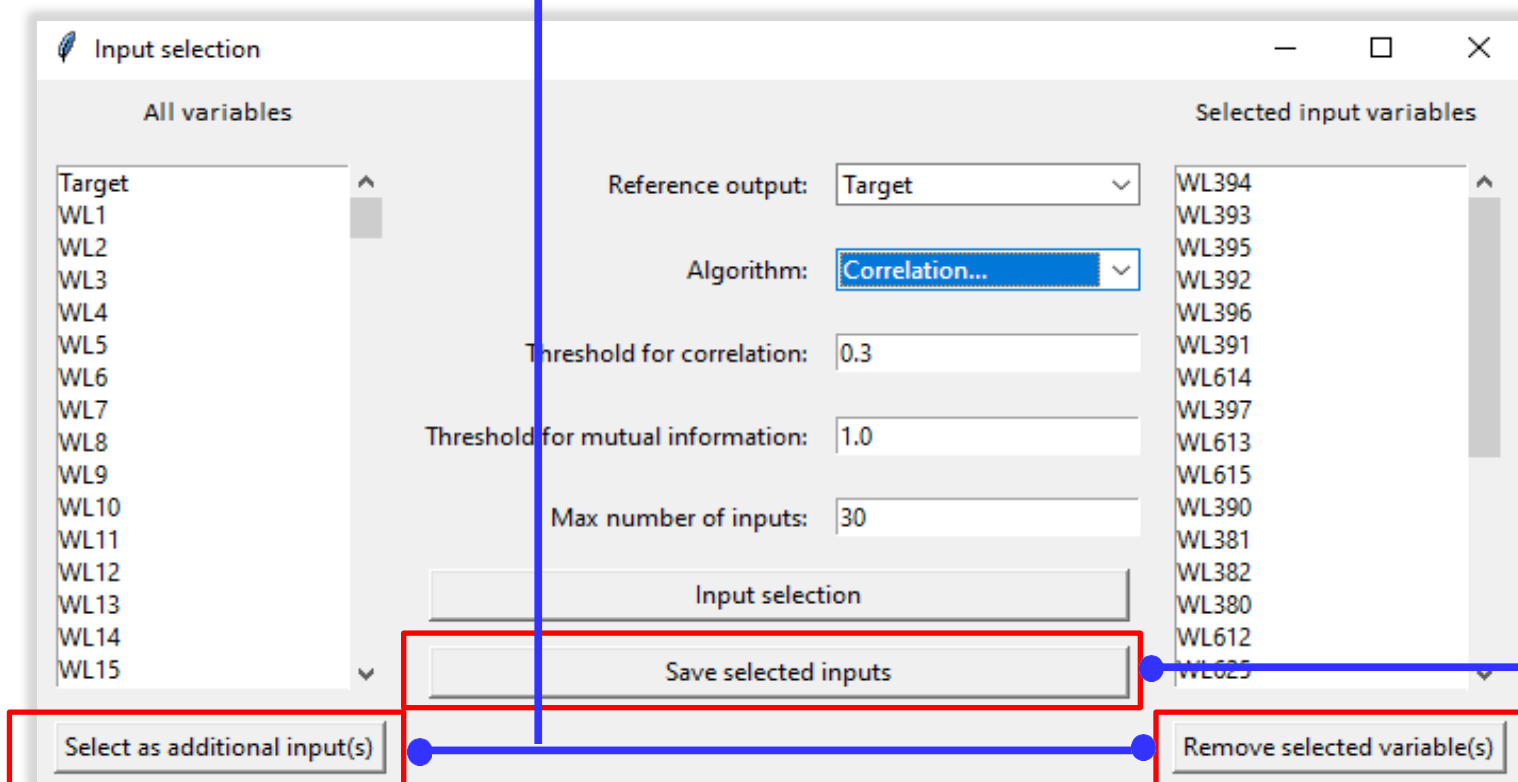
Make changes after automatic selection

- “**Select as additional input(s)**” can add more to the right list
- “**Remove selected variable(s)**” can remove variables from the right list

5

Save inputs selected automatically

- Click “**Save selected inputs**”
These inputs will be made available to modeling algorithms



Input selection

All variables

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15

Reference output: Target

Algorithm: Correlation...

Threshold for correlation: 0.3

Threshold for mutual information: 1.0

Max number of inputs: 30

Input selection

Save selected inputs

Selected input variables

WL394
WL393
WL395
WL392
WL396
WL391
WL614
WL397
WL613
WL615
WL390
WL381
WL382
WL380
WL612
WL625

Select as additional input(s)

Remove selected variable(s)

Soft Sensor Analytics



Modeling – ordinary least squares

- 1 Choose dataset and modeling method
- Click the “Modeling” tab
 - Click Data3_RM_RO_[1:549] (training data)
 - Go to “Modeling-->”, Click “Ordinary least squares...”, a window pops up as below

- 2 Selected inputs and output are loaded automatically
- If any change is necessary, the highlighted buttons can be used to add/remove variables

The screenshot shows the 'Modeling' tab of a software interface. The 'Dataset' list includes 'Data3', 'Data3_RM', 'Data3_RM_RO', 'Data3_RM_RO_[1:549]', and 'Data3_RM_RO_[550:1049]'. The 'Available variables' list includes 'Target' and 'WL1' through 'WL19'. The 'Working data' section shows 'Data3_RM_RO_[1:549]' selected. The 'Modeling' dropdown menu is open, showing 'Ordinary least squares...' selected. The 'Ordinary least Squares' dialog box is open, showing 'Available variables' (Target, WL1-WL13) and 'Input' (WL394-WL626). The 'Output' is set to '<<Target>>'. The 'Intercept' and 'Normalize' checkboxes are checked. The dialog box has buttons for 'Inputs', 'Output', and 'Remove selected variable(s)'. The 'Inputs' button is highlighted with a red box. The 'Output' button is highlighted with a red box. The 'Remove selected variable(s)' button is highlighted with a red box. The 'Save and close window' and 'Discard model and close window' buttons are at the bottom.

This provides different modeling algorithms

Soft Sensor Analytics



Modeling – ordinary least squares (select inputs manually)

4
Modeling
Click “-->Modeling” to perform regression

Available variables	Input
Target	WL394
WL1	WL395
WL2	WL614
WL3	WL393
WL4	WL396
WL5	WL392
WL6	WL615
WL7	WL625
WL8	WL382
WL9	WL613
WL10	WL381
WL11	WL397
WL12	WL383
WL13	WL626

3
Options
Tick “Intercept” or/and “Normalize”.

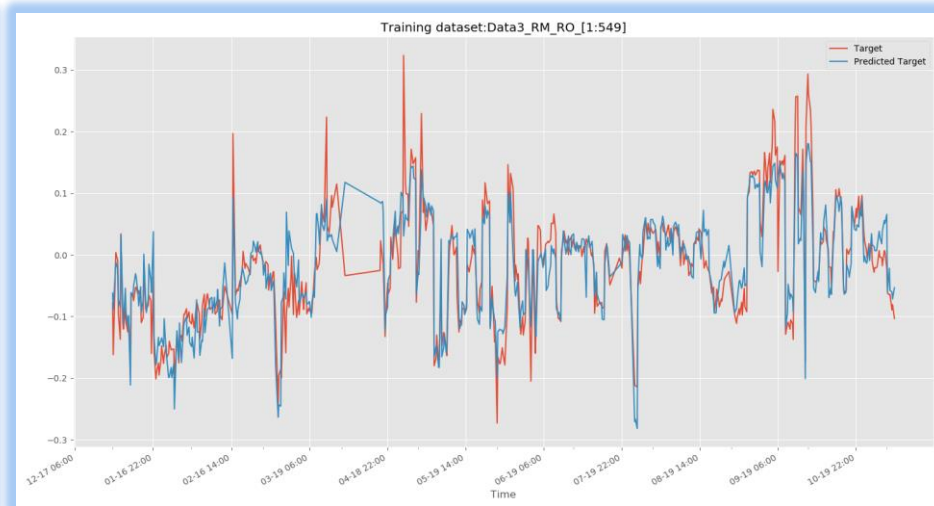
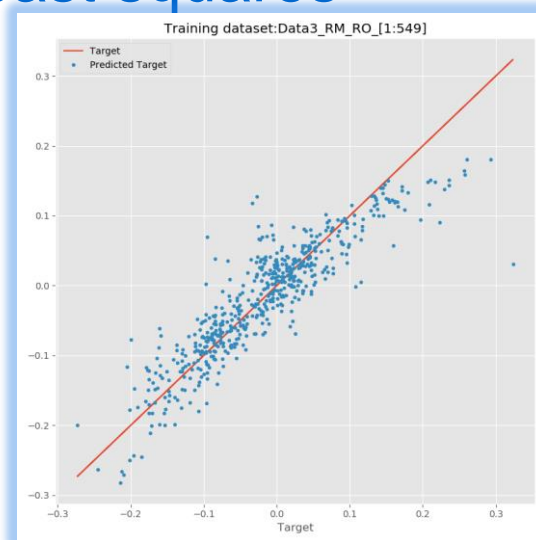
Soft Sensor Analytics



Results – ordinary least squares

Results

- Scatter plot and time trend are generated
- Model information is shown in “Detail” column



Ordinary least Squares

Available variables: Target, WL1, WL2, WL3, WL4, WL5, WL6, WL7, WL8, WL9, WL10, WL11, WL12, WL13

Input: WL394, WL395, WL614, WL393, WL396, WL392, WL615, WL625, WL382, WL613, WL381, WL397, WL383, WL626

Output: <<Target>>

Method: Ordinary least Squares

Structure: $y = ax + c$

Modeling: $x(0)$ WL394, $x(1)$ WL395, $x(2)$ WL614, $x(3)$ WL393, $x(4)$ WL396, $x(5)$ WL392, $x(6)$ WL615

Intercept: Normalize:

Buttons: Save and close window, Discard model and close window

5 Save model

- Close all figures
- Click “Save” to save this model for validation.

Soft Sensor Analytics



Modeling – Save current or Load previous set of inputs/output

7

If previous set of inputs/output is better, it can be recovered by clicking “**Load latest selected inputs & output**”

Then, modeling can be conducted again

Ordinary least Squares

Available variables

Input

Output: <<Target>>

-->Modeling

Intercept

Normalize

Method

Structure

Value

Ordinary least Squares

$y = ax + c$

x

x(0)

x(1)

x(2)

x(3)

x(4)

x(5)

x(6)

WL394

WL395

WL614

WL393

WL396

WL392

WL615

WL625

WL382

WL613

WL381

WL397

WL383

WL626

Inputs

Output

Remove selected variable(s)

Select as additional input(s)

Load latest selected inputs & output

Save selected inputs & output

Save and close window

Discard model and close window

6

If any changes are made to the inputs or/and output, and if the performance is good, the set of the inputs/output variables can be saved by clicking “**Save selected inputs & output**”.

Soft Sensor Analytics



Modeling – partial least squares

- 1 Choose dataset and modeling method
- Click the “Modeling” tab
 - Click Data3_RM_RO_[1:549] (training data)
 - Go to “Modeling-->”, Click “Partial least squares...”, a window pops up as below

- 2 Selected inputs and output are loaded automatically
- If any change is necessary, the highlighted buttons can be used to add/remove variables

The screenshot displays the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Modeling' tab is active, showing a 'Dataset' list with 'Data3_RM_RO_[1:549]' selected. The 'Available variables' list includes 'Target' and 'WL1' through 'WL19'. The 'Working data' section shows 'Data3_RM_RO_[1:549]' with options for 'Time Trend' and 'Histogram'. The 'Data Preprocess' section has a 'Customized Preprocess-->' dropdown and an 'Automatic Preprocessing' button. The 'Modeling' section has a 'Modeling-->' dropdown with 'Partial least squares...' selected. A 'Partial least squares regression' dialog box is open, showing 'Available variables' (Target, WL1-WL13) and 'Input' (WL394-WL626). The 'Output' field is empty. The dialog has buttons for 'Inputs', 'Output', and 'Remove selected variable(s)'. The 'Number of latent variables' is set to 5. The dialog also has buttons for 'Save and close window' and 'Discard model and close window'. A text box says 'This provides different modeling algorithms'.

Partial least squares regression

Available variables	Input
Target	WL394
WL1	WL395
WL2	WL614
WL3	WL393
WL4	WL396
WL5	WL392
WL6	WL615
WL7	WL625
WL8	WL382
WL9	WL613
WL10	WL381
WL11	WL397
WL12	WL383
WL13	WL626

Output: <<Target

--> Analysis

--> Modeling

Normalize

Number of latent variables: 5

Inputs Output Remove selected variable(s)

Select as additional input(s) Load latest selected inputs & output

Save selected inputs & output

Save and close window

Discard model and close window

This provides different modeling algorithms

Soft Sensor Analytics



Modeling – Partial least squares (PLS)

4

Analysis

Click “-->Analysis” to find a good number of latent variables

A figure will pop up as show in next slide

Partial least squares regression

Available variables	Input
Target	WL394
WL1	WL395
WL2	WL614
WL3	WL393
WL4	WL396
WL5	WL392
WL6	WL615
WL7	WL625
WL8	WL382
WL9	WL613
WL10	WL381
WL11	WL397
WL12	WL383
WL13	WL626

Output <<Target>>

--> Analysis

--> Modeling

Normalize

Number of latent variables:

5

Value

Save and close window

Discard model and close window

3

Options

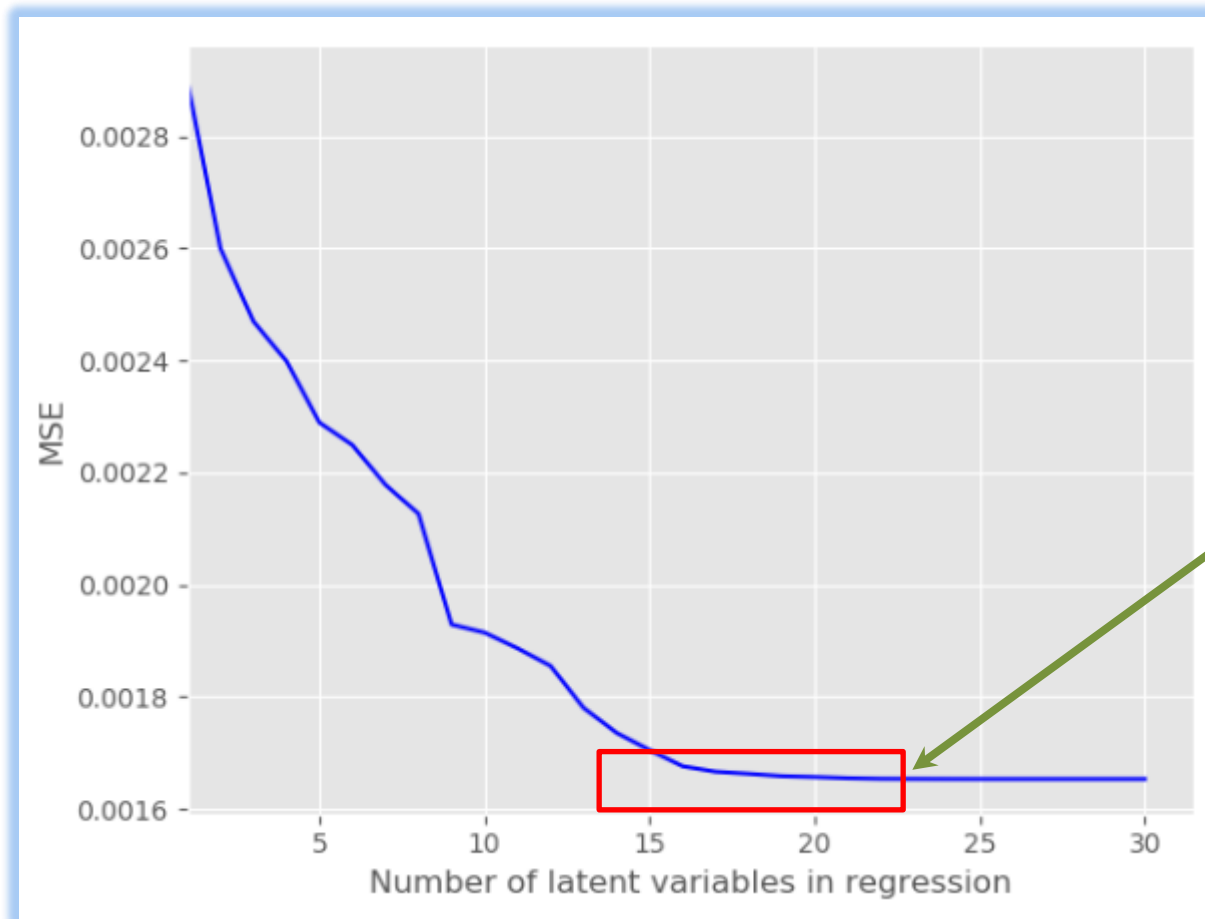
Tick “Normalize” if this is needed.

We **don't** Tick this time

Soft Sensor Analytics



Modeling – PLS (determine # of latent variables)



Choose number of latent variables

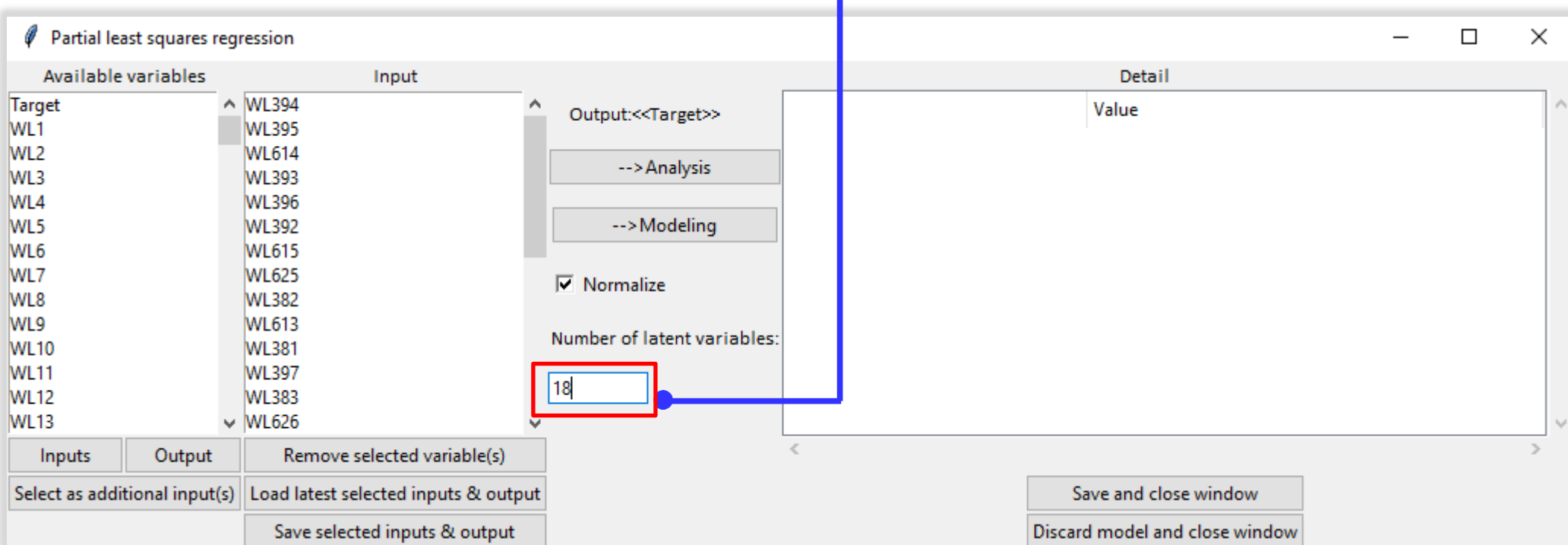
Number of latent variables is a key hyper-parameter of PLS. One optional analysis is provided to help you.

By clicking 'Analysis', the left figure pops up. Based on MSE values, a number between 16 and 20 is a good choice.

Modeling – PLS (determine # of latent variables)

5

- Determine # of latent variables
- Close all figures
- Enter “18” in the “number of latent variables” box



Partial least squares regression

Available variables

Input

Output: <<Target>>

--> Analysis

--> Modeling

Normalize

Number of latent variables:

18

Save and close window

Discard model and close window

Soft Sensor Analytics



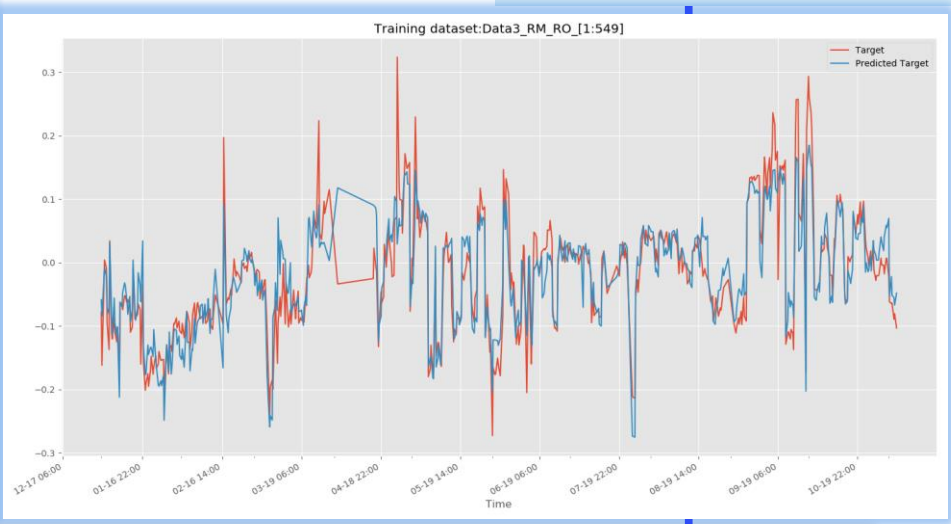
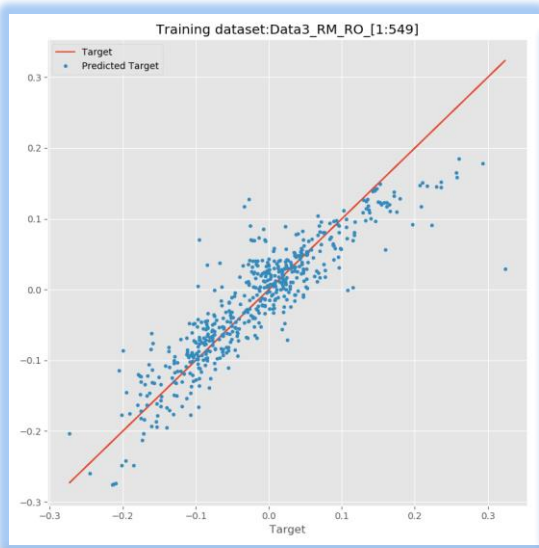
7

- Close figures
- Click “Save” to save this model

Results – PLS

6

- Regression**
Click “-->Modeling”
- scatter plot and time trend are generated
 - model information is shown in “Detail” column



Partial least squares regression

Available variables	Input	Output<<Target>>	Detail
Target	WL394		Method
WL1	WL395		Structure
WL2	WL614		y = ax
WL3	WL393		...
WL4	WL396		x(0) WL394
WL5	WL392		x(1) WL395
WL6	WL615		x(2) WL614
WL7	WL625		x(3) WL393
WL8	WL382		x(4) WL396
WL9	WL613		x(5) WL392
WL10	WL381		x(6) WL615
WL11	WL397		
WL12	WL383		
WL13	WL626		

Buttons: Inputs, Output, Remove selected variable(s), Select as additional input(s), Load latest selected inputs & output, Save selected inputs & output

Buttons: -->Analysis, -->Modeling, Normalized, Number of latent variables: 18

Buttons: Save and close window, Discard model and close window

Automatic modeling using Gaussian Process Regression

1

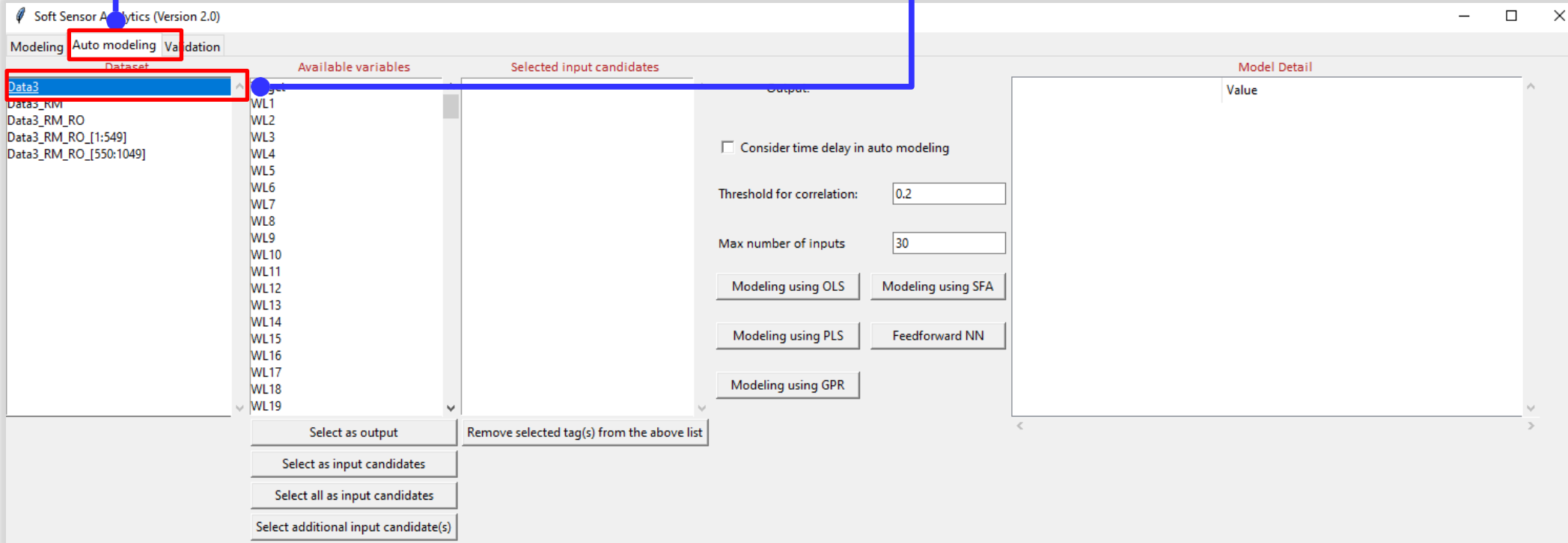
Automatic modeling

- Click the “Auto modeling” tab

2

Choose dataset to work on

- Choose original data “Data3” on the left list
The variables in this dataset will appear in the second list



Soft Sensor Analytics (Version 2.0)

Modeling **Auto modeling** Validation

Dataset: Data3

Available variables: WL1, WL2, WL3, WL4, WL5, WL6, WL7, WL8, WL9, WL10, WL11, WL12, WL13, WL14, WL15, WL16, WL17, WL18, WL19

Selected input candidates: (empty)

Modeling using OLS | Modeling using SFA

Modeling using PLS | Feedforward NN

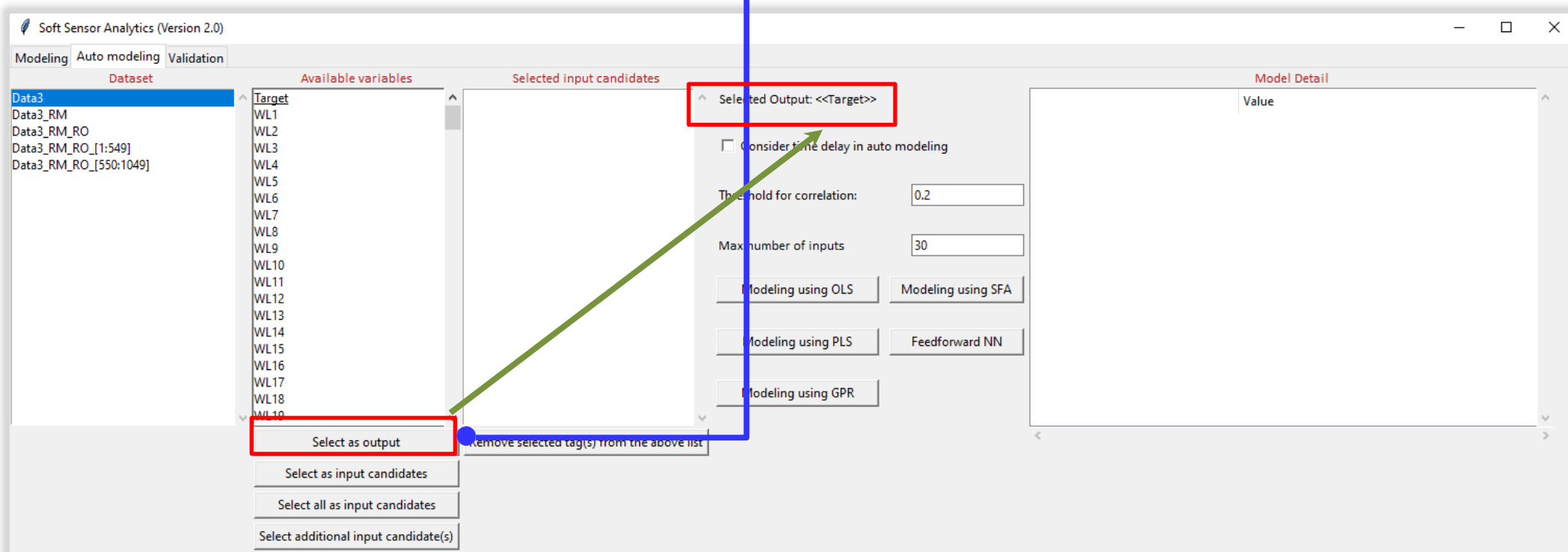
Modeling using GPR

Model Detail: Value

Automatic modeling – inputs/output selection

3

- Previously selected output is shown on this window
- Select another output for modelling if that needs to be changed



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Available variables

Selected input candidates

Selected Output: <<Target>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Model Detail

Value

Select as output

Select as input candidates

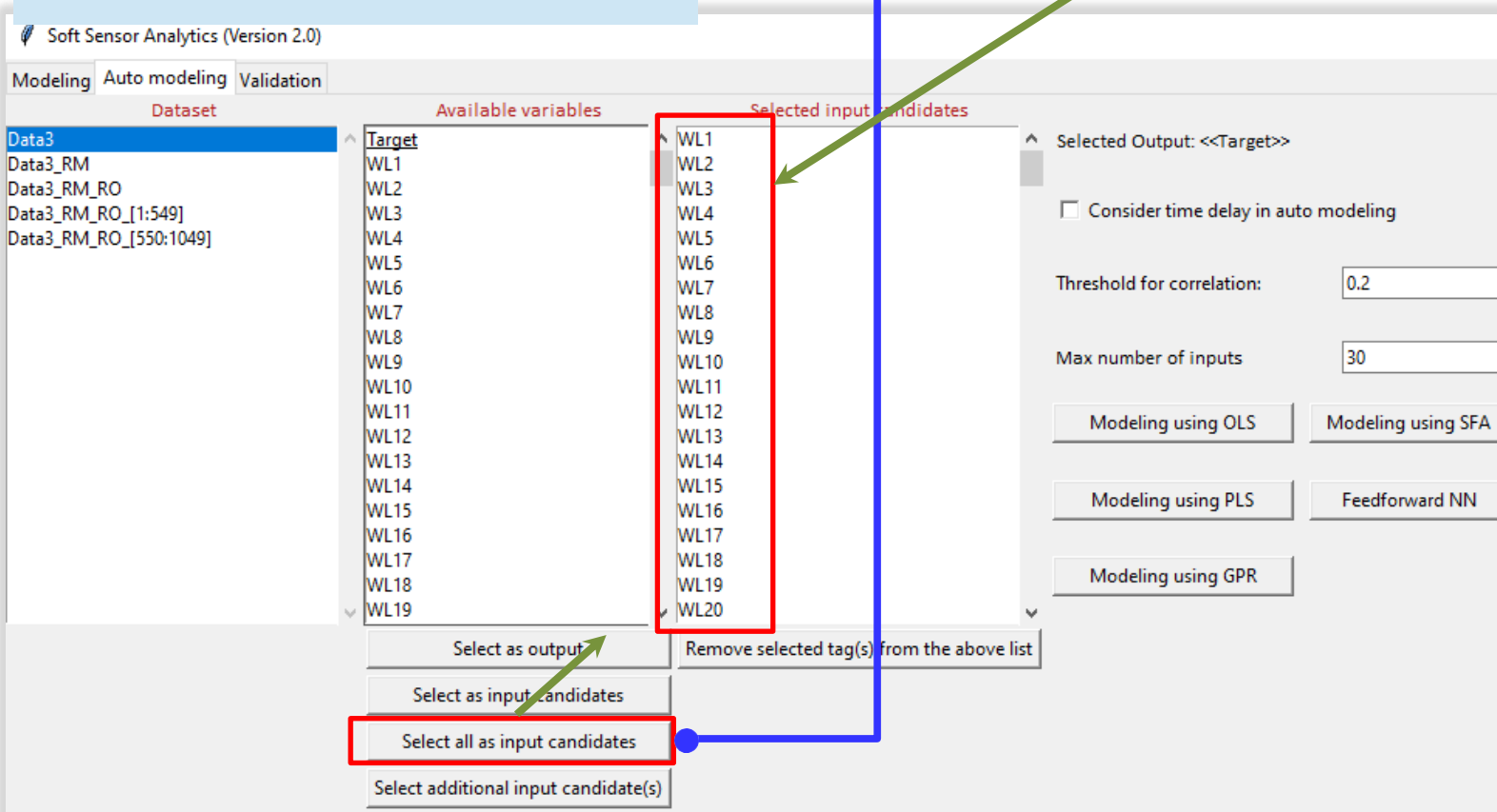
Select all as input candidates

Select additional input candidate(s)

Automatic modeling – inputs/output selection

- 4 Select inputs that may be used for modelling
- Click “Select all as input candidates” in the second list

The already selected output (**Target**) will not appear on this list when selecting all as inputs



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset

Available variables

Selected input candidates

Selected Output: <<Target>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output

Select as input candidates

Select all as input candidates

Select additional input candidate(s)

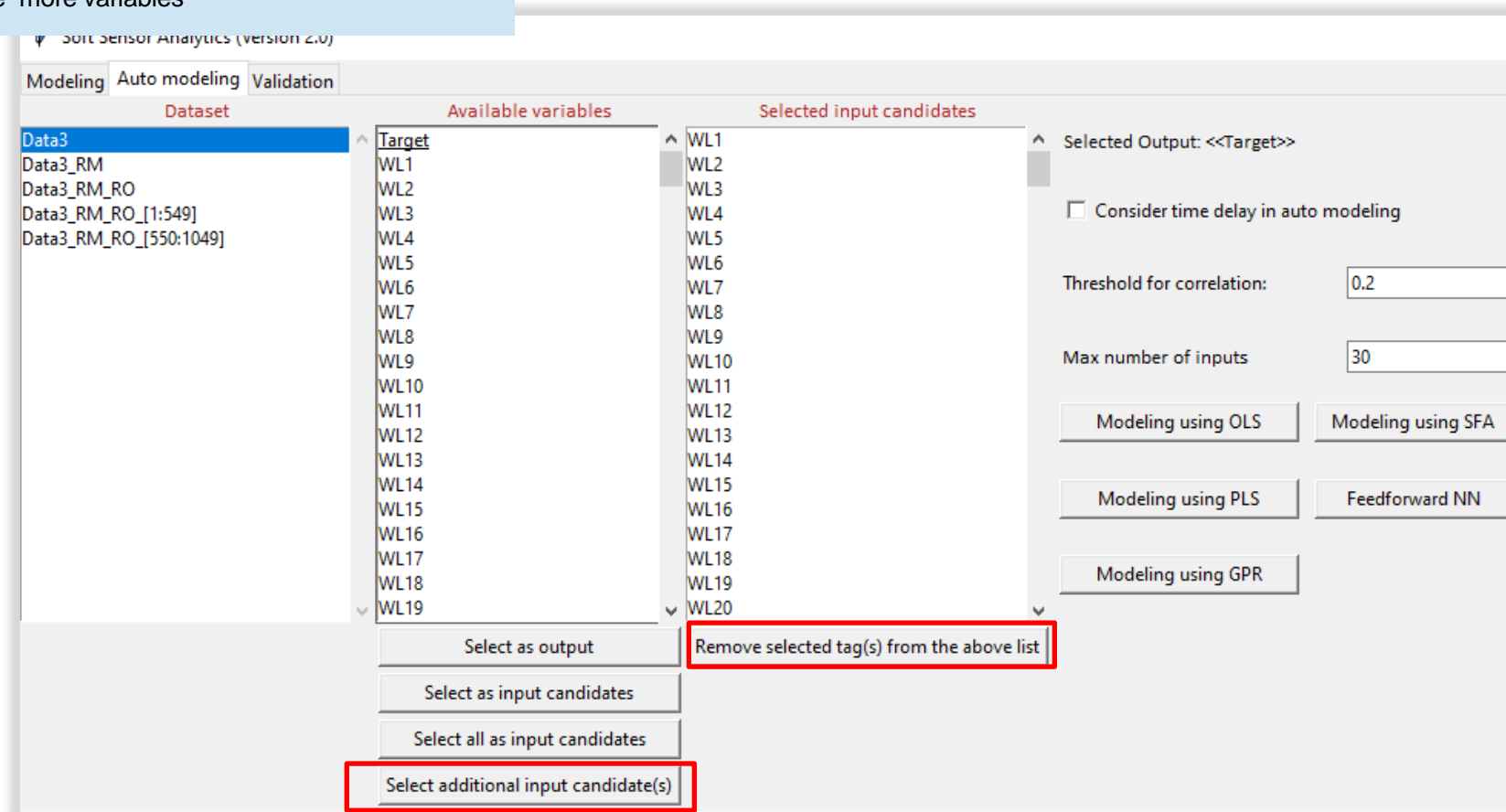
Remove selected tag(s) from the above list

Note: We can also use “Select as input candidates” button to select potential inputs from all variables based on available information/knowledge

Automatic modeling – inputs/output selection

The selected inputs can be modified

- Click **“Remove selected tag(s) from the above list”** to remove unnecessary variables
- Click **“Select additional input candidates”** to include more variables



Soft Sensor Analytics (version 2.0)

Modeling Auto modeling Validation

Dataset

Available variables

Selected input candidates

Selected Output: <<Target>>

Consider time delay in auto modeling

Threshold for correlation: 0.2

Max number of inputs: 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

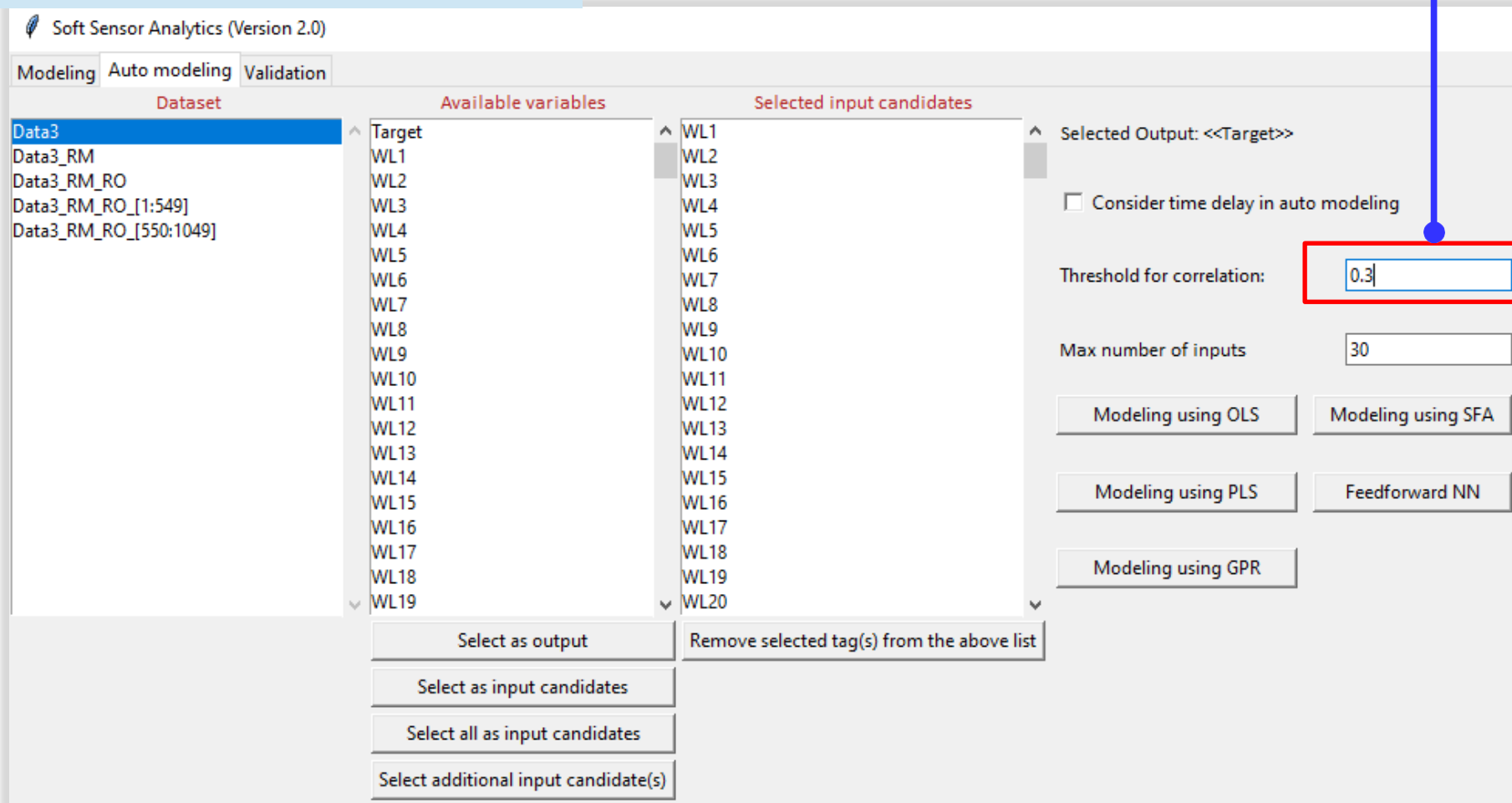
Select as input candidates

Select all as input candidates

Select additional input candidate(s)

Automatic modeling

- 5
- Set threshold on correlation coefficient
- This is a lower bound on correlation coefficients for selecting input variables
 - Set **“0.3”** as the threshold



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset Available variables Selected input candidates

Data3
Data3_RM
Data3_RM_RO
Data3_RM_RO_[1:549]
Data3_RM_RO_[550:1049]

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19

WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19
WL20

Selected Output: <<Target>>

Consider time delay in auto modeling

Threshold for correlation: 0.3

Max number of inputs 30

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

Select as input candidates

Select all as input candidates

Select additional input candidate(s)



Automatic modeling

- 6 Set max number of inputs
- Set "50" as the max number of inputs

- 7 One-click modeling using GPR
- Click "Modeling using GPR" to build a model

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset Available variables Selected input candidates

Data3
Data3_RM
Data3_RM_RO
Data3_RM_RO_[1:549]
Data3_RM_RO_[550:1049]

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19

WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19
WL20

Selected Output: <<Target>>

Consider time delay in auto modeling

Threshold for correlation: 0.3

Max number of inputs: 50

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output Remove selected tag(s) from the above list

Select as input candidates

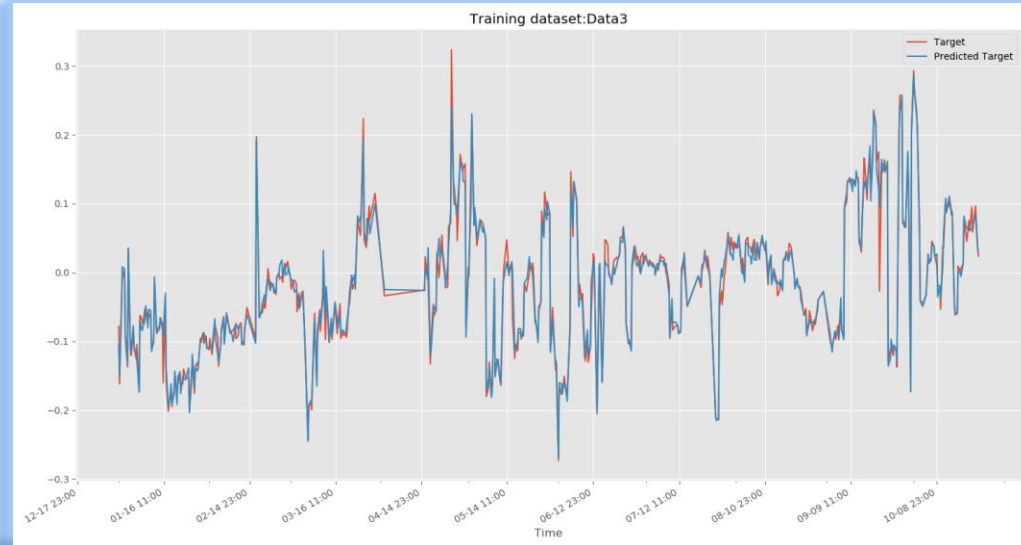
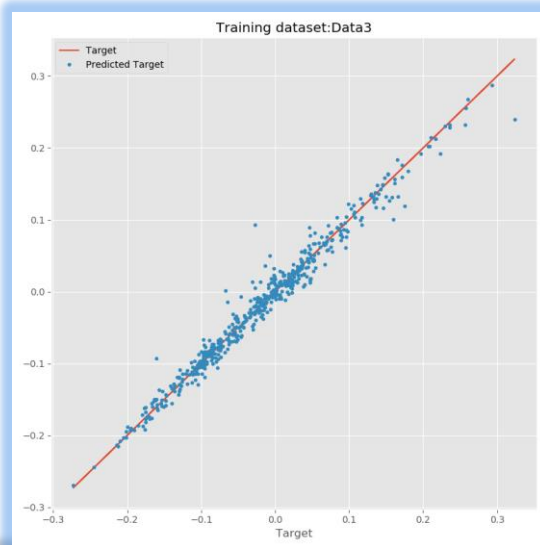
Select all as input candidates

Select additional input candidate(s)

Automatic modeling using GPR - Results

Results

- Scatter plot and time trend are generated
- Model information is shown in "Detail" column



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation

Dataset Available variables Selected input candidates

Data3
Data3_RM
Data3_RM_RO
Data3_RM_RO_[1:549]
Data3_RM_RO_[550:1049]
Data3_for_training
Data3_for_validation

Target
WL1
WL2
WL3
WL4
WL5
WL6
WL7
WL8
WL9
WL10
WL11
WL12
WL13
WL14
WL15
WL16
WL17
WL18
WL19
WL20

Selected Output: <<Target>>

Consider time delay in auto modeling

Threshold for correlation: 0.3

Max number of inputs: 50

Modeling using OLS Modeling using SFA

Modeling using PLS Feedforward NN

Modeling using GPR

Select as output
Select as input candidates
Select all as input candidates
Select additional input candidate(s)

Remove selected tag(s) from the above list

Model Detail

Method	Value
Gaussian process regression	
x	...
y	Target
alpha	0.0001
Performance	...

Soft Sensor Analytics



Automatic modeling - Results

The screenshot shows the "Soft Sensor Analytics" software interface. The "Dataset" list on the left includes "Data3", "Data3_RM", "Data3_RM_RO", "Data3_RM_RO [1:549]", "Data3_RM_RO [550:1049]", "Data3_for_training", and "Data3_for_validation". The "Data3_for_training" and "Data3_for_validation" entries are highlighted with a red box. A blue line connects this box to a text box at the bottom of the slide. The "Available variables" list includes "Target" and "WL1" through "WL19". The "Selected input candidates" list includes "WL1" through "WL20". The "Selected Output" is set to "<<Target>>". The "Threshold for correlation" is set to 0.3, and the "Max Number of inputs" is set to 50. The "Modeling" section includes buttons for "Modeling using OLS", "Modeling using SFA", "Modeling using PLS", "Feedforward NN", and "Modeling using GPR". The "Model Detail" section shows the "Method" as "Gaussian process regression" and the "Value" as "Target".

Two datasets are generated and saved

- The first half "Data3_for_training" is used for modeling
- The first half "Data3_for_validation" can be used for validation

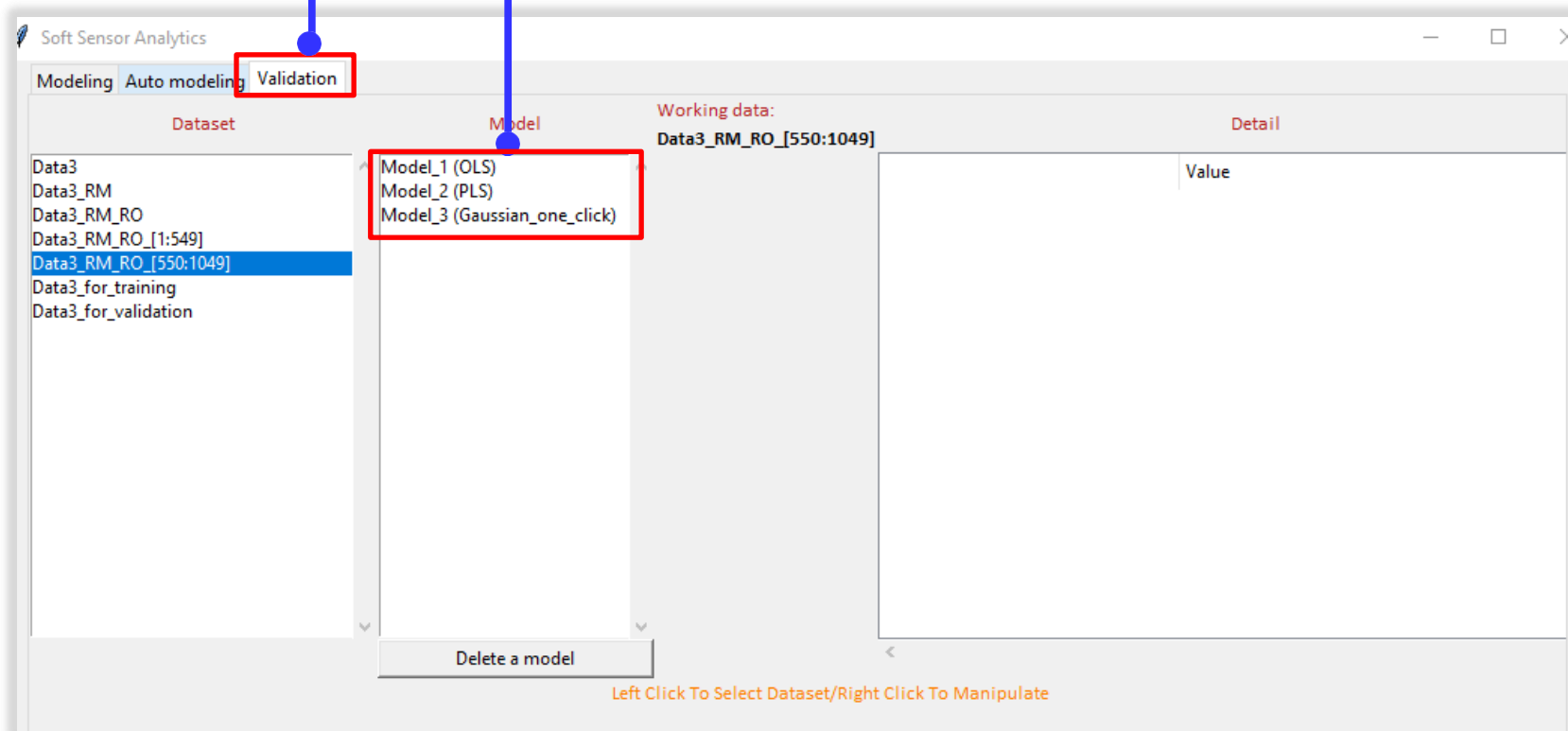
Validation of the models

Model Validation: the models based on different methods can be validated based on the second dataset.

1

Select models for validation

- Click “Validation” tab on the top
- Click sequentially to select the three models established for Data3



Soft Sensor Analytics

Modeling Auto modeling **Validation**

Dataset	Model	Working data:	Detail
Data3	Model_1 (OLS)	Data3_RM_RO_[550:1049]	Value
Data3_RM	Model_2 (PLS)		
Data3_RM_RO	Model_3 (Gaussian_one_click)		
Data3_RM_RO_[1:549]			
Data3_RM_RO_[550:1049]			
Data3_for_training			
Data3_for_validation			

Delete a model

Left Click To Select Dataset/Right Click To Manipulate

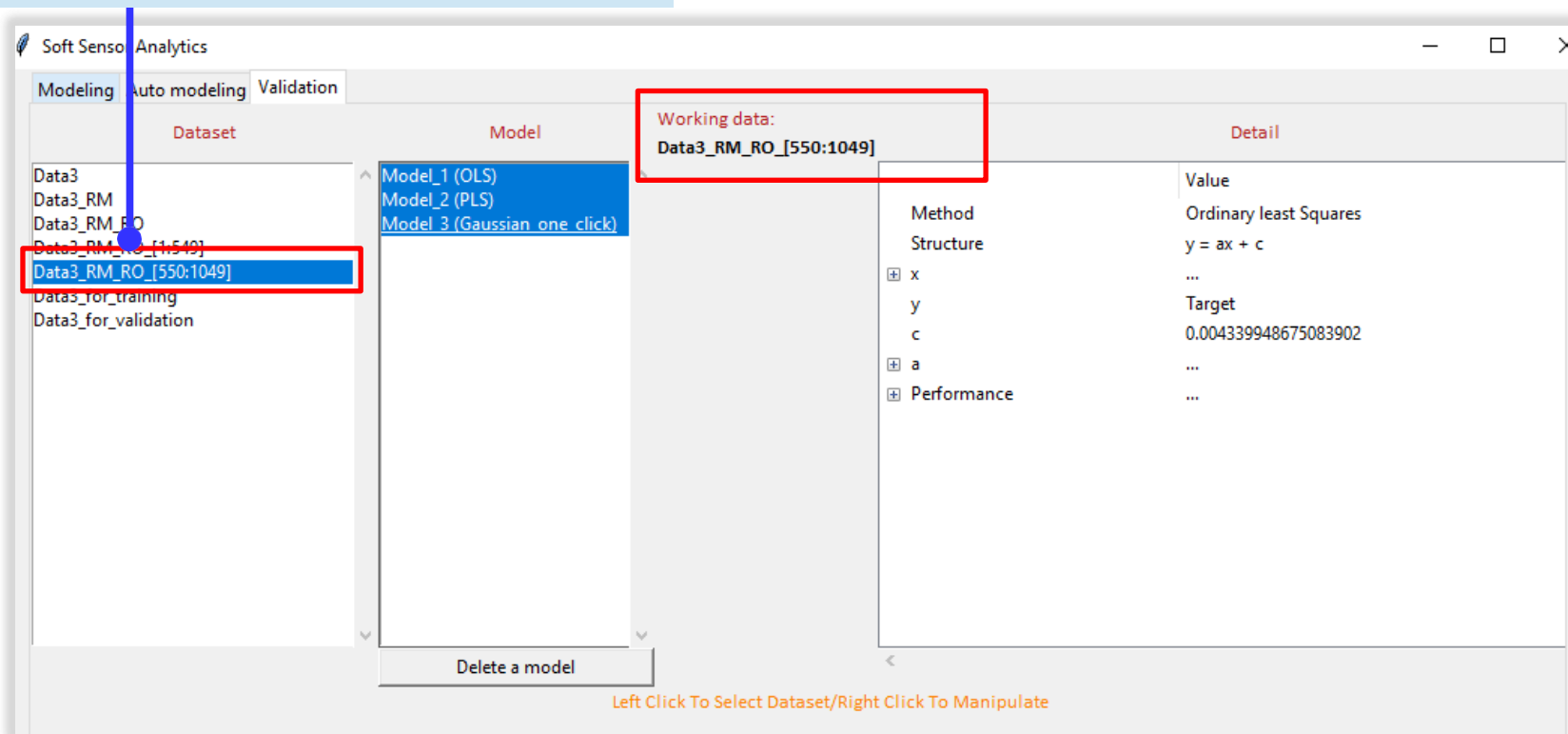
Validation of the models

Model Validation: the models based on different methods can be validated based on the second dataset.

2

Select a data set for validation

- Click “Data3_for_validation” for validation



The screenshot shows the 'Soft Sensor Analytics' application window with three tabs: 'Modeling', 'Auto modeling', and 'Validation'. The 'Validation' tab is active. The interface is divided into four main sections: 'Dataset', 'Model', 'Working data', and 'Detail'.

- Dataset:** A list of datasets including 'Data3', 'Data3_RM', 'Data3_RM_RO', 'Data3_RM_RO_[1:519]', 'Data3_RM_RO_[550:1049]', 'Data3_for_training', and 'Data3_for_validation'. A red box highlights 'Data3_RM_RO_[550:1049]', and a blue arrow points to it from the instruction box above.
- Model:** A list of models including 'Model_1 (OLS)', 'Model_2 (PLS)', and 'Model_3 (Gaussian one click)'. 'Model_1 (OLS)' is selected and highlighted in blue.
- Working data:** A red box highlights 'Data3_RM_RO_[550:1049]'.
- Detail:** A table showing model details for the selected model and dataset.

	Value
Method	Ordinary least Squares
Structure	$y = ax + c$
x	...
y	Target
c	0.004339948675083902
a	...
Performance	...

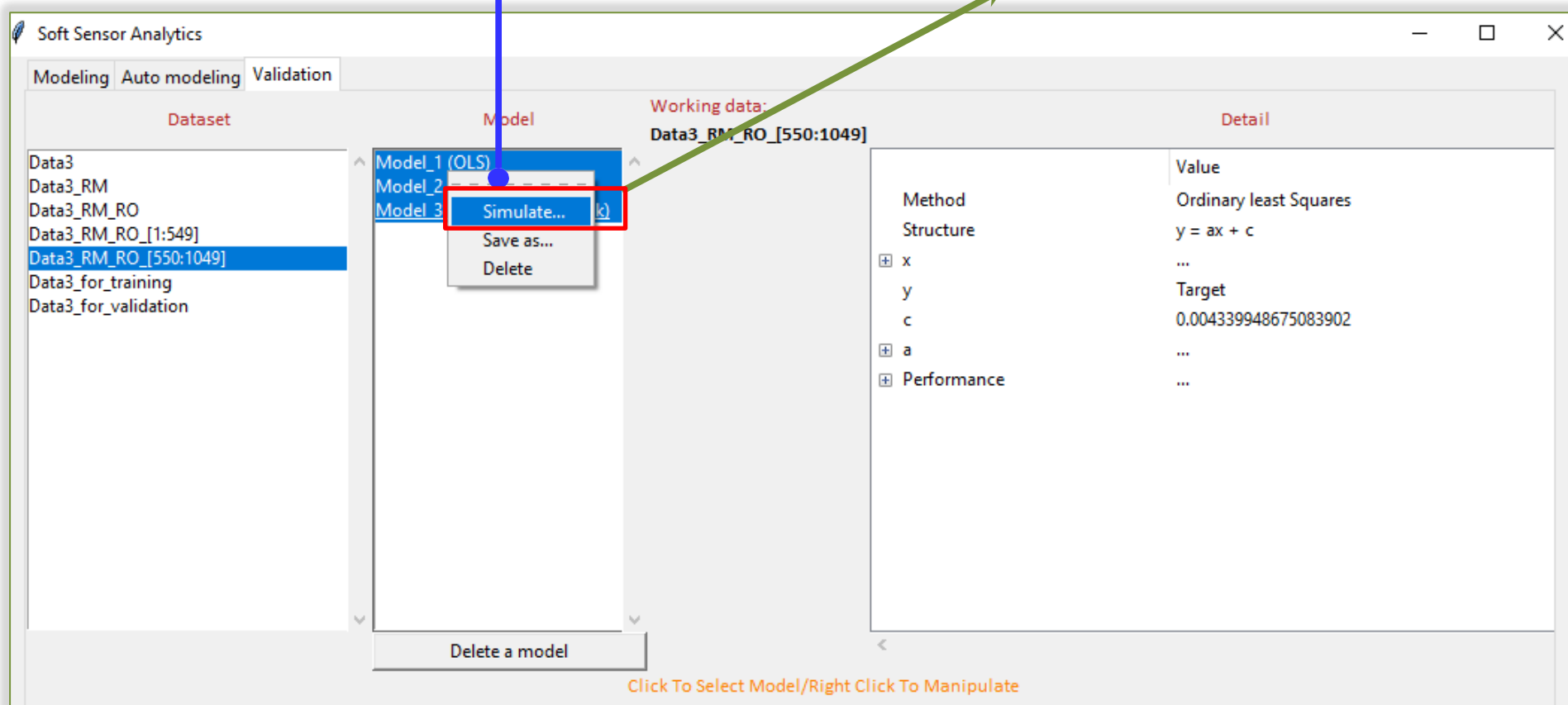
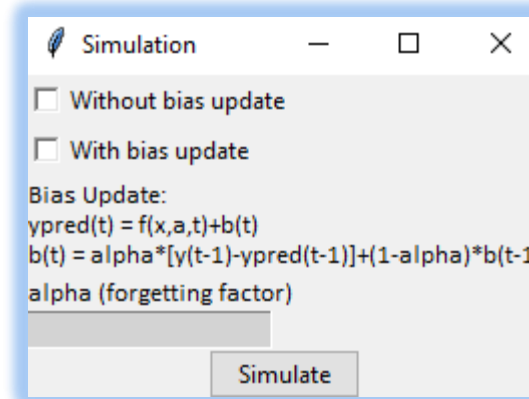
At the bottom of the 'Model' section, there is a 'Delete a model' button. At the bottom of the window, a note reads: 'Left Click To Select Dataset/Right Click To Manipulate'.

Validation of the models

3

Choose “Simulate” to continue

- **Right-Click** on the two models
- **Click “Simulate...”** button, a window will pop up



Soft Sensor Analytics

Modeling Auto modeling Validation

Dataset

Model

Working data: Data3_RM_RO_[550:1049]

Detail

Method	Value
Structure	Ordinary least Squares
$y = ax + c$	
x	...
y	Target
c	0.004339948675083902
a	...
Performance	...

Click To Select Model/Right Click To Manipulate

Soft Sensor Analytics

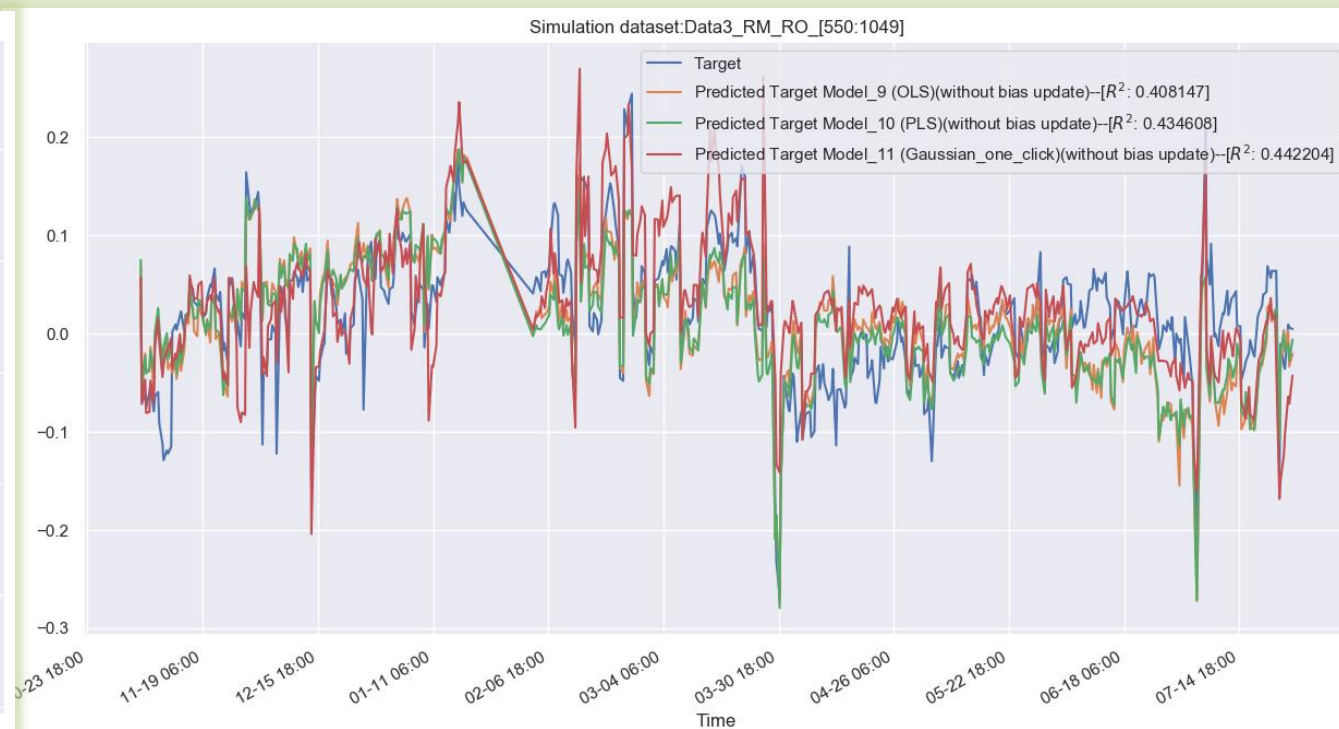
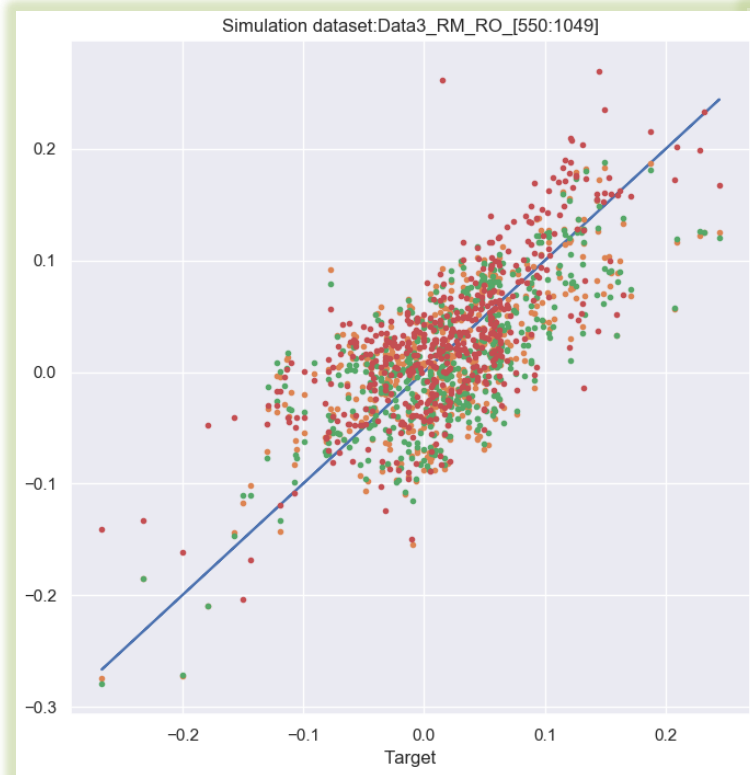
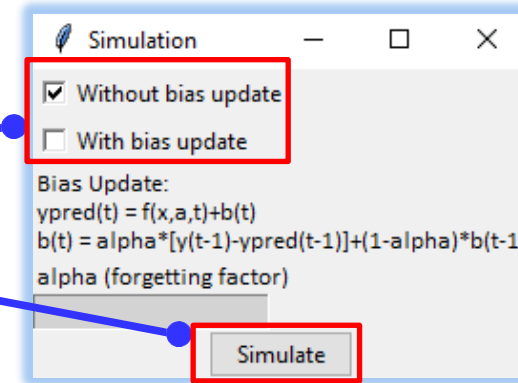


Validation of the models

4

Model validation

- Tick “Without bias update” box
- Click “Simulate” button



Soft Sensor Analytics

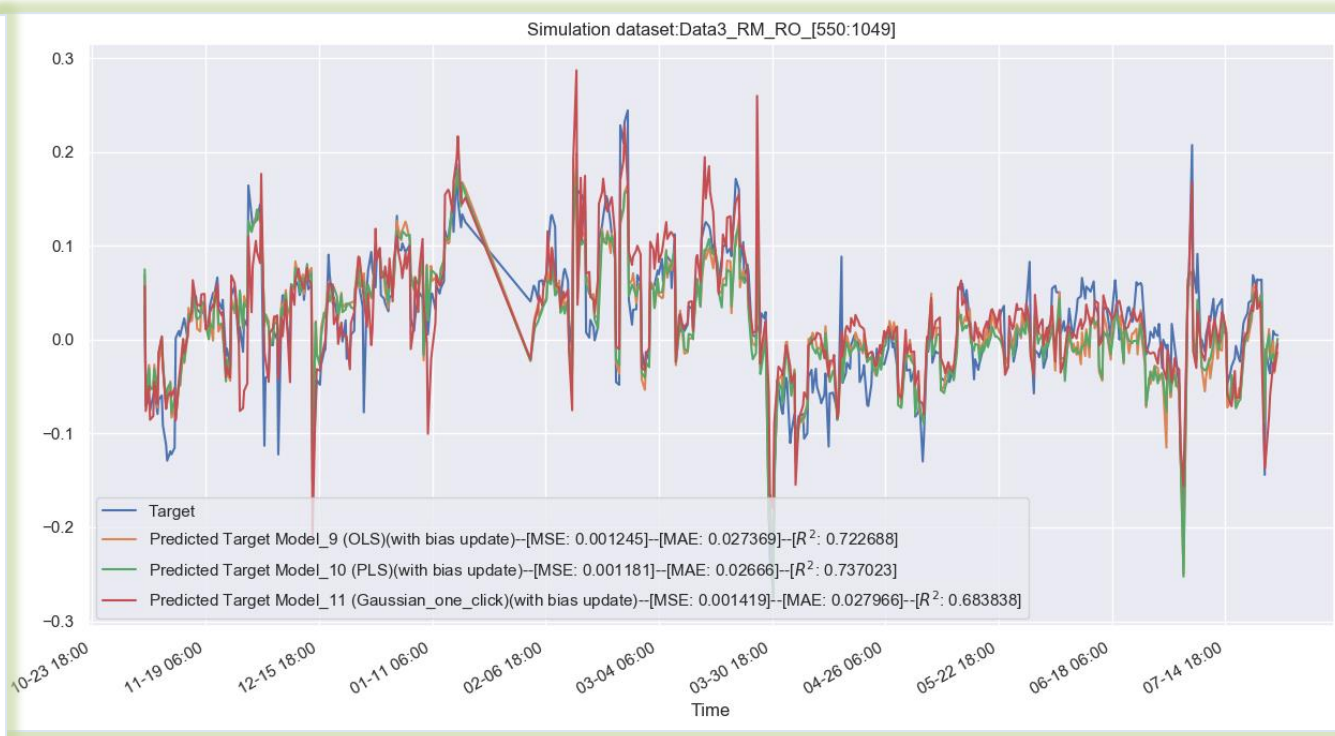
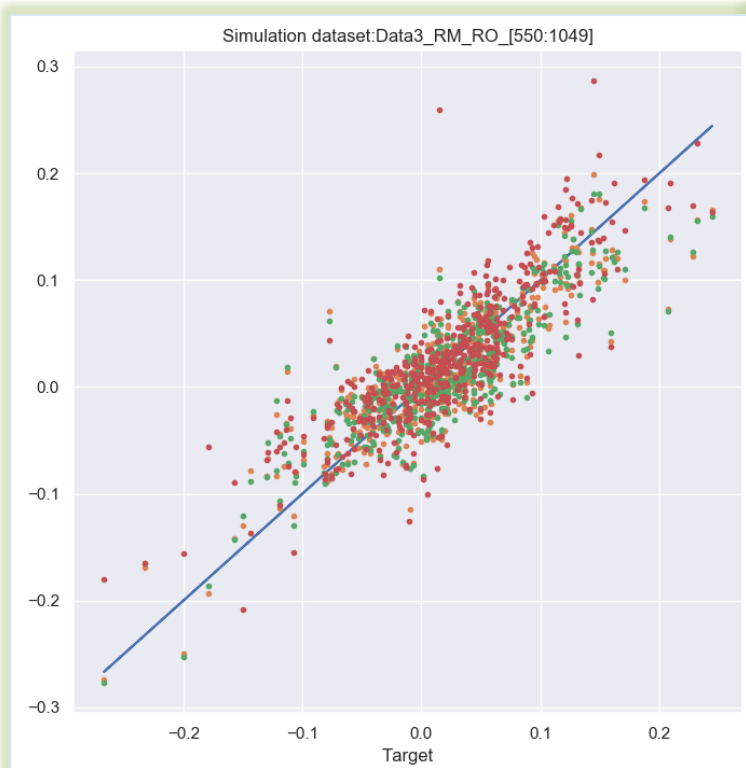
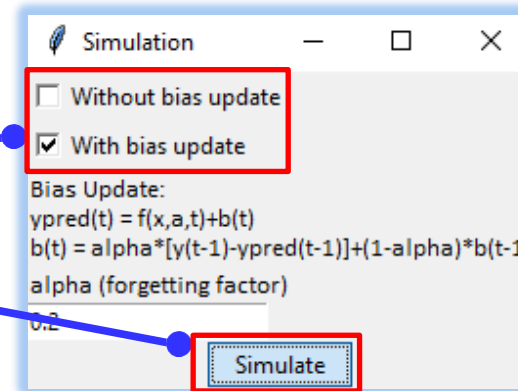


Validation of the models

5

Model validation

- Tick **“With bias update”** box
- Click **“Simulate”** button





Part III: Case study on Hot Lime Softener

Hot lime softener process



Hot Lime Softener

Hot lime softener is used to remove hardness (Ca, Mg), silica (SiO_2) and alkalinity before going through filtration and ion exchange process to meet feed water quality for once through steam generator.



Figure from www.tiwsteelplatework.ca

Objective



Hot Lime Softener Components Concentration Prediction

Objective: to predict the concentrations of **alkalinity** and **silica** based on inlet water quality, HLS operating data and chemical dosage.

Soft Sensor Analytics



Data preparation

Make sure the data are prepared following the guidelines below:

- Prepare the dataset using Excel
- the first column of a spreadsheet always contains timestamps
- the first row of a spreadsheet always contains the name of variables
- The data for each variable are recorded in each corresponding column
- For the created spreadsheet, select the entire timestamp column, right-click on this column and go to **“format cells”**, then go to **“Time”** or **“Custom”**, change the format of time information to a format similar to: **“mm/dd/yyyy h:mm”** or **“yyyy-mm-dd h:mm:ss”**.
- Save the dataset in **.xlsx** format

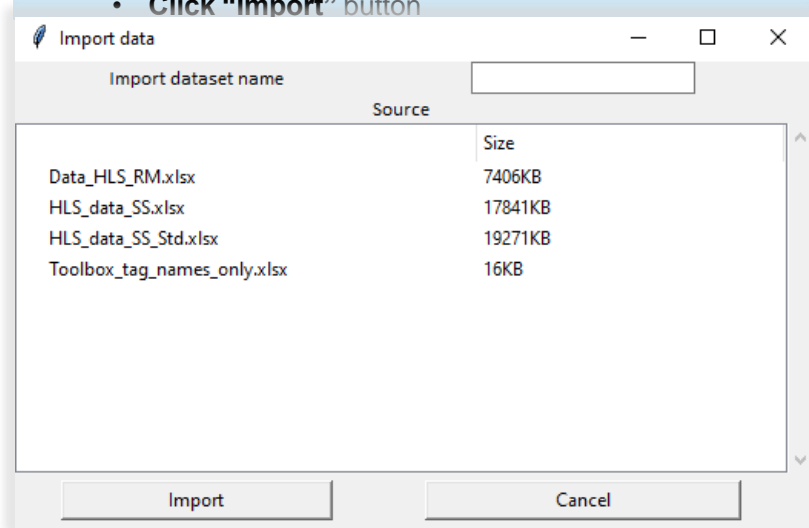
314	1/14/2016	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-1.469	0.65641	-0.1531	0.90432	-0.873
315	1/14/2016 1:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-1.6997	0.66893	-0.2204	0.88176	-0.8914
316	1/14/2016 2:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.1131	0.66263	-0.2204	0.88911	-0.9449
317	1/14/2016 3:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2294	0.67204	-2.0738	0.94238	-0.9986
318	1/14/2016 4:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2489	-0.3601	-2.0738	0.98012	-1.0233
319	1/14/2016 5:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2444	-0.3632	-2.0738	0.84389	-1.0195
320	1/14/2016 6:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.3782	-0.3757	-0.2261	0.86605	-1.0041
321	1/14/2016 7:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.3899	-0.8902	-0.226	0.72998	-1.0062
322	1/14/2016 8:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.3743	-0.8871	-2.0738	0.69228	-1.0582
323	1/14/2016 9:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.3912	-0.8871	-2.0738	0.76048	-1.1013
324	1/14/2016 10:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.3788	-0.3726	-2.0738	0.67712	-1.0576
325	1/14/2016 11:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2409	-0.3694	-2.0738	0.79852	-1.0393
326	1/14/2016 12:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2444	-0.3694	-2.0738	0.89692	-1.0302
327	1/14/2016 13:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2477	-0.3663	-2.0738	0.80923	-0.9735
328	1/14/2016 14:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.2483	-0.3695	-2.0738	0.82126	-1.013
329	1/14/2016 15:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.1141	-0.366	-2.0738	0.81599	-1.0103
330	1/14/2016 16:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.1083	-0.3726	-2.0738	0.81341	-0.983
331	1/14/2016 17:00	-0.1437	-0.2204	-1.3804	2.38931	-1.204	-0.7399	-1.2126	1.42912	-0.7477	-0.0923	0.06973	-1.1488	-0.2455	-0.8405	1.15259	-0.2197	-0.0529	1.03408	-2.1283	-0.3694	-2.0738	0.78357	-1.0134

Import data sets

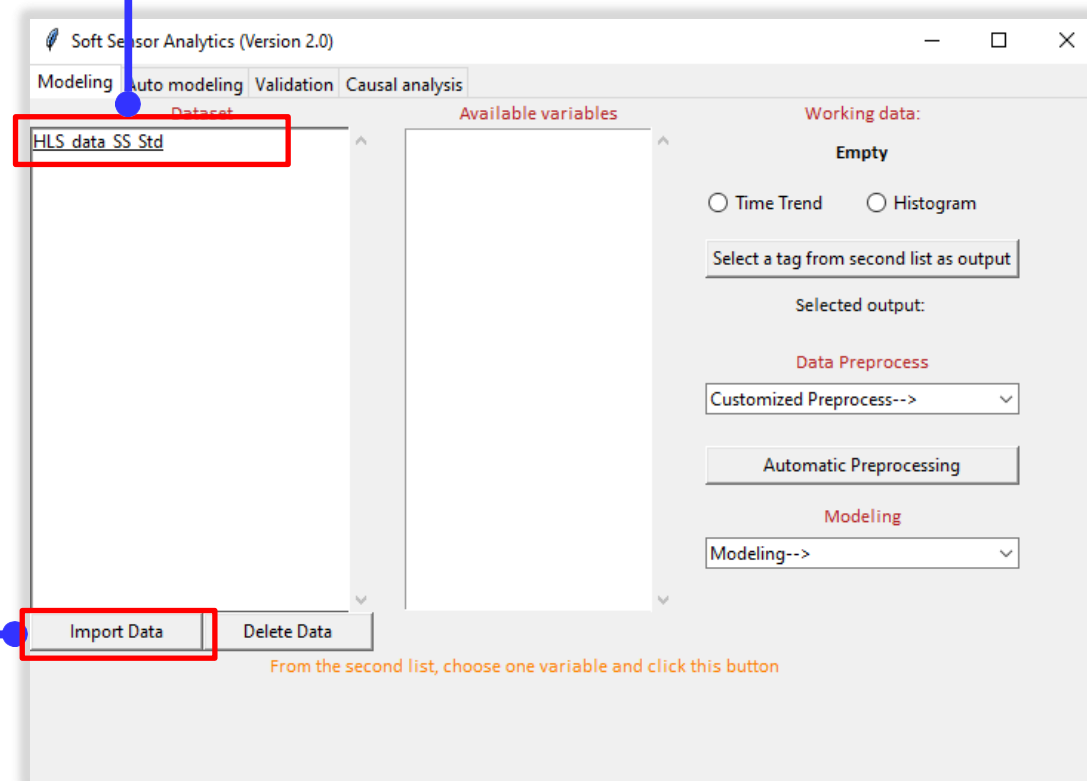
Note: Please save data sets in the “.xlsx” format

1

- Click “Import Data” button
- Select data file “HLS_data_SS_Std.xlsx”
- Click “Import” button



“HLS_data_SS_Std” has been imported



Soft Sensor Analytics



Current window with HLS data imported

- 1
- Click on the “Data” tab.
 - “HLS_data_SS_Std” has already been imported as highlighted in the right window“

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' application window. The 'Modeling' tab is selected, and the 'Data' sub-tab is active. The 'Dataset' list on the left contains one entry, 'HLS_data_SS_Std', which is highlighted with a red box. The 'Available variables' list on the right is empty. The 'Working data' section on the right is also empty. The 'Data Preprocess' section has a dropdown menu set to 'Customized Preprocess-->'. The 'Modeling' section has a dropdown menu set to 'Modeling-->'. At the bottom of the window, there are two buttons: 'Import Data' and 'Delete Data'. A blue arrow points from the 'Data' tab to the 'HLS_data_SS_Std' entry, and a green arrow points from the text in the first list item to the same entry. Below the 'Delete Data' button, there is a note: 'From the second list, choose one variable and click this button'.

Soft Sensor Analytics



Variables in the dataset

2

Choose Data3 to work on it

- Click on "HLS_data_SS_Std" from left workspace (i.e., column "Dataset")

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Dataset' column contains 'HLS_data_SS_Std'. The 'Available variables' column lists the following variables: TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, Super_Ca, Super_Mg, Super_Total, Super_Alk, Super_pH, PW_IN, and Super_IN. The right panel shows 'Working data: HLS_data_SS_Std' with radio buttons for 'Time Trend' and 'Histogram'. Below this is a text box 'Select a tag from second list as output' and 'Selected output:'. The 'Data Preprocess' section has a dropdown menu set to 'Customized Preprocess-->' and an 'Automatic Preprocessing' button. The 'Modeling' section has a dropdown menu set to 'Modeling-->'. At the bottom, there are 'Import Data' and 'Delete Data' buttons, and a red text label 'Delete Selected Dataset'.

The variables contained in the data set are shown in the right column. Scroll down to view all the variables.

Soft Sensor Analytics



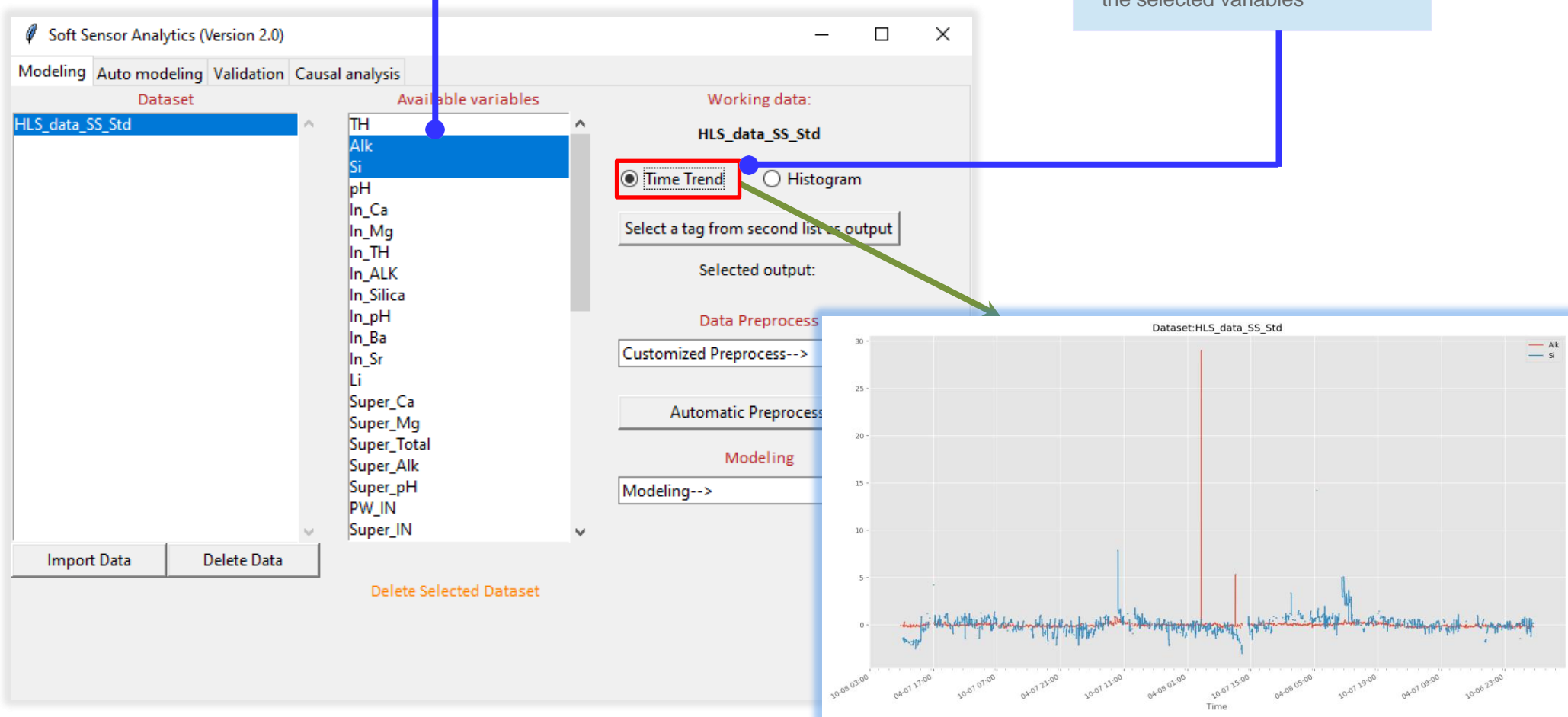
Visualize some of the variables

3

Click to select "ALK" and "Si" for visualization

4

Click "Time Trend" to visualize the selected variables



Soft Sensor Analytics



Visualize some of the variables

5

Click to select variables for visualization

6

Click "Time Trend" to visualize the selected variables

The screenshot displays the 'Soft Sensor Analytics (Version 2.0)' interface. On the left, the 'Dataset' list contains 'HLS_data_SS_Std'. The 'Available variables' list includes TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, Super_Ca, Super_Mg, Super_Total, Super_Alk, Super_pH, PW_IN, and Super_IN. A blue box highlights the 'Alk' and 'Si' variables. Below the list are 'Import Data' and 'Delete Data' buttons, and a 'Delete Selected Dataset' button at the bottom.

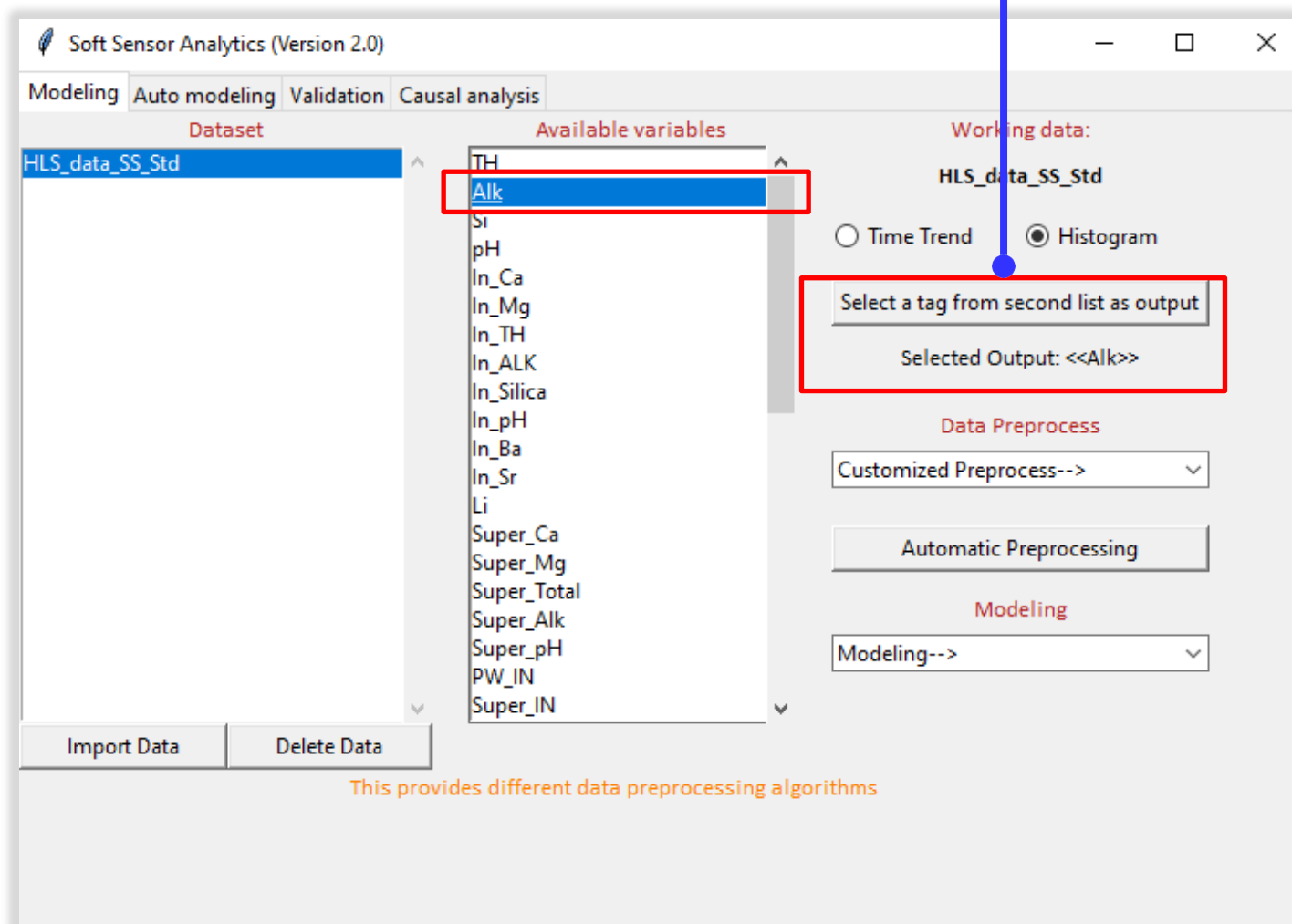
On the right, the 'Working data:' section shows 'HLS_data_SS_Std'. The 'Time Trend' radio button is unselected, and the 'Histogram' radio button is selected and highlighted with a red box. Below this is a text box 'Select a tag from second list as output' and a 'Selected output:' label. A green arrow points from the 'Histogram' button to a histogram plot.

The histogram plot, titled 'Dataset:HLS_data_SS_Std', shows the frequency distribution for 'Alk' (red bars) and 'Si' (blue bars). The x-axis represents the variable value (0 to 30), and the y-axis represents the frequency (0 to 20000). The 'Alk' distribution is highly peaked around 0, while the 'Si' distribution is broader and also centered near 0.

Select an output for modeling

7 Select output from second list

- Choose a variable “Alk” from the second list
- Click “Select a tag from second list as output” button to select this variable as the output for modeling



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset HLS_data_SS_Std

Available variables

- TH
- Alk
- Sr
- pH
- In_Ca
- In_Mg
- In_TH
- In_ALK
- In_Silica
- In_pH
- In_Ba
- In_Sr
- Li
- Super_Ca
- Super_Mg
- Super_Total
- Super_Alk
- Super_pH
- PW_IN
- Super_IN

Working data: HLS_data_SS_Std

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Alk>>

Data Preprocess

Customized Preprocess-->

Automatic Preprocessing

Modeling

Modeling-->

Import Data Delete Data

This provides different data preprocessing algorithms

Soft Sensor Analytics



Data preprocessing – Remove missing data

8

This step is to remove missing data from the dataset

- Select “HLS_data_SS_Std” from Dataset column
- Click “Customized Preprocess -->” on the right
- Click “Remove missing data”; new data is “Data3_RM”

The screenshot displays the 'Soft Sensor Analytics (Version 2.0)' application window. The 'Dataset' dropdown menu is highlighted with a red box and contains 'HLS_data_SS_Std' and 'HLS_data_SS_Std_RM'. The 'Available variables' list includes 'In_Ca', 'In_Mg', 'In_TH', 'In_ALK', 'In_Silica', 'In_pH', 'In_Ba', 'In_Sr', 'Li', 'Super_Ca', 'Super_Mg', 'Super_Total', 'Super_Alk', 'Super_pH', 'PW_IN', 'Super_IN', 'DB_INLET', 'Sludge_rec', 'Mix_LEVEL', and 'DEAR_LEVEL'. The 'Data Preprocess' dropdown menu is also highlighted with a red box and shows 'Remove missing data' selected. A green arrow points from the 'Remove missing data' option to the 'HLS_data_SS_Std' dataset selection. The 'Working data' section shows 'HLS_data_SS_Std' and 'Histogram' selected. The 'Selected Output' is '<<Alk>>'. The 'Data Preprocess' section lists various options: 'Customized Preprocess-->', 'Remove outlier', 'Remove tags of low variance', 'Filter', 'Normalize', 'Resample/Downsample', 'Select range', 'Select multiple ranges', and 'Correlation analysis'.

This provides different data preprocessing algorithms

Soft Sensor Analytics



Data preprocessing – Downsampling

9

The dataset contains both offline lab samples and online measurements. This step is to downsample the data.

- Select “HLS_data_SS_Std_RM” from Dataset column
- Click “Customized Preprocess -->” on the right
- Click “Resample/Downsample”; a window will pop up

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Dataset' column contains 'HLS_data_SS_Std' and 'HLS_data_SS_Std_RM', with the latter highlighted. The 'Available variables' list includes TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, Super_Ca, Super_Mg, Super_Total, Super_Alk, Super_pH, PW_IN, and Super_IN. The 'Data Preprocess' menu is open, with 'Resample/Downsample' selected. A 'Resample/Downsample' dialog box is also shown, with 'Choose to resample' selected, a frequency of 1 hour, and a 'mean' resample method. The 'Selected Output' is '<<Alk>>'. A green arrow points from the 'Resample/Downsample' option in the menu to the dialog box.

This provides different data preprocessing algorithms

Data preprocessing – Downsampling

10

- Select “Choose to manually downsample considering delay” radio-button
- Choose “Alk” as the reference variable (lab sample)
- Assign “Delayed samples between online and lab” to be “5” (This means that each lab sample is aligned with the 6th fast sample within a slow sampling period)

Resample/Downsample

Dataset: HLS_data_SS_Std_RM , Average sampling interval: 0 days 02:02:55.072900

Choose to resample

Please select resample frequency: 1

Please select resample method: [dropdown]

Choose to downsample

Downsample reference variable: [dropdown]

Percentage of interval for averaging (*100%): 0.15

Direction for averaging: Backward...

Choose to manually downsample considering delay

Downsample reference variable: Alk

Delayed samples between online and lab (counting forwards): 5

Percentage of interval for averaging (*100%): 0.15

Choose to automatically downsample considering delay

Downsample reference variable: [dropdown]

Percentage of interval for averaging (*100%): 0.15

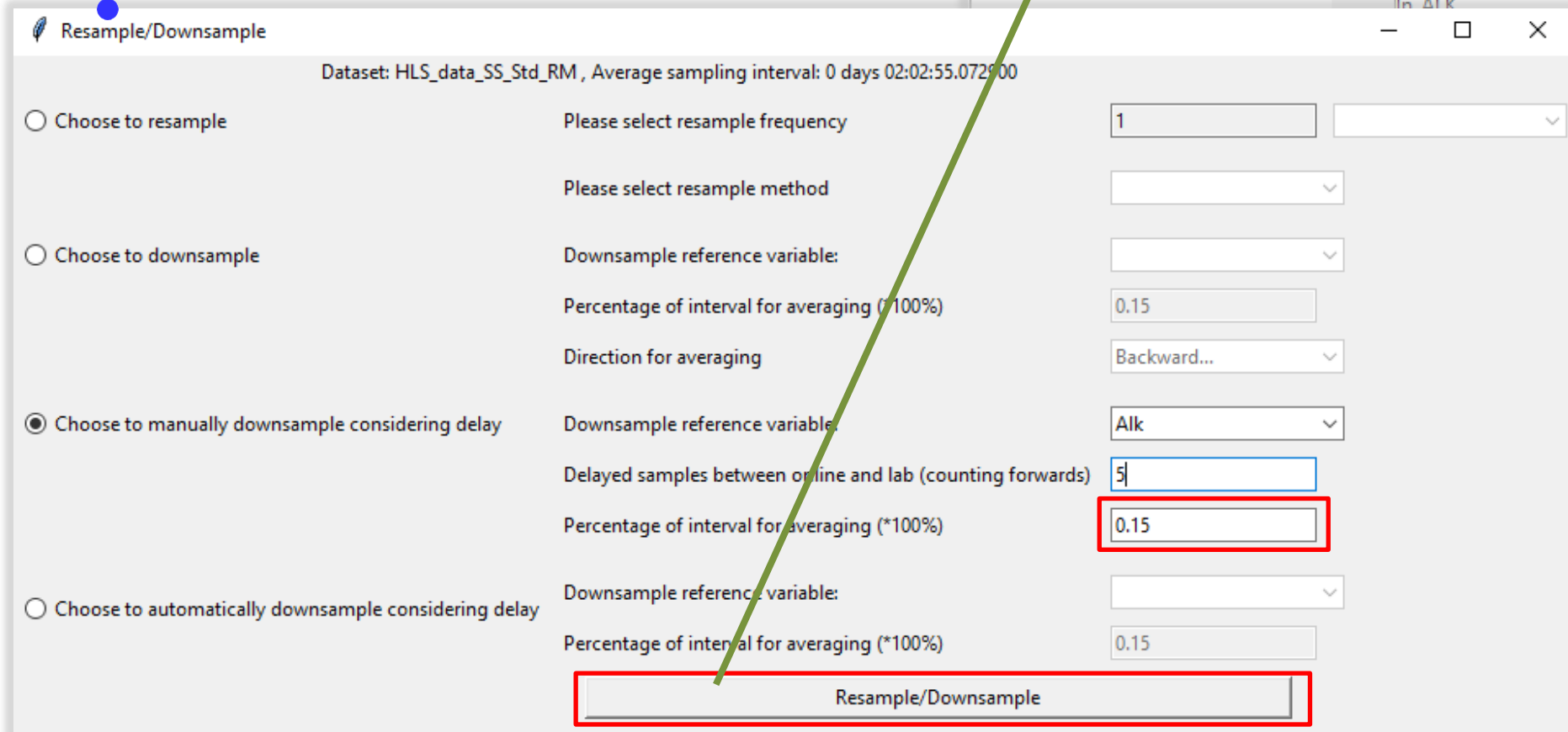
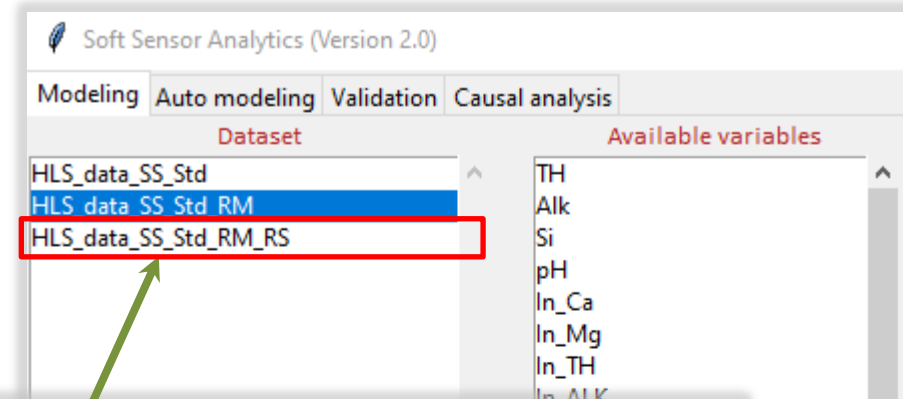
Resample/Downsample

Soft Sensor Analytics



Data preprocessing – Downsampling

- 11
- Assign “Percentage of interval for averaging” to be “0.15”
 - Click “Resample/Downsample” button, then a new dataset “HLS_data_SS_Std_RM_RS” is generated and shown under the “Modeling” tab



Soft Sensor Analytics



Data preprocessing – Remove outliers

12

This step is to remove outliers in the dataset

- Select “HLS_data_SS_Std_RM_RS” from Dataset column
- Click “Remove outlier“ on the right; then a new dataset “HLS_data_SS_Std_RM_RS_RO” will be generated and shown as in the screen capture below

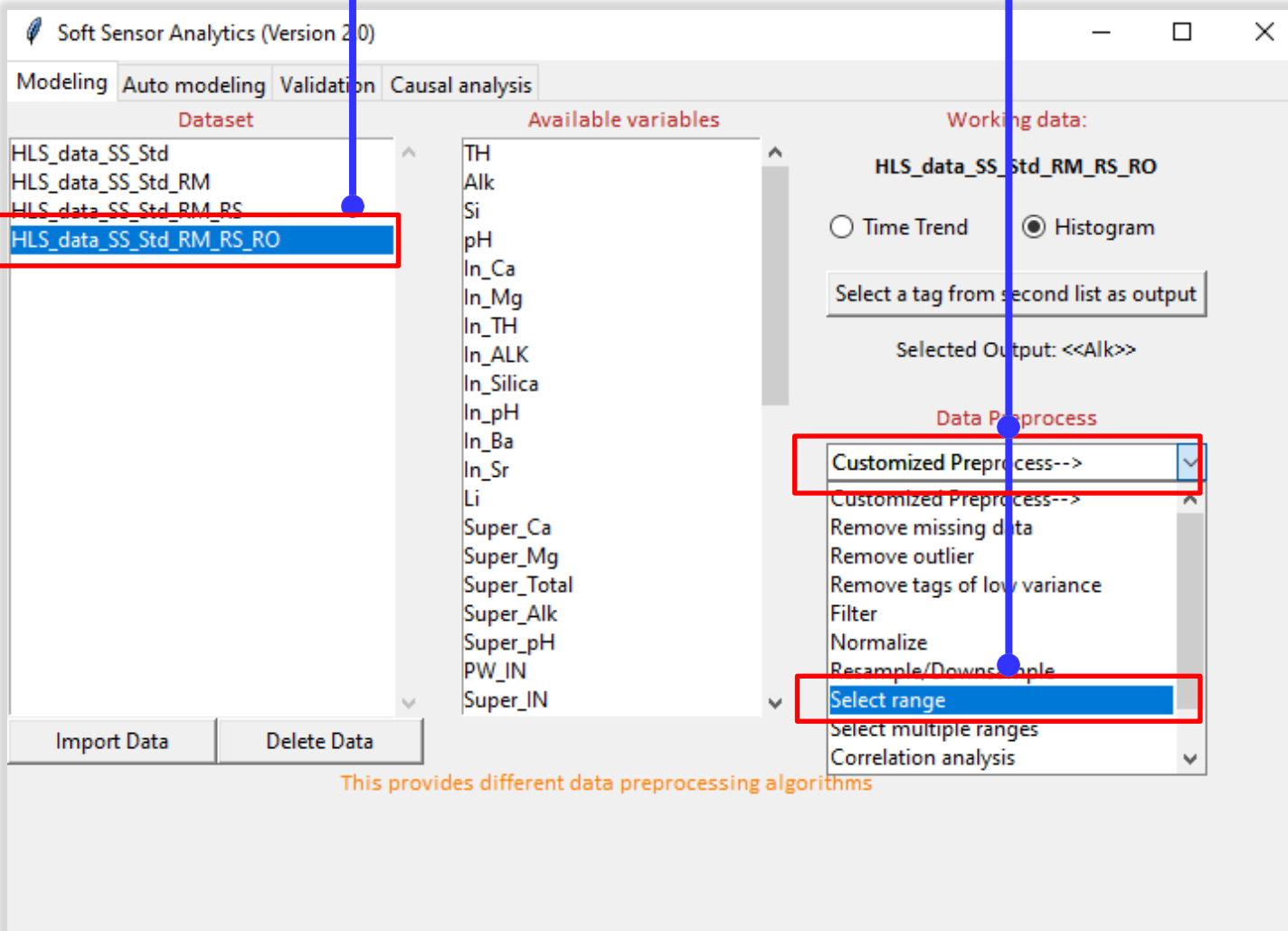
The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. The 'Dataset' column on the left contains a list of datasets, with 'HLS_data_SS_Std_RM_RS' selected and 'HLS_data_SS_Std_RM_RS_RO' highlighted with a red box. The 'Available variables' column in the center lists various chemical and physical properties. The 'Working data' section on the right shows 'HLS_data_SS_Std_RM_RS' and a 'Histogram' selected. The 'Data Preprocess' section on the right contains a list of preprocessing algorithms, with 'Remove outlier' highlighted in blue and enclosed in a red box. A blue arrow points from the 'Remove outlier' option to the 'HLS_data_SS_Std_RM_RS' dataset, and a green arrow points from the 'Remove outlier' option to the 'HLS_data_SS_Std_RM_RS_RO' dataset. At the bottom, a caption reads: 'This provides different data preprocessing algorithms'.

Data preprocessing – Split data into two parts for training and validation

13

Get a smaller dataset for training

- Select “HLS_data_SS_Std_RM_RS_RO”
- Click “Customized Preprocess-->”, then Click “Select range”; a window will pop up, as shown in the next slide



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

HLS_data_SS_Std
HLS_data_SS_Std_RM
HLS_data_SS_Std_RM_RS
HLS_data_SS_Std_RM_RS_RO

Available variables

TH
Alk
Si
pH
In_Ca
In_Mg
In_TH
In_ALK
In_Silica
In_pH
In_Ba
In_Sr
Li
Super_Ca
Super_Mg
Super_Total
Super_Alk
Super_pH
PW_IN
Super_IN

Working data:

HLS_data_SS_Std_RM_RS_RO

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Alk>>

Data Preprocess

Customized Preprocess-->
Customized Preprocess-->
Remove missing data
Remove outlier
Remove tags of low variance
Filter
Normalize
Resample/Downsample
Select range
Select multiple ranges
Correlation analysis

Import Data Delete Data

This provides different data preprocessing algorithms

Data preprocessing – Split data into smaller sets

14

To split data enter the range of dataset for training, do the following:

- Click “Alk” on the left column
- In “Start”, Enter “1”
- In “End”; Enter “300”
- Click “Insert”

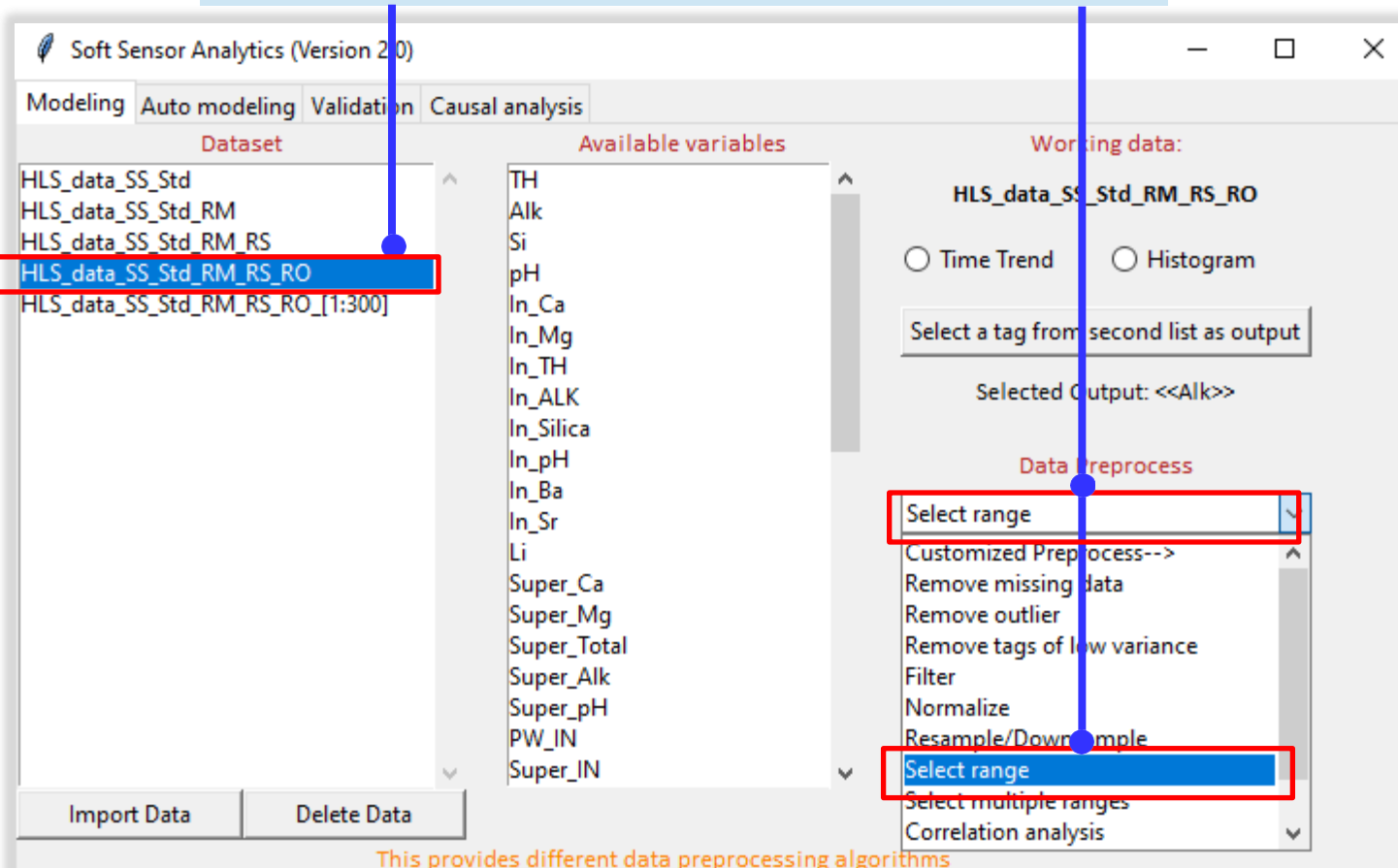


Data preprocessing – Divide data into two parts for training and validation

15

Get a smaller dataset for validation

- Select “HLS_data_SS_Std_RM_RS_RO” again
- Click “Customized Preprocess-->”, then Click “Select range”; a window will pop up, as shown in the next slide



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

- HLS_data_SS_Std
- HLS_data_SS_Std_RM
- HLS_data_SS_Std_RM_RS
- HLS_data_SS_Std_RM_RS_RO**
- HLS_data_SS_Std_RM_RS_RO_[1:300]

Available variables

- TH
- Alk
- Si
- pH
- In_Ca
- In_Mg
- In_TH
- In_ALK
- In_Silica
- In_pH
- In_Ba
- In_Sr
- Li
- Super_Ca
- Super_Mg
- Super_Total
- Super_Alk
- Super_pH
- PW_IN
- Super_IN

Working data:

HLS_data_SS_Std_RM_RS_RO

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Alk>>

Data preprocess

- Select range**
- Customized Preprocess-->
- Remove missing data
- Remove outlier
- Remove tags of low variance
- Filter
- Normalize
- Resample/Down sample
- Select range**
- Select multiple ranges
- Correlation analysis

Import Data Delete Data

This provides different data preprocessing algorithms

Soft Sensor Analytics

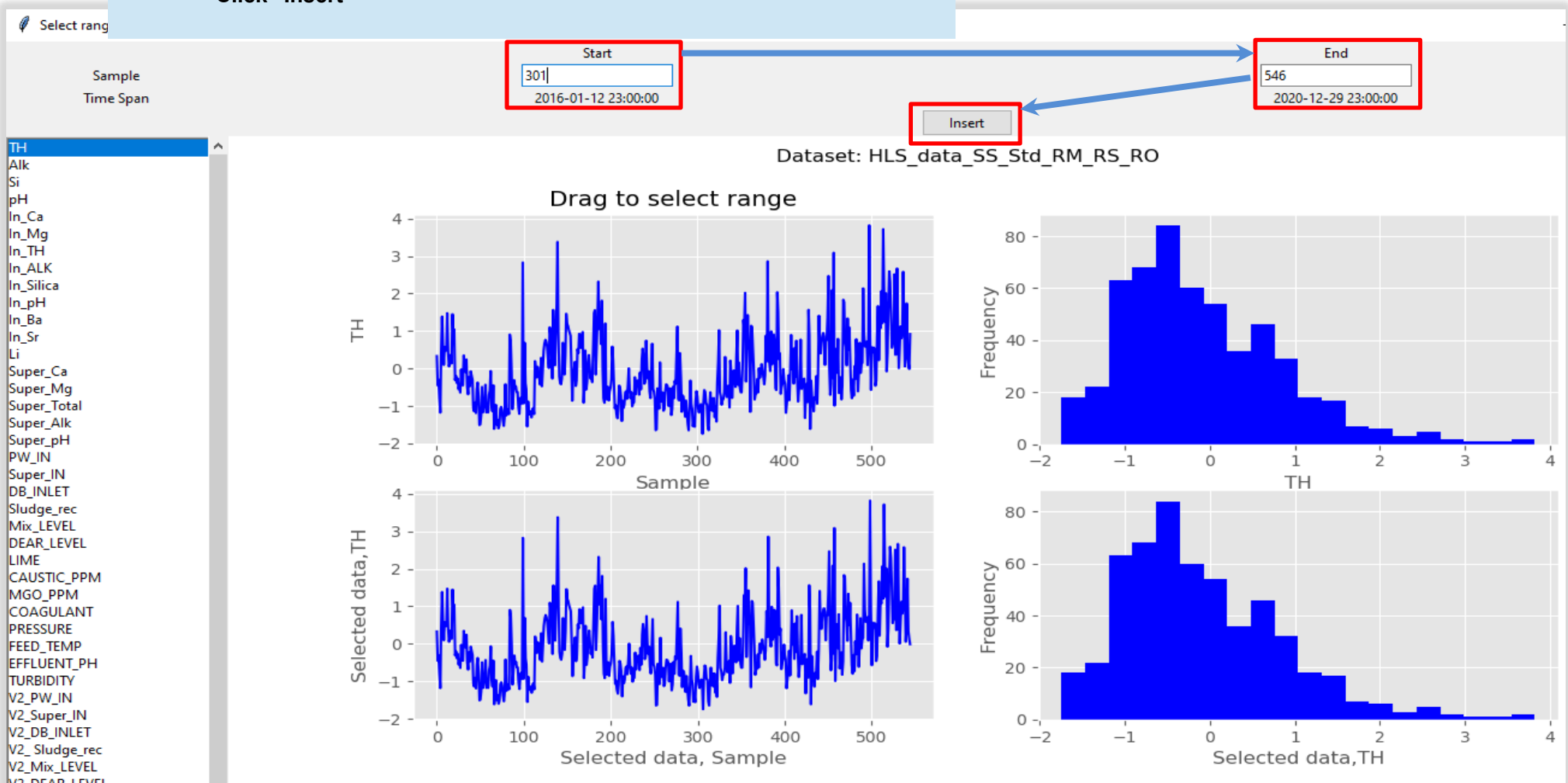


Data preprocessing – Divide data into two parts

16

To split data enter the range of dataset for validation, do the following:

- In “**Start**”, Enter “**301**”
- In “**End**”; Enter “**546**”
- Click “**Insert**”



Note: Please close the current popup figures before taking next action

Soft Sensor Analytics



Generated data sets after pre-processing

Updated Dataset list

After preprocess two new datasets are generated
One can be used for training, other for validation

The screenshot shows the 'Soft Sensor Analytics (Version 2.0)' interface. It has four tabs: 'Modeling', 'Auto modeling', 'Validation', and 'Causal analysis'. The 'Modeling' tab is active. The interface is divided into three main sections: 'Dataset', 'Available variables', and 'Working data:'.
- The 'Dataset' list contains: HLS_data_SS_Std, HLS_data_SS_Std_RM, HLS_data_SS_Std_RM_RS, HLS_data_SS_Std_RM_RS_RO, HLS_data_SS_Std_RM_RS_RO_[1:300], and HLS_data_SS_Std_RM_RS_RO_[301:546]. The last two items are highlighted with a red box.
- The 'Available variables' list contains: TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, Super_Ca, Super_Mg, Super_Total, Super_Alk, Super_pH, PW_IN, and Super_IN.
- The 'Working data:' section shows 'HLS_data_SS_Std_RM_RS_RO' selected. There are radio buttons for 'Time Trend' and 'Histogram'. A button says 'Select a tag from second list as output'. Below that, 'Selected Output: <<Alk>>' is displayed.
- The 'Data Preprocess' section has a dropdown menu set to 'Customized Preprocess-->' and a button for 'Automatic Preprocessing'.
- The 'Modeling' section has a dropdown menu set to 'Modeling-->'.
- At the bottom left, there are 'Import Data' and 'Delete Data' buttons.

From the second list, choose one variable and click this button



Part III-A: Prediction of Alkalinity

Soft Sensor Analytics



Modeling – Lasso regression

- 1 Choose dataset and modeling method
 - Click the “Modeling” tab
 - Click `HLS_data_SS_Std_RM_RS_RO [1:300]` (training data)
 - Go to “Modeling-->”, Click “Ordinary least squares...”, a window pops up as below

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

Available variables

Working data:

HLS_data_SS_Std_RM_RS_RO [1:300]

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Alk>>

Data Preprocess

Customized Preprocess-->

Automatic Preprocessing

Modeling

Modeling-->

Ordinary least squares...

Ridge regression...

Kernel ridge regression...

Lasso...

This provides different modeling algorithms

LASSO

Available variables

Input

Output: <<Alk>>

--> Modeling

Intercept

Normalize

Alpha 0.0001

Precision 0.001

Max iteration 1000

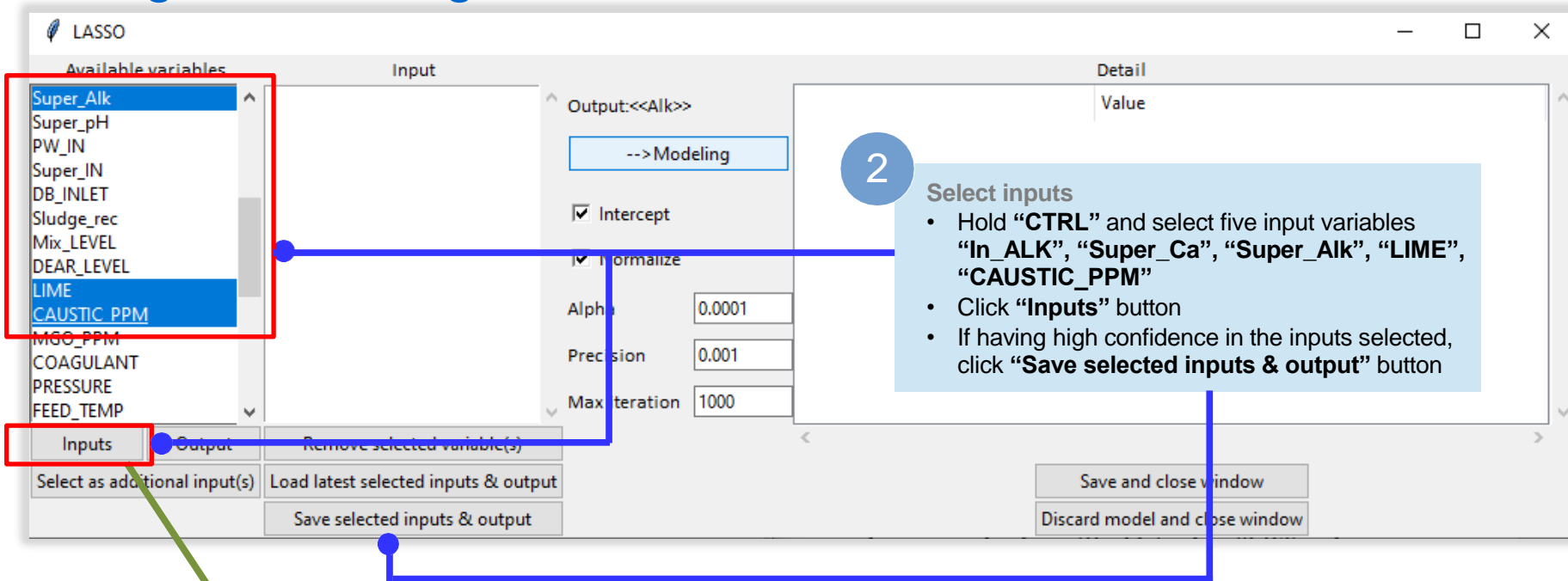
Detail

Value

Save and close window

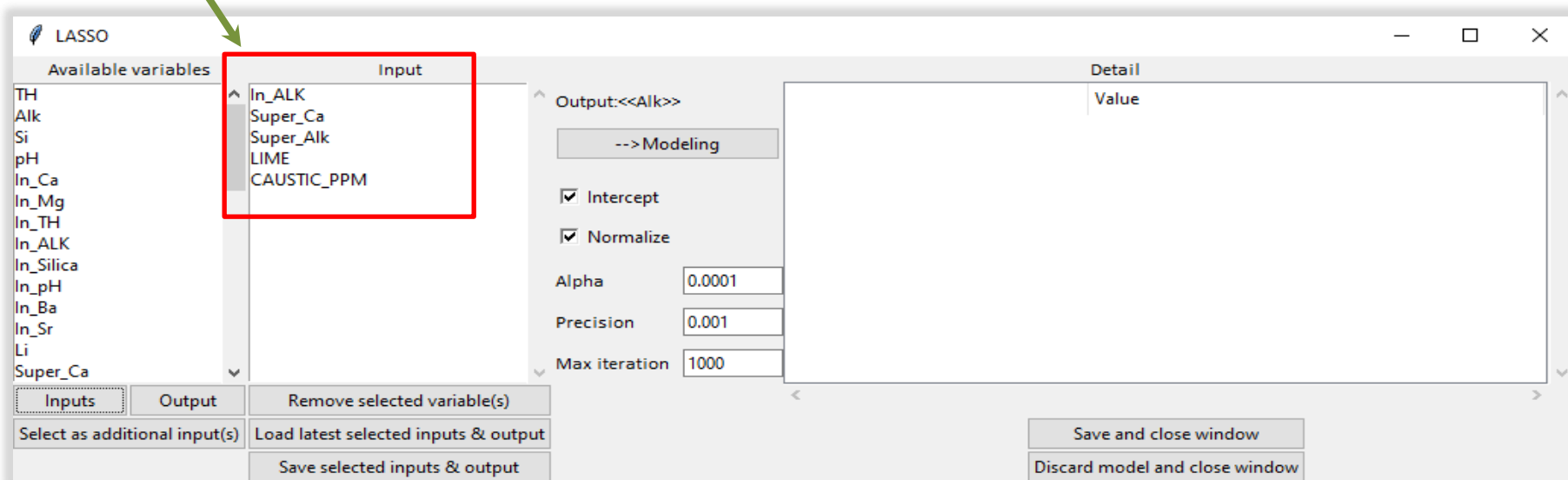
Discard model and close window

Modeling – Lasso regression



2 **Select inputs**

- Hold “CTRL” and select five input variables “In_ALK”, “Super_Ca”, “Super_Alk”, “LIME”, “CAUSTIC_PPM”
- Click “Inputs” button
- If having high confidence in the inputs selected, click “Save selected inputs & output” button



Soft Sensor Analytics



Modeling – Lasso regression

Selected inputs and output are loaded automatically
If any change is necessary, the highlighted buttons
can be used to add/remove variables

The screenshot shows the LASSO modeling software interface. It is divided into several sections:

- Available variables:** A list of variables including TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, and Super_Ca.
- Input:** A list of selected input variables: In_ALK, Super_Ca, Super_Alk, LIME, and CAUSTIC_PPM.
- Output:** The output variable is set to <<Alk>>.
- Modeling Options:** Includes checkboxes for Intercept and Normalize, and input fields for Alpha (0.0001), Precision (0.001), and Max iteration (1000).
- Detail:** A section for viewing the model's value.
- Buttons:** A row of buttons at the bottom includes "Inputs", "Output", "Remove selected variable(s)", "Select as additional input(s)", "Load latest selected inputs & output", "Save selected inputs & output", "Save and close window", and "Discard model and close window".

Two blue vertical lines with circular endpoints point from the text box above to the "Inputs" and "Output" buttons. A red rectangular box highlights the "Inputs", "Output", and "Remove selected variable(s)" buttons.

Soft Sensor Analytics



Modeling – Lasso regression

4

Modeling
Click "-->Modeling" to perform Lasso regression

The screenshot shows the LASSO software interface. On the left, there are two lists: 'Available variables' and 'Input'. The 'Input' list contains: In_ALK, Super_Ca, Super_Alk, LIME, and CAUSTIC_PPM. Below these lists are buttons for 'Inputs', 'Output', and 'Remove selected variable(s)'. In the center, there are checkboxes for 'Intercept' and 'Normalize', both of which are checked. Below these are input fields for 'Alpha' (set to 0.0001), 'Precision' (set to 0.001), and 'Max iteration' (set to 1000). A red box highlights the '-->Modeling' button. On the right, there is a 'Detail' section with a 'Value' field. At the bottom right, there are buttons for 'Save and close window' and 'Discard model and close window'. A red box highlights the 'Alpha' input field.

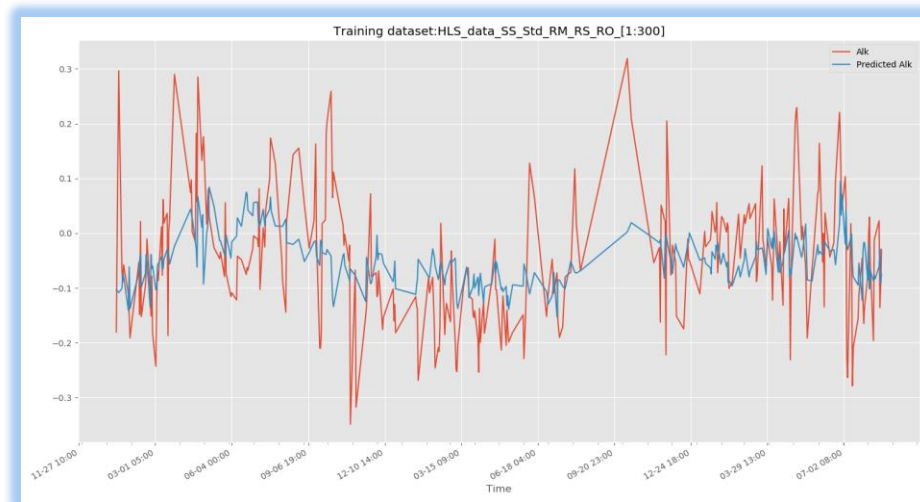
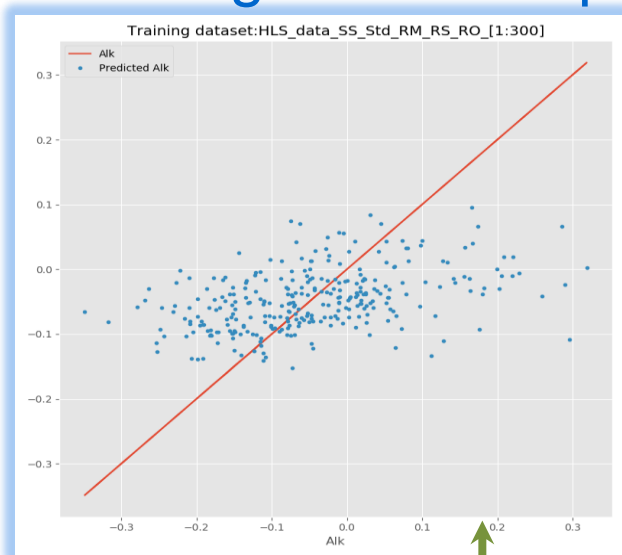
3

Options
Set "Alpha" as "0.0005".

Soft Sensor Analytics



Results – Lasso regression for predicting alkalinity



LASSO

Available variables	Input	Output <<Alk>>	Detail
TH Alk Si pH In_Ca In_Mg In_TH In_ALK In_Silica In_pH In_Ba In_Sr Li Super_Ca	In_ALK Super_Ca Super_Alk LIME CAUSTIC_PPM	-->Modeling <input checked="" type="checkbox"/> Intercept <input checked="" type="checkbox"/> Normalize Alpha: 0.0001 Precision: 0.001 Max iteration: 1000	Value 1 sample(s) 0 sample(s) Lasso y = ax + c ... Alk -0.05203576393196815 ... alpha 0.0001 ... Performance ...

5 Save model

- Close all figures
- Click "Save" to save this model for validation.

Save and close window

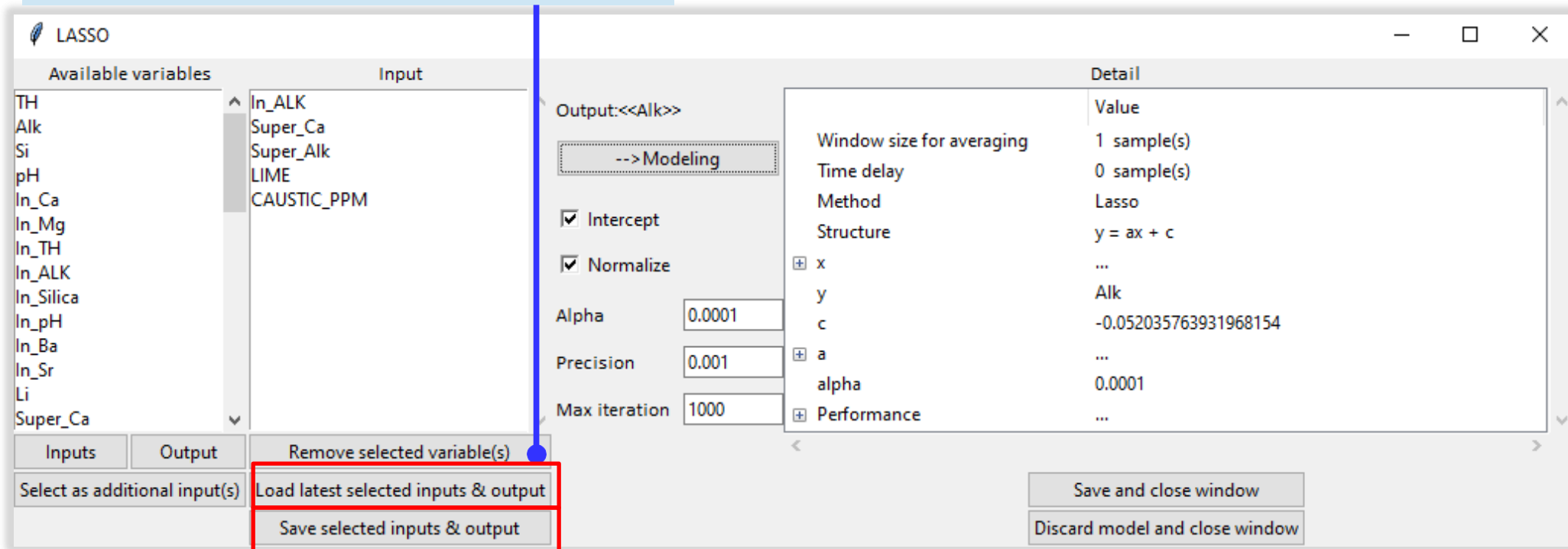
Discard model and close window

Modeling – Save current or Load previous set of inputs/output

7

If previous set of inputs/output is better, it can be recovered by clicking “Load latest selected inputs & output”

Then, modeling can be conducted again



The screenshot shows the LASSO software interface. On the left, there are two columns: 'Available variables' and 'Input'. The 'Input' column contains: In_ALK, Super_Ca, Super_Alk, LIME, and CAUSTIC_PPM. Below these columns are buttons for 'Inputs', 'Output', 'Remove selected variable(s)', 'Load latest selected inputs & output', and 'Save selected inputs & output'. The 'Load latest selected inputs & output' button is highlighted with a red box. In the center, there are checkboxes for 'Intercept' and 'Normalize', and input fields for 'Alpha' (0.0001), 'Precision' (0.001), and 'Max iteration' (1000). A '--> Modeling' button is also present. On the right, a 'Detail' panel shows the following parameters and values:

Parameter	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Lasso
Structure	$y = ax + c$
x	...
y	Alk
c	-0.052035763931968154
a	...
alpha	0.0001
Performance	...

At the bottom right, there are buttons for 'Save and close window' and 'Discard model and close window'.

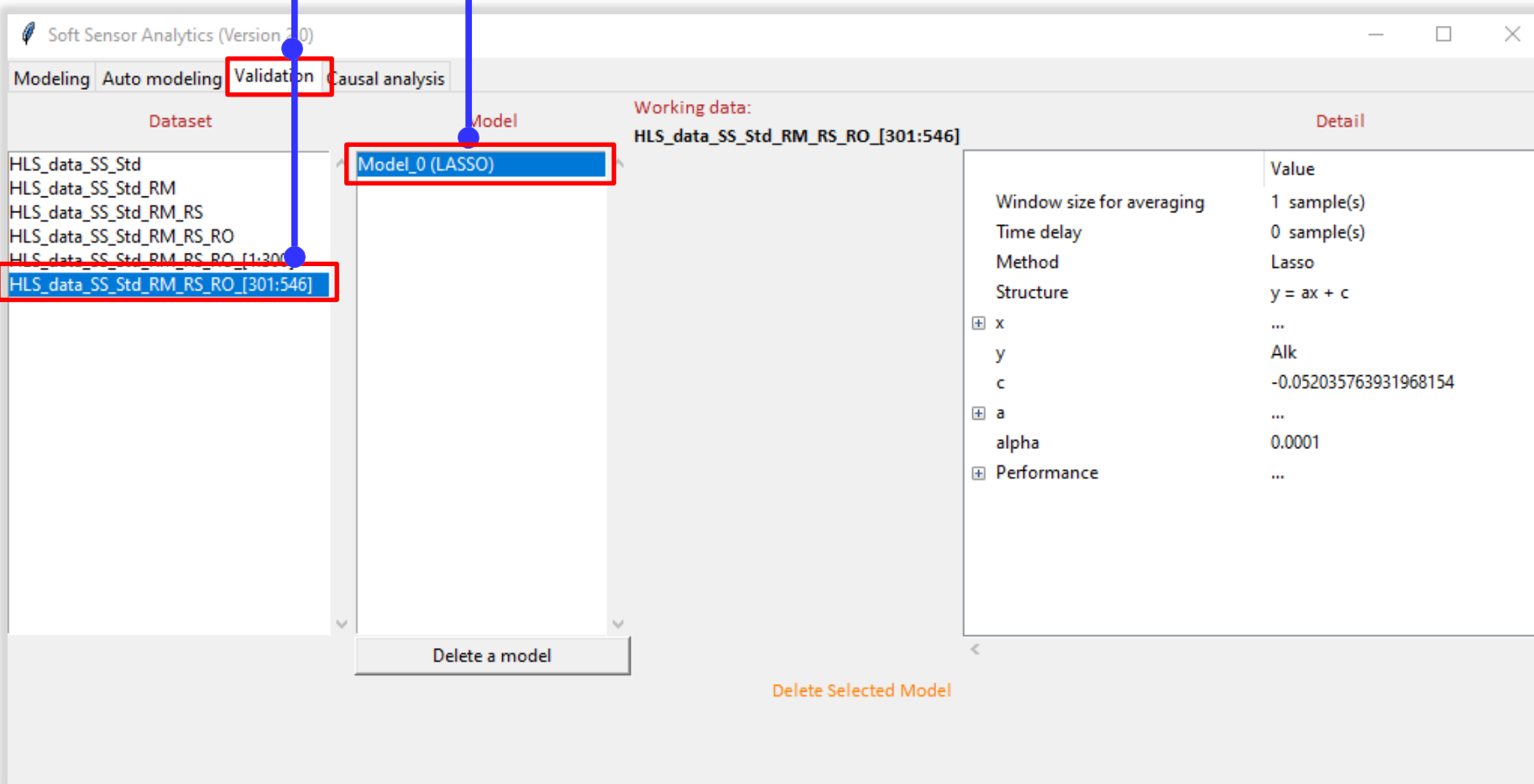
6

If any changes are made to the inputs or/and output, and if the performance is good, the set of the inputs/output variables can be saved by clicking “Save selected inputs & output”.

Validation of the models

Model Validation: the models based on different methods can be validated based on the second dataset.

- 1 **Select model for validation**
 - Click “Validation” tab on the top
 - Choose “HLS_data_SS_Std_RM_RS_RO[301:546]”
 - Click “Model_0 (Lasso)” to validate this model



The screenshot shows the 'Soft Sensor Analytics (Version 7.0)' interface. The 'Validation' tab is selected. The 'Dataset' list on the left includes 'HLS_data_SS_Std_RM_RS_RO [301:546]', which is highlighted with a red box. The 'Model' list in the center includes 'Model_0 (LASSO)', also highlighted with a red box. The 'Working data' section shows 'HLS_data_SS_Std_RM_RS_RO [301:546]'. The 'Detail' panel on the right displays the following parameters:

Parameter	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Lasso
Structure	$y = ax + c$
x	...
y	Alk
c	-0.052035763931968154
a	...
alpha	0.0001
Performance	...

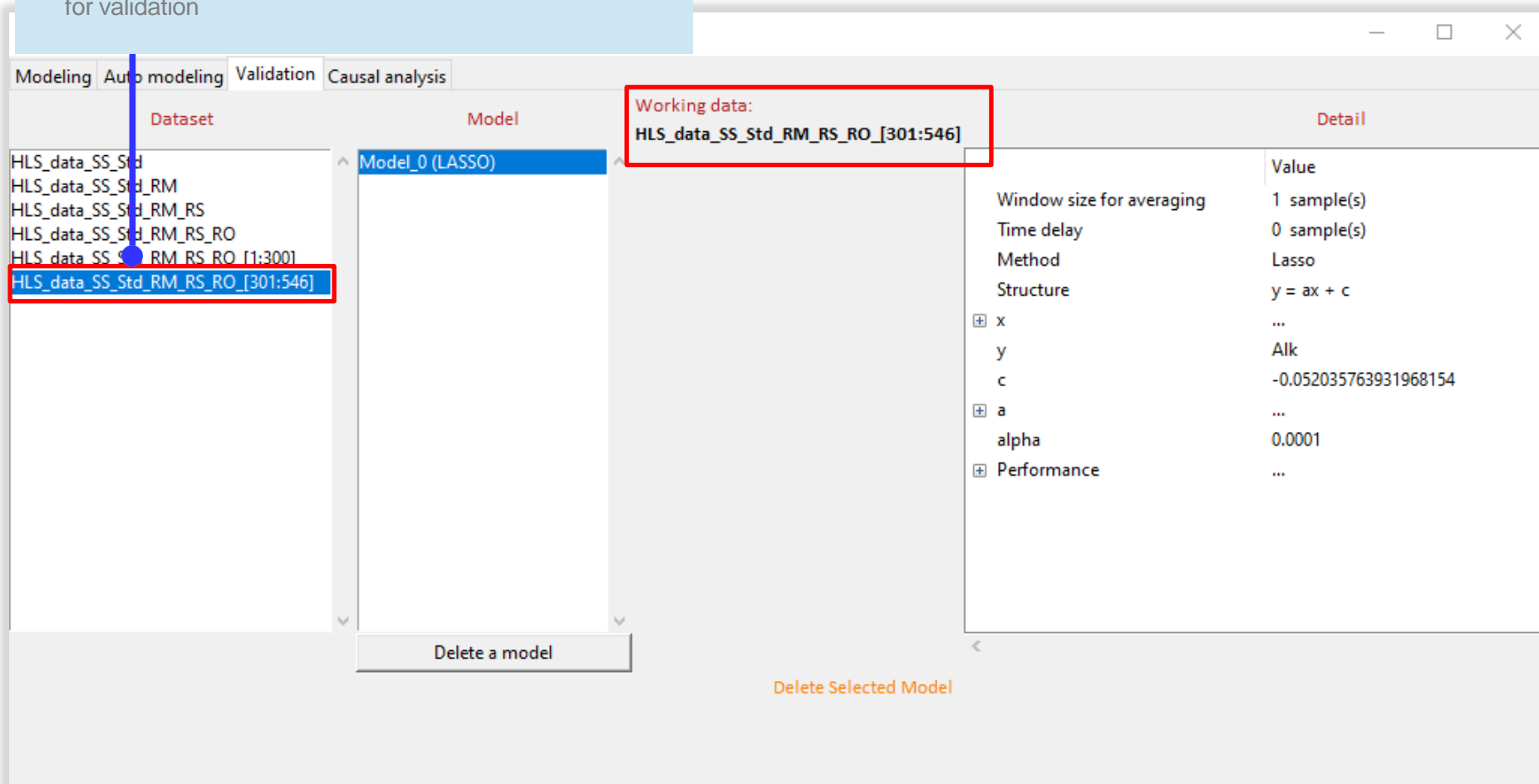
Validation of the model

Model Validation: the models based on different methods can be validated based on the second dataset.

2

Select a data set for validation

- Choose “HLS_data_SS_Std_RM_RS_RO[301:546]” for validation



Parameter	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Lasso
Structure	$y = ax + c$
x	...
y	Alk
c	-0.052035763931968154
a	...
alpha	0.0001
Performance	...

Validation of the model

3

Choose “Simulate” to continue

- **Right-Click** on the two models
- **Click “Simulate...”** button, a window will pop up

Simulation

Without bias update

With bias update

Bias Update:
 $y_{pred}(t) = f(x, a, t) + b(t)$
 $b(t) = \alpha * [y(t-1) - y_{pred}(t-1)] + (1 - \alpha) * b(t-1)$
alpha (forgetting factor)

Simulate

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset Model Working data: Detail

HLS_data_SS_Std
HLS_data_SS_Std_RM
HLS_data_SS_Std_RM_RS
HLS_data_SS_Std_RM_RS_RO
HLS_data_SS_Std_RM_RS_RO [1:300]
HLS_data_SS_Std_RM_RS_RO [301:546]

Model_0 (LASSO)

Simulate...
Save as...
Delete

Delete a model

Parameter	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Lasso
Structure	$y = ax + c$
x	...
y	Alk
c	-0.052035763931968154
a	...
alpha	0.0001
Performance	...

Click To Select Model/Right Click To Manipulate

Soft Sensor Analytics

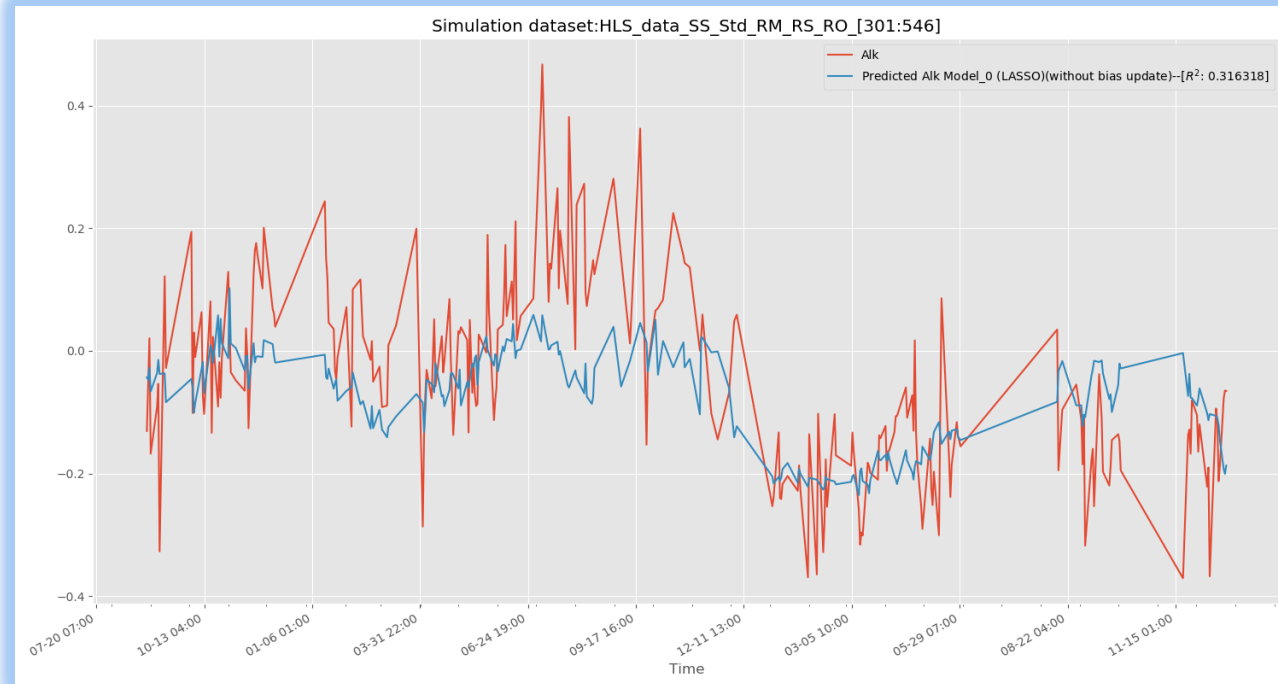
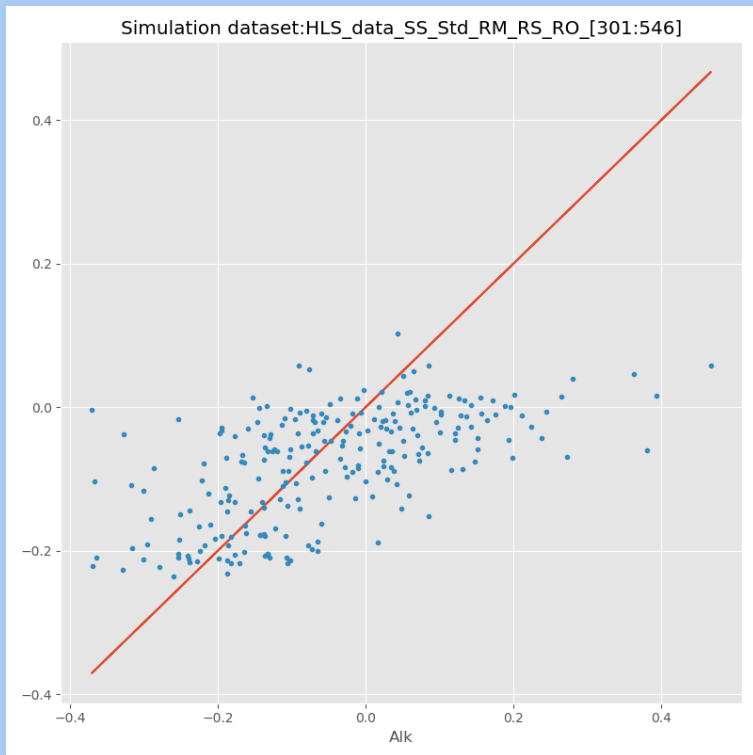
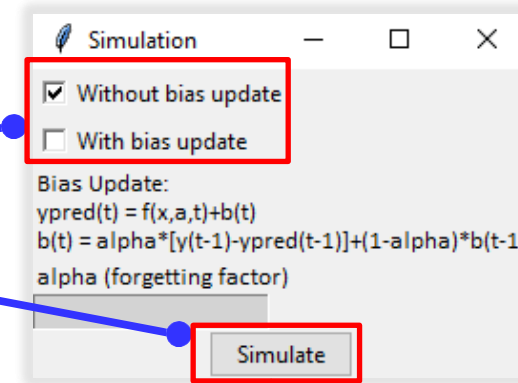


Validation of the model

4

Model validation

- Tick “Without bias update” box
- Click “Simulate” button



Soft Sensor Analytics

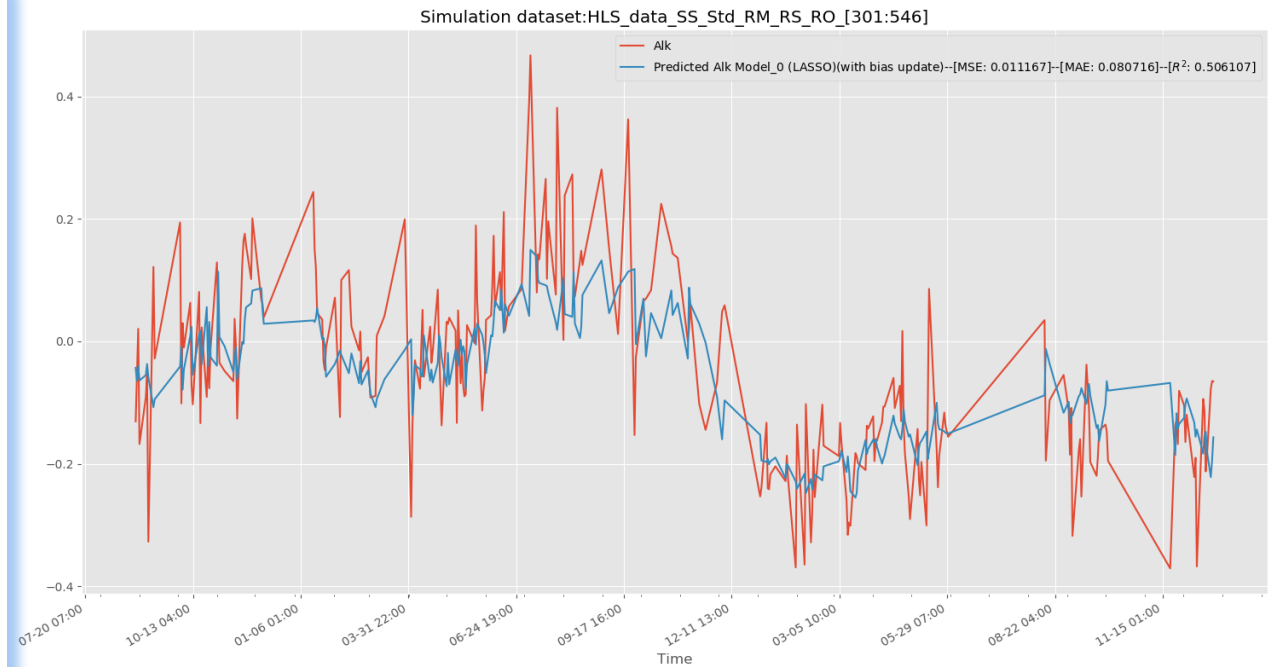
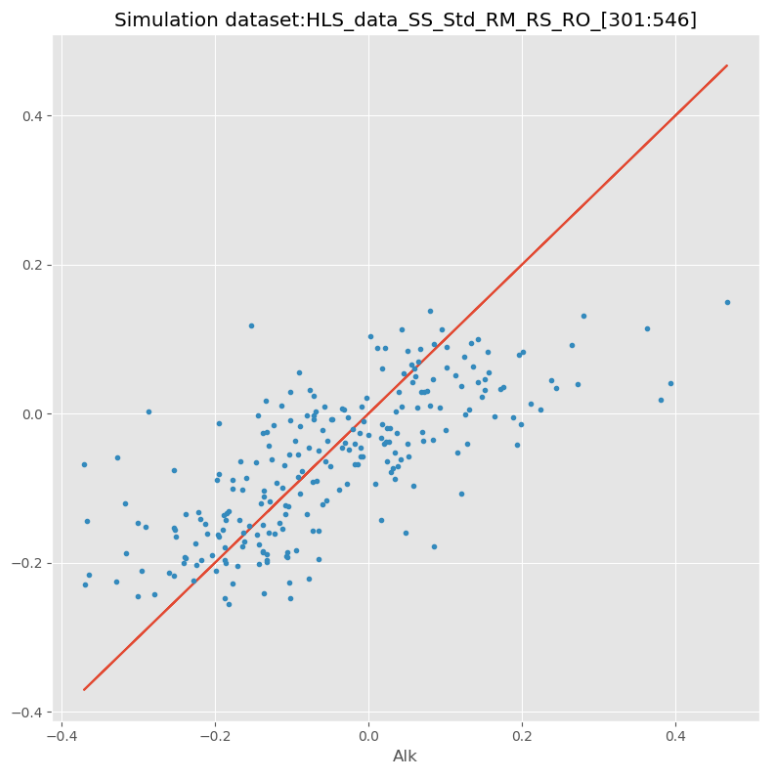
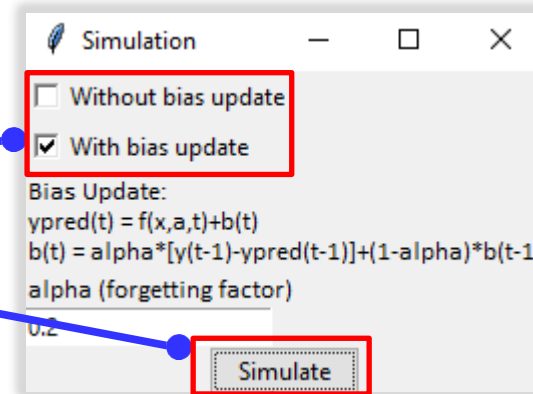


Validation of the model

5

Model validation

- Tick **“With bias update”** box
- Click **“Simulate”** button





Part III-B: Prediction of Silica

Soft Sensor Analytics



Change output to be predicted

- 1 Change the output to be predicted
 - Go to the “**Modeling**” tab
 - Choose “**HLS_data_SS_Std_RM_RS_RO[1:300]**” as the dataset
 - Click “**Si**” in the second list
 - Click “**Select a tag from second list as output**” button

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

- HLS_data_SS_Std
- HLS_data_SS_Std_RM
- HLS_data_SS_Std_RM_RS
- HLS_data_SS_Std_RM_RS_RO [1:300]**
- HLS_data_SS_Std_RM_RS_RO [301:546]

Available variables

- TH
- Alk
- Si**
- pH
- In_Ca
- In_Mg
- In_TH
- In_ALK
- In_Silica
- In_pH
- In_Ba
- In_Sr
- Li
- Super_Ca
- Super_Mg
- Super_Total
- Super_AlK
- Super_pH
- PW_IN
- Super_IN

Working data:

HLS_data_SS_Std_RM_RS_RO [1:300]

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Alk>>

Data Preprocess

Customized Preprocess-->

Automatic Preprocessing

Modeling

Modeling-->

Import Data Delete Data

From the second list, choose one variable and click this button

Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

- HLS_data_SS_Std
- HLS_data_SS_Std_RM
- HLS_data_SS_Std_RM_RS
- HLS_data_SS_Std_RM_RS_RO [1:300]**
- HLS_data_SS_Std_RM_RS_RO [301:546]

Available variables

- TH
- Alk
- Si**
- pH
- In_Ca
- In_Mg
- In_TH
- In_ALK
- In_Silica
- In_pH
- In_Ba
- In_Sr
- Li
- Super_Ca
- Super_Mg
- Super_Total
- Super_AlK
- Super_pH
- PW_IN
- Super_IN

Working data:

HLS_data_SS_Std_RM_RS_RO [1:300]

Time Trend Histogram

Select a tag from second list as output

Selected Output: <<Si>>

Soft Sensor Analytics



Modeling – Lasso regression

- 2
- Choose dataset and modeling method
- Click the “Modeling” tab
 - Click **HLS_data_SS_Std_RM_RS_RO [1:300]** (training data)
 - Go to “Modeling-->”, Click “Ordinary least squares...”, a window pops up as below

This provides different modeling algorithms

Soft Sensor Analytics



Modeling – Lasso regression

The screenshot shows the LASSO software interface. On the left, a list of 'Available variables' includes DB_INLET, Sludge_rec, Mix_LEVEL, DEAR_LEVEL, LIME, CAUSTIC_PPM, MGO_PPM, COAGULANT, PRESSURE, FEED_TEMP, EFFLUENT_PH, TURBIDITY, V2_PW_IN, and V2_Super_IN. The 'Input' list contains In_ALK, Super_Ca, Super_Alk, LIME, and CAUSTIC_PPM. The 'Output' is set to <<Si>>. A blue circle with the number '2' is placed over the 'Inputs' button. A blue callout box with the title 'Select inputs' contains the following instructions:

- Hold “CTRL” and select five input variables “In_ALK”, “Super_Ca”, “Super_Alk”, “LIME”, “CAUSTIC_PPM”
- Click “Inputs” button
- If having high confidence in the inputs selected, click “Save selected inputs & output” button

Buttons at the bottom include 'Inputs', 'Output', 'Remove selected variable(s)', 'Select as additional input(s)', 'Load latest selected inputs & output', 'Save selected inputs & output', 'Save and close window', and 'Discard model and close window'.

The screenshot shows the LASSO software interface after the selection process. The 'Available variables' list now includes TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, and Super_Ca. The 'Input' list contains In_Ca, In_Mg, In_Silica, Super_Mg, Super_Alk, MGO_PPM, and EFFLUENT PH. The 'Output' is still <<Si>>. The 'Inputs' button is highlighted with a red box. The same set of buttons as in the previous screenshot is visible at the bottom.

Soft Sensor Analytics



Modeling – Lasso regression

Selected inputs and output are loaded automatically

If any change is necessary, the highlighted buttons can be used to add/remove variables

The screenshot shows the LASSO modeling software interface. It is divided into several sections:

- Available variables:** A list of variables including TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, and Super_Ca.
- Input:** A list of selected input variables including In_Ca, In_Mg, In_Silica, Super_Mg, Super_Alk, MGO_PPM, and EFFLUENT_PH.
- Output:** The output variable is set to <<Si>>.
- Modeling Options:** Includes a "--> Modeling" button, checkboxes for "Intercept" and "Normalize", and input fields for "Alpha" (0.0001), "Precision" (0.001), and "Max iteration" (1000).
- Detail:** A section for viewing the model's value.
- Control Buttons:** A set of buttons at the bottom, including "Inputs", "Output", "Remove selected variable(s)", "Select as additional input(s)", "Load latest selected inputs & output", "Save selected inputs & output", "Save and close window", and "Discard model and close window".

Blue vertical lines indicate the flow of data from the "Available variables" list to the "Input" list and from the "Output" list to the "Modeling" section. A red box highlights the "Inputs", "Output", and "Remove selected variable(s)" buttons.

Soft Sensor Analytics



Modeling – Lasso regression

4

Modeling
Click “-->Modeling” to perform Lasso regression

The screenshot shows the LASSO software interface. On the left, there are two lists: 'Available variables' and 'Input'. The 'Input' list contains: In_Ca, In_Mg, In_Silica, Super_Mg, Super_Alkal, MGO_PPM, and EFFLUENT_PH. Below these lists are buttons for 'Inputs', 'Output', and 'Remove selected variable(s)'. In the center, there is a section for 'Output: <<Si>>' with a red box around the '--> Modeling' button. Below this are checkboxes for 'Intercept' and 'Normalize', both checked. There are also input fields for 'Alpha' (0.0001), 'Precision' (0.001), and 'Max iteration' (1000). At the bottom right, there are buttons for 'Save and close window' and 'Discard model and close window'. A red box highlights the 'Alpha' input field, and a blue arrow points from it to callout 3.

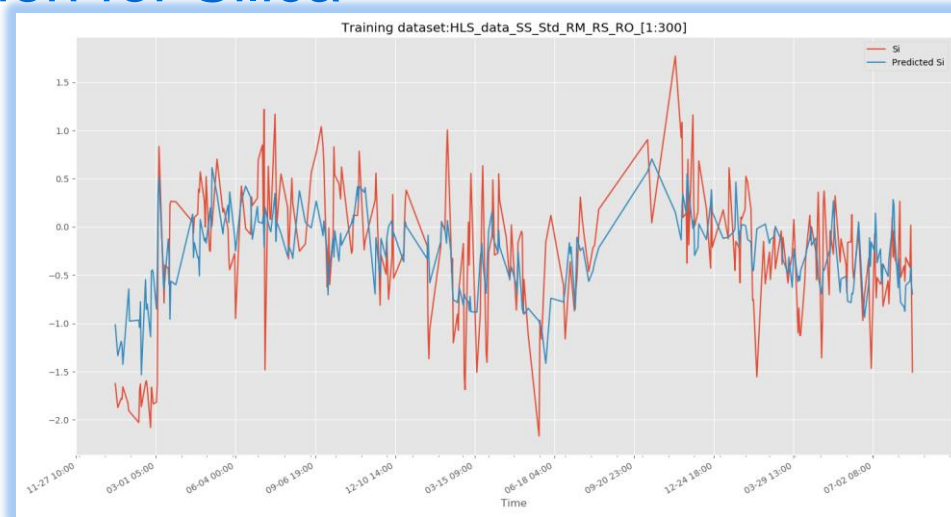
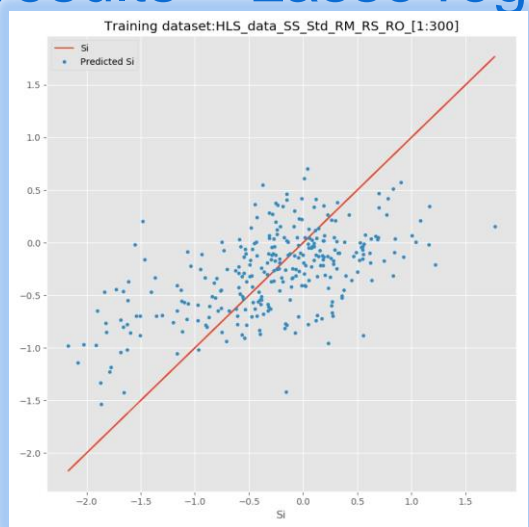
3

Options
Set “Alpha” as “0.0005”.

Soft Sensor Analytics



Modeling results – Lasso regression for Silica



LASSO

Available variables: TH, Alk, Si, pH, In_Ca, In_Mg, In_TH, In_ALK, In_Silica, In_pH, In_Ba, In_Sr, Li, Super_Ca

Input: In_ALK, Super_Ca, Super_Alk, LIME, CAUSTIC_PPM

Output <<Alk>>

--> Modeling

Intercept

Normalize

Alpha: 0.0001

Precision: 0.001

Max iteration: 1000

Window size for averaging: 1 sample(s)

Time delay: 0 sample(s)

Method: Lasso

Structure: $y = ax + c$

x: ...

y: Alk

c: -0.05203576393196815

a: ...

alpha: 0.0001

Performance: ...

5 Save model

- Close all figures
- Click "Save" to save this model for validation.

Save and close window

Discard model and close window

Inputs | Output | Remove selected variable(s)

Select as additional input(s) | Load latest selected inputs & output

Save selected inputs & output

Soft Sensor Analytics



Modeling – Save current or Load previous set of inputs/output

7

If previous set of inputs/output is better, it can be recovered by clicking “**Load latest selected inputs & output**”

Then, modeling can be conducted again

The screenshot shows the LASSO software interface. On the left, there are two lists: 'Available variables' and 'Input'. The 'Input' list contains In_ALK, Super_Ca, Super_Alk, LIME, and CAUSTIC_PPM. Below these lists are buttons for 'Inputs', 'Output', and 'Remove selected variable(s)'. A red box highlights the 'Load latest selected inputs & output' and 'Save selected inputs & output' buttons. In the center, there are checkboxes for 'Intercept' and 'Normalize', and input fields for 'Alpha' (0.0001), 'Precision' (0.001), and 'Max iteration' (1000). A '--> Modeling' button is also present. On the right, a 'Detail' panel shows the current configuration: 'Window size for averaging' (1 sample(s)), 'Time delay' (0 sample(s)), 'Method' (Lasso), and 'Structure' (y = ax + c). The 'Value' column shows parameters: x, y, c, a, alpha, and Performance.

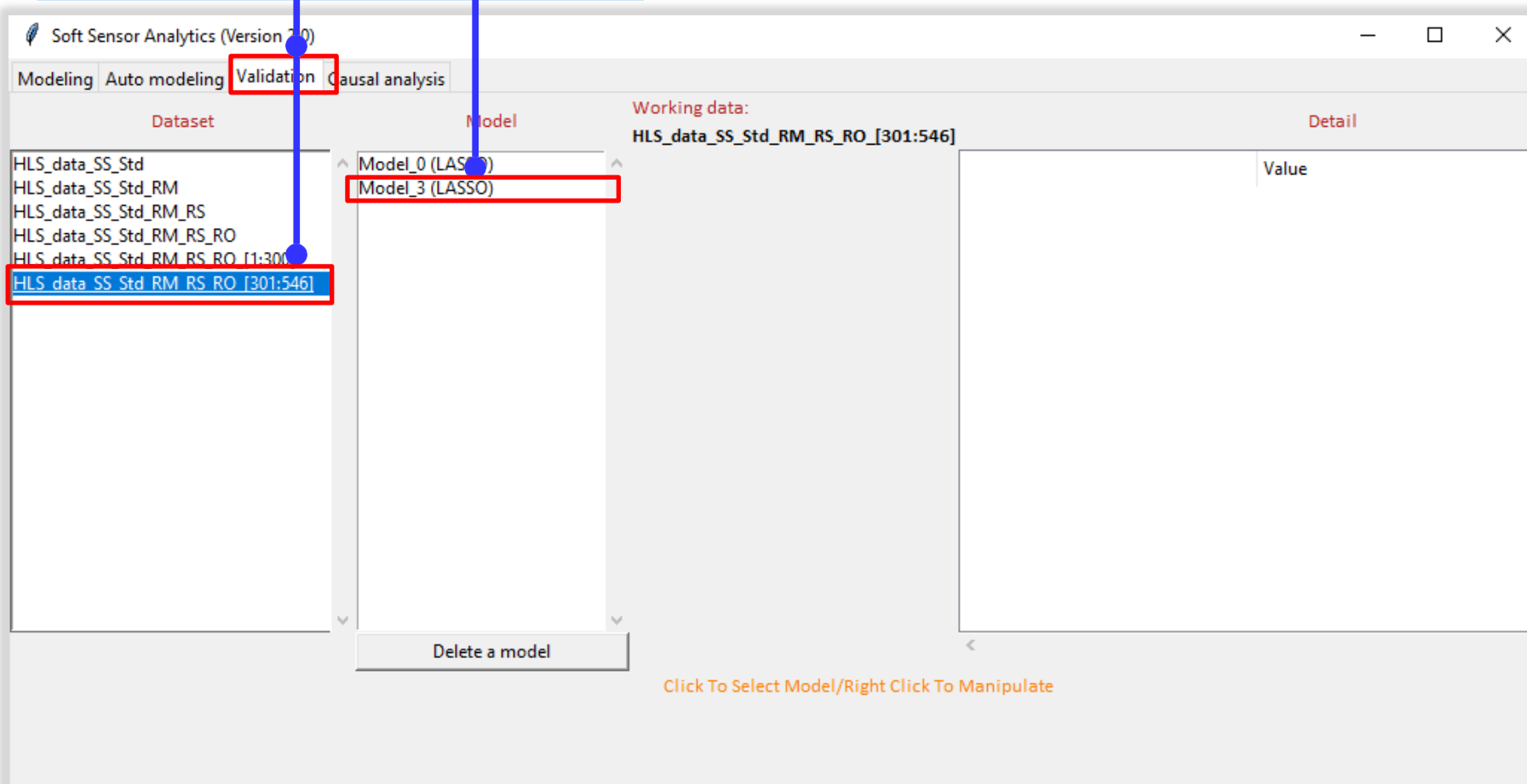
6

If any changes are made to the inputs or/and output, and if the performance is good, the set of the inputs/output variables can be saved by clicking “**Save selected inputs & output**”.

Validation of the model

Model Validation: the models based on different methods can be validated based on the second dataset.

- 1 **Select models for validation**
 - Click "Validation" tab on the top
 - Choose "HLS_data_SS_Std_RM_RS_RO[301:546]"
 - Click "Model_3 (Lasso)" to validate this model



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling **Validation** Causal analysis

Dataset Model Working data: Detail

HLS_data_SS_Std
HLS_data_SS_Std_RM
HLS_data_SS_Std_RM_RS
HLS_data_SS_Std_RM_RS_RO
HLS_data_SS_Std_RM_RS_RO [1:300]
HLS_data_SS_Std_RM_RS_RO [301:546]

Model_0 (LASSO)
Model_3 (LASSO)

HLS_data_SS_Std_RM_RS_RO [301:546]

Delete a model

Value

Click To Select Model/Right Click To Manipulate

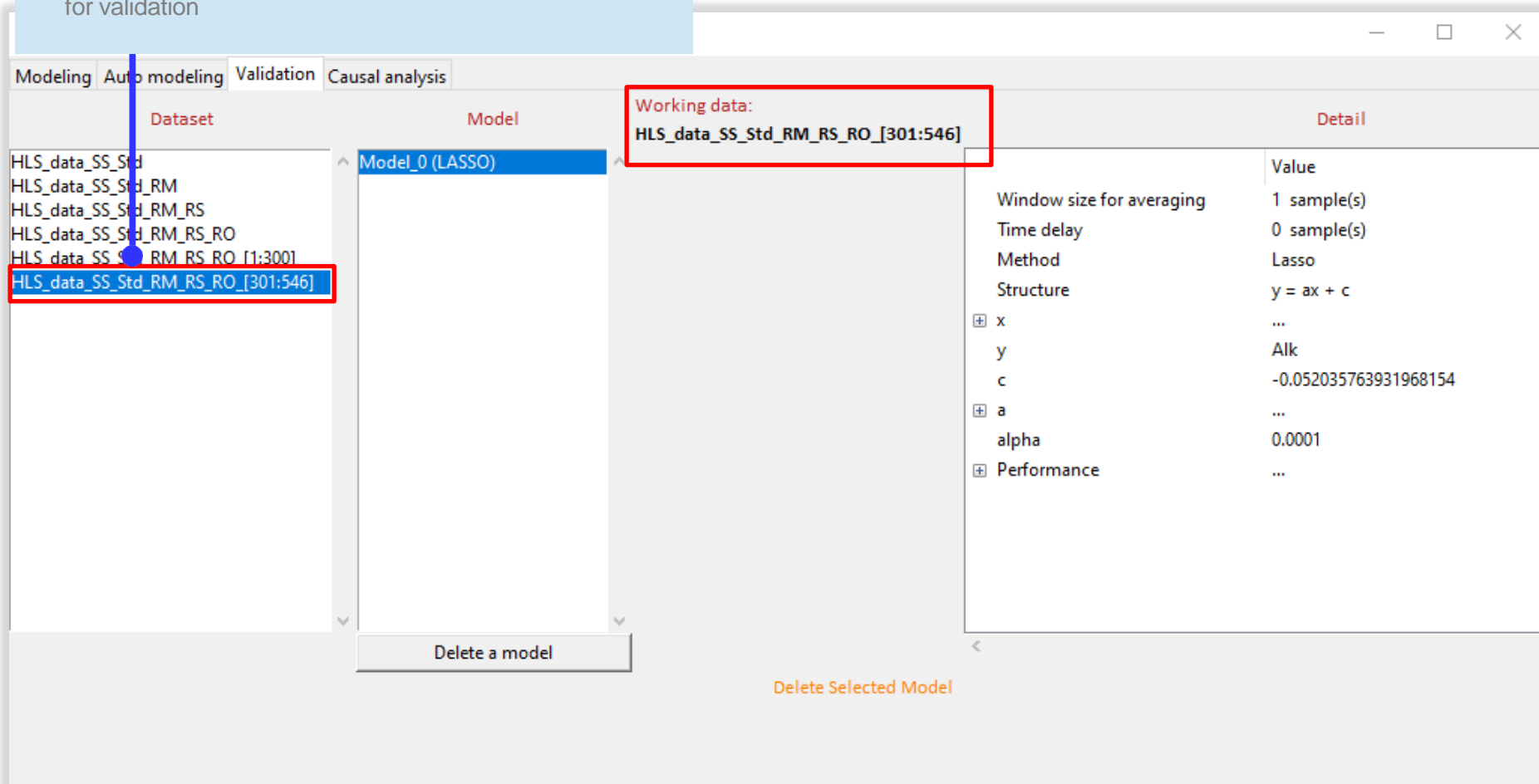
Validation of the model

Model Validation: the models based on different methods can be validated based on the second dataset.

2

Select a data set for validation

- Choose “HLS_data_SS_Std_RM_RS_RO[301:546]” for validation



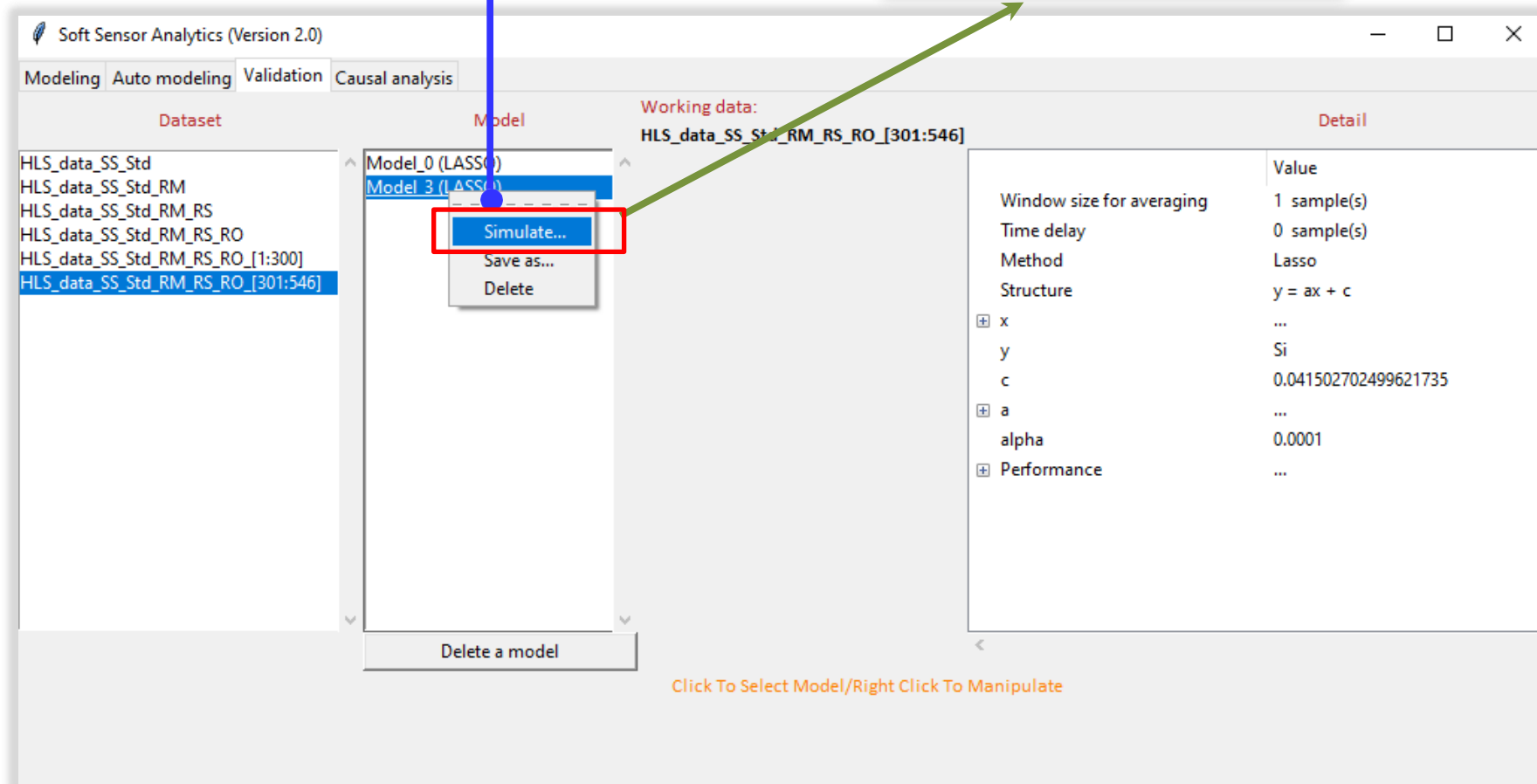
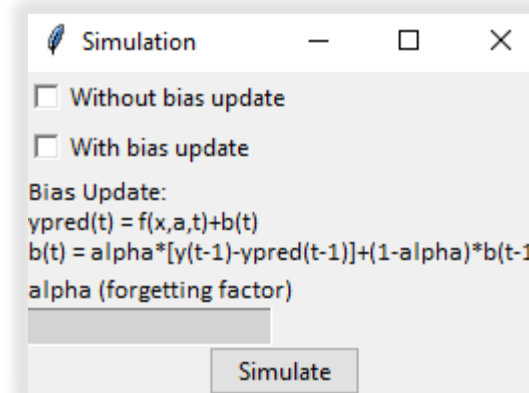
Dataset	Model	Working data:	Detail	
HLS_data_SS_Std	Model_0 (LASSO)	HLS_data_SS_Std_RM_RS_RO [301:546]	Value	
HLS_data_SS_Std_RM			Window size for averaging	1 sample(s)
HLS_data_SS_Std_RM_RS			Time delay	0 sample(s)
HLS_data_SS_Std_RM_RS_RO			Method	Lasso
HLS_data_SS_Std_RM_RS_RO [1:300]			Structure	$y = ax + c$
HLS_data_SS_Std_RM_RS_RO [301:546]			+ x	...
			y	Alk
			c	-0.052035763931968154
			+ a	...
			alpha	0.0001
			+ Performance	...

Validation of the model

3

Choose “Simulate” to continue

- **Right-Click** on the two models
- **Click “Simulate...”** button, a window will pop up



Soft Sensor Analytics (Version 2.0)

Modeling Auto modeling Validation Causal analysis

Dataset

- HLS_data_SS_Std
- HLS_data_SS_Std_RM
- HLS_data_SS_Std_RM_RS
- HLS_data_SS_Std_RM_RS_RO
- HLS_data_SS_Std_RM_RS_RO_[1:300]
- HLS_data_SS_Std_RM_RS_RO_[301:546]

Model

- Model_0 (LASSO)
- Model_3 (LASSO)

Working data: HLS_data_SS_Std_RM_RS_RO_[301:546]

Detail

	Value
Window size for averaging	1 sample(s)
Time delay	0 sample(s)
Method	Lasso
Structure	$y = ax + c$
x	...
y	Si
c	0.041502702499621735
a	...
alpha	0.0001
Performance	...

Delete a model

Click To Select Model/Right Click To Manipulate

Soft Sensor Analytics

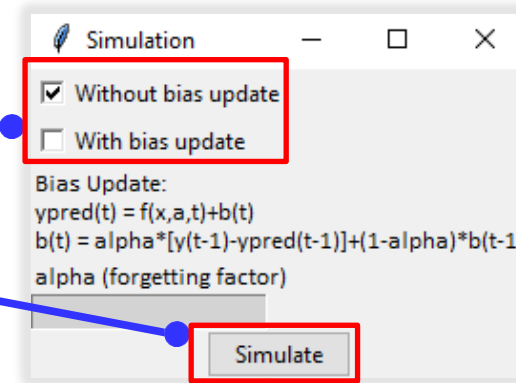


Validation of the models

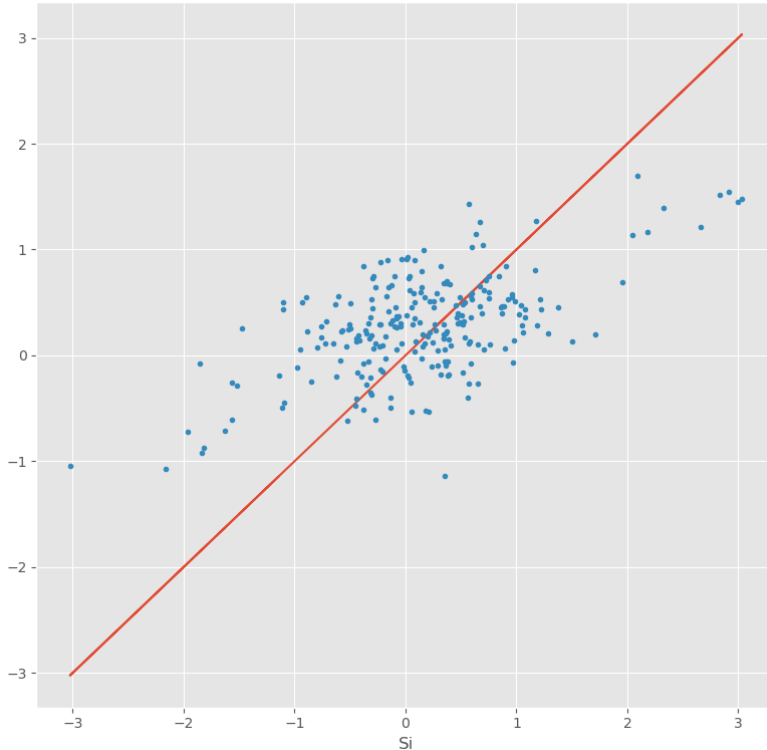
4

Model validation

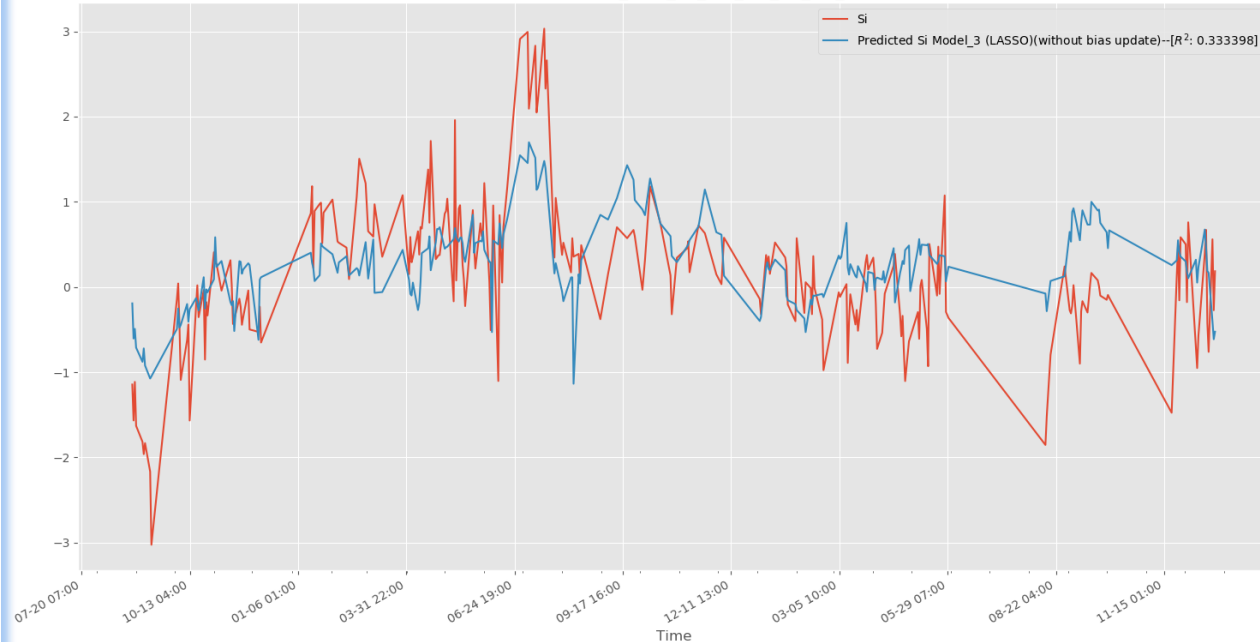
- Tick “Without bias update” box
- Click “Simulate” button



Simulation dataset:HLS_data_SS_Std_RM_RS_RO_[301:546]



Simulation dataset:HLS_data_SS_Std_RM_RS_RO_[301:546]



Soft Sensor Analytics

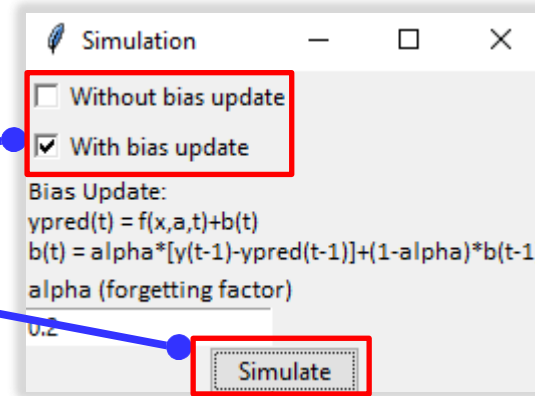


Validation of the models

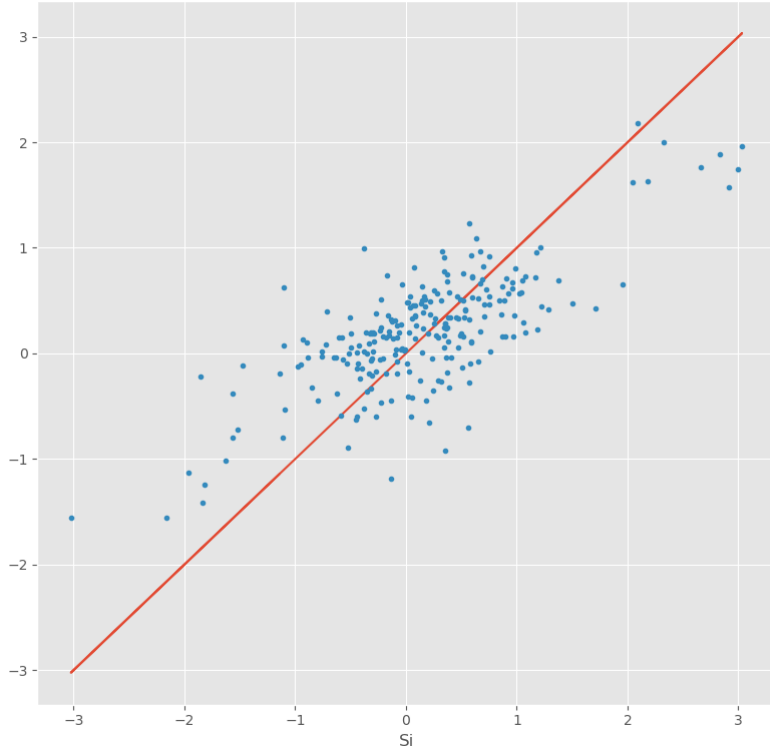
5

Model validation

- Tick “With bias update” box
- Click “Simulate” button



Simulation dataset:HLS_data_SS_Std_RM_RS_RO_[301:546]



Simulation dataset:HLS_data_SS_Std_RM_RS_RO_[301:546]

