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

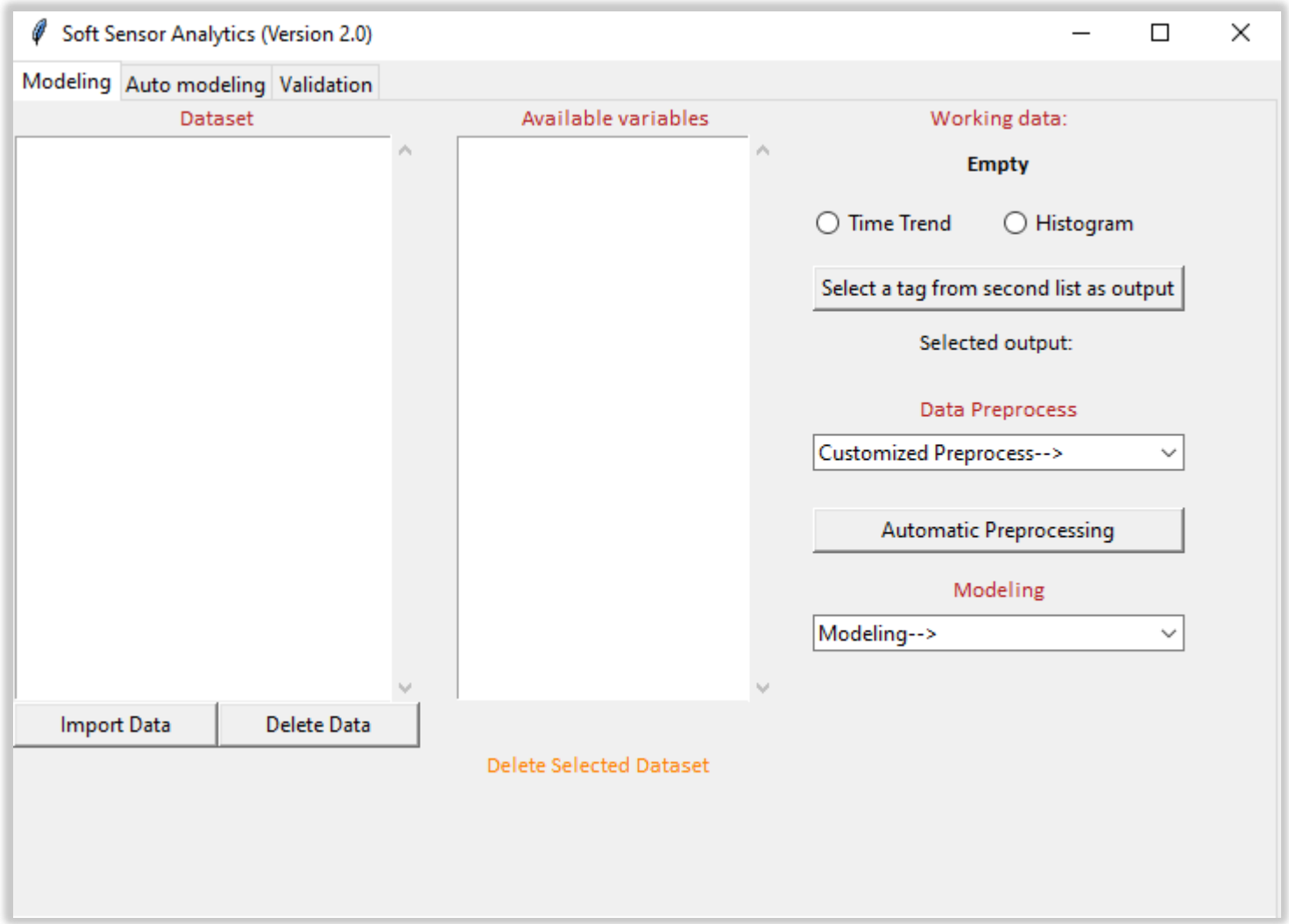
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Chemical	FeedT	FeedA	FeedB	FeedC	FeedD	Underflov	Underflov	Underflov	Underflov	Lab	DenN	DenFr
2	1/1/2016 0:00	0.022112	0.006653	-0.06091	-0.06707	0.127711	0.097012	-0.0143	-0.0069	0.016138	0.049496	-0.03981	0.202639	0.051406
3	1/1/2016 0:10	0.017209	0.006962	-0.06285	-0.06516	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



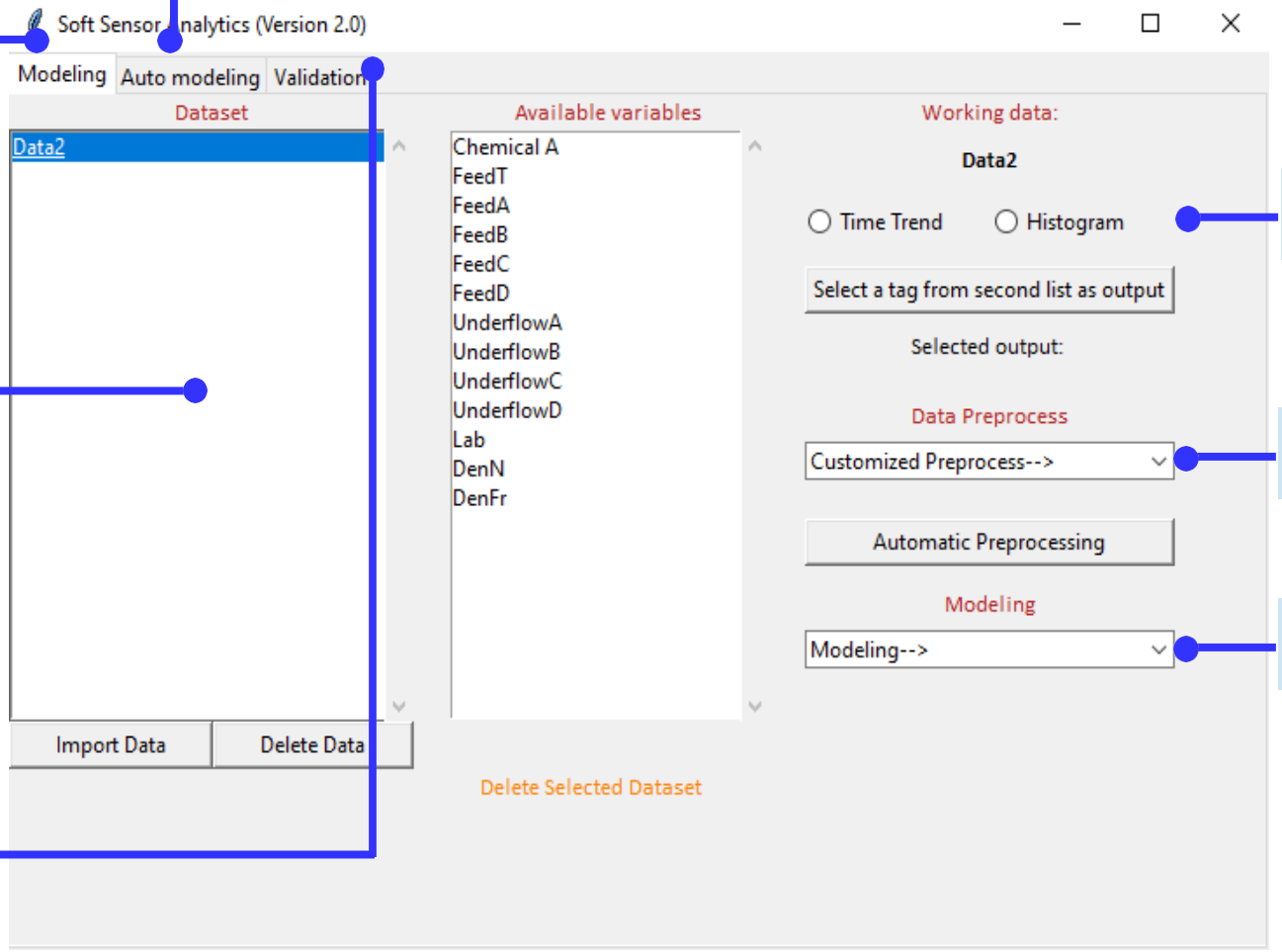
Soft Sensor Analytics

An overview of the toolbox

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

One-click modeling
Click to choose PLS or Gaussian Process Regression to build a model automatically

Data modeling



Data visualization

Preprocessing methods

Modeling algorithms

Data management
Import and **delete** data.

Model management
Compare the performance of different models

Part I: Start Guide

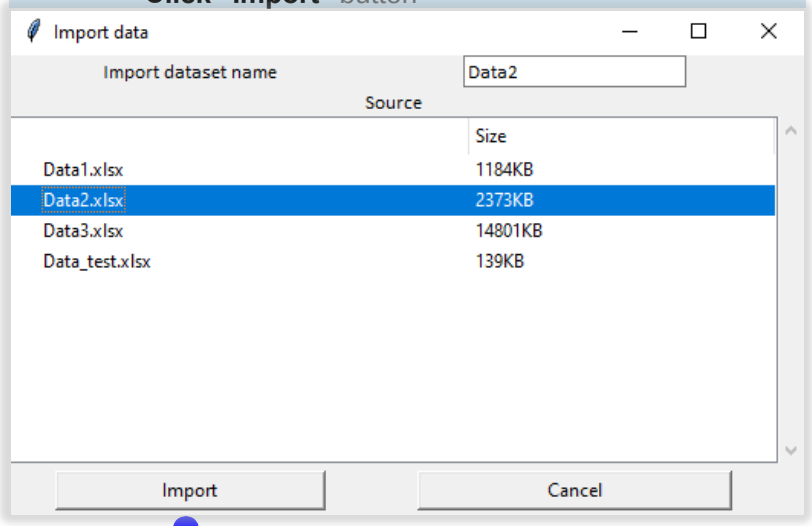
Soft Sensor Analytics

Quick Start Guide – Import data sets

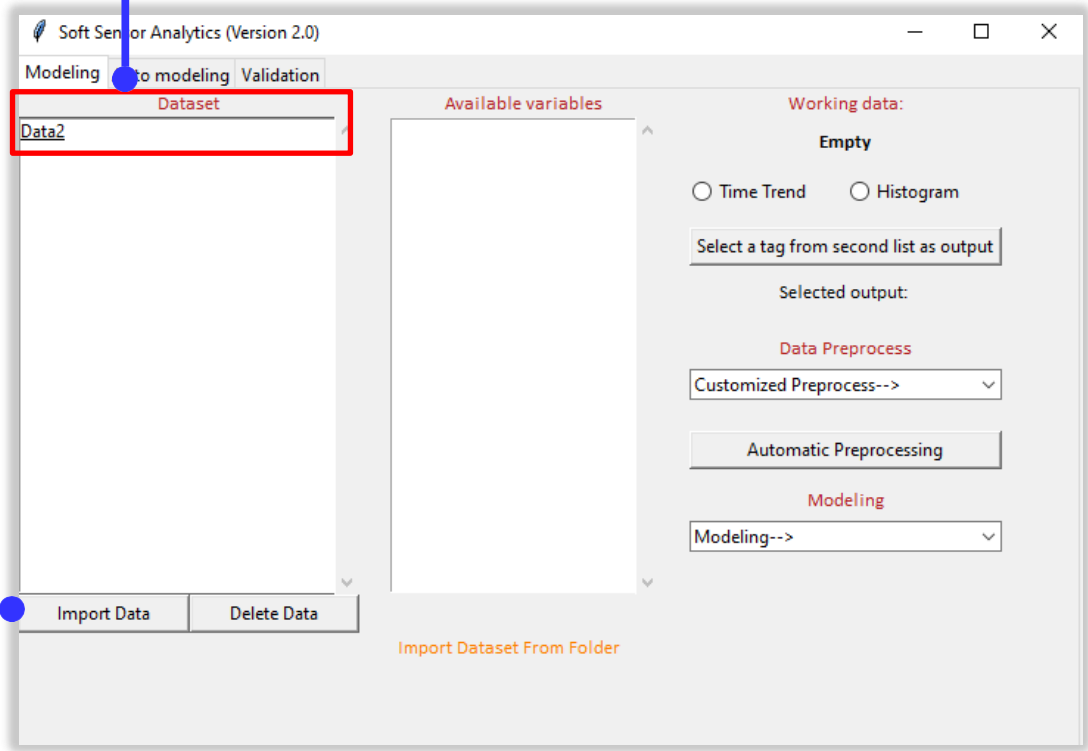
Note: Please save data sets in the ".xlsx" format

1

- **Click "Import Data"** button
- **Select** data file **"Data2.xlsx"**
- **Click "Import"** button



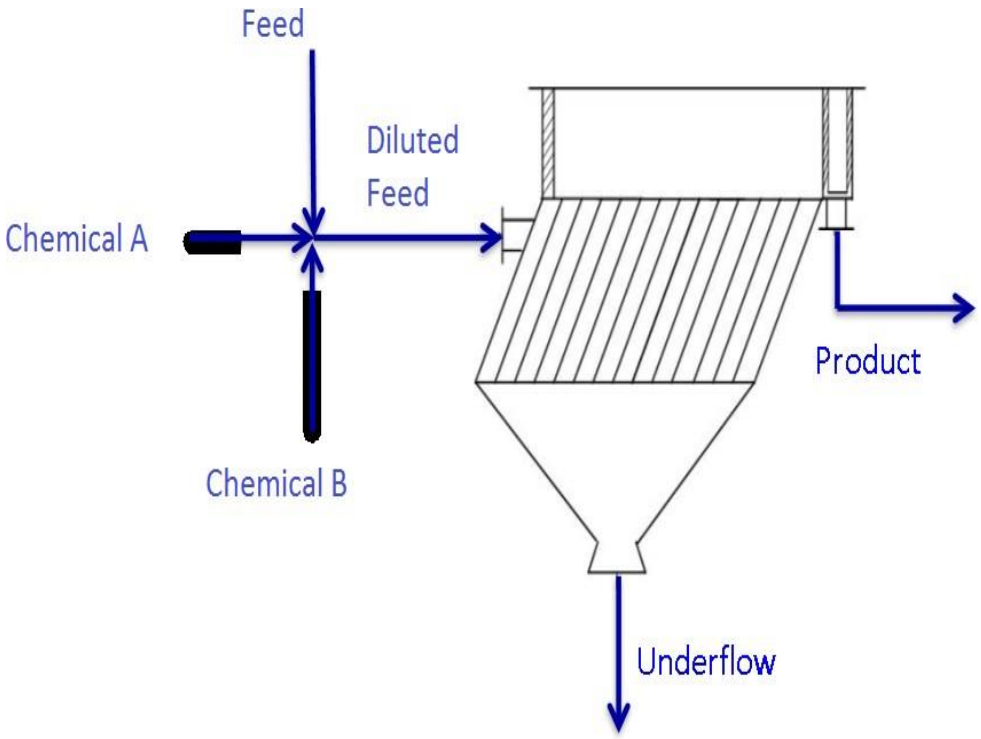
"**Data2**" has been imported



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".



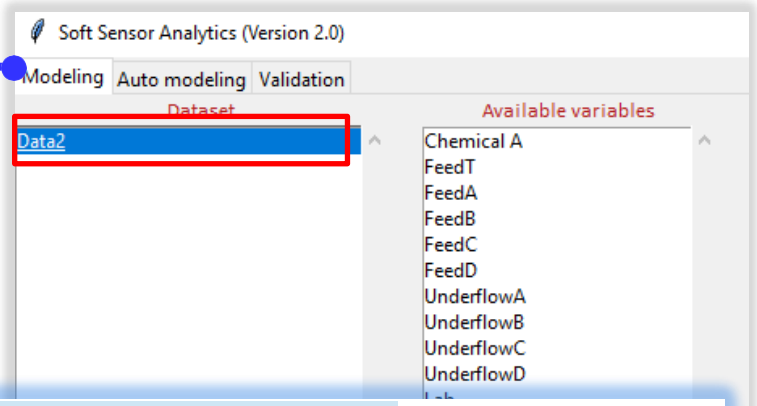
Symbol	Description
Chemical A	Chemical A flowrate
FeedA	Unit A Feed
FeedB	Unit B Feed
FeedC	Unit C Feed
FeedD	Unit D Feed
UnderflowA	Unit A underflow
UnderflowB	Unit B underflow
UnderflowC	Unit C underflow
UnderflowD	Unit D underflow
Lab	Chemical ratio value
DenN	Chemical A Density
DenFr	Feed Density

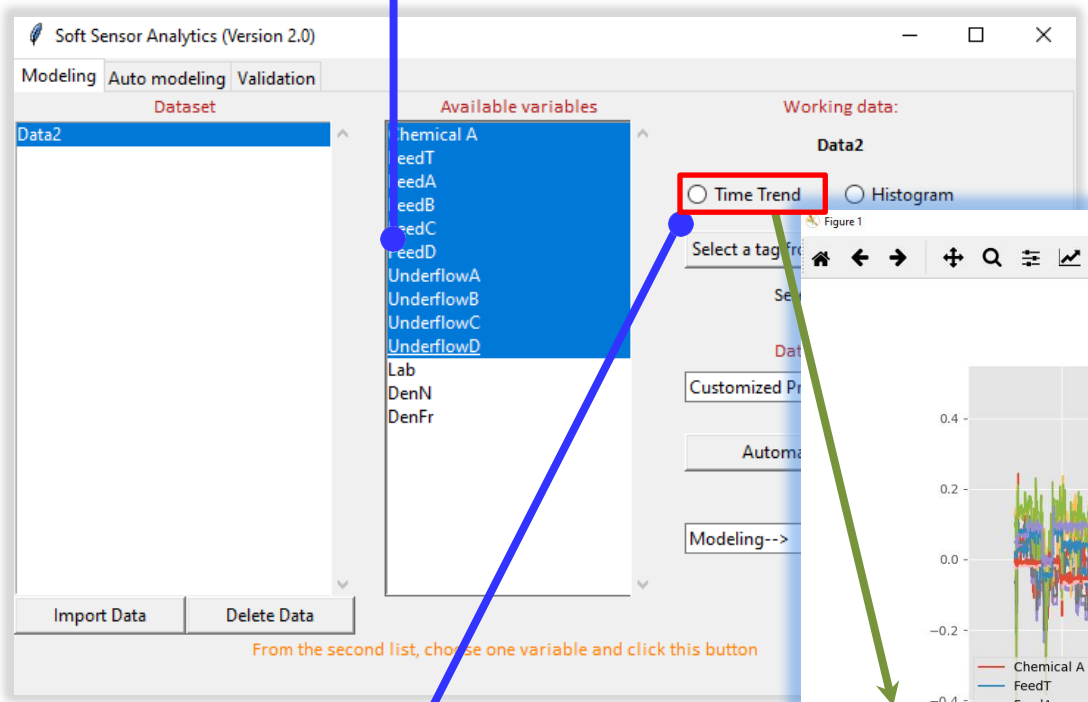
Soft Sensor Analytics

Visualize the selected data

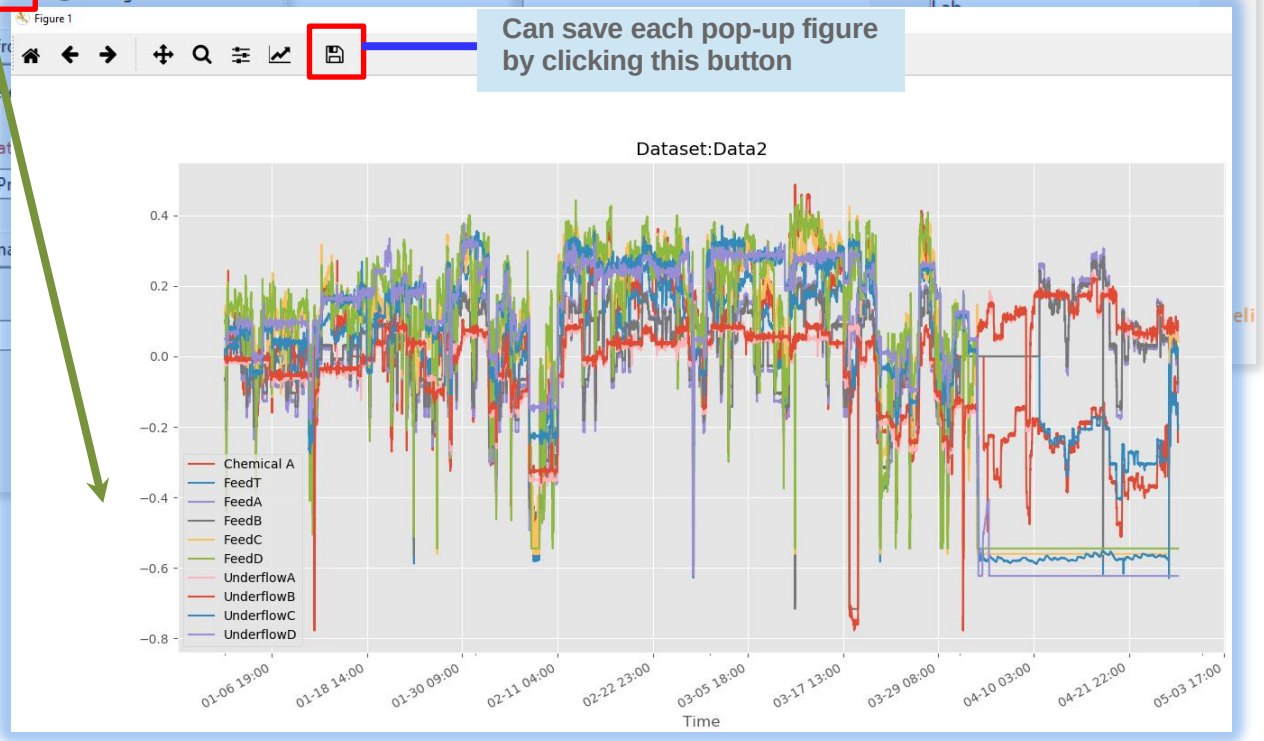
3 **Drag-select** to select variables for visualization
Hold "CTRL" to multi-select if they are not consecutive

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

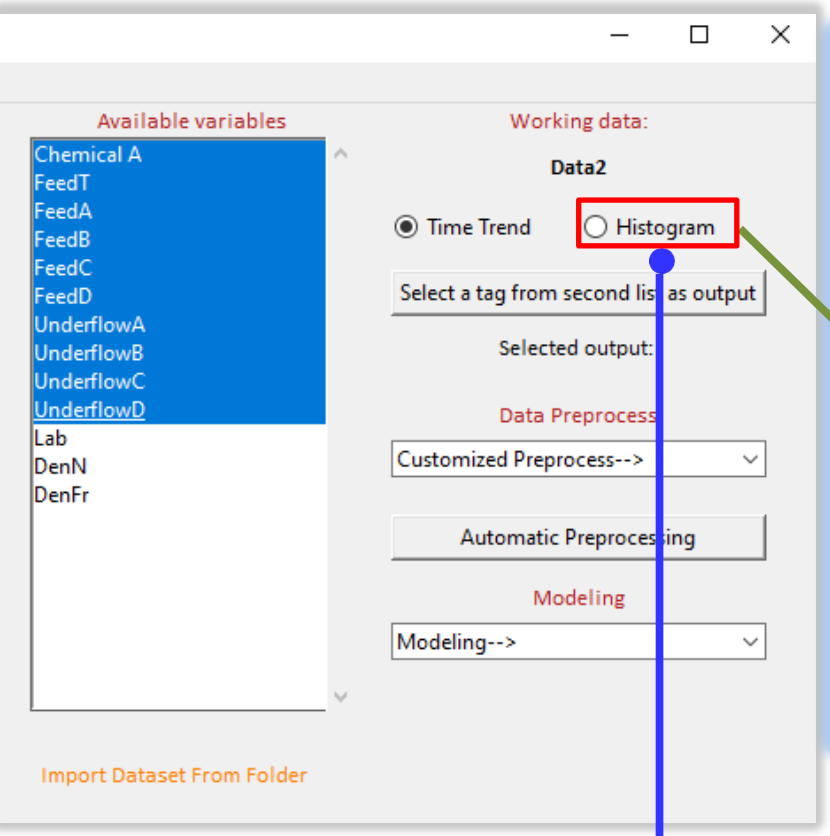


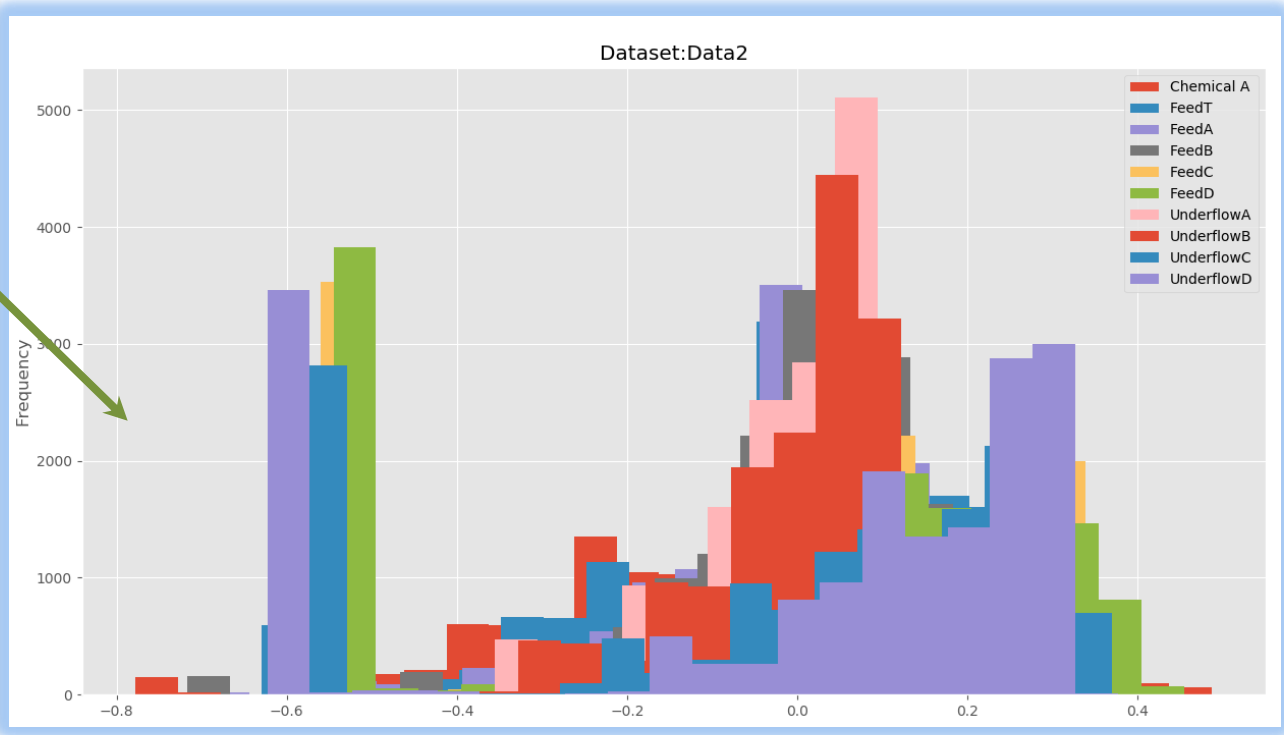
4 **Click "Time Trend"** to visualize the selected variables

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

Soft Sensor Analytics

Visualize the selected data





6

Click "Histogram" to generate histogram for selected variables

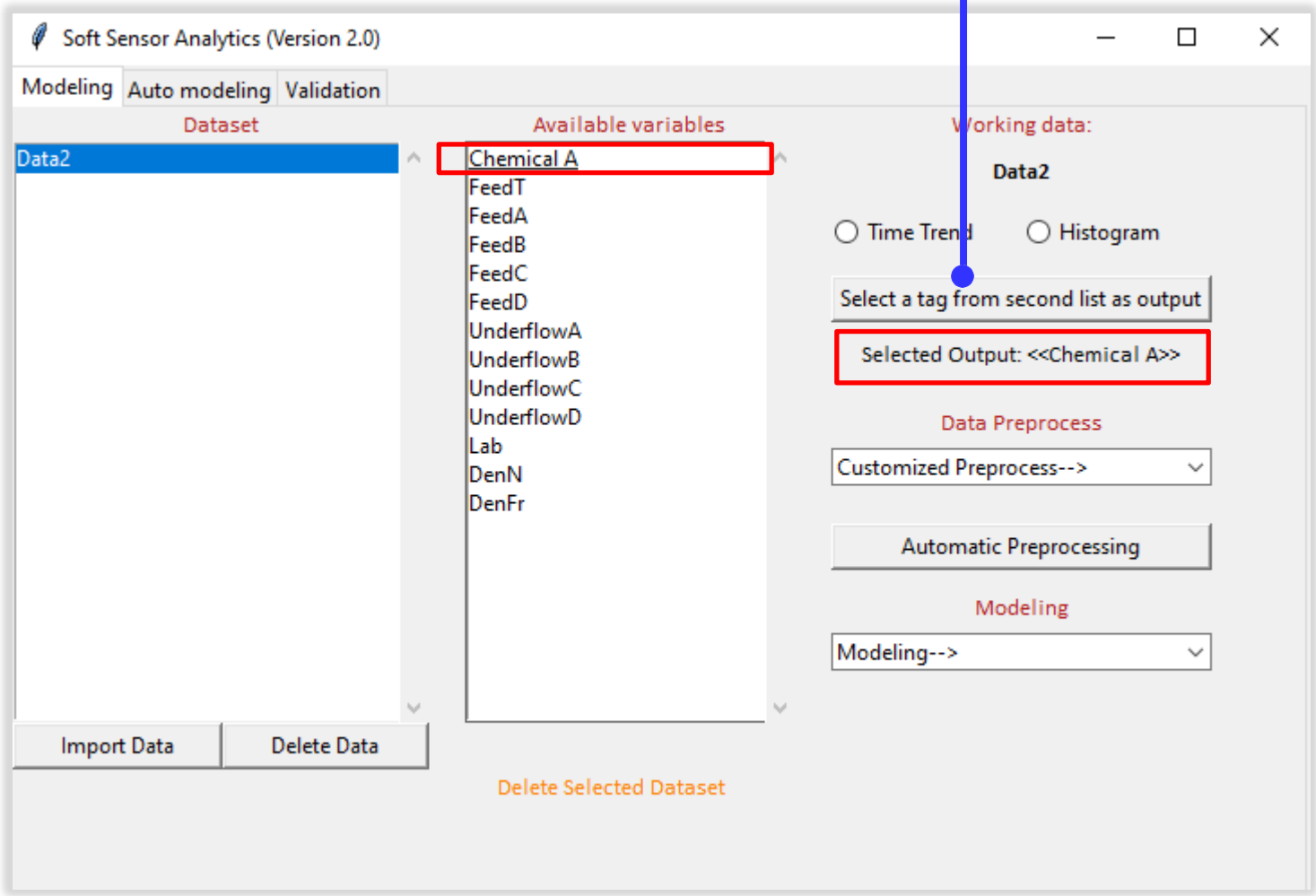
Soft Sensor Analytics

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



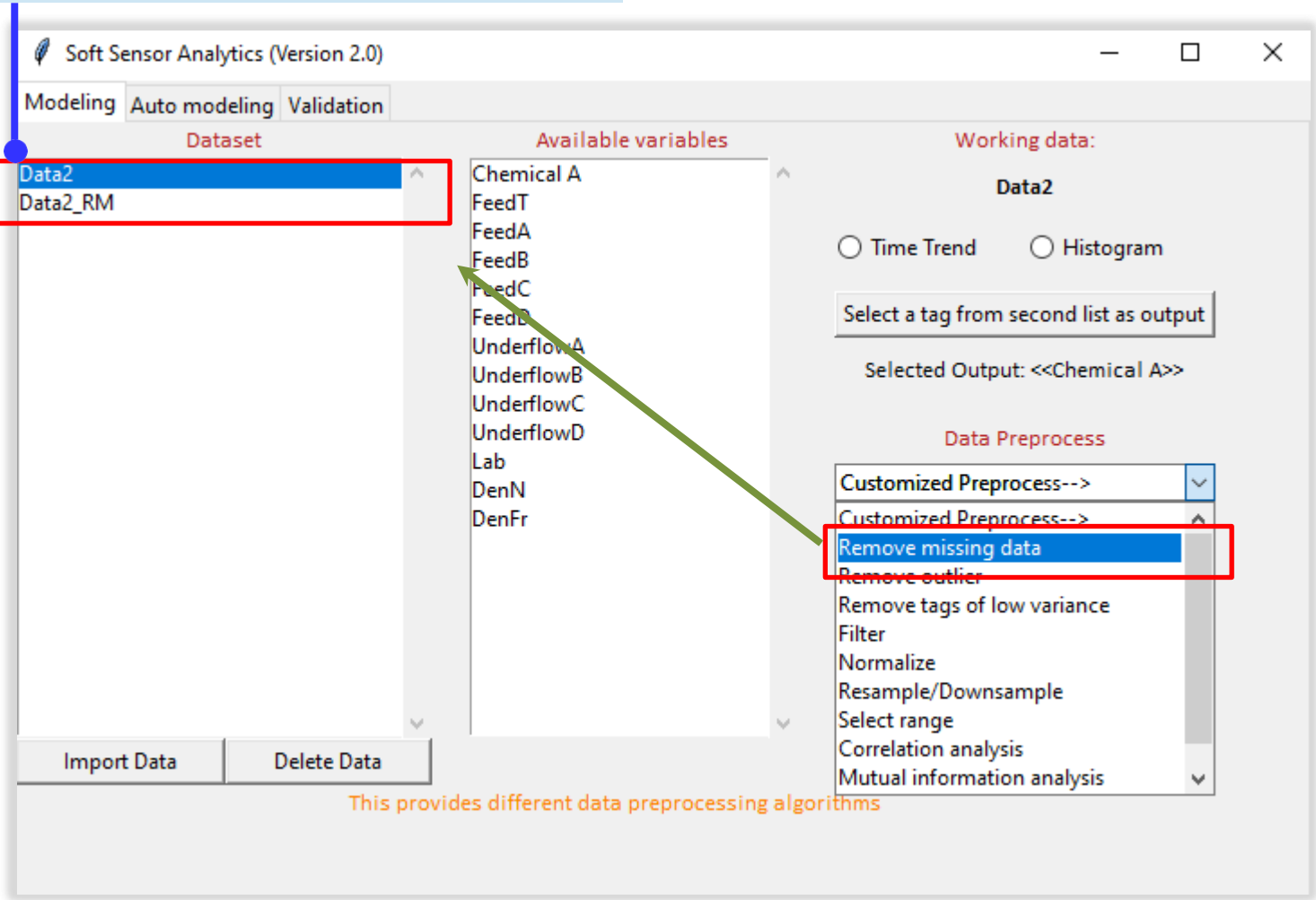
Soft Sensor Analytics

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"



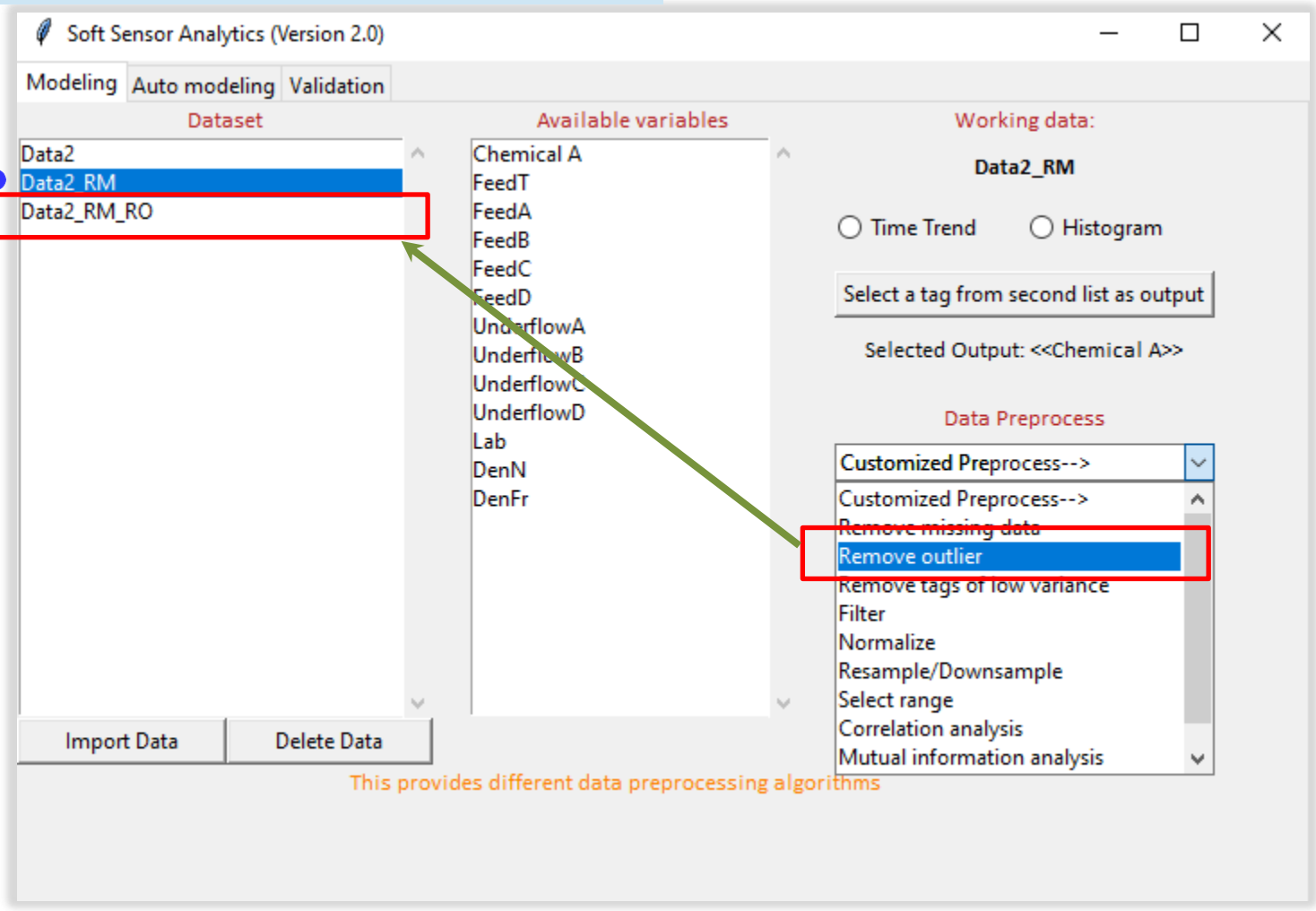
Soft Sensor Analytics

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"



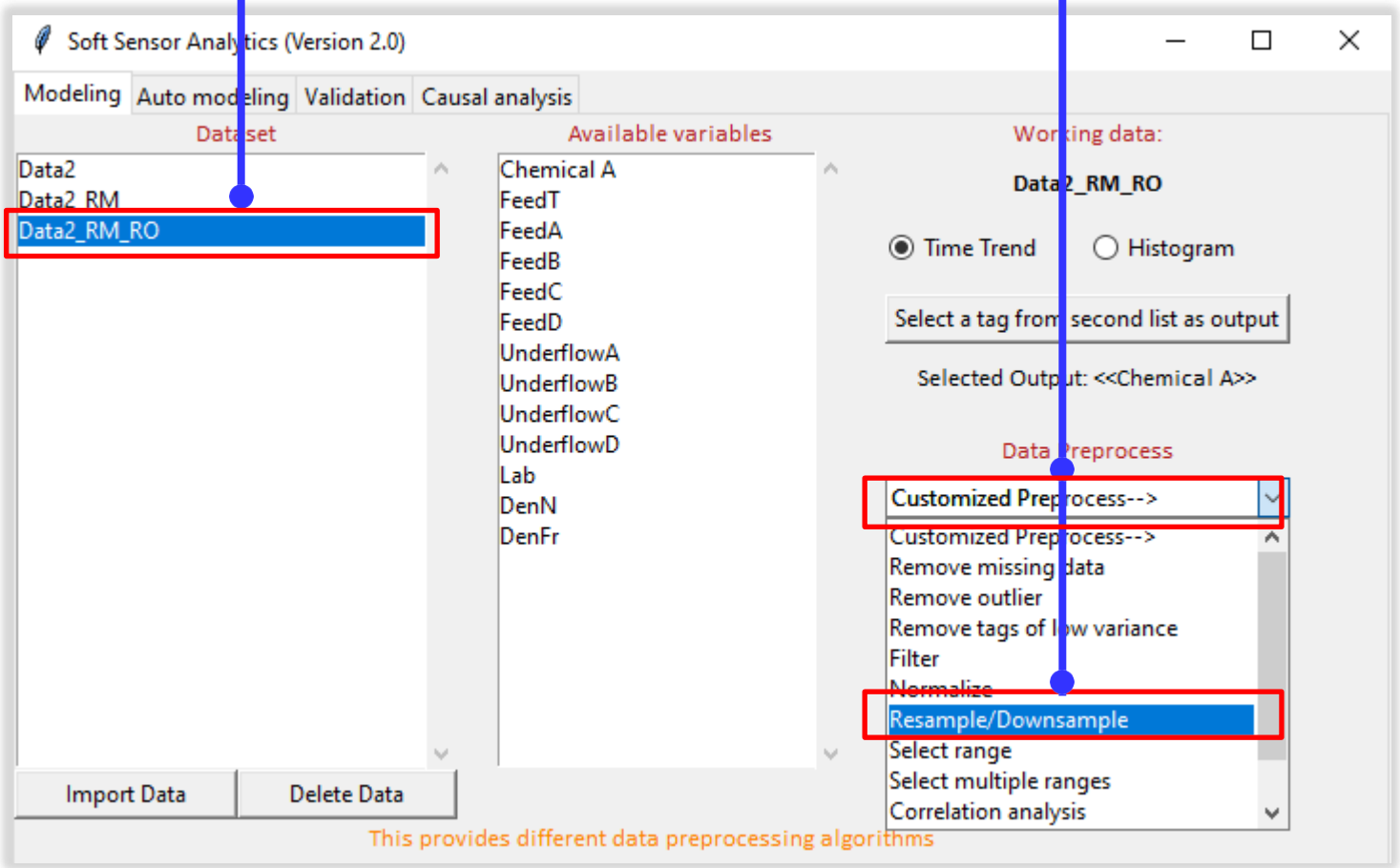
Soft Sensor Analytics

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

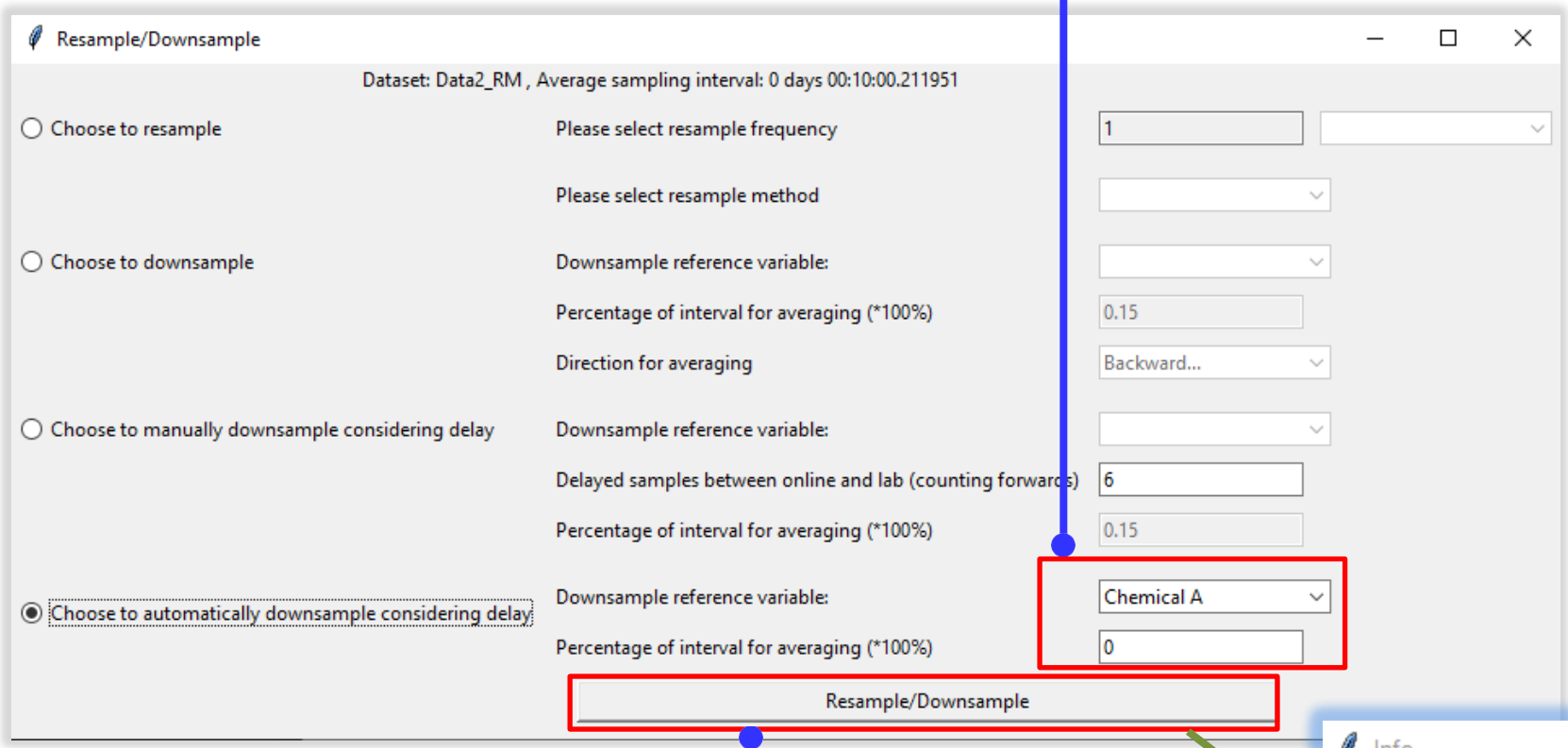


Soft Sensor Analytics

Data preprocessing – Resample and estimate output delay

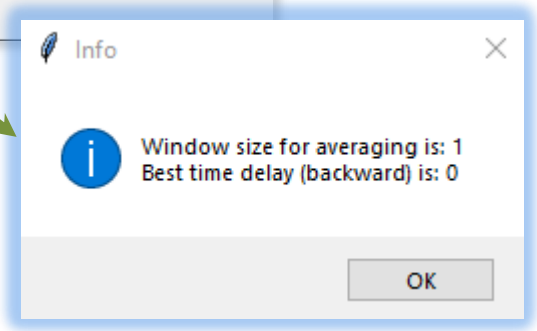
11

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



12

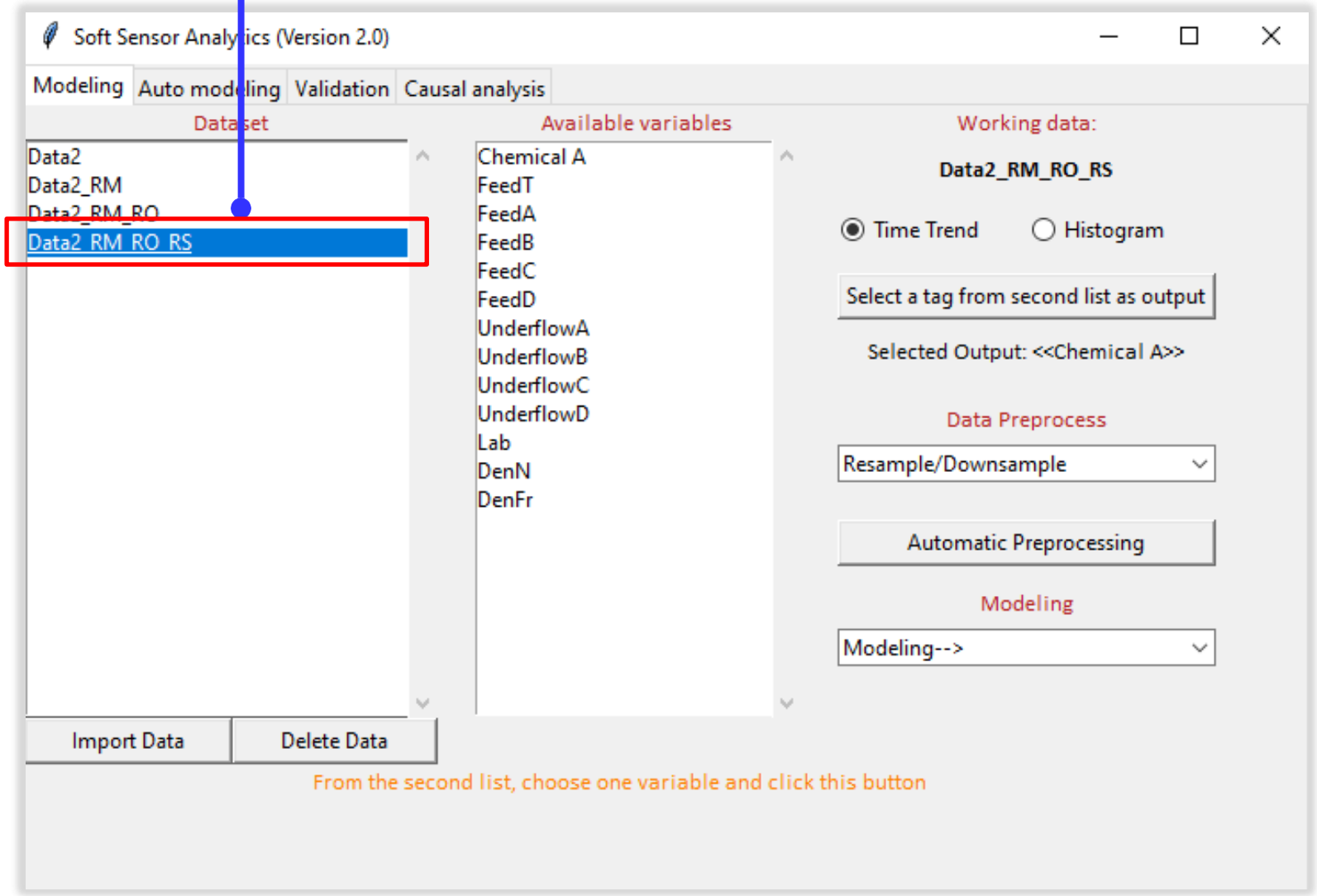
- **Click** "Resample/Downsample" button



Soft Sensor Analytics

Data preprocessing – Resample and estimate output delay

A new dataset **"Data2_RM_RO_RS"** is generated for modeling



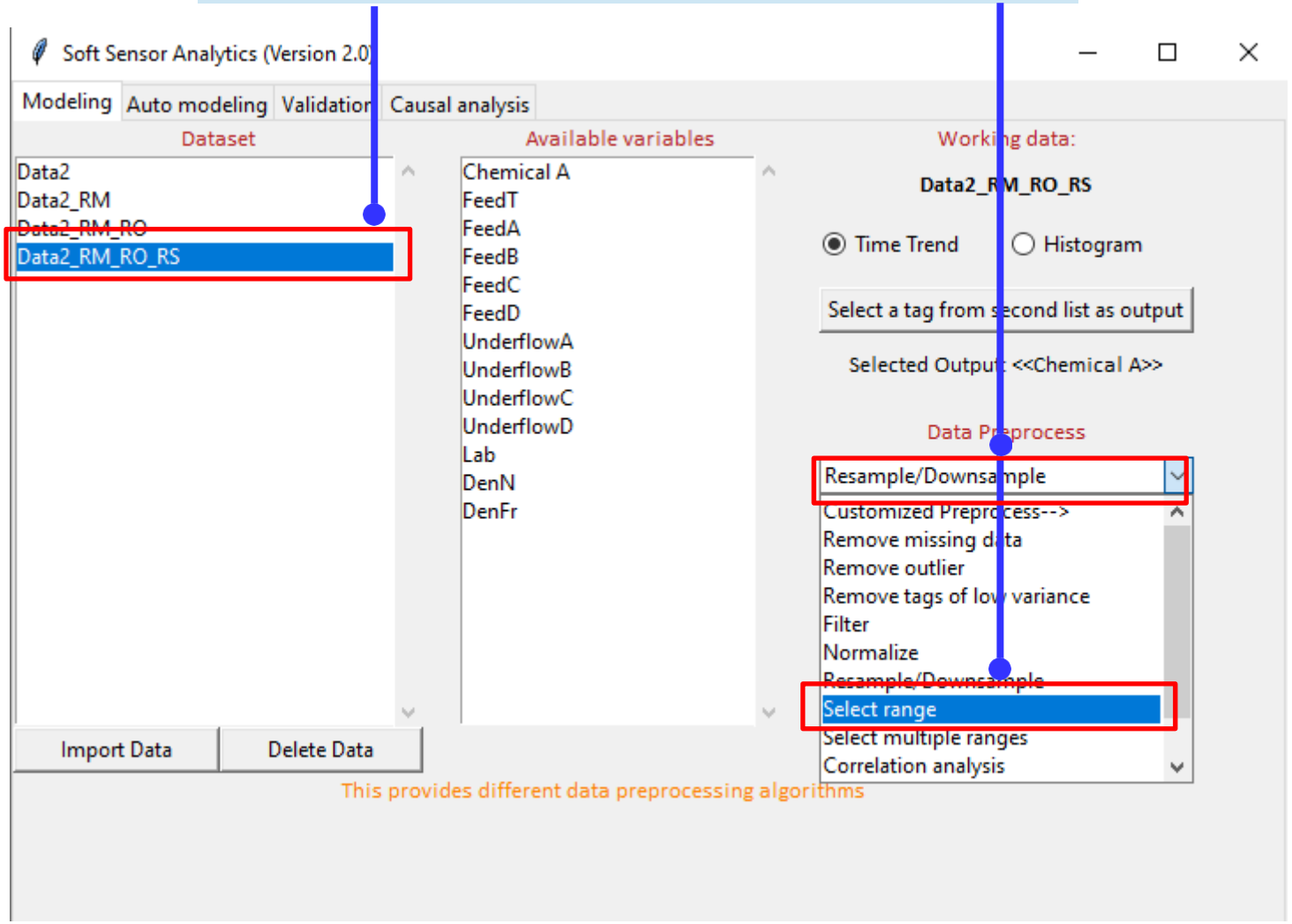
Soft Sensor Analytics

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



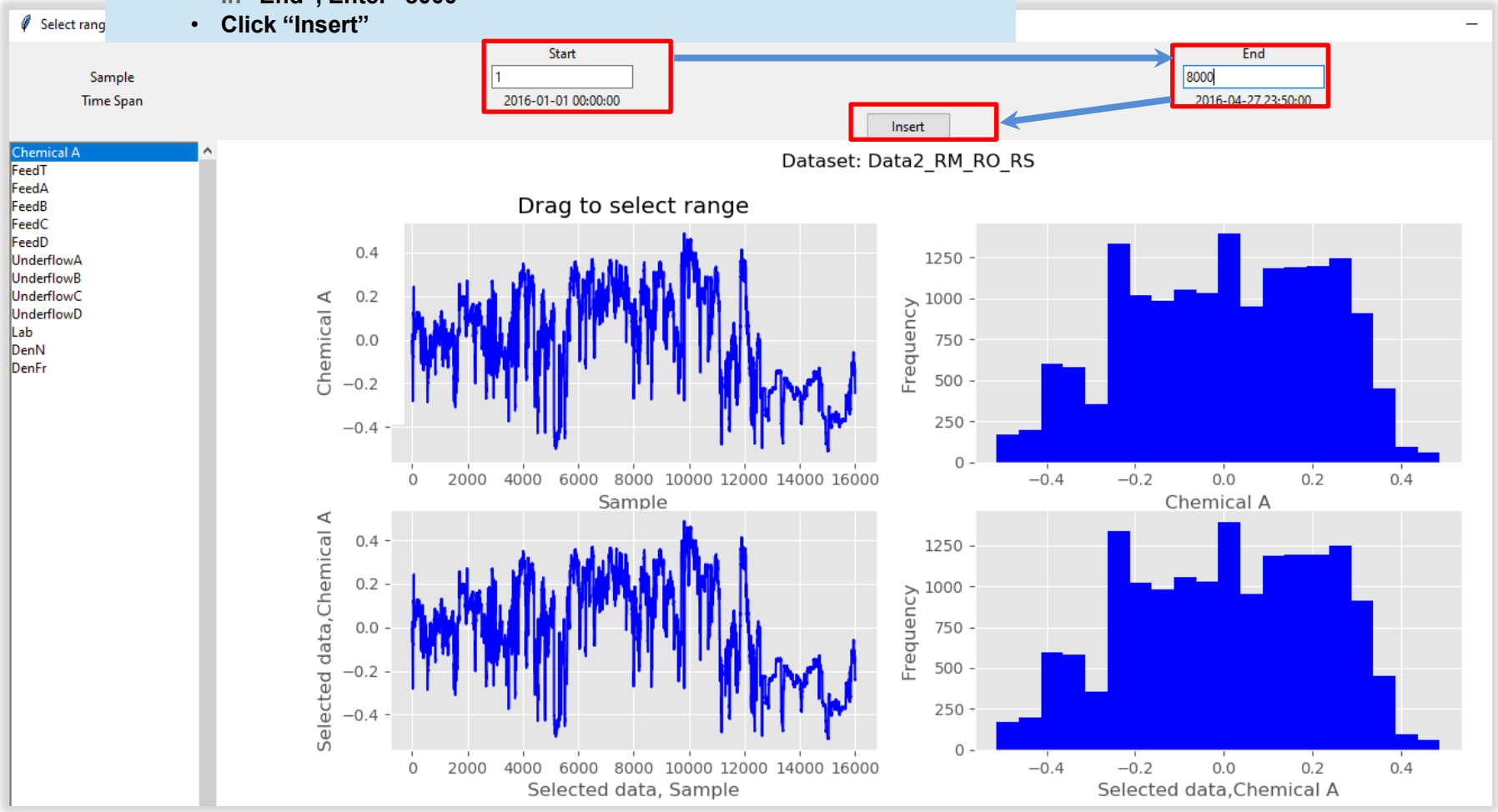
Soft Sensor Analytics

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"**



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

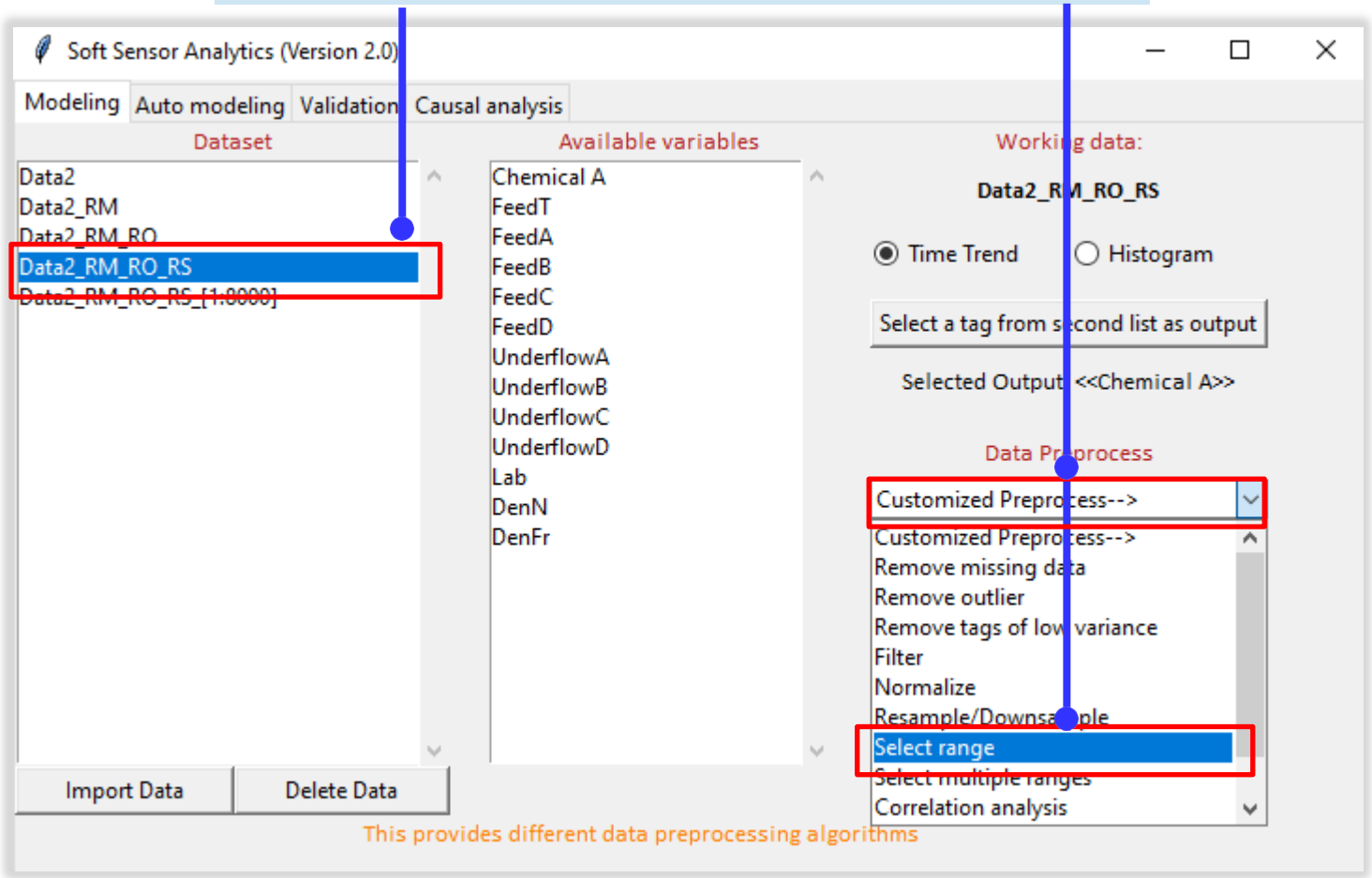
Soft Sensor Analytics

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



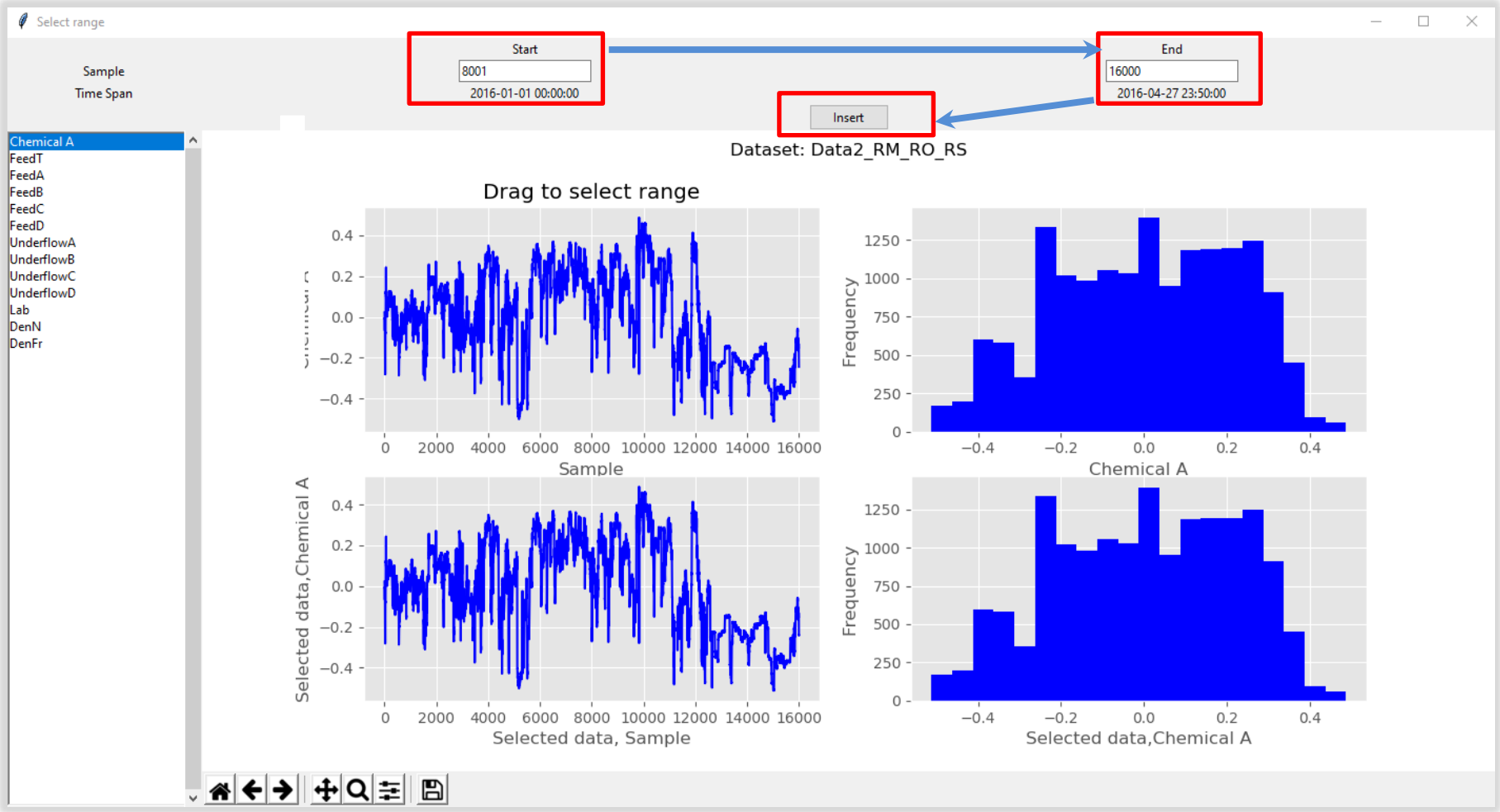
Soft Sensor Analytics

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"**

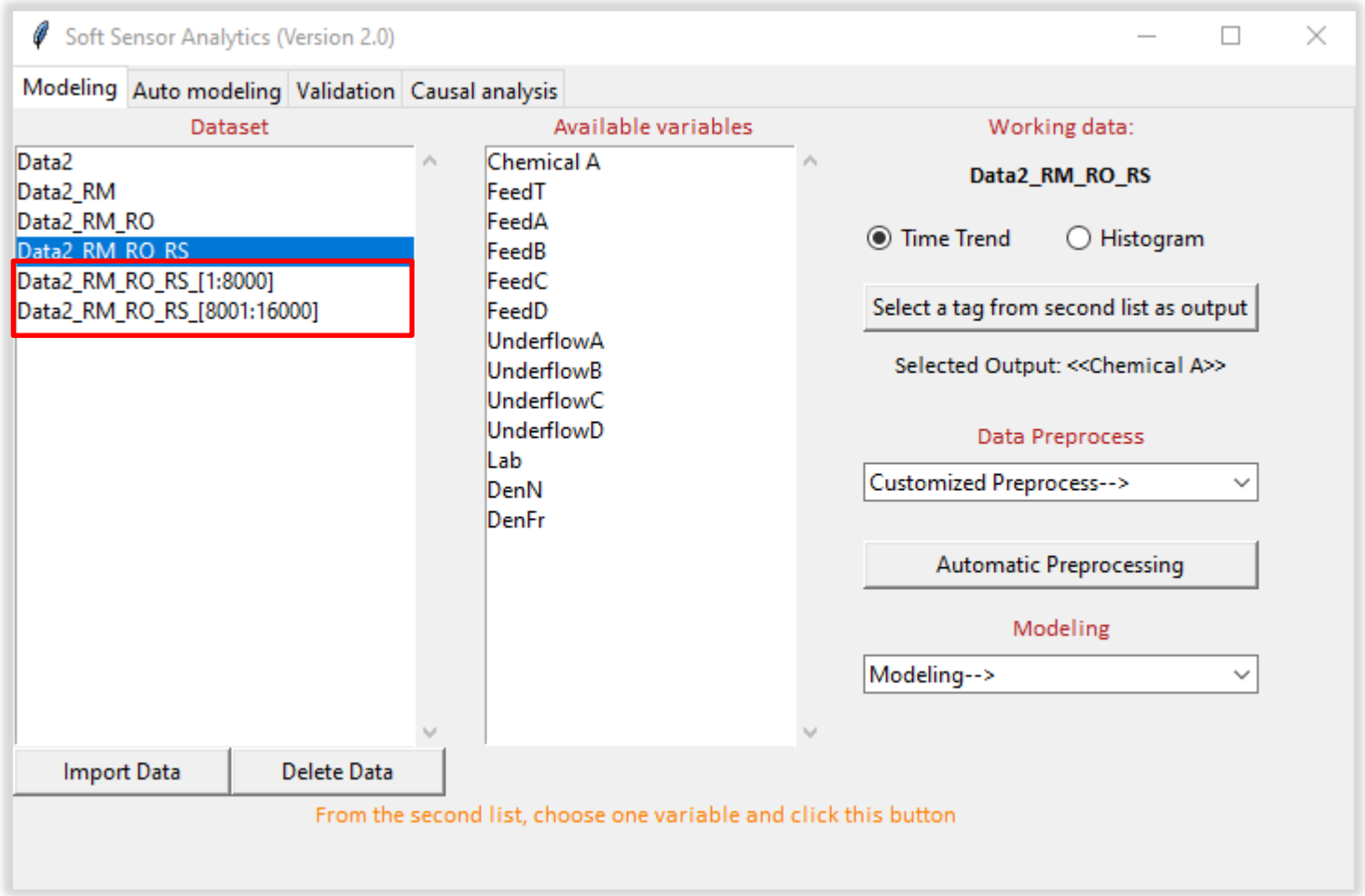


Note: Please close the current popup windows before taking next action

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



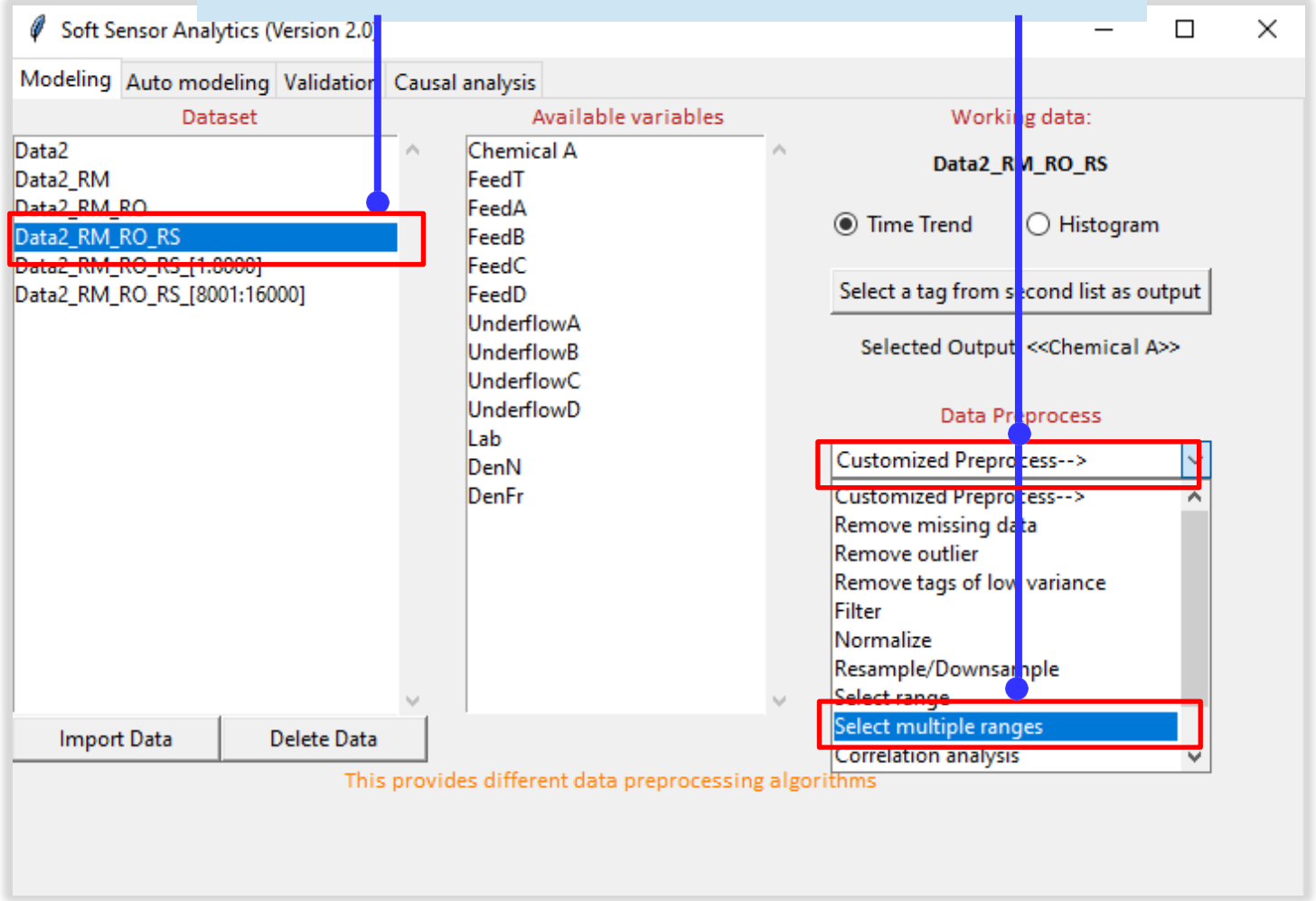
Soft Sensor Analytics

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.**

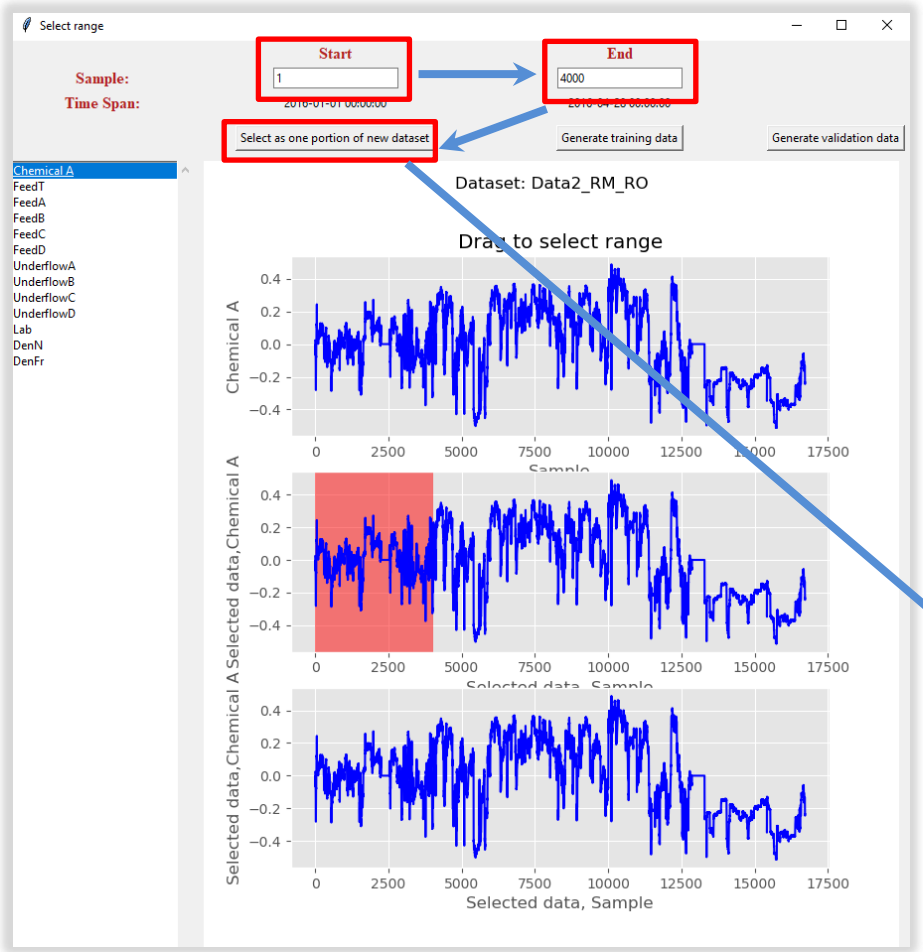


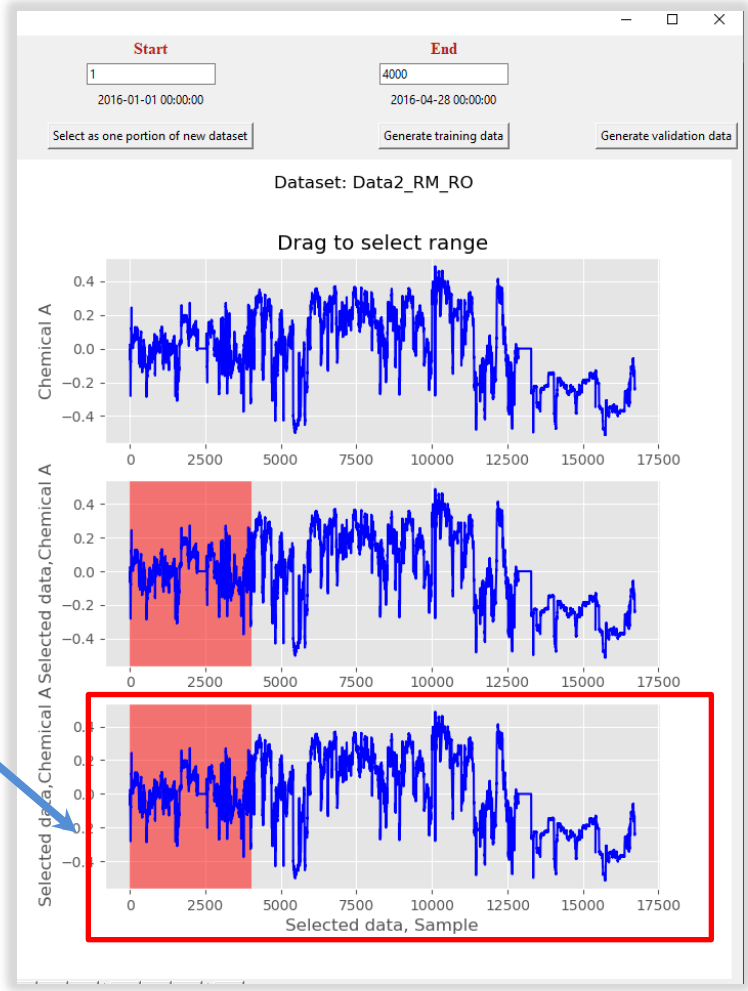
Soft Sensor Analytics

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"**



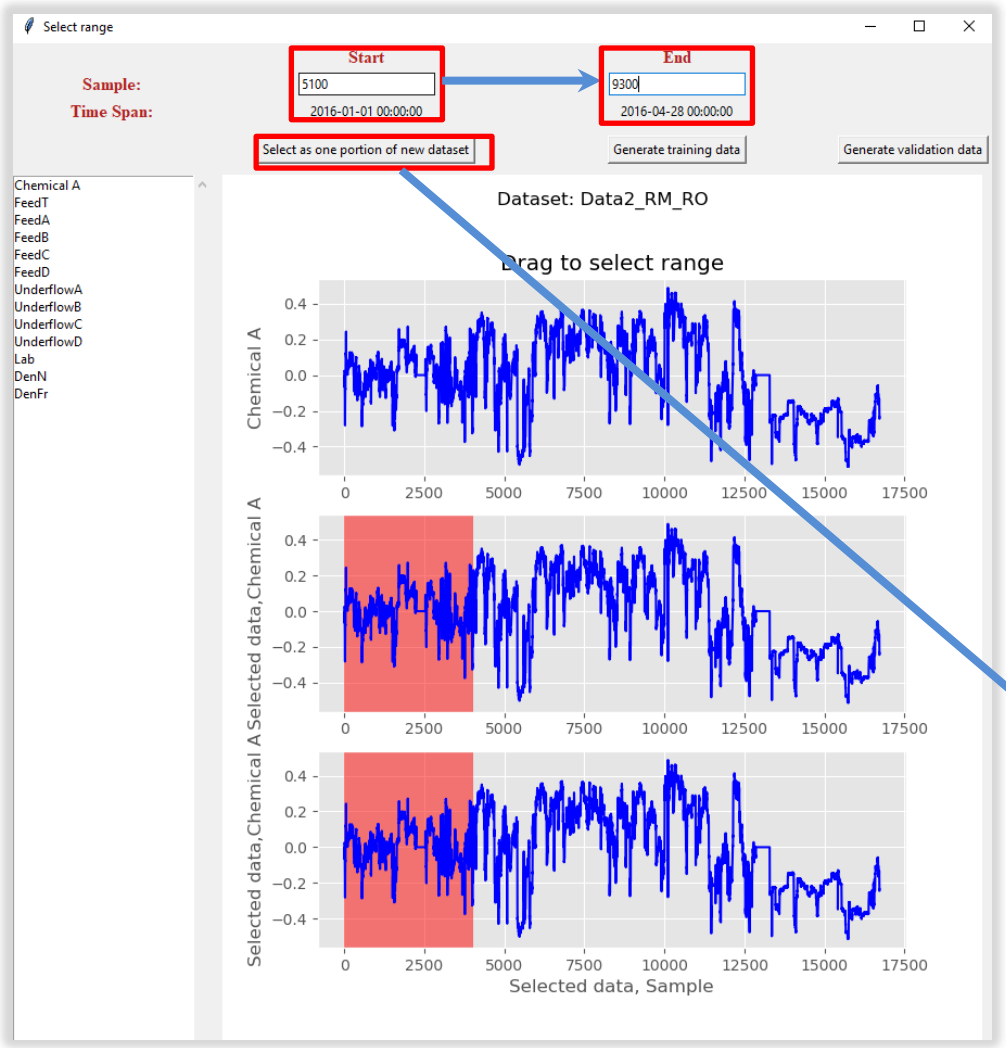


Soft Sensor Analytics

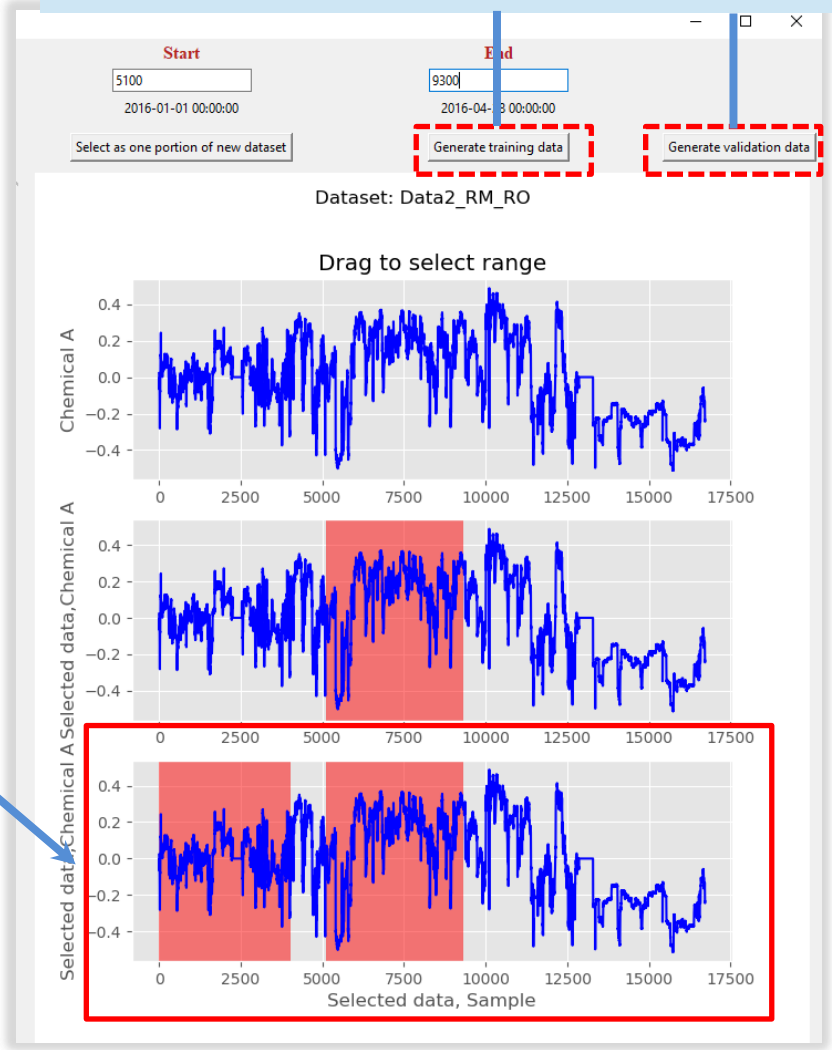
A more flexible way to generate datasets for training/validation

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.



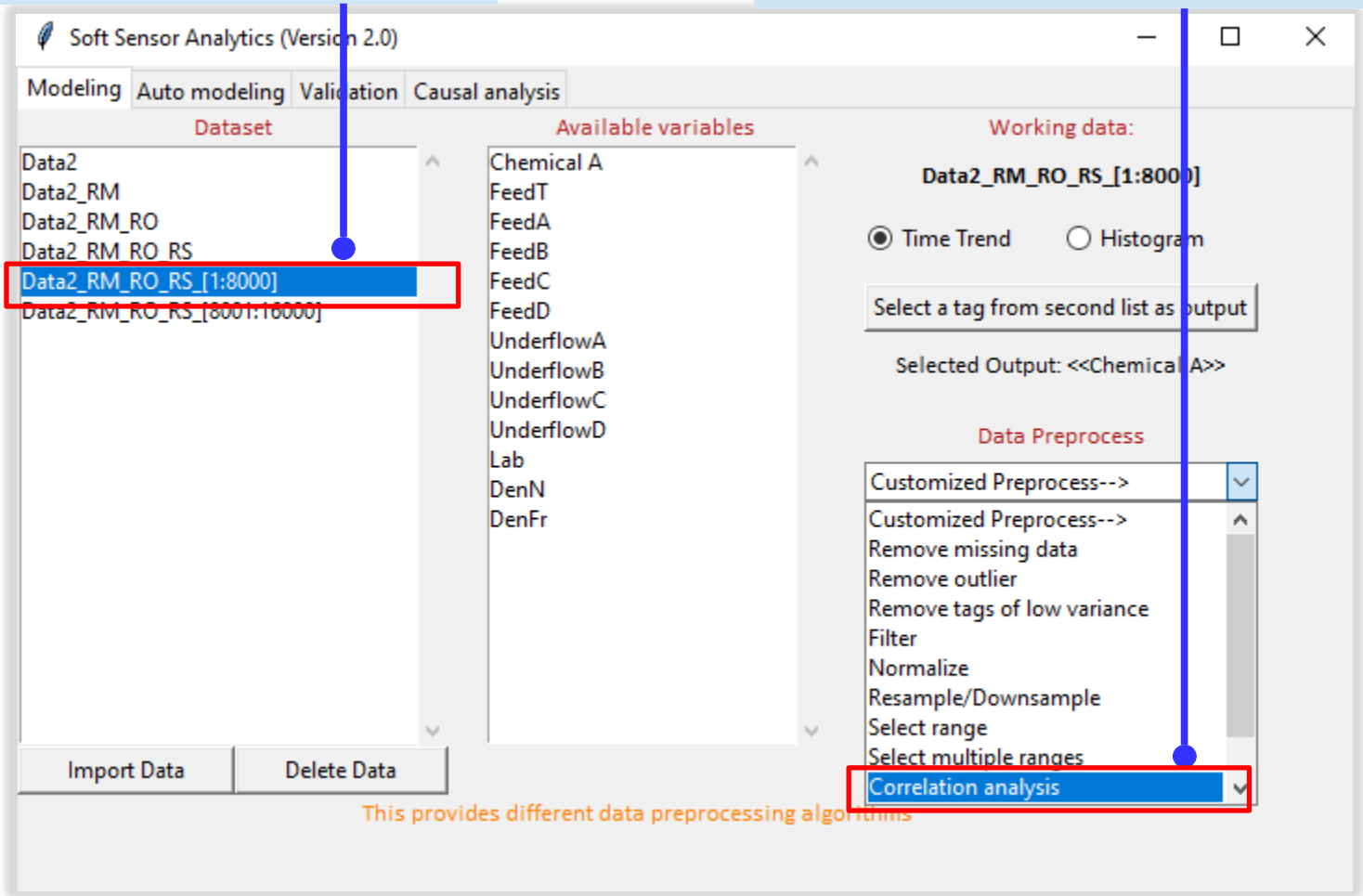
Soft Sensor Analytics

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



Soft Sensor Analytics

Preprocessing - Correlation analysis

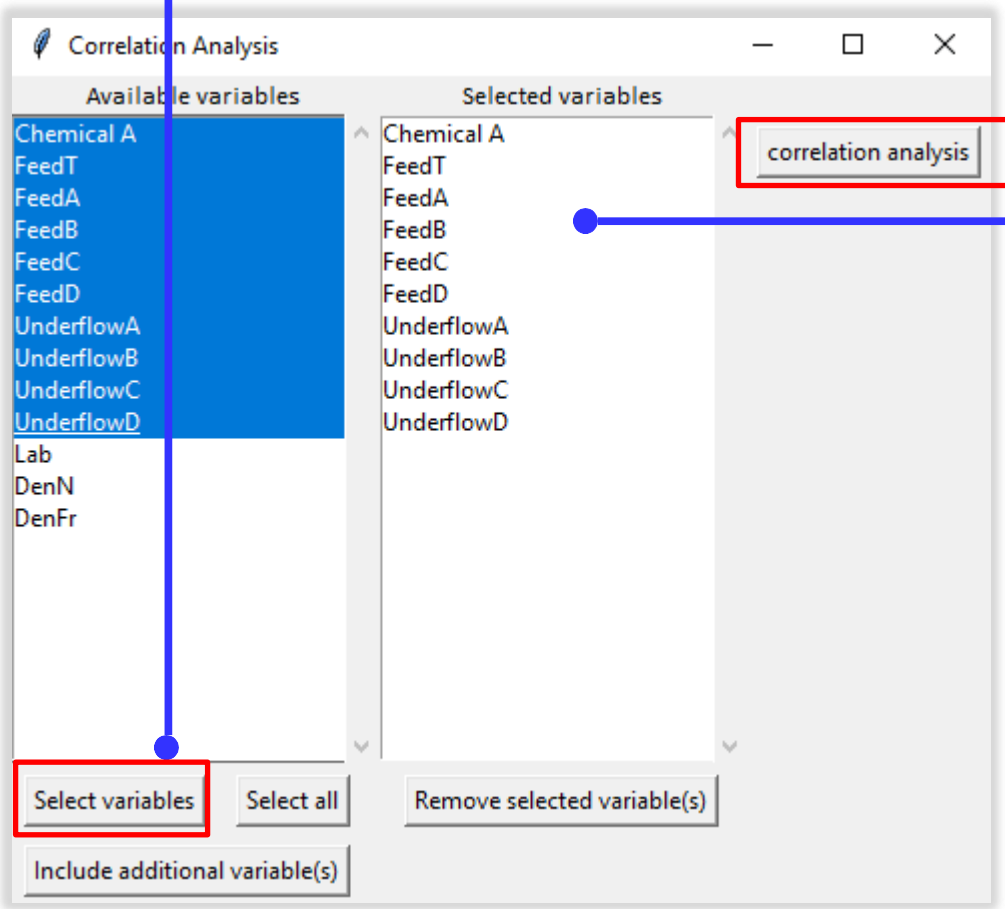
3

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

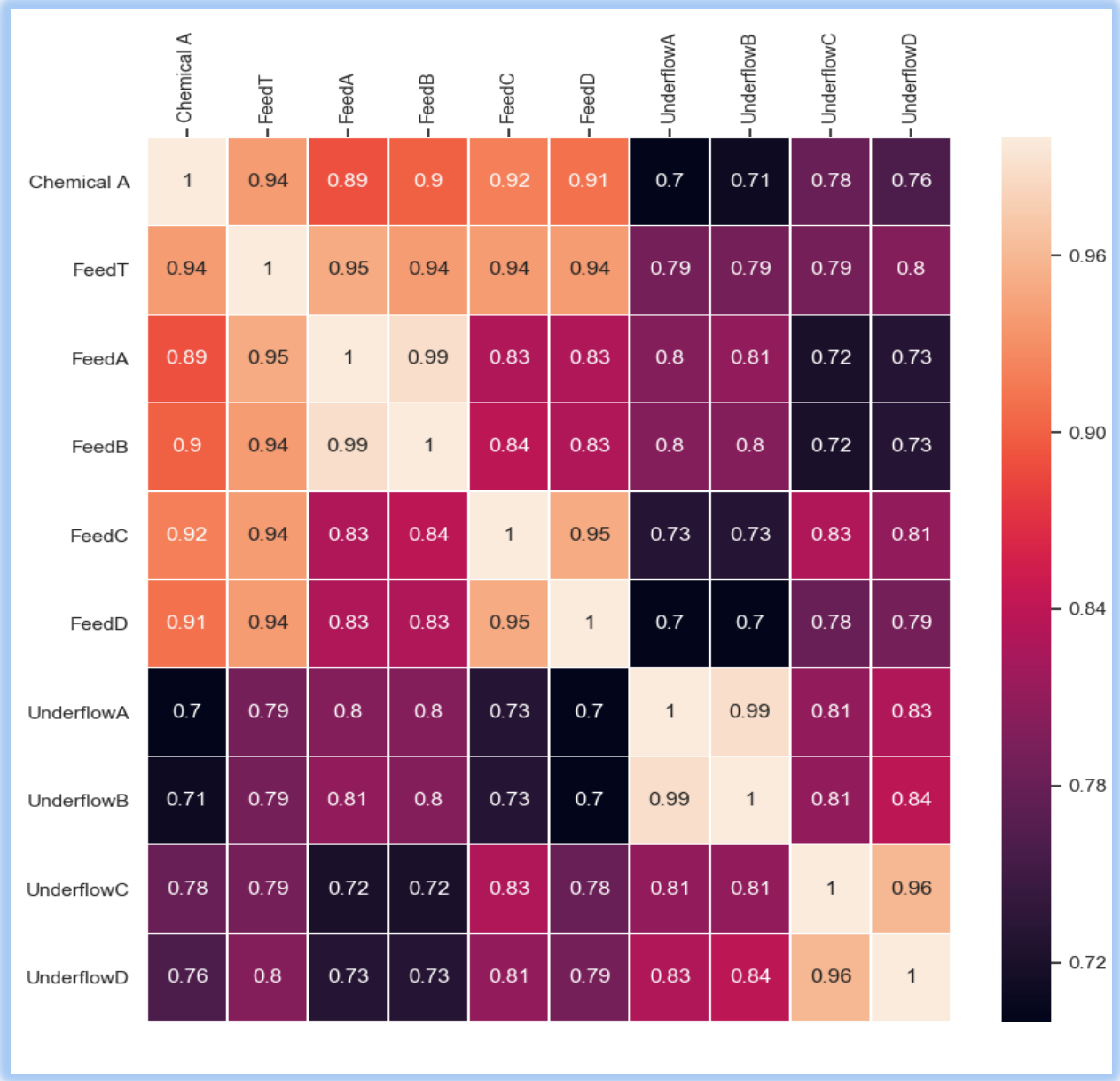
Soft Sensor Analytics

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

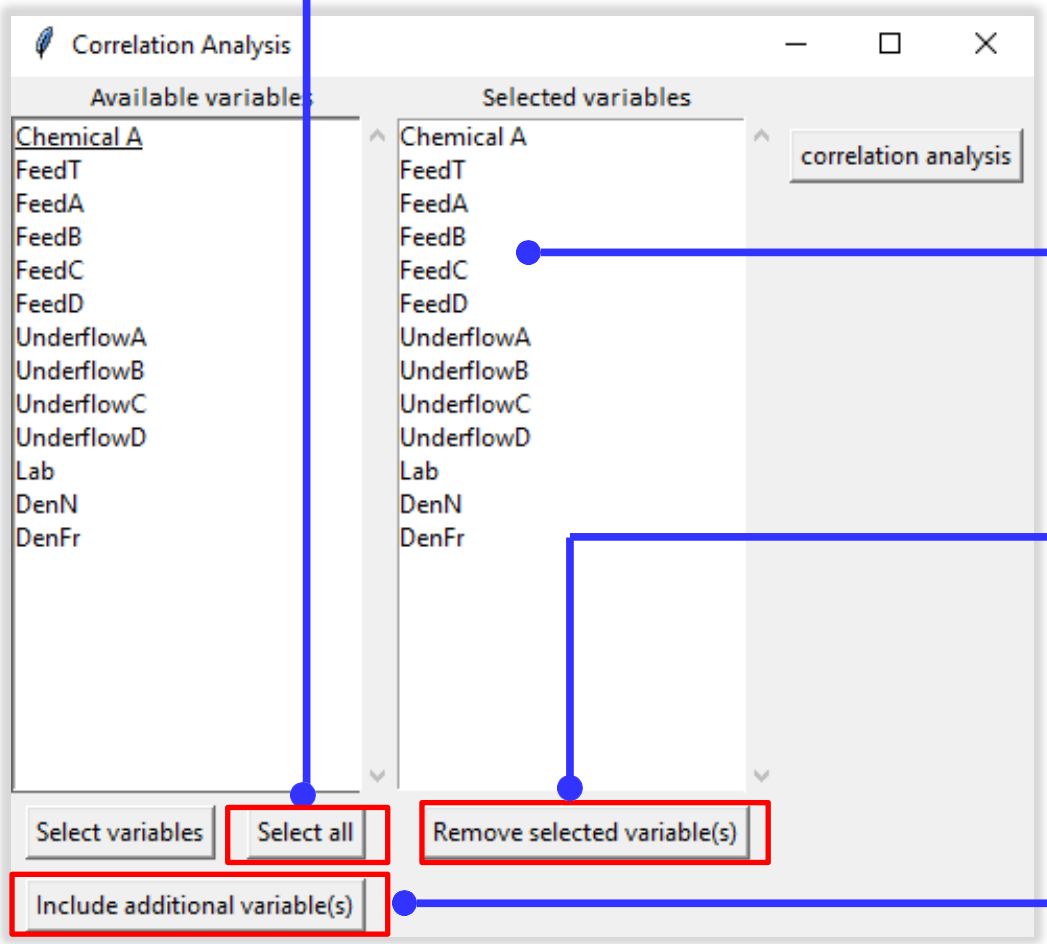
Soft Sensor Analytics

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.

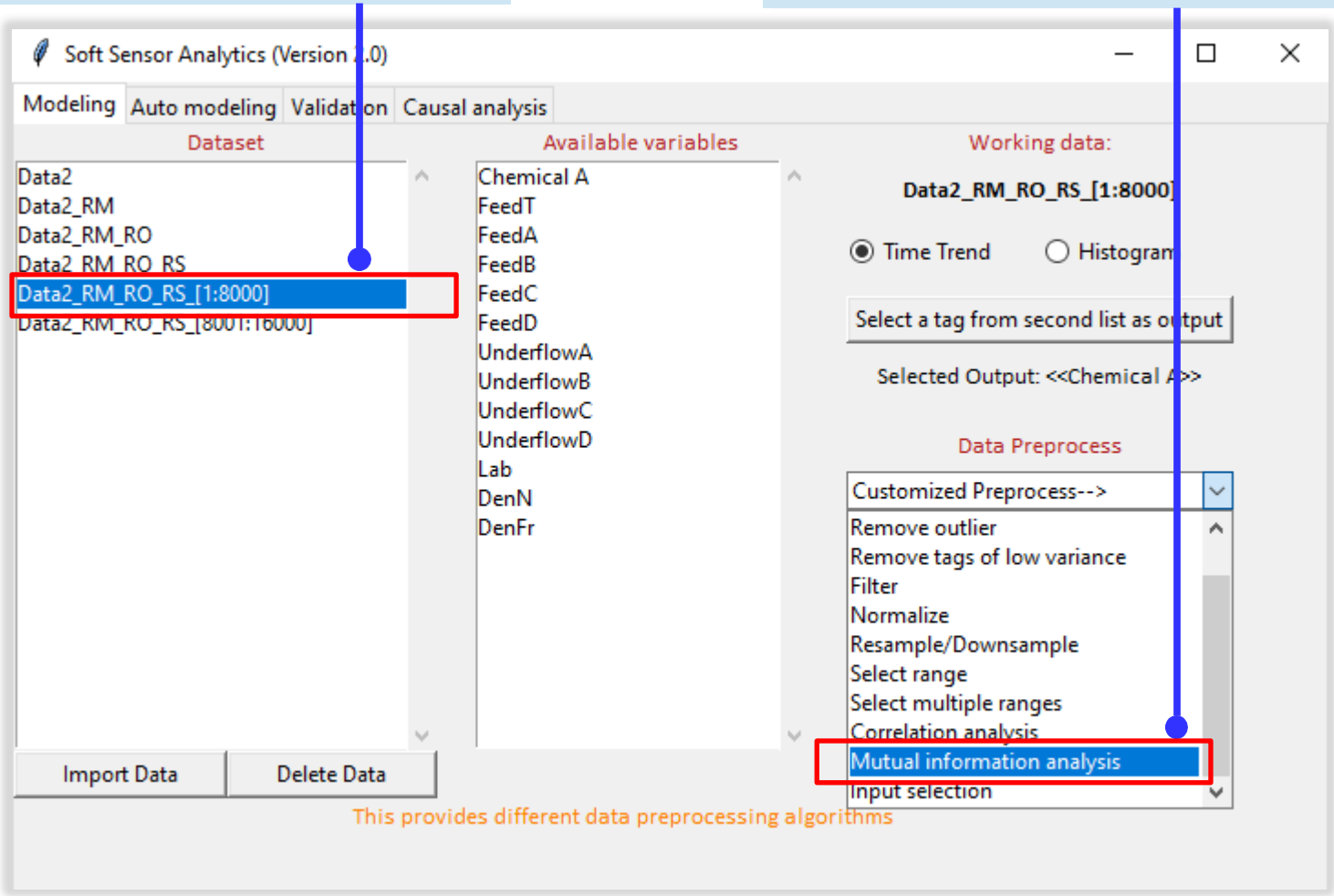
Soft Sensor Analytics

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

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.

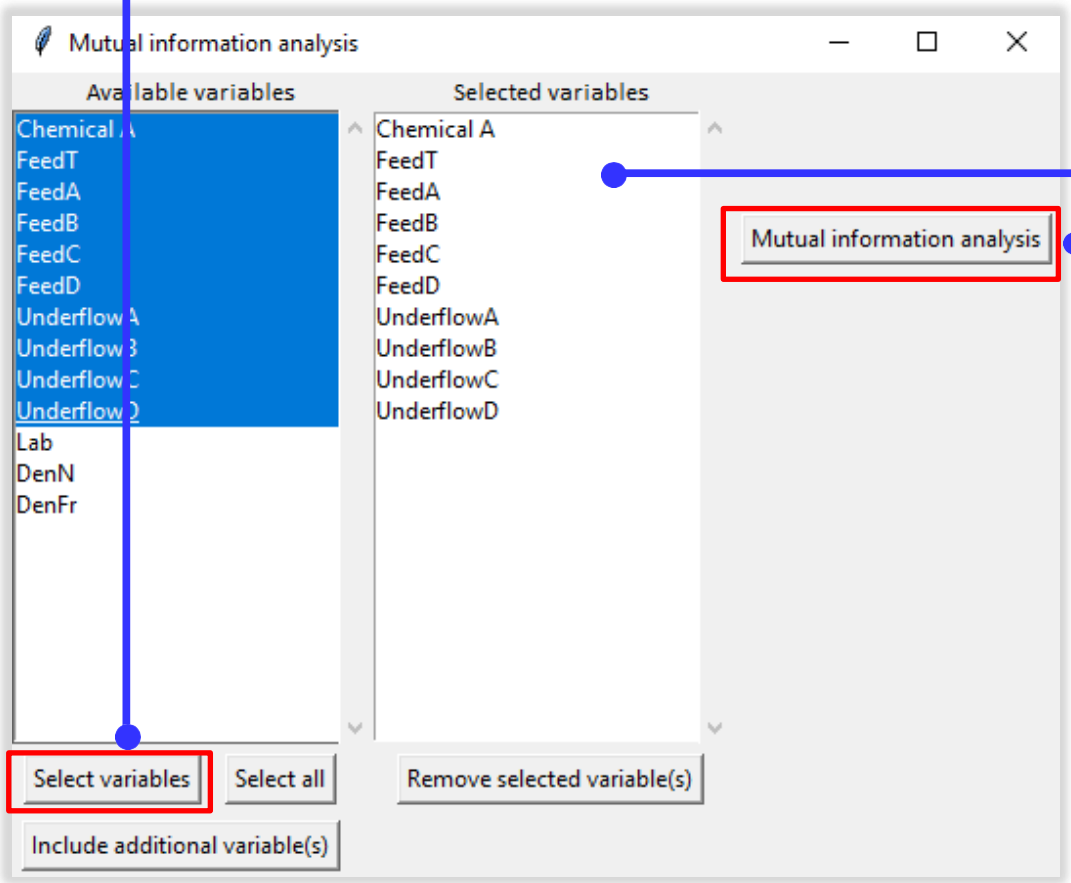
3

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



4

Mutual information analysis

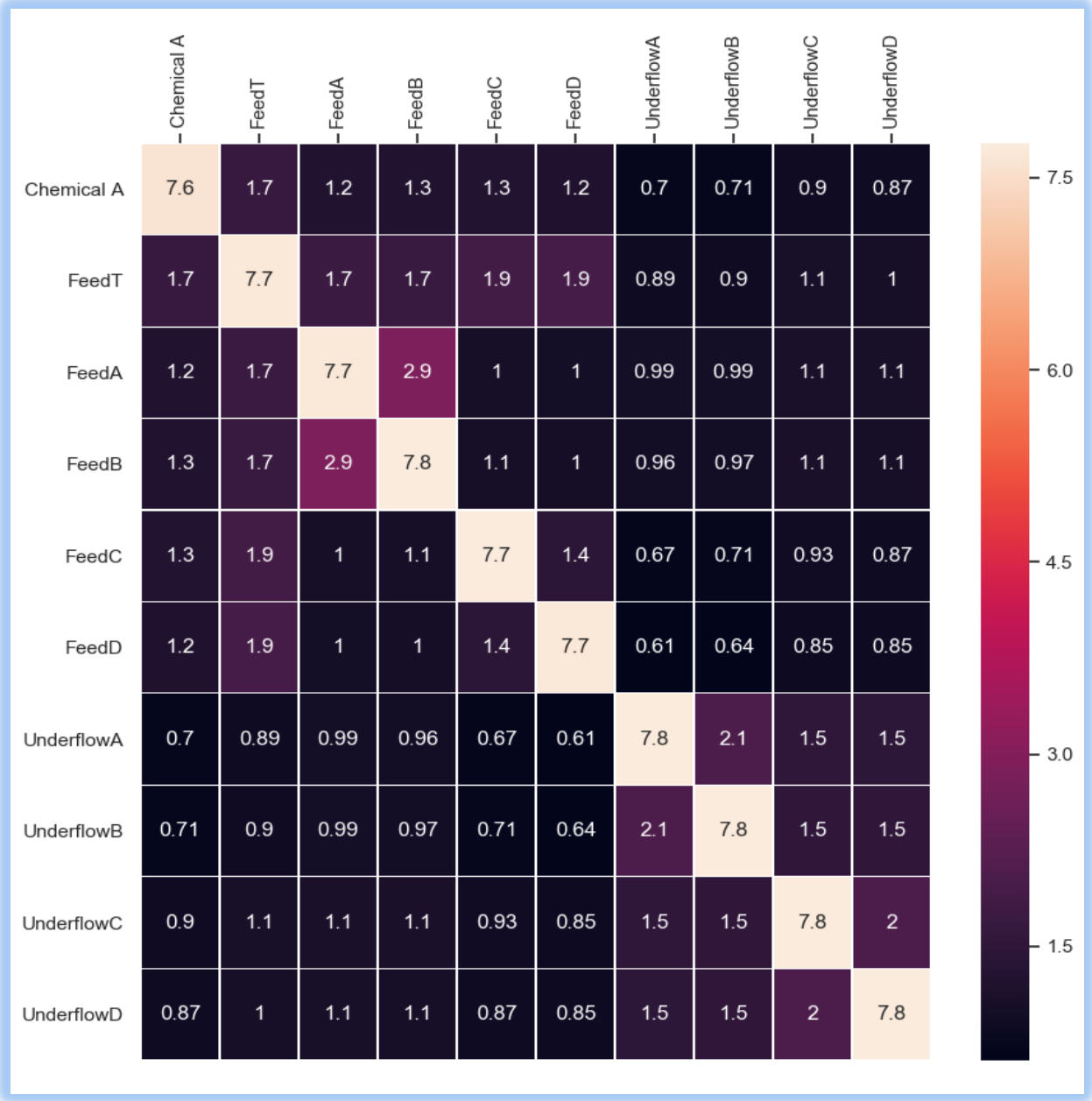
Then, Click "Mutual information analysis"

Soft Sensor Analytics

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

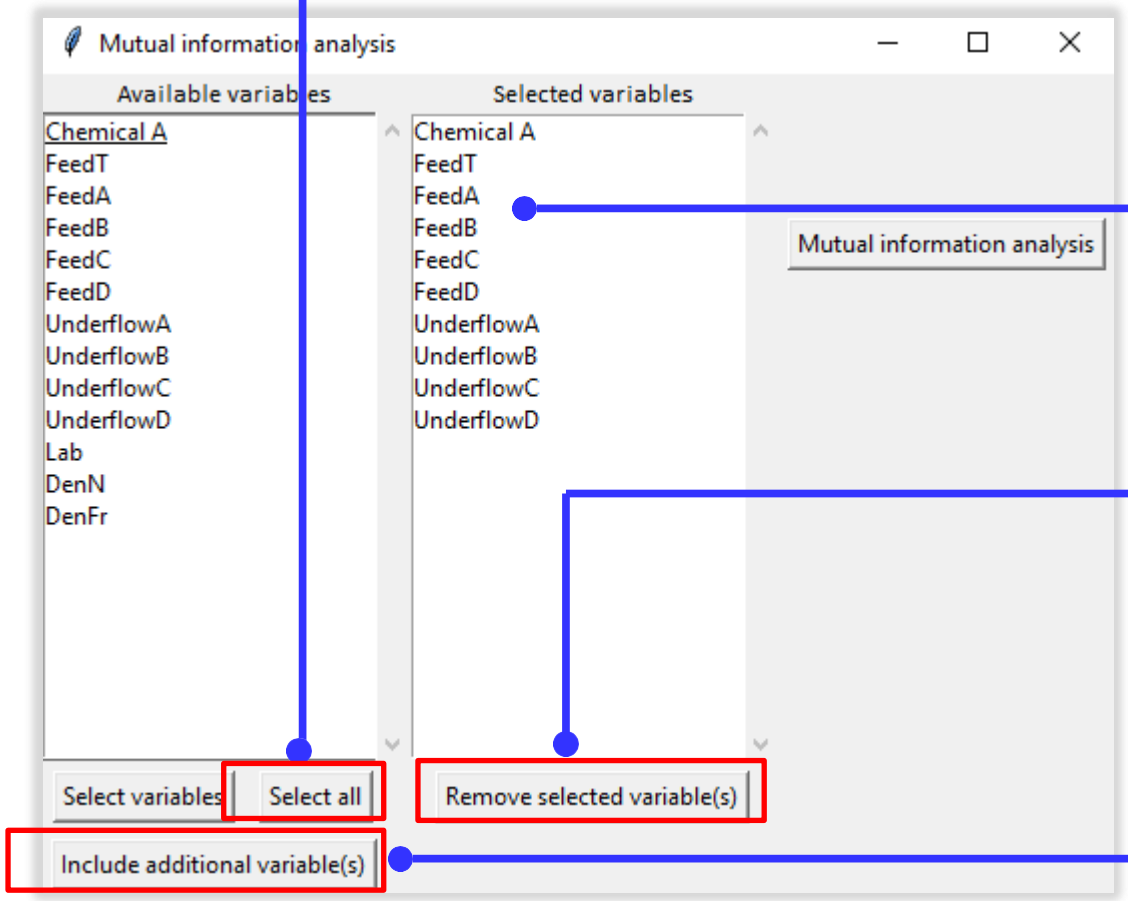
Soft Sensor Analytics

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.

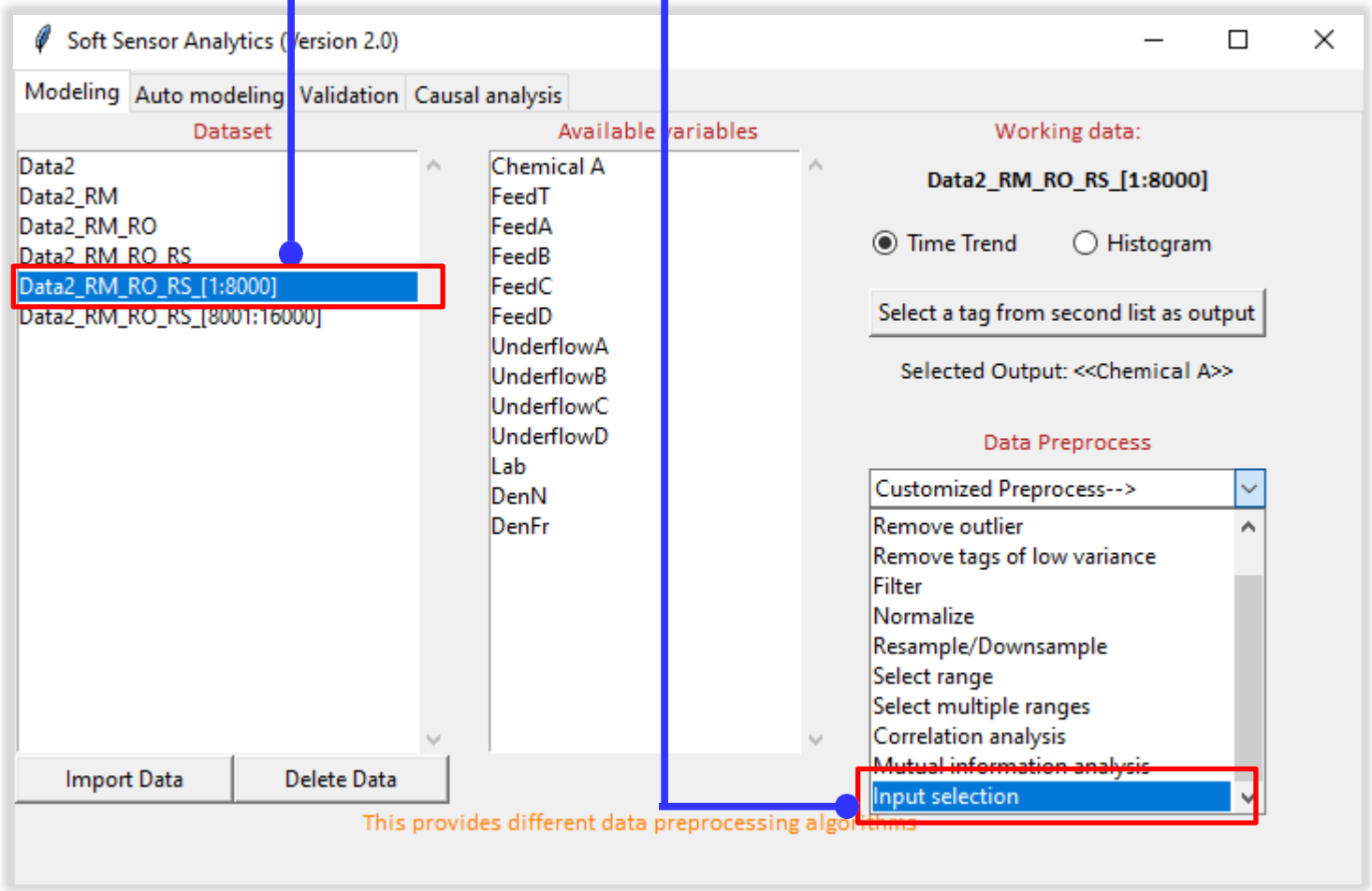
Soft Sensor Analytics

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**

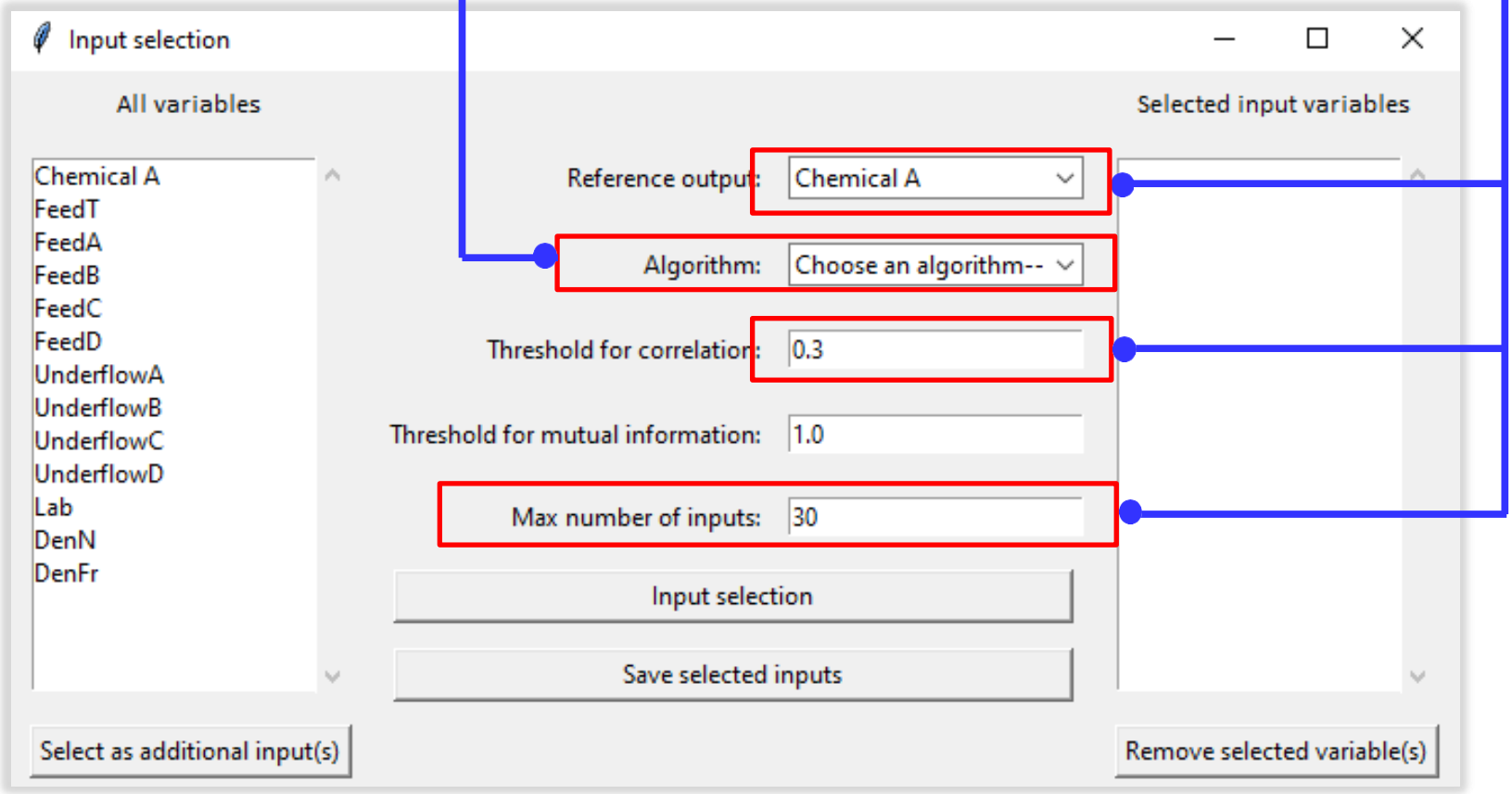


Soft Sensor Analytics

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



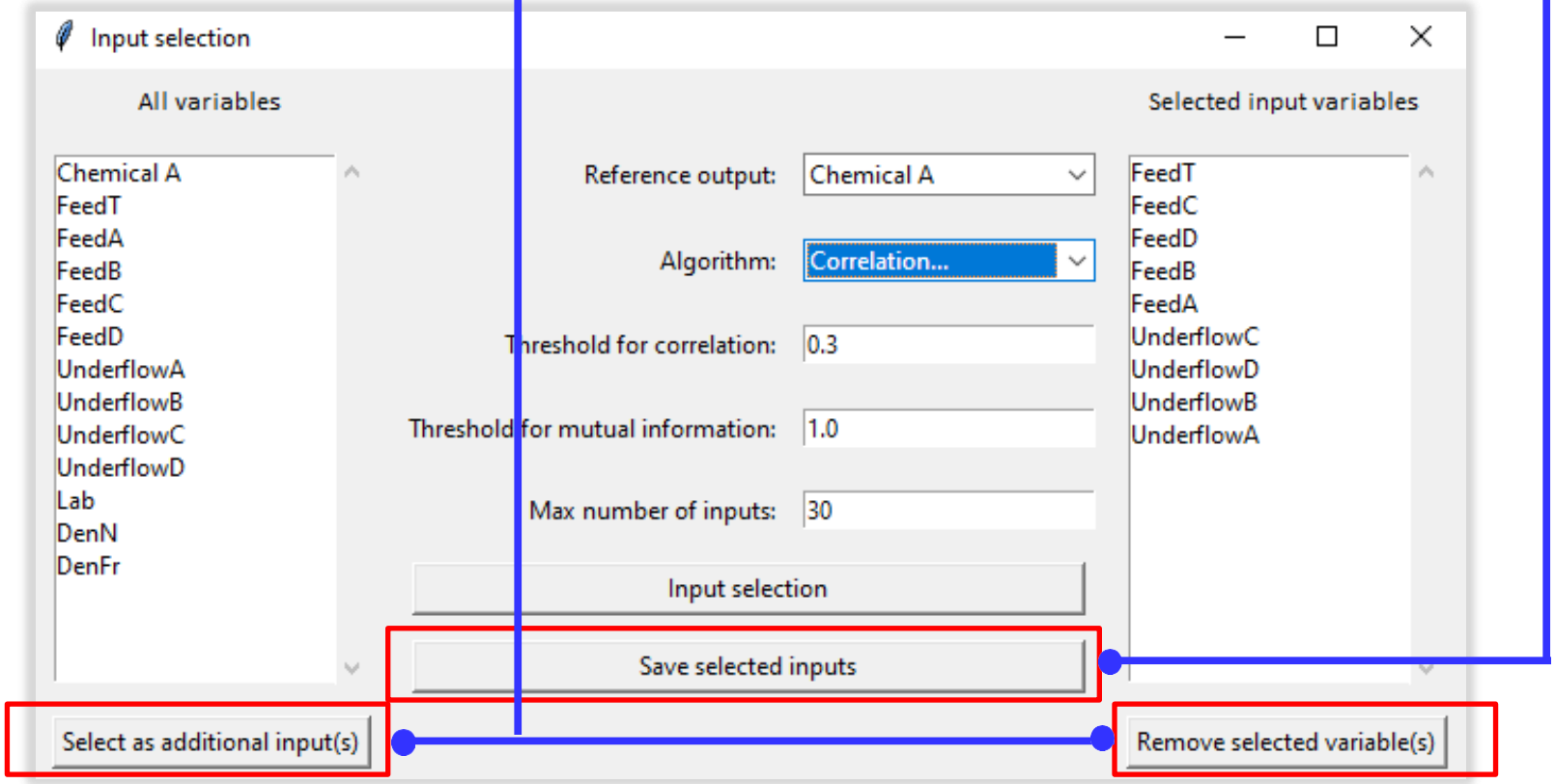
Soft Sensor Analytics

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

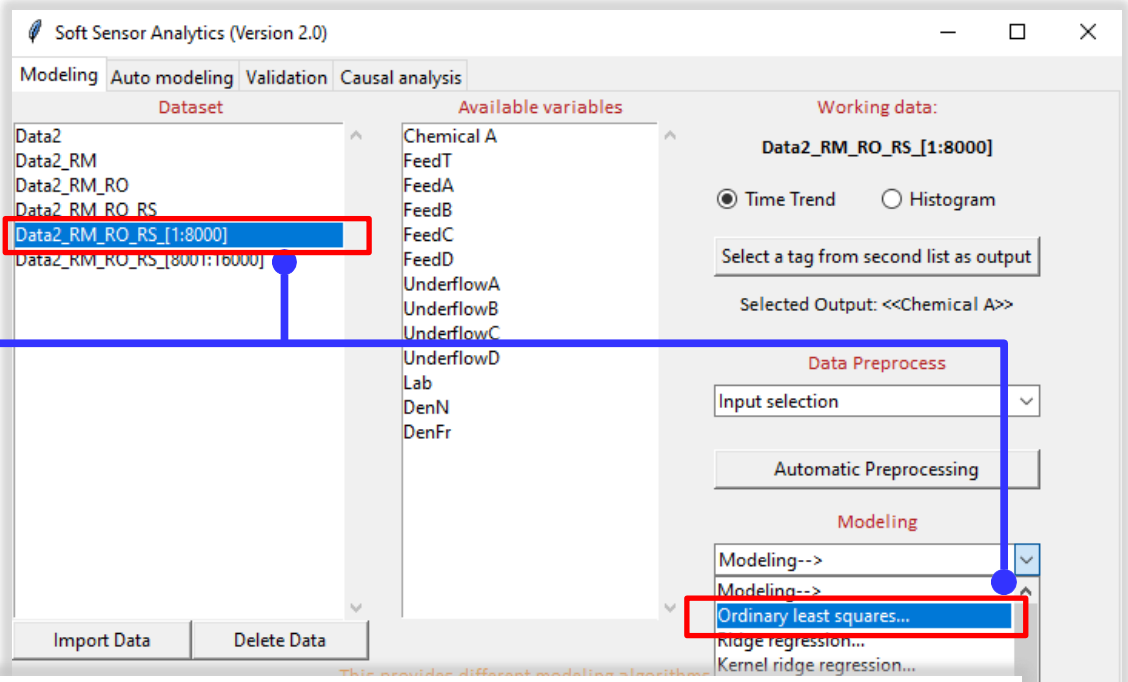
Modeling – ordinary least squares

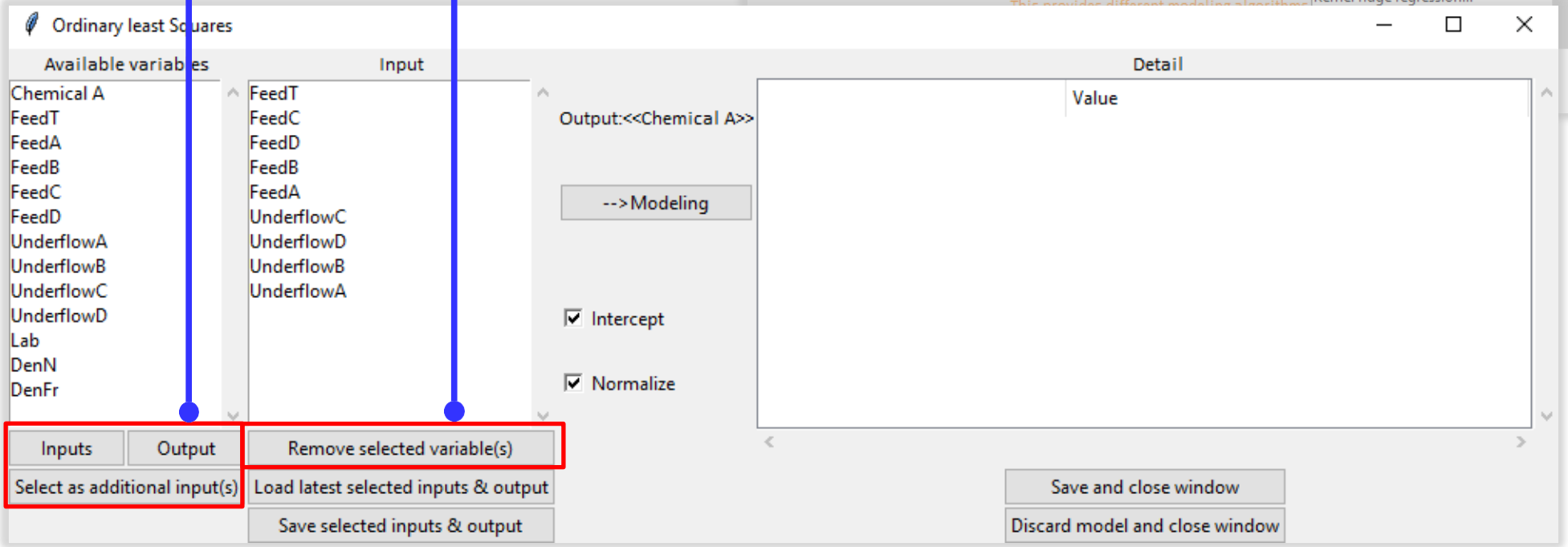
1 **Choose dataset and modeling method**

- **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



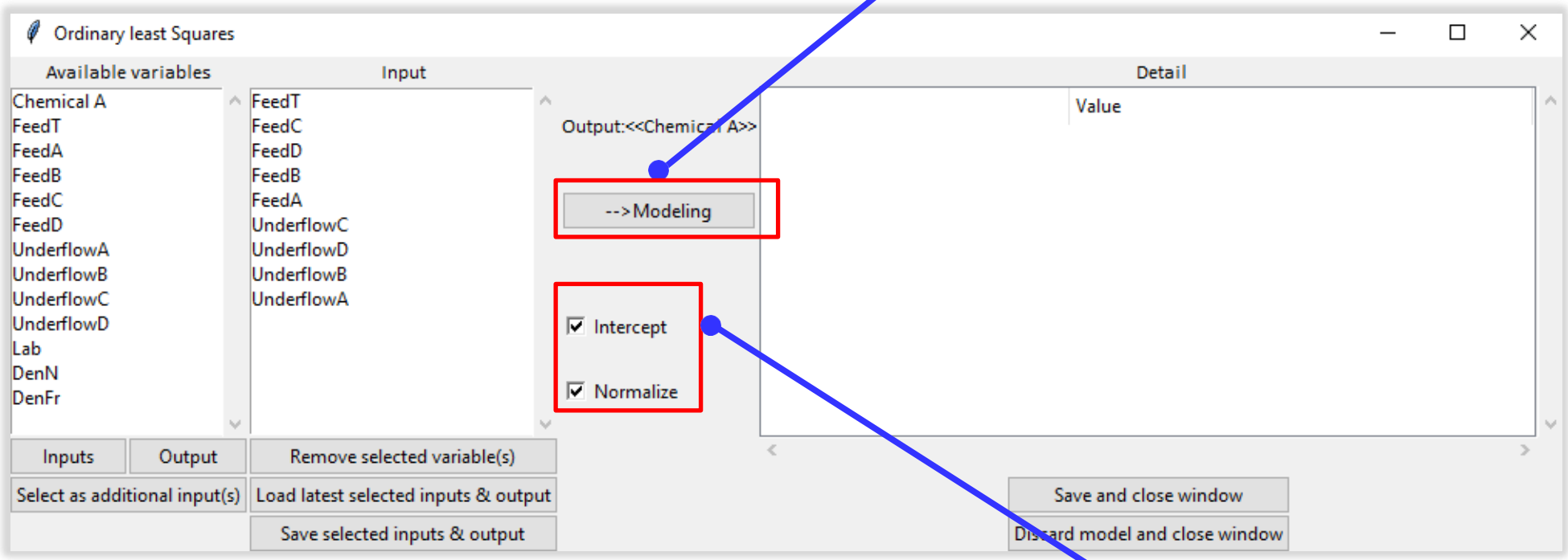


Soft Sensor Analytics

Modeling – ordinary least squares (select inputs manually)

4

Modeling
Click "-->Modeling" to perform regression



3

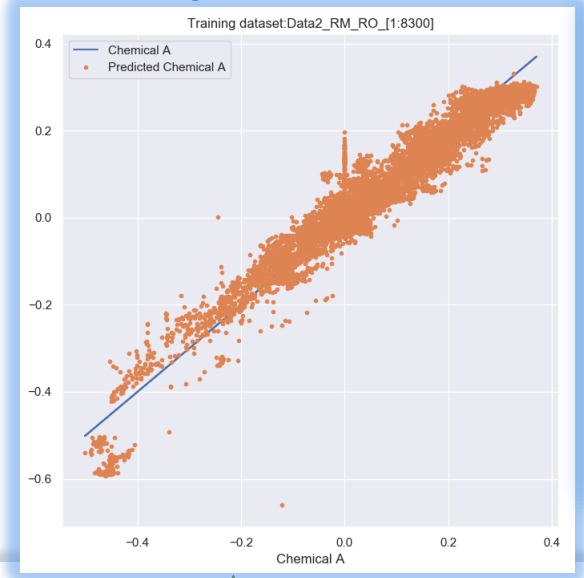
Options
Tick "Intercept" or/and **"Normalize"**.

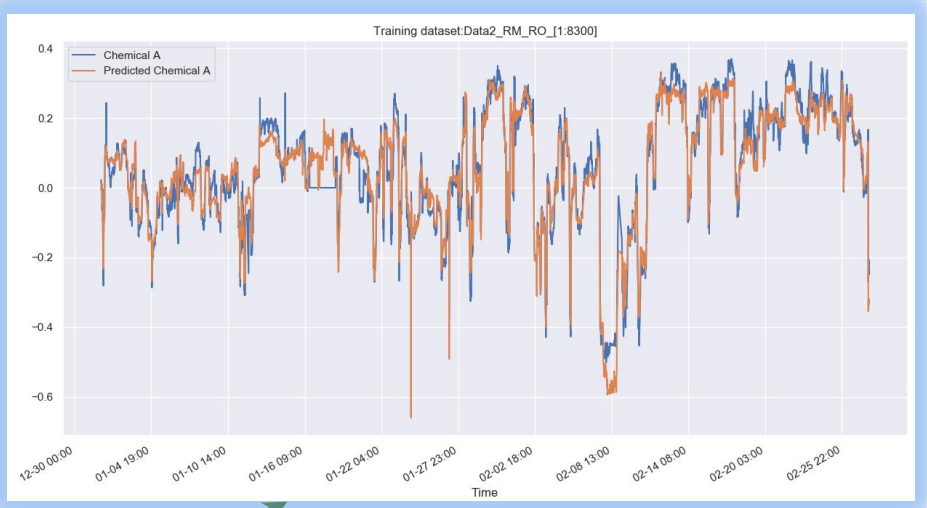
Soft Sensor Analytics

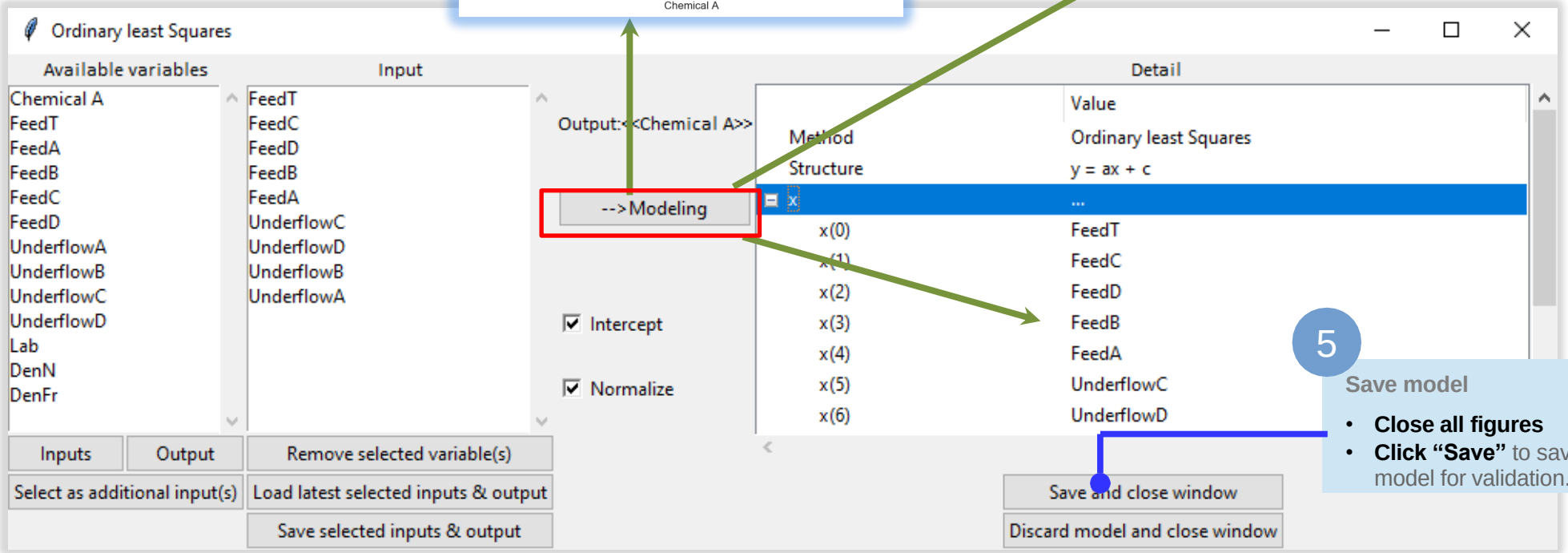
Results – ordinary least squares

Results

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







5

Save model

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

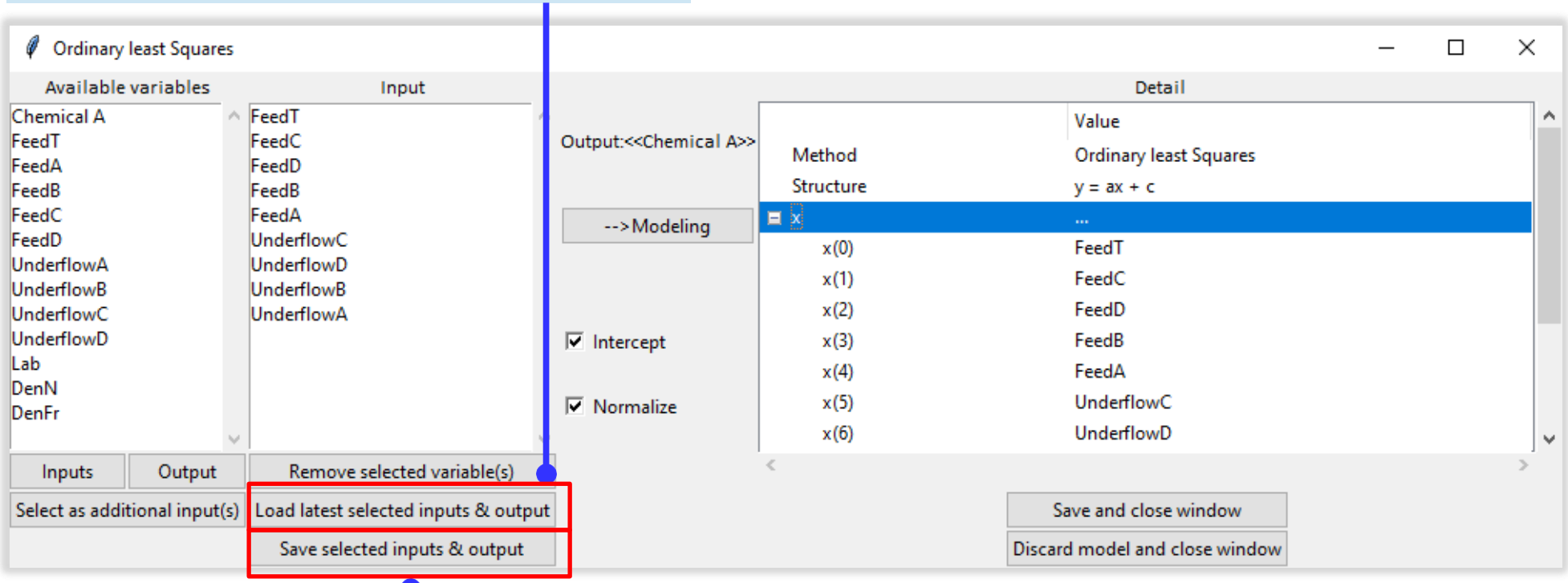
Soft Sensor Analytics

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



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

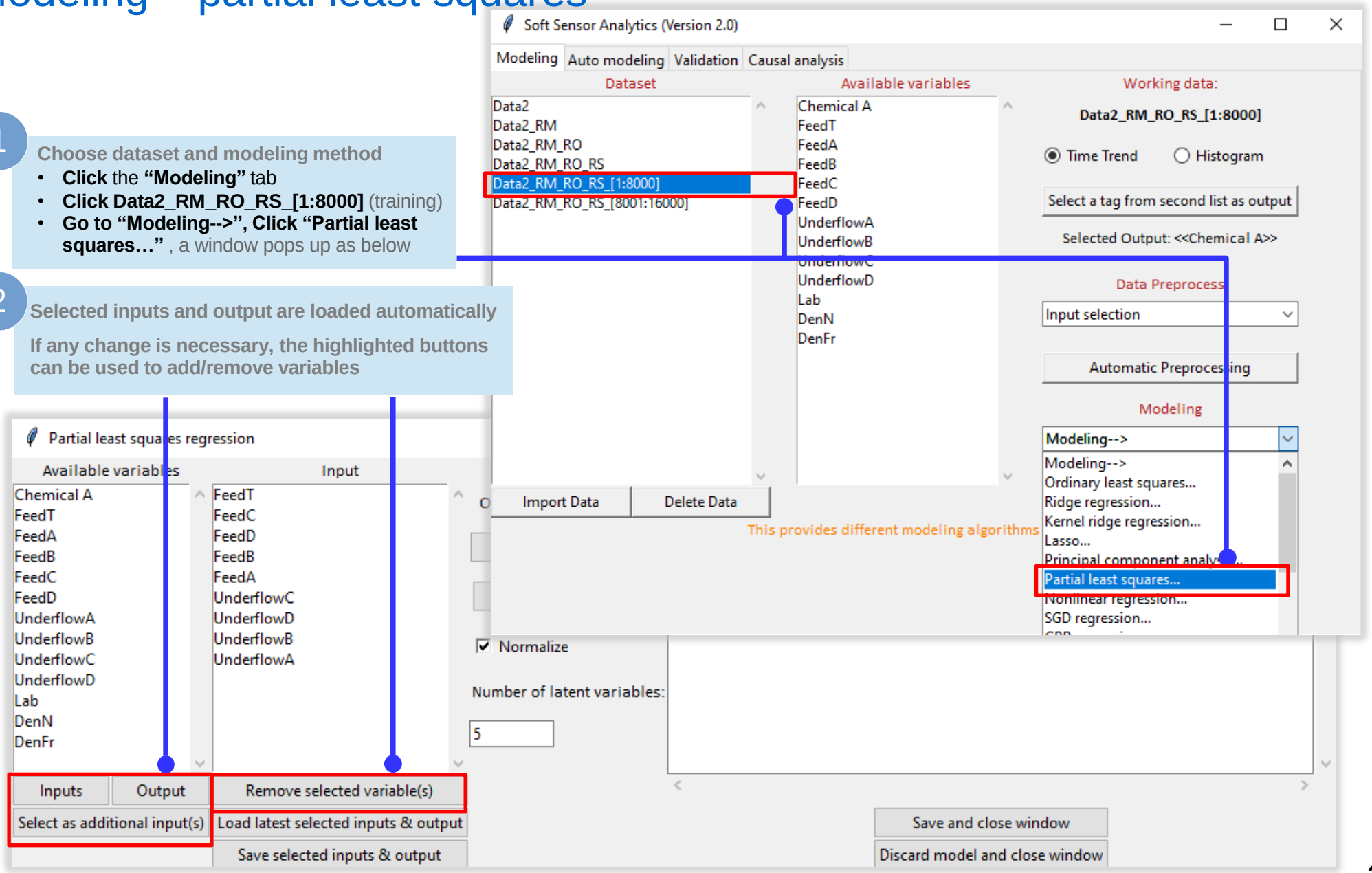
Choose dataset and modeling method

- **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



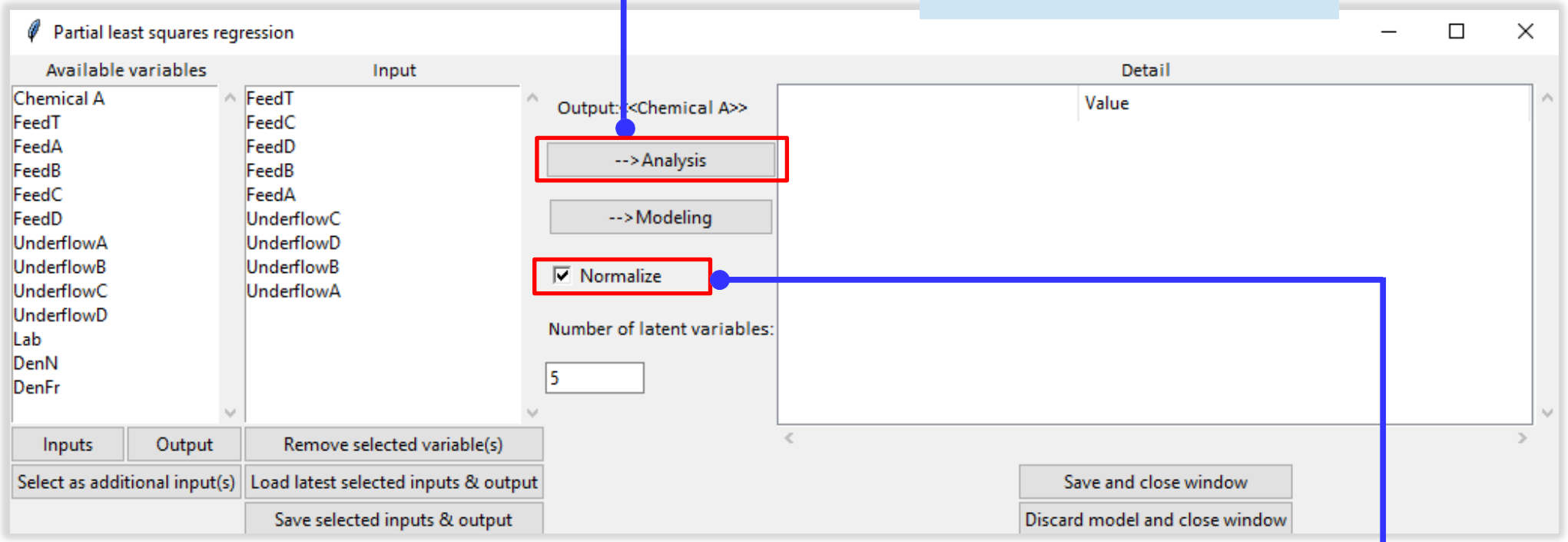
Soft Sensor Analytics

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

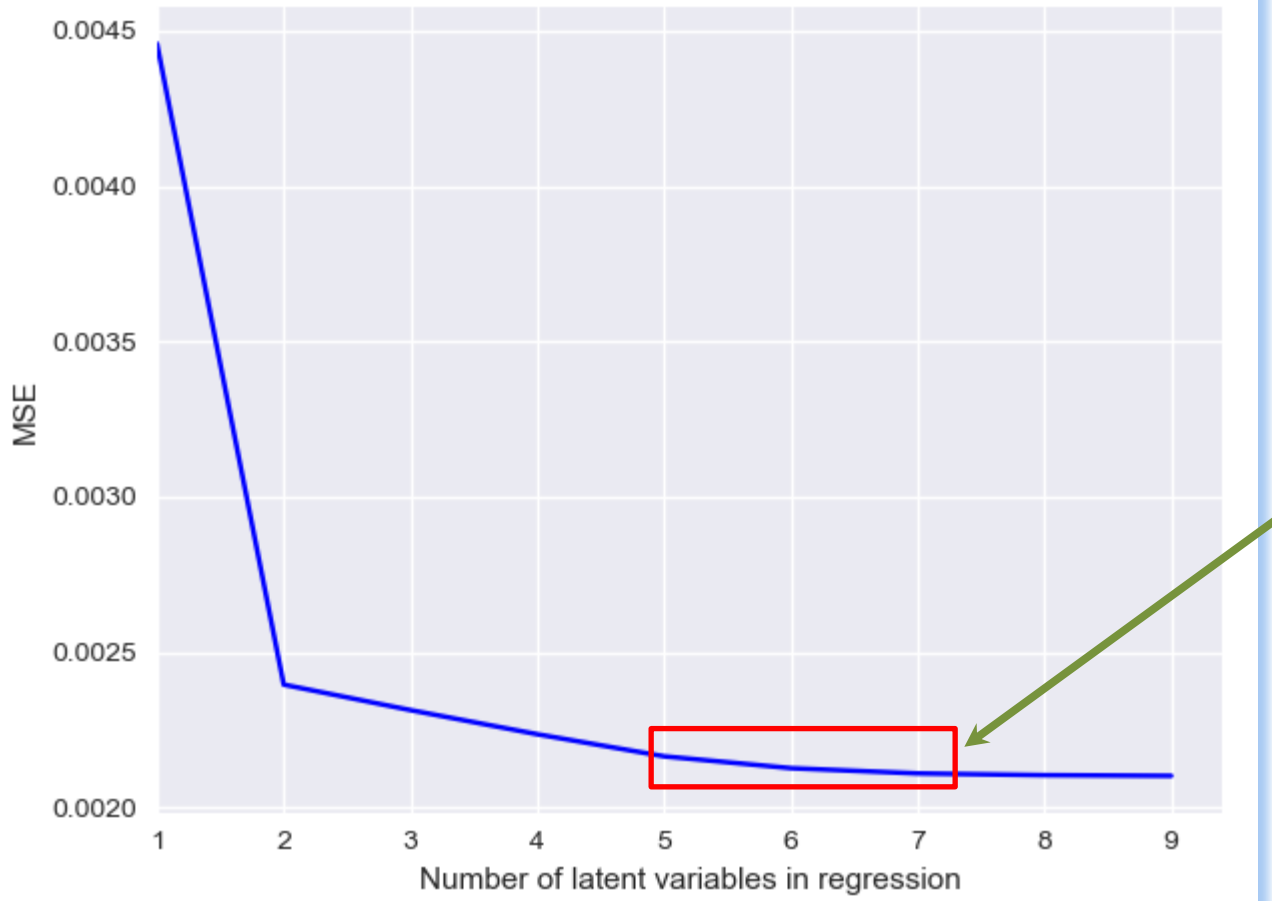


3

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.

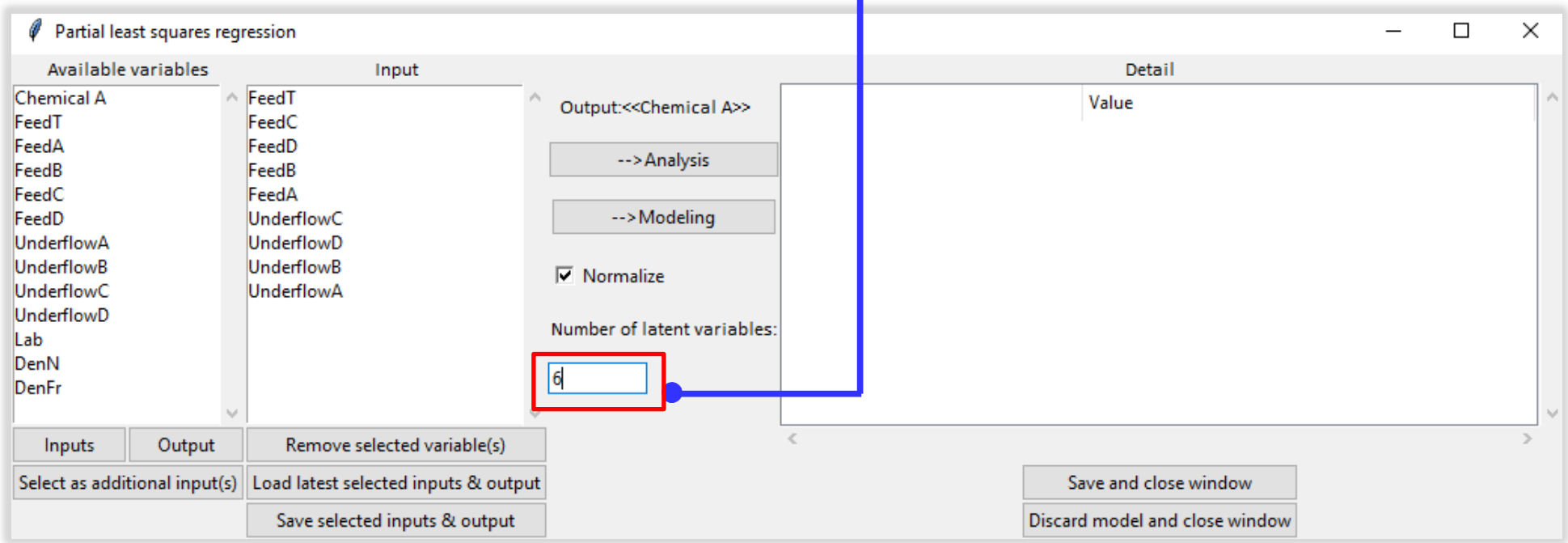
Soft Sensor Analytics

Modeling – PLS (determine # of latent variables)

5

Determine # of latent variables

- **Close all figures**
- **Enter "6"** in the **"number of latent variables"** box

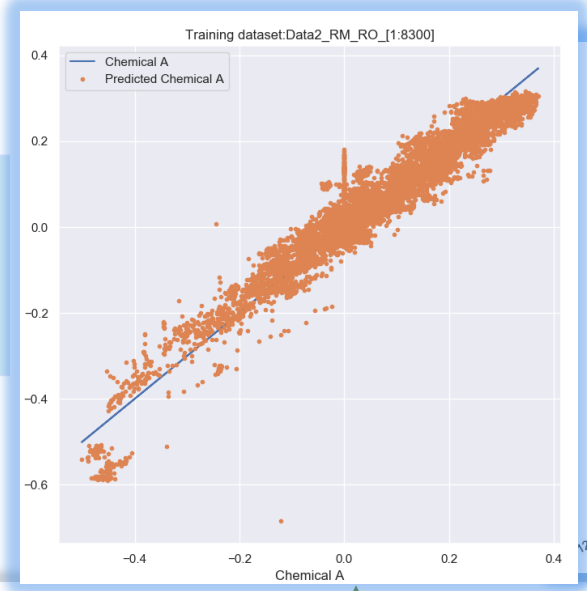


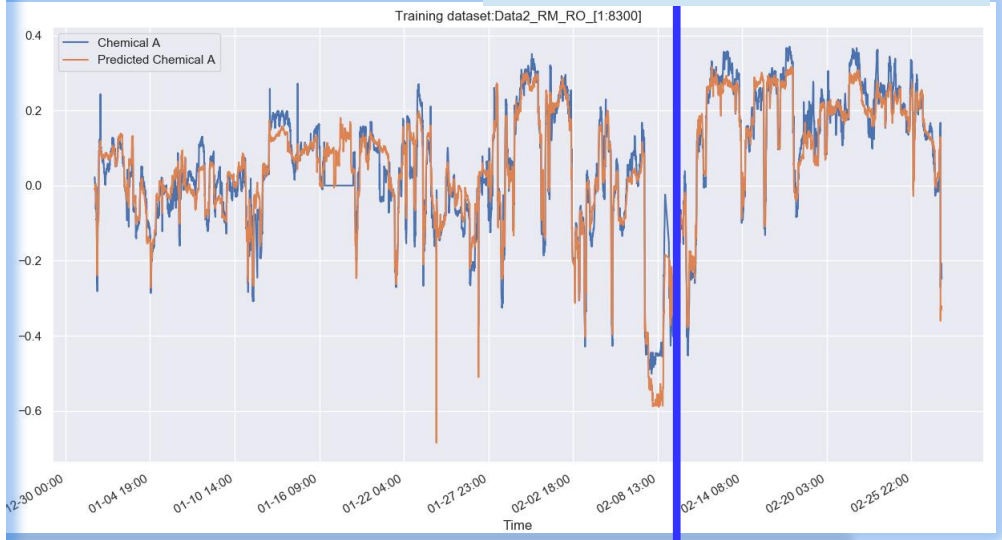
Soft Sensor Analytics

Results – PLS

6

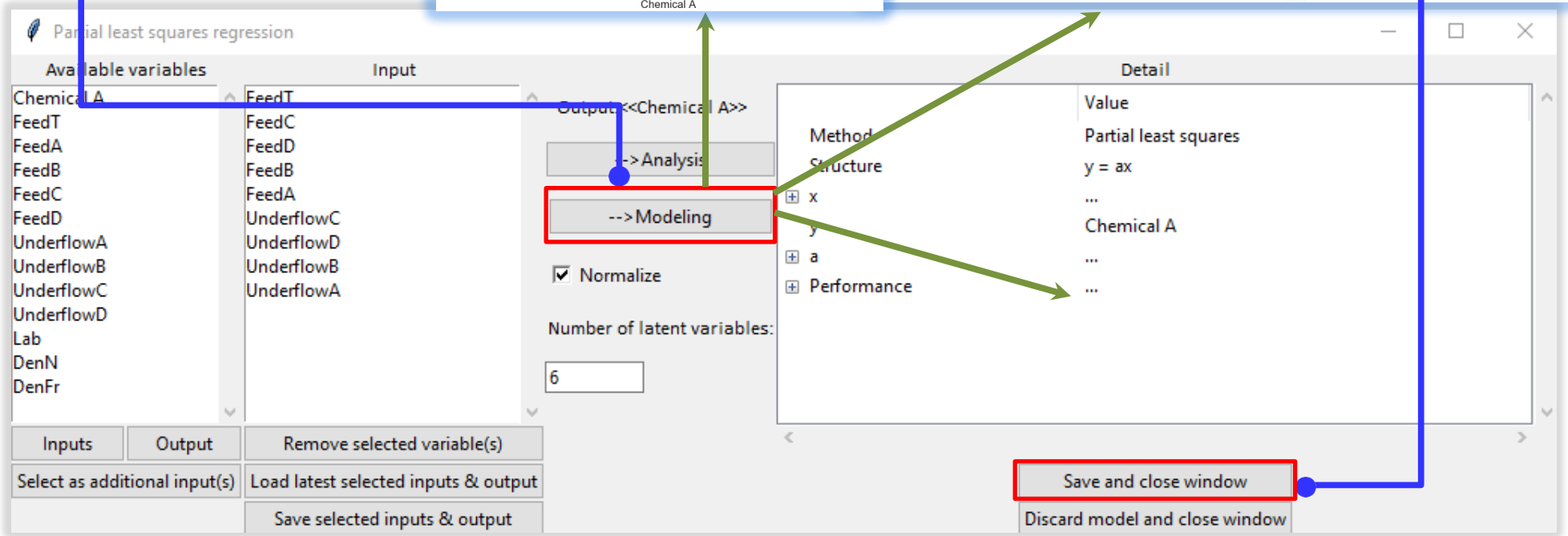
Regression
Click "-->Modeling"
- scatter plot and time trend are generated
- model information is shown in **"Detail"** column





7

- **Close figures**
- **Click "Save"** to save this model



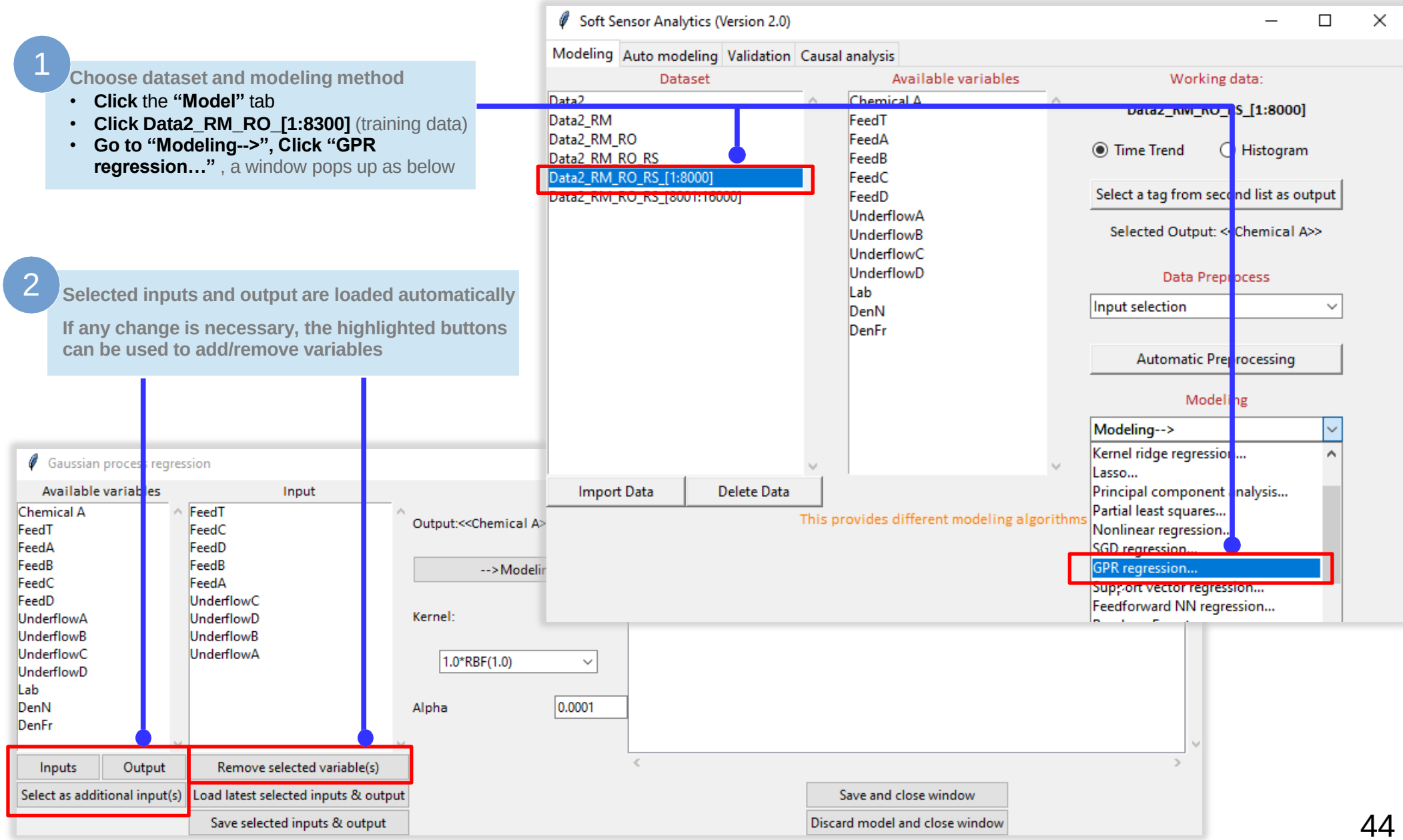
Soft Sensor Analytics

Modeling – Gaussian regression

1. **Choose dataset and modeling method**
   - **Click** the **"Model"** tab
   - **Click Data2_RM_RO_[1:8300]** (training data)
   - **Go to "Modeling-->", Click "GPR regression…"** , a window pops up as below

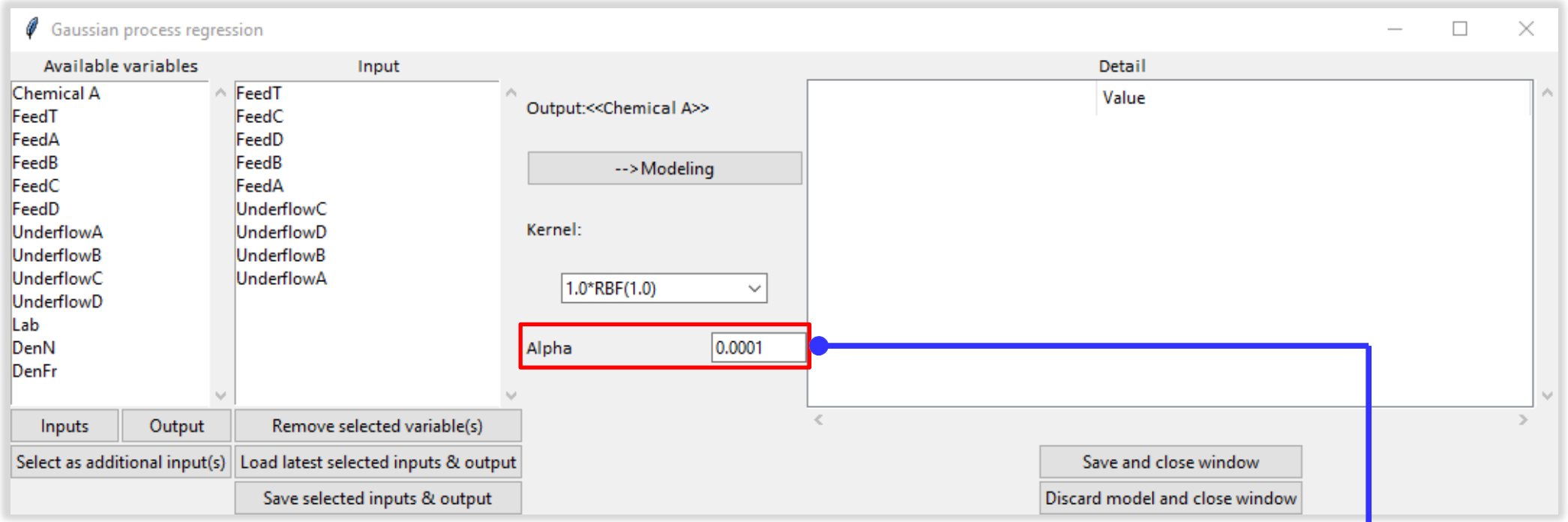
2. **Selected inputs and output are loaded automatically**

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



Soft Sensor Analytics

Modeling – Gaussian process regression



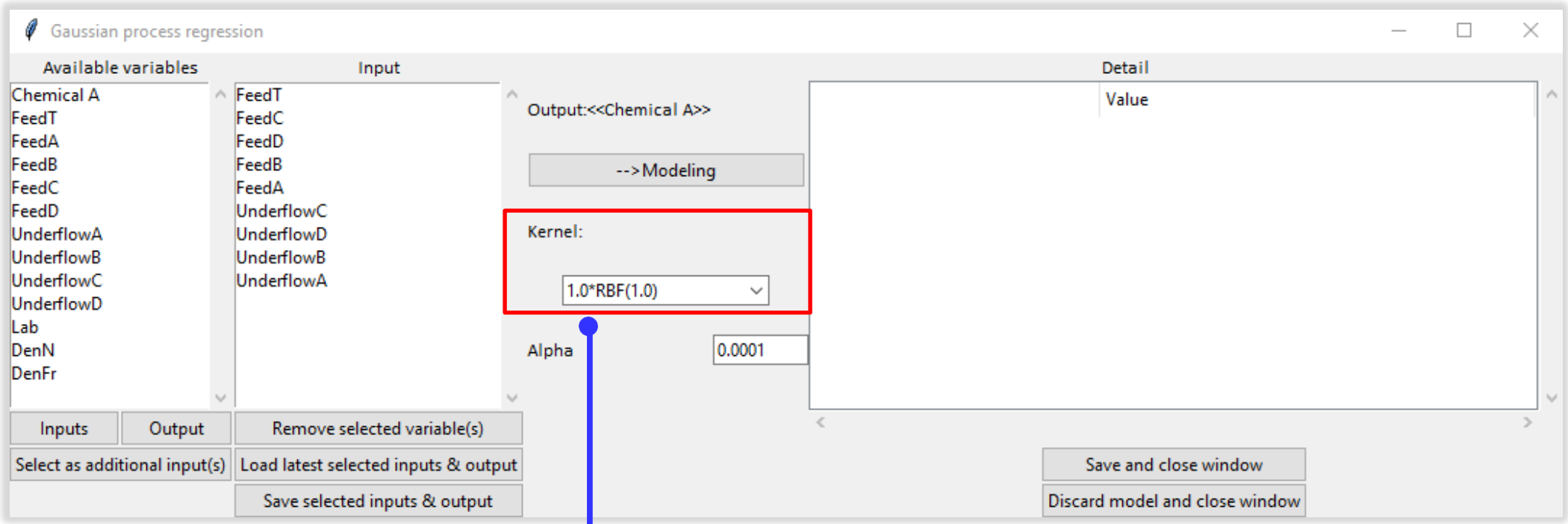
3

Options

Enter different values if necessary, **default is "0.0001"**

Soft Sensor Analytics

Modeling – Gaussian process regression



4

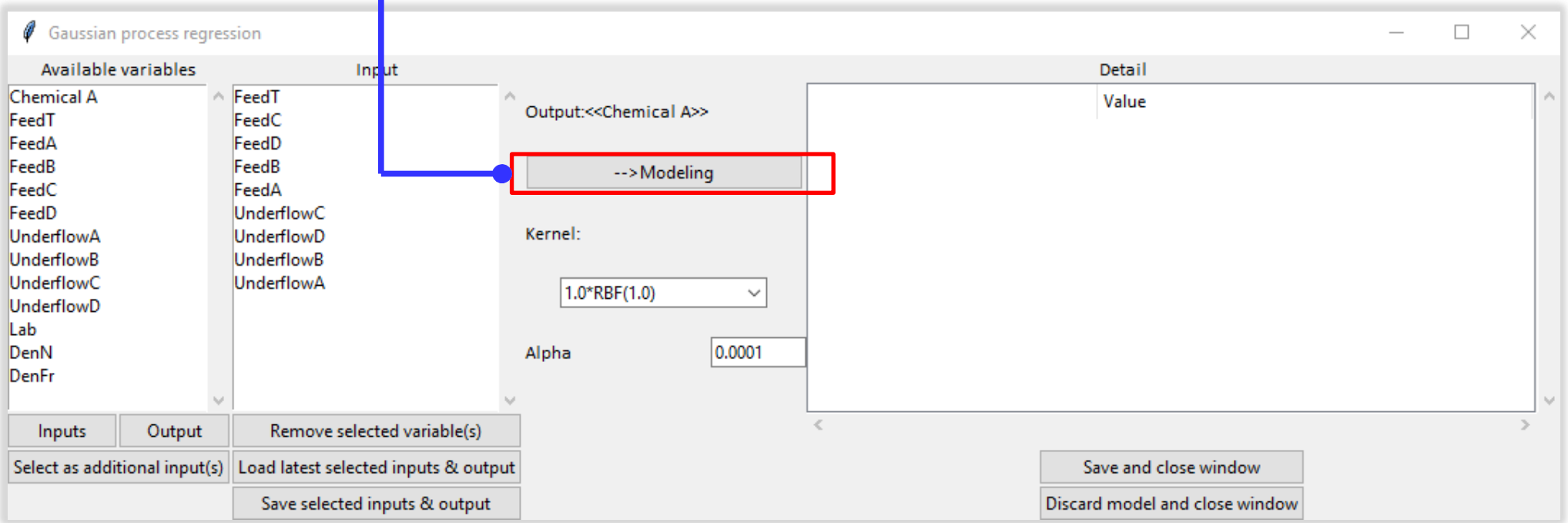
Kernel function selection
"1.0*RBF(1.0)" is selected as the kernel function

Soft Sensor Analytics

Modeling – Gaussian process regression

5

Regression
Click "-->Modeling"
- scatter plot and time trend are generated as shown in the next slides

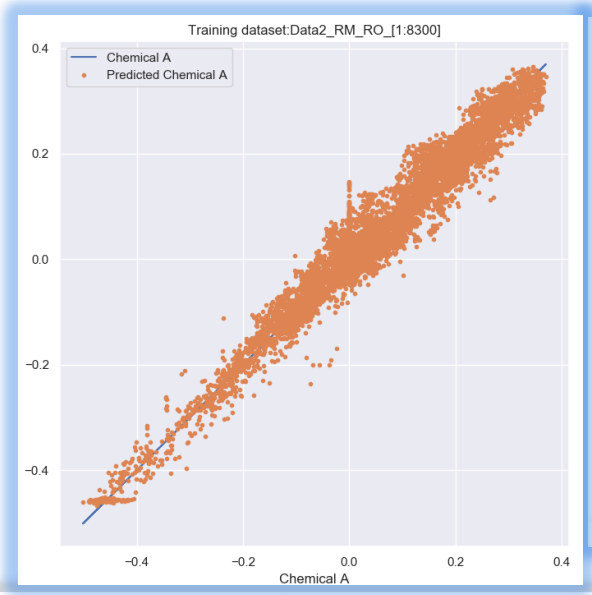


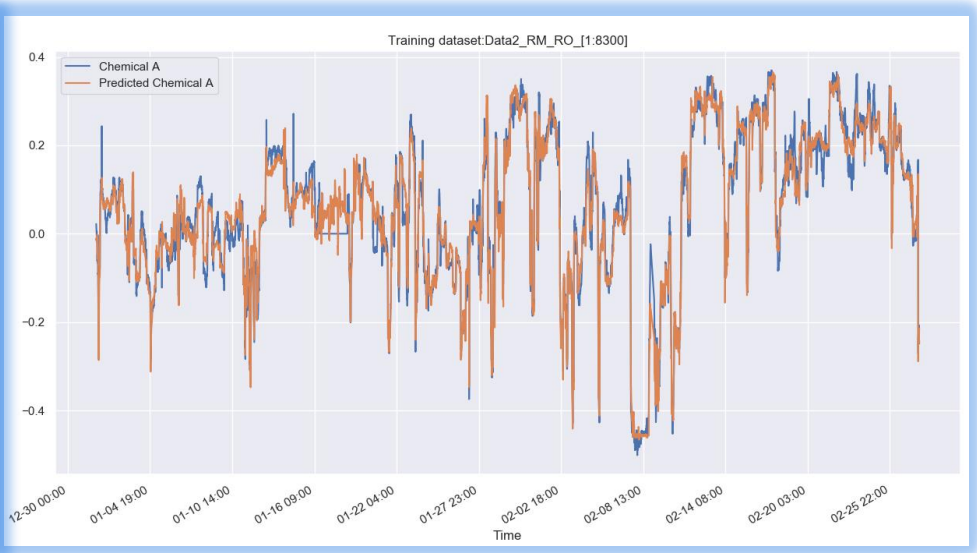
Soft Sensor Analytics

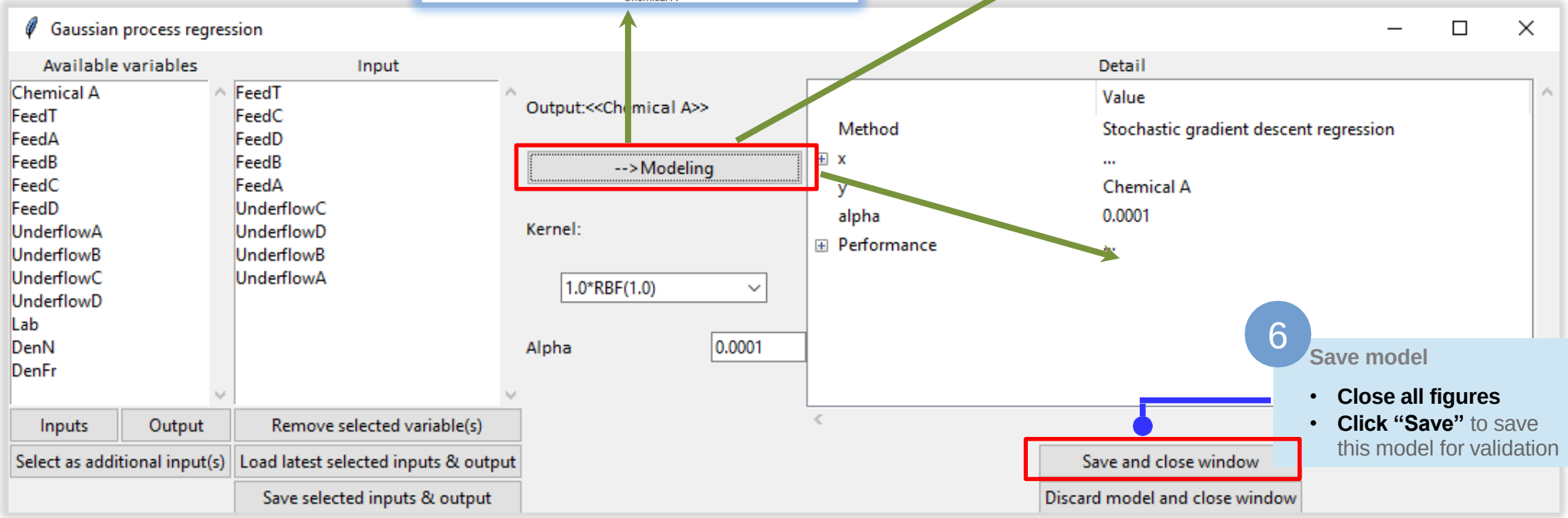
Results – Gaussian process regression

Results

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







6

Save model

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

Soft Sensor Analytics

Modeling – nonlinear regression

1

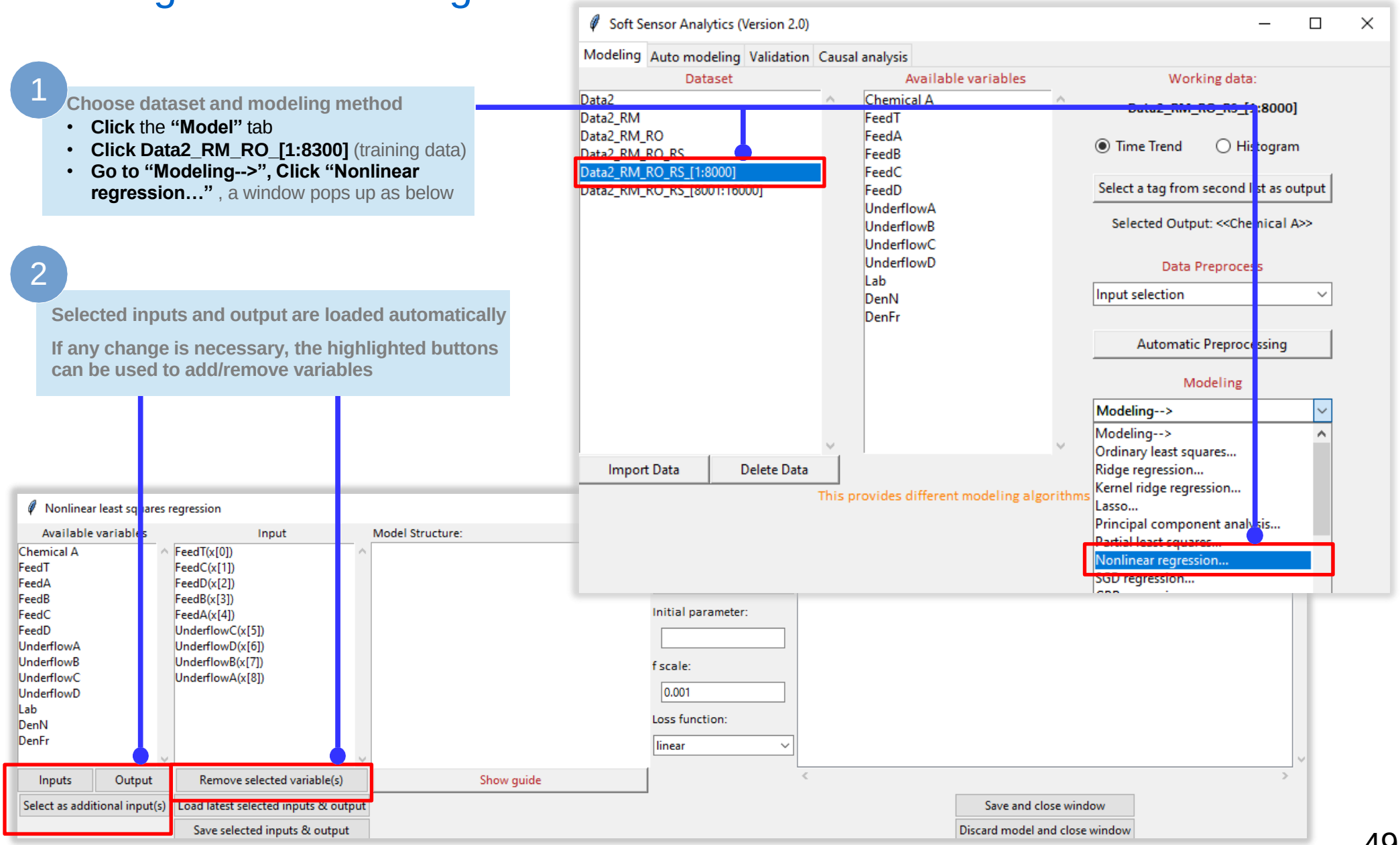
Choose dataset and modeling method

- **Click** the **"Model"** tab
- **Click Data2_RM_RO_[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

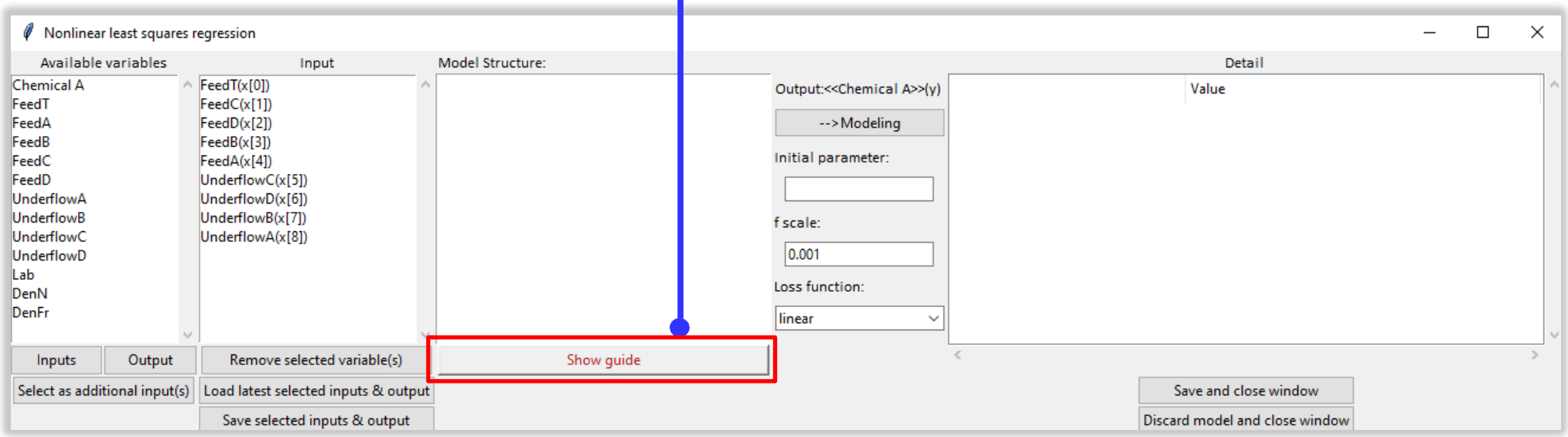


Soft Sensor Analytics

Modeling – nonlinear regression

3

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



Soft Sensor Analytics

Modeling – nonlinear regression

Note: this model structure is used as a simple example

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

4

Set nonlinear model structure

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

$a_1 \longrightarrow$ a[0]

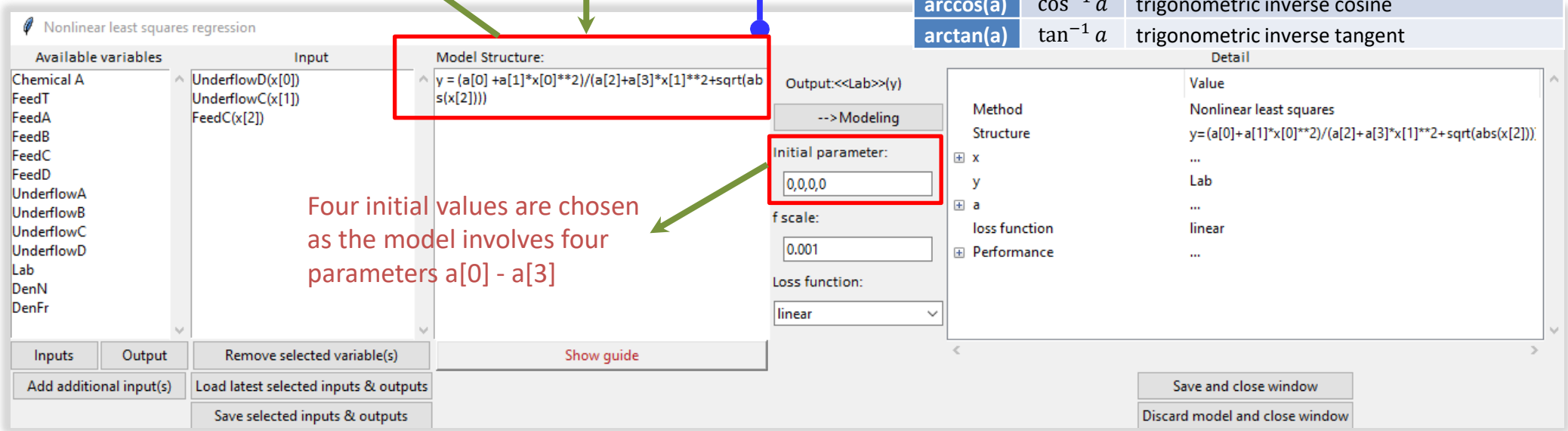
$a_3 \longrightarrow$ a[2]

$x_1 \longrightarrow$ x[0]

$x_3 \longrightarrow$ x[2]

y=(a[0]+a[1]*x[0]**2)/(a[2]+a[3]*x[1]**2+sqrt(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	$a \bmod b$	modulo
-a	$-a$	negation
abs(a)	$\|a\|$	absolute value
a**b	$a \hat{} b$	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



Soft Sensor Analytics

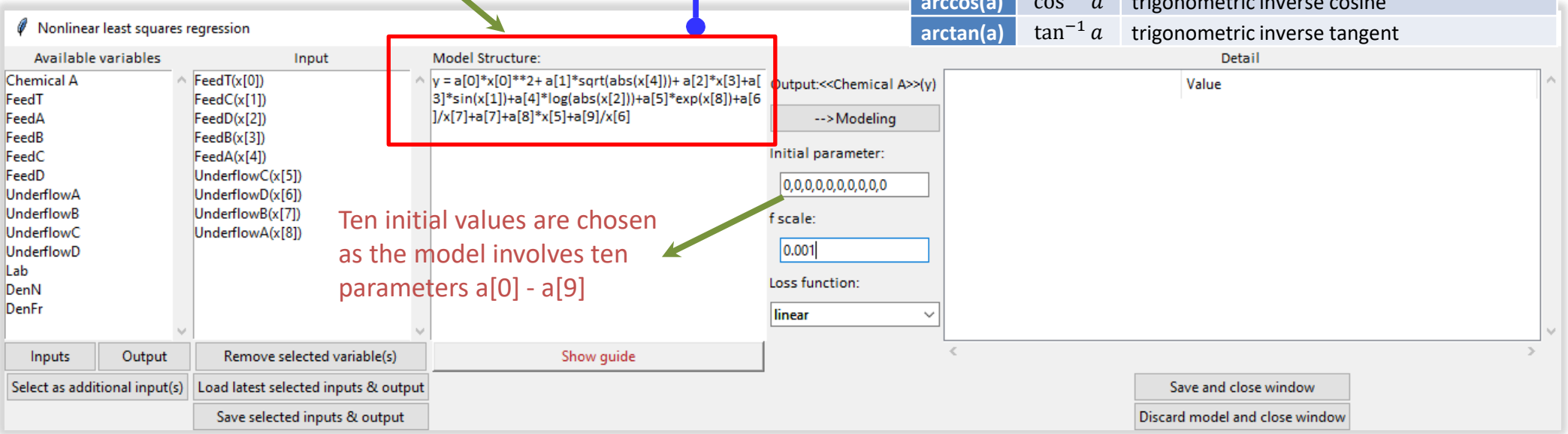
Modeling – nonlinear regression

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]*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]

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	$a \ mod \ b$	modulo
-a	$-a$	negation
abs(a)	$\|a\|$	absolute value
a**b	$a \hat{} b$	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



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

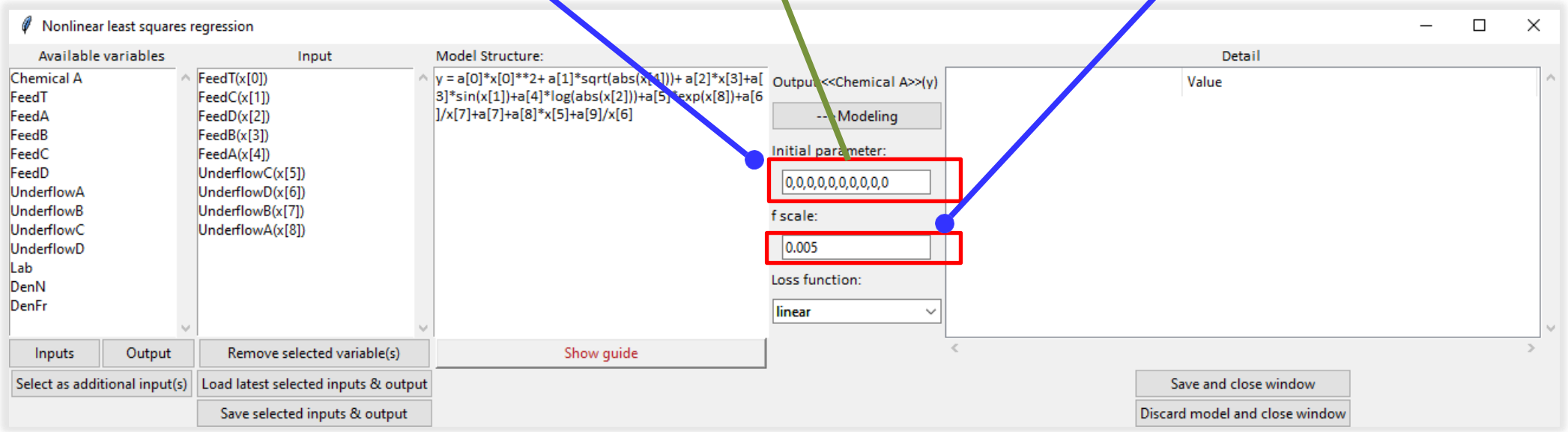
Soft Sensor Analytics

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



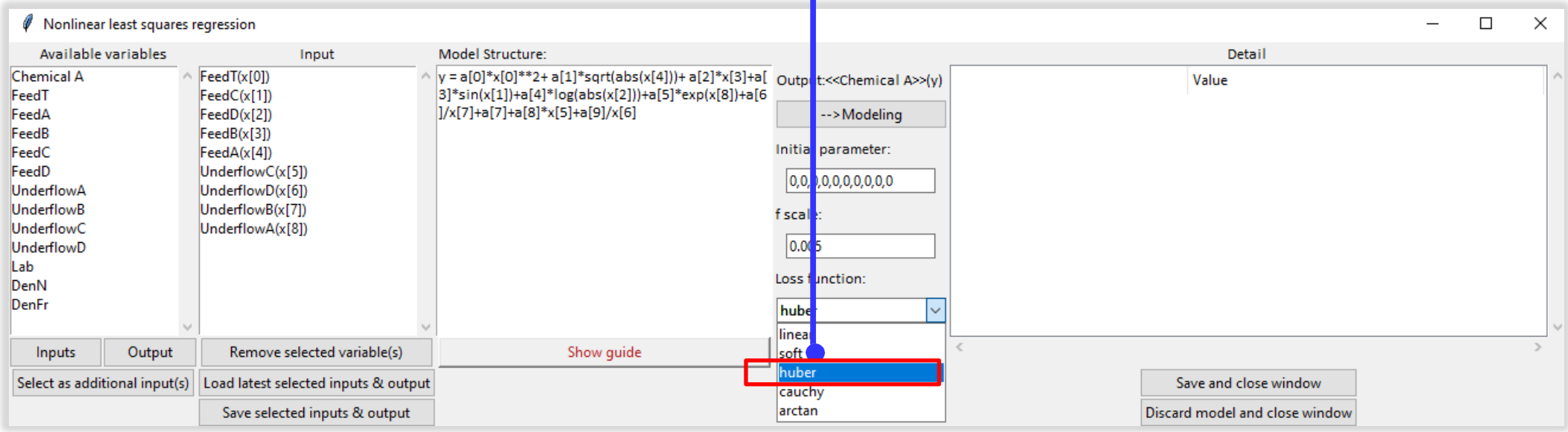
Soft Sensor Analytics

Modeling – nonlinear regression

7

Choose loss function

- Select **“huber”**

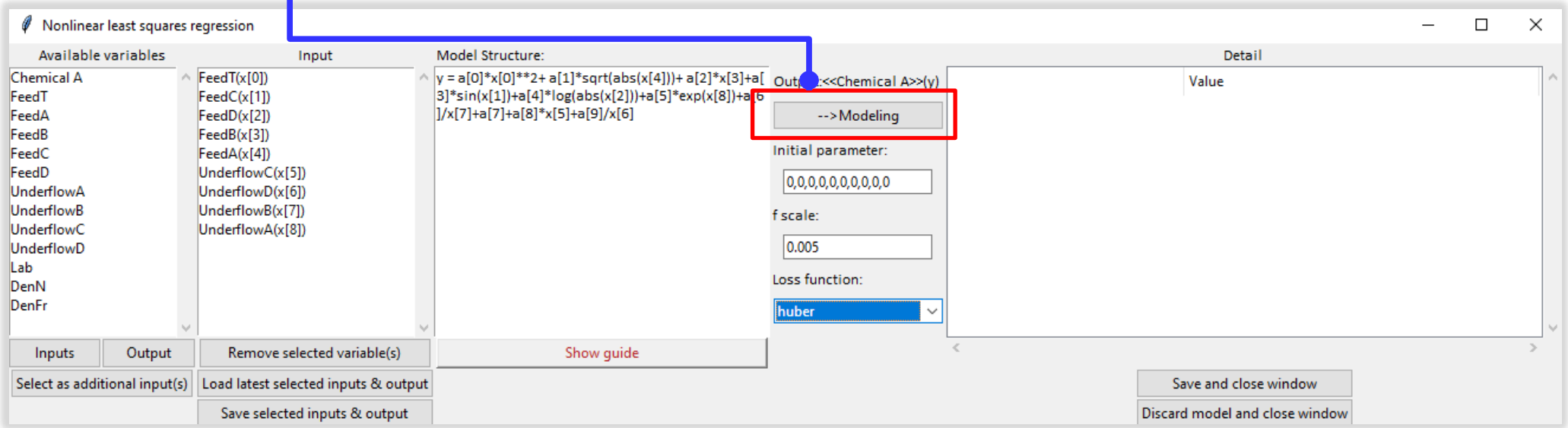


Soft Sensor Analytics

Modeling – nonlinear regression

8

Regression
Click "-->Modeling"
- scatter plot and time trend are generated as shown in the next slides

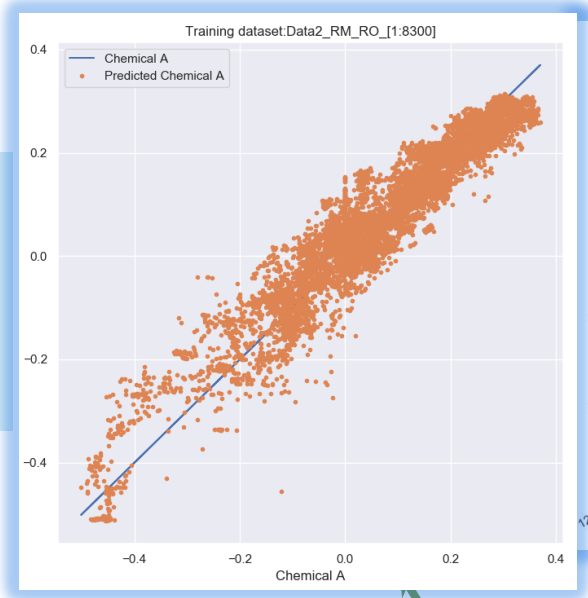


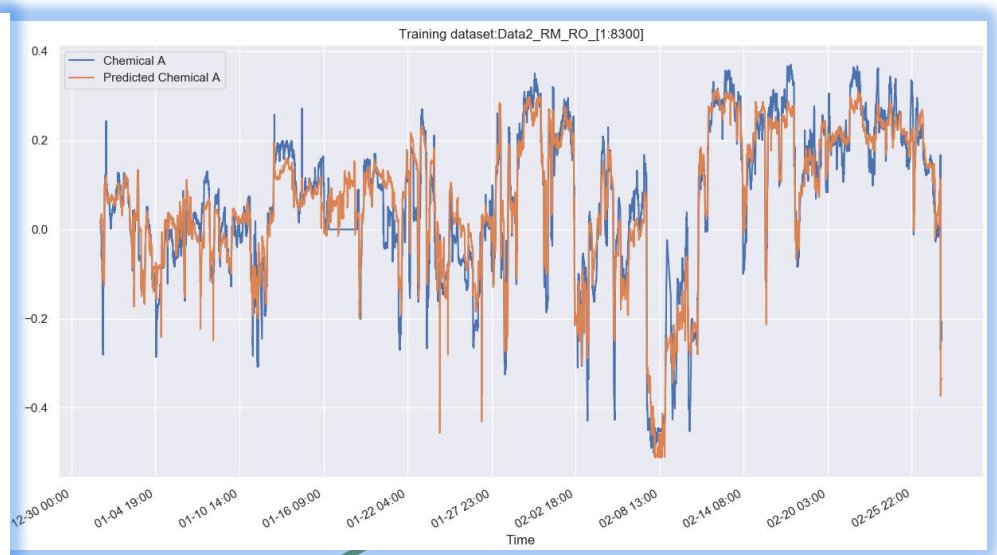
Soft Sensor Analytics

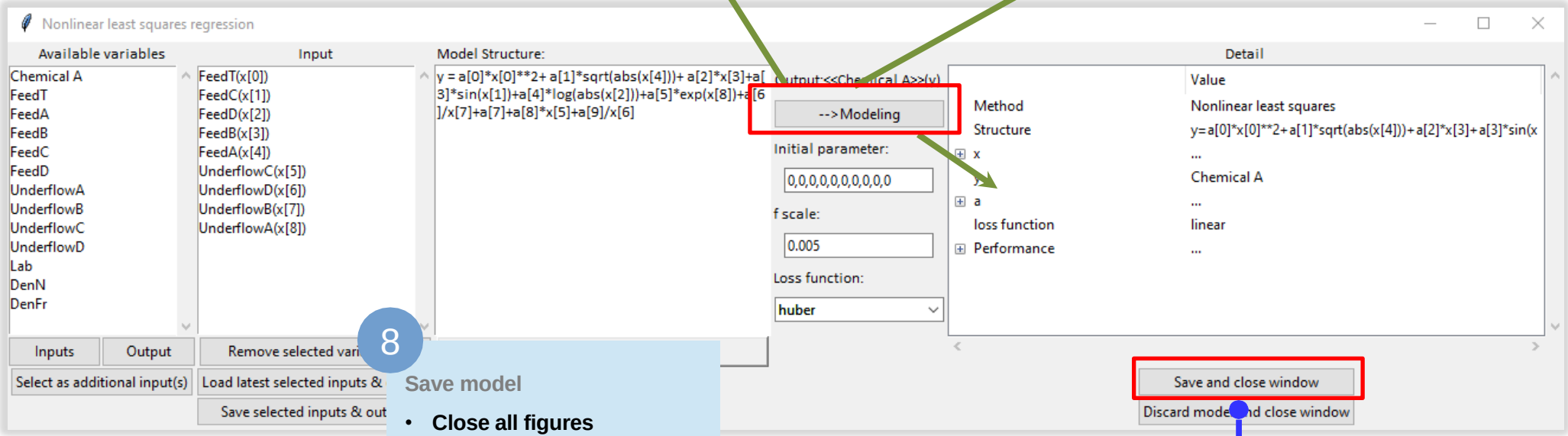
Results – nonlinear regression

Results

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







8

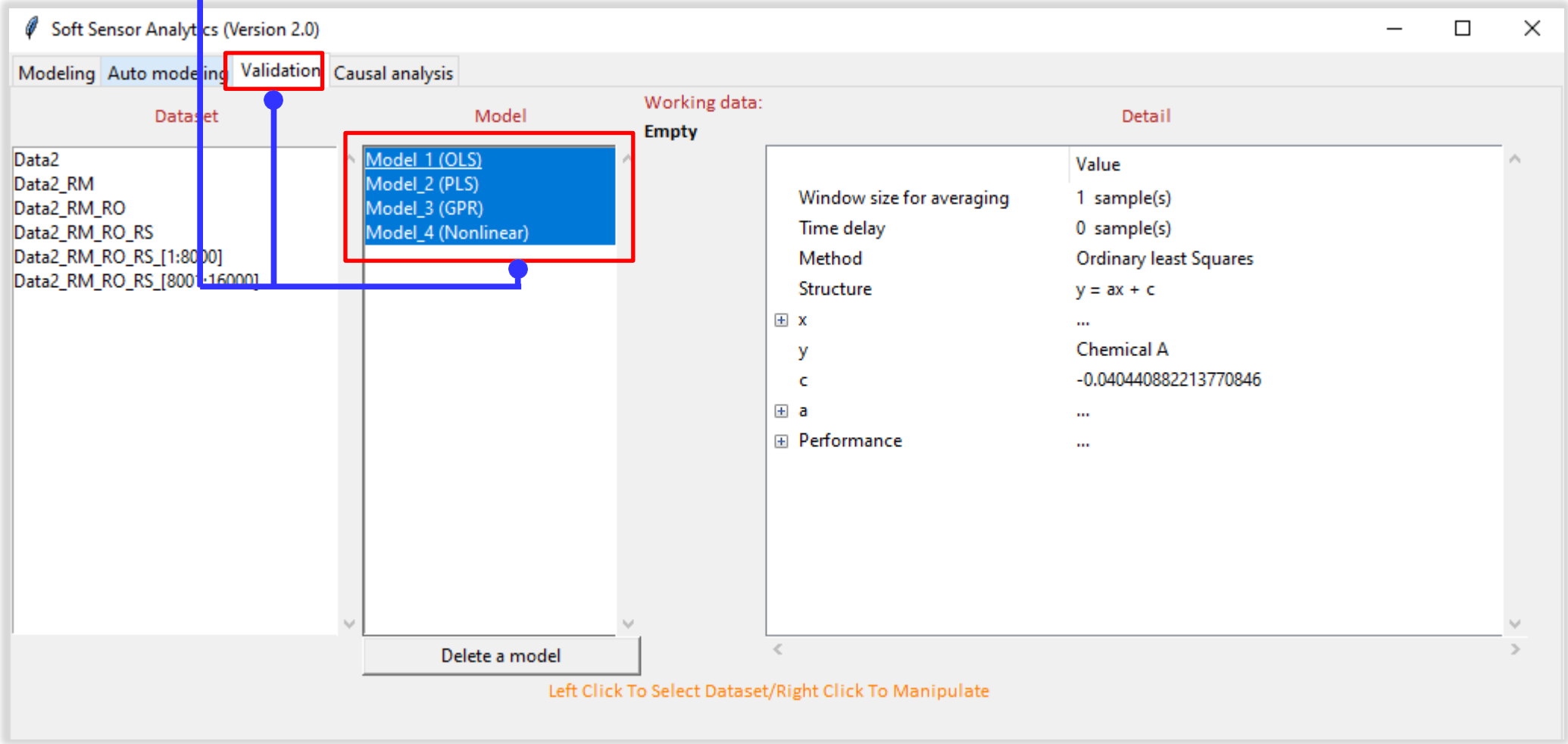
Save model

- **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

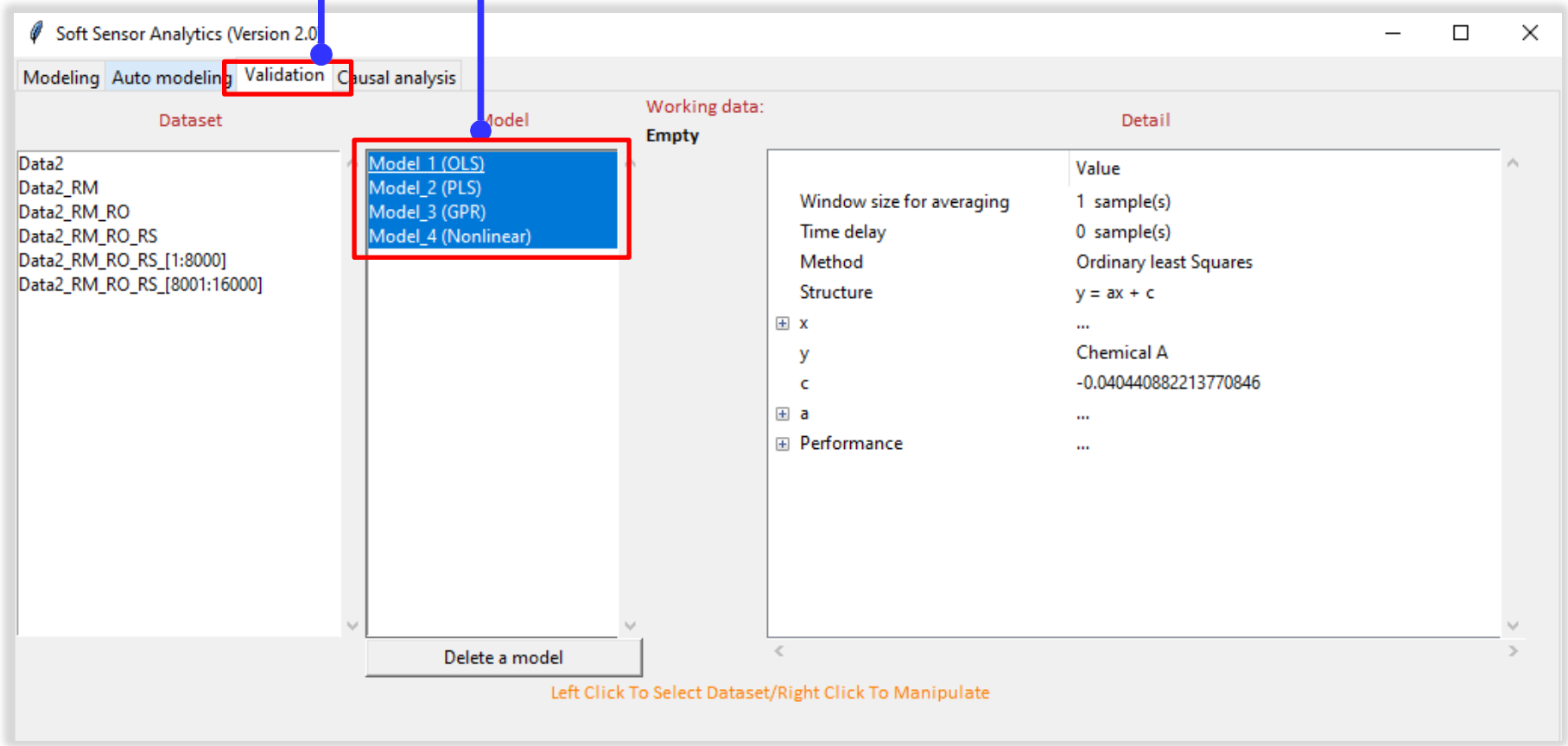
Validation of the models

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

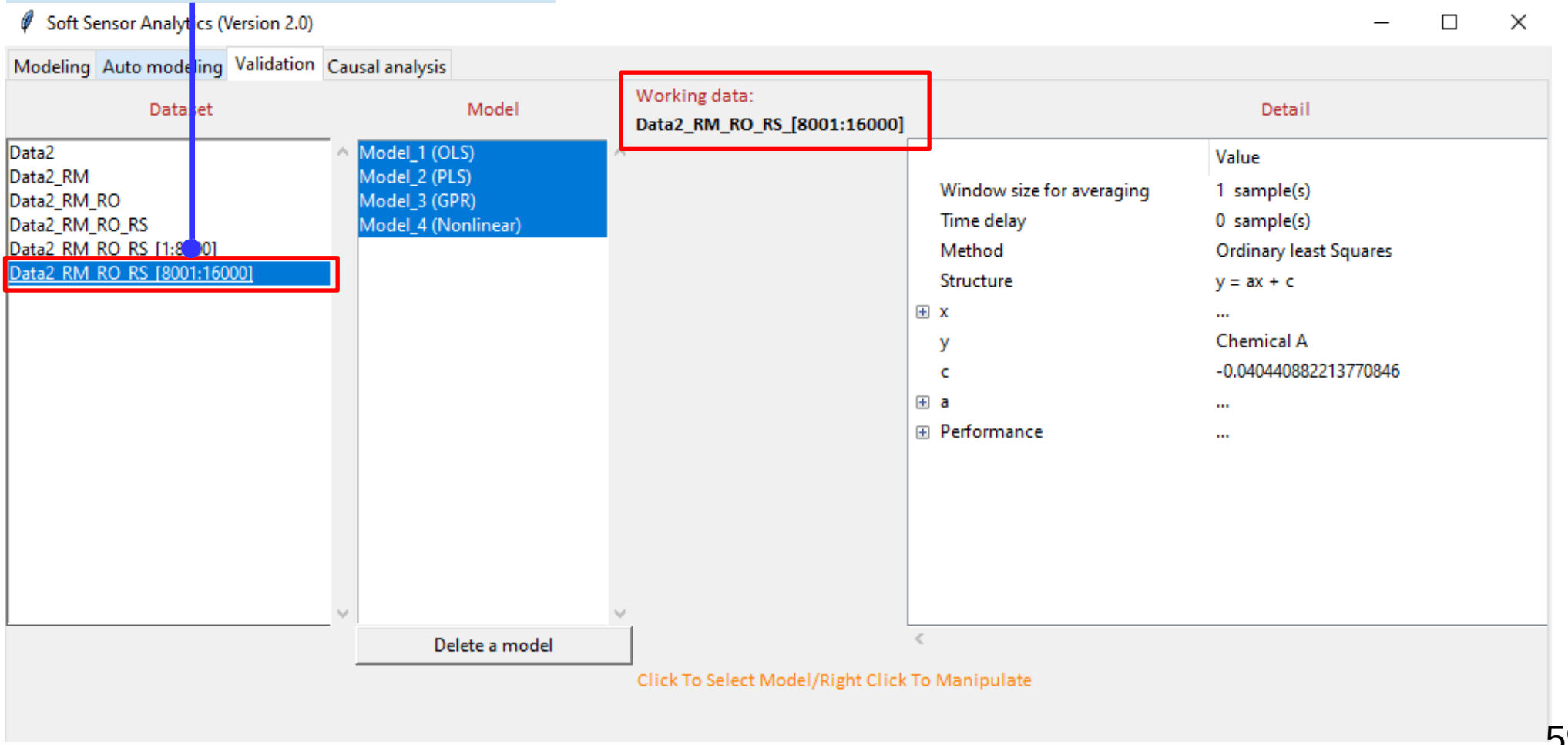
Validation of the models

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**



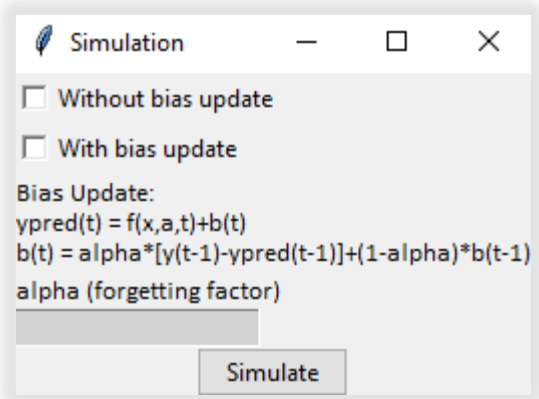
Soft Sensor Analytics

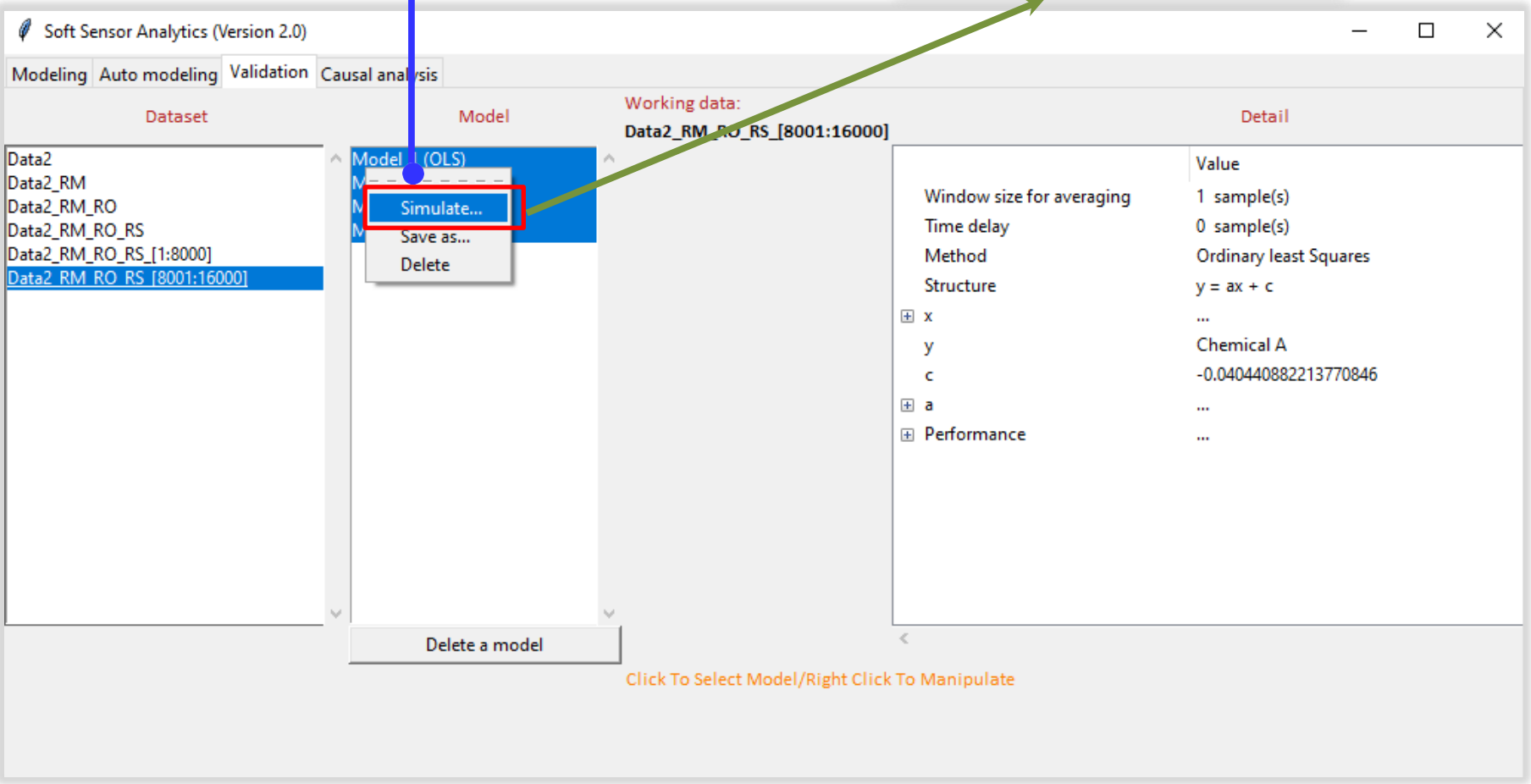
Validation of the models

3

Choose "Simulate" to continue

- **Right-Click** on the four models
- **Click "Simulate…"** button, a window will pop up





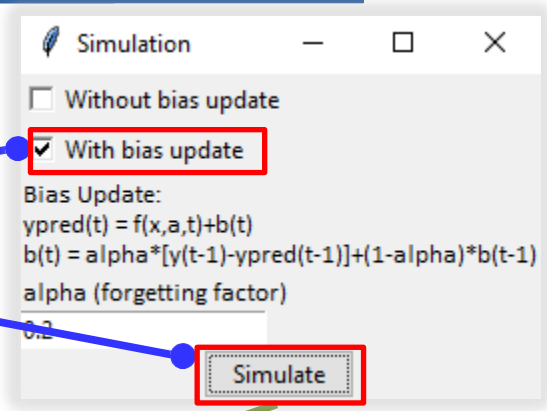
Soft Sensor Analytics

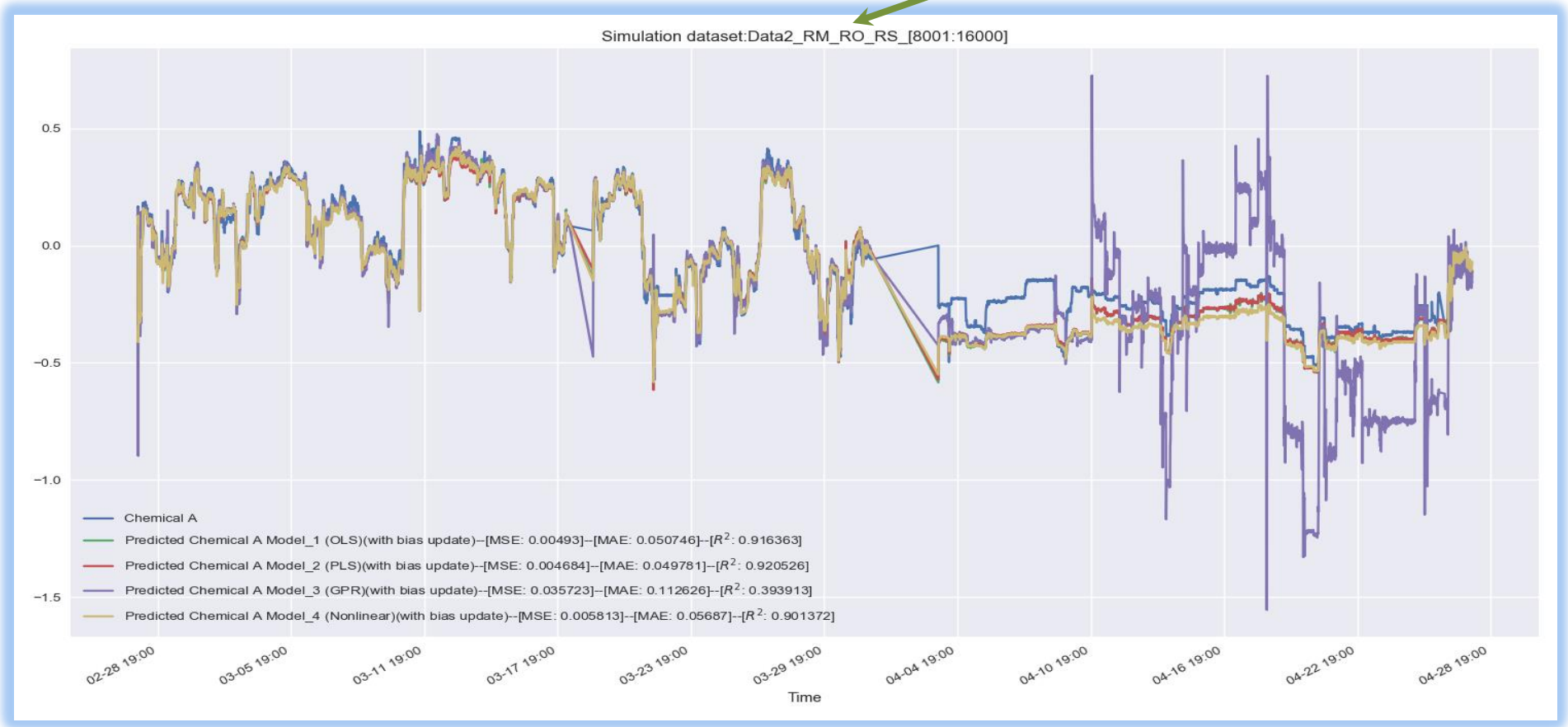
Validation of the models

4

Model validation

- **Tick "With bias update"** box
- **Click "Simulate"** button





Soft Sensor Analytics

Alternative way to process data quickly -- Automatic Preprocessing

1

Under the "Modeling" tab, choose dataset to work on

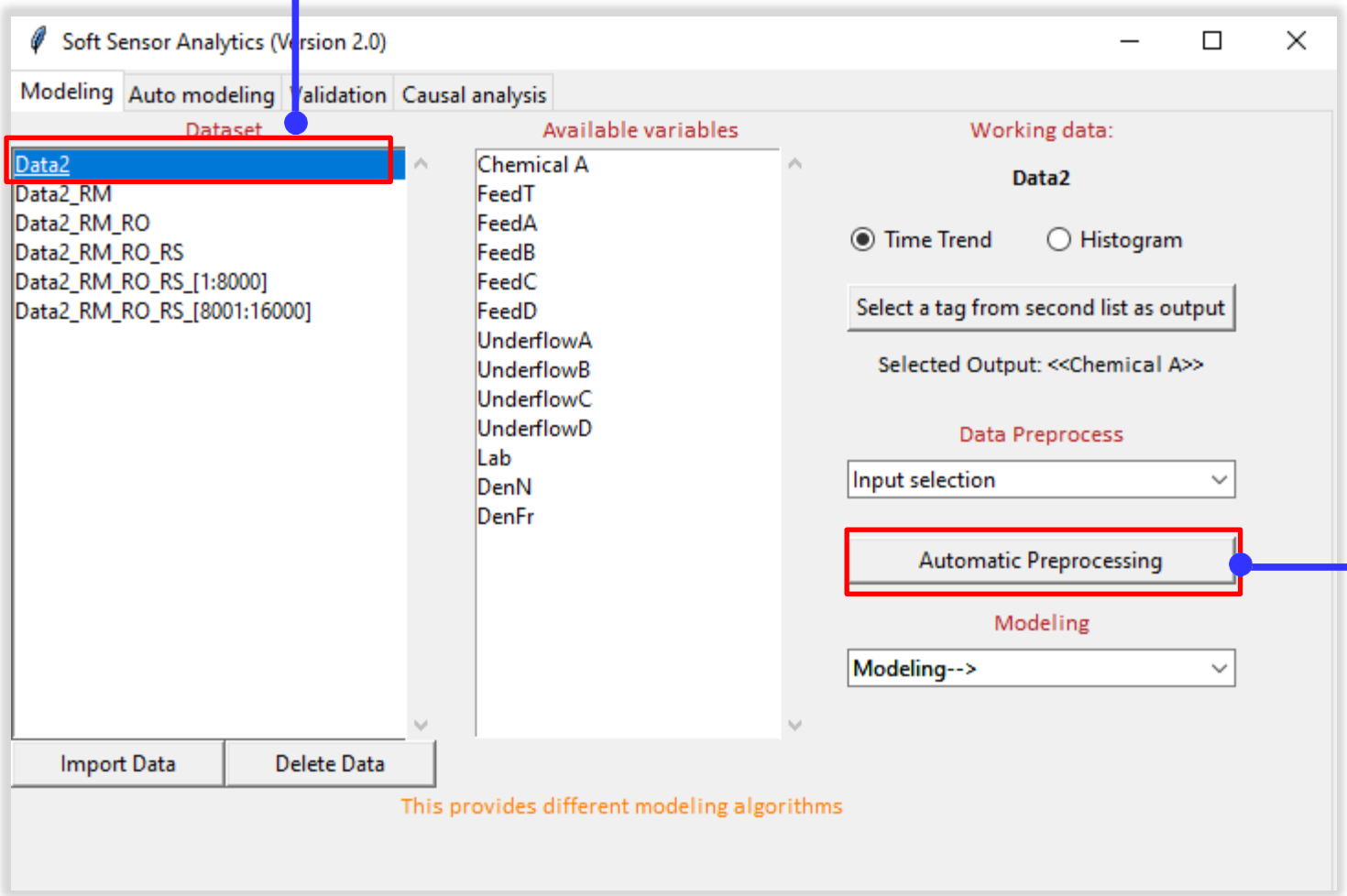
- Choose original data **"Data2"** on the left list

The variables in this dataset will appear in the second list

2

Automatic Preprocessing

- **Click "Automatic Preprocessing"** button



Soft Sensor Analytics

Automatic Preprocessing

3

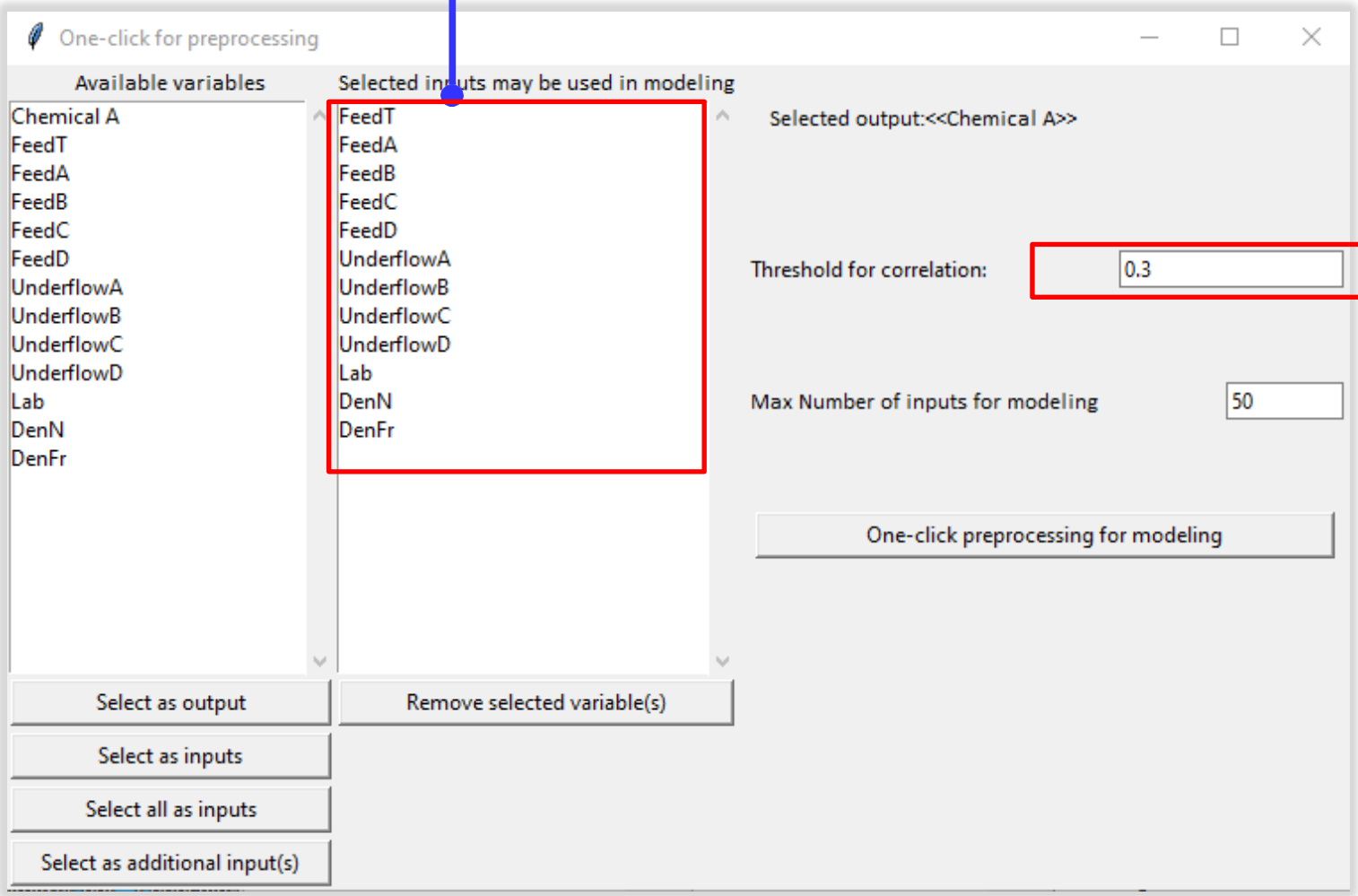
- **Click "Select all as inputs"** button

All the variables except the output will be shown in the second column as candidate inputs

4

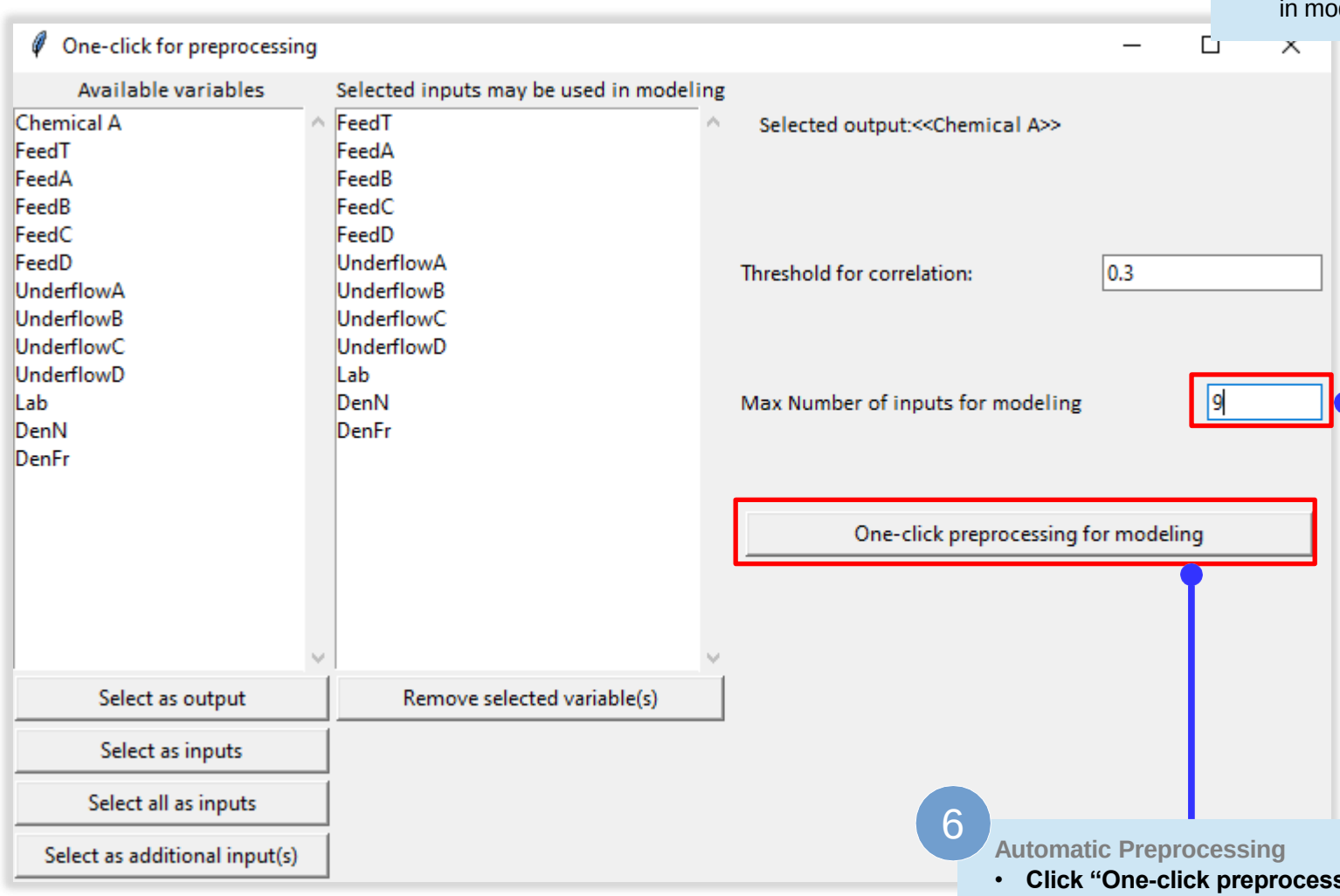
Threshold selection

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



Soft Sensor Analytics

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

6

Automatic Preprocessing

- **Click "One-click preprocessing for modeling"** button

Soft Sensor Analytics

Automatic Preprocessing

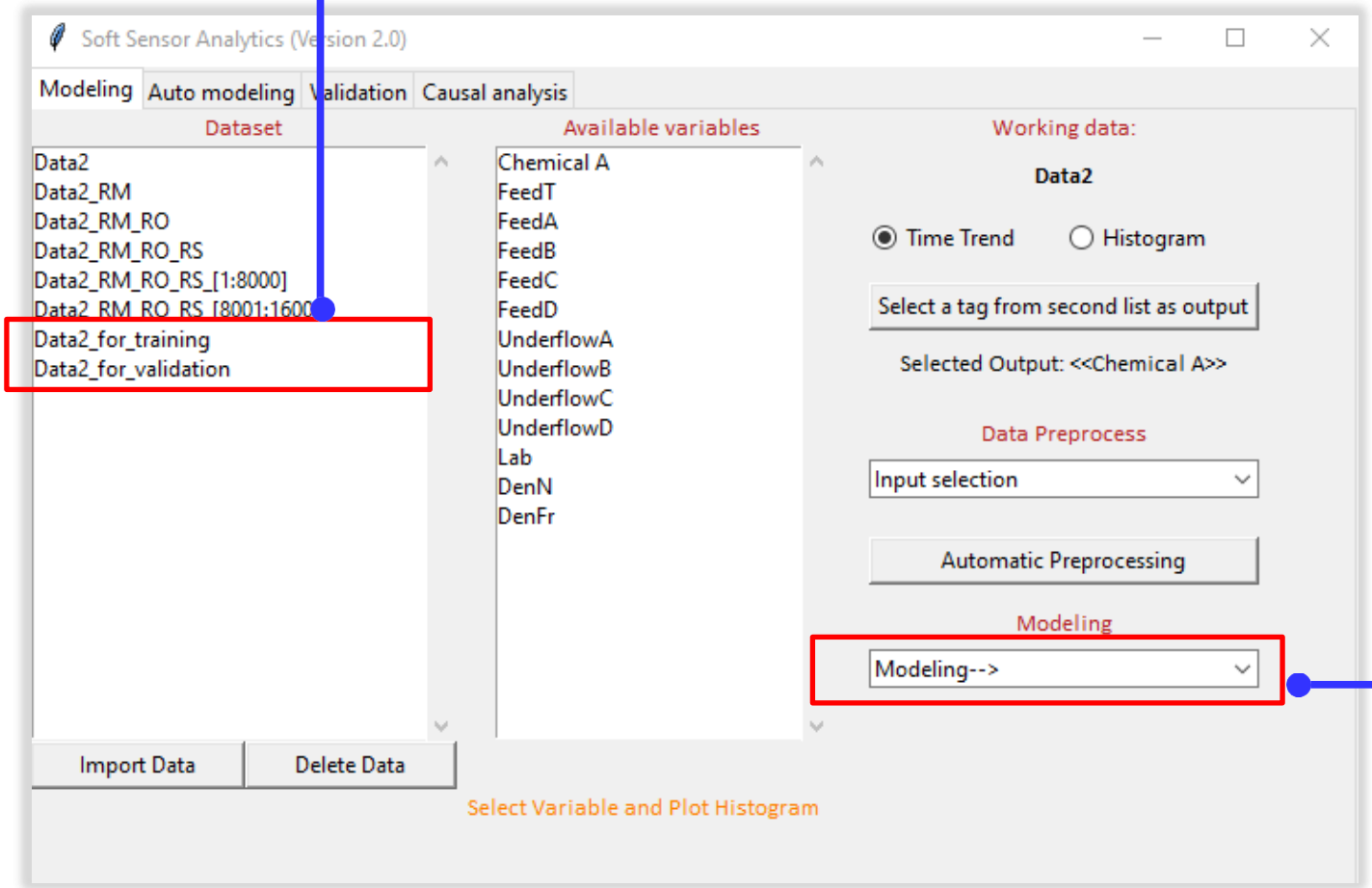
Results

- 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



Soft Sensor Analytics

Another way to model quickly -- Automatic modeling

1

Automatic modeling

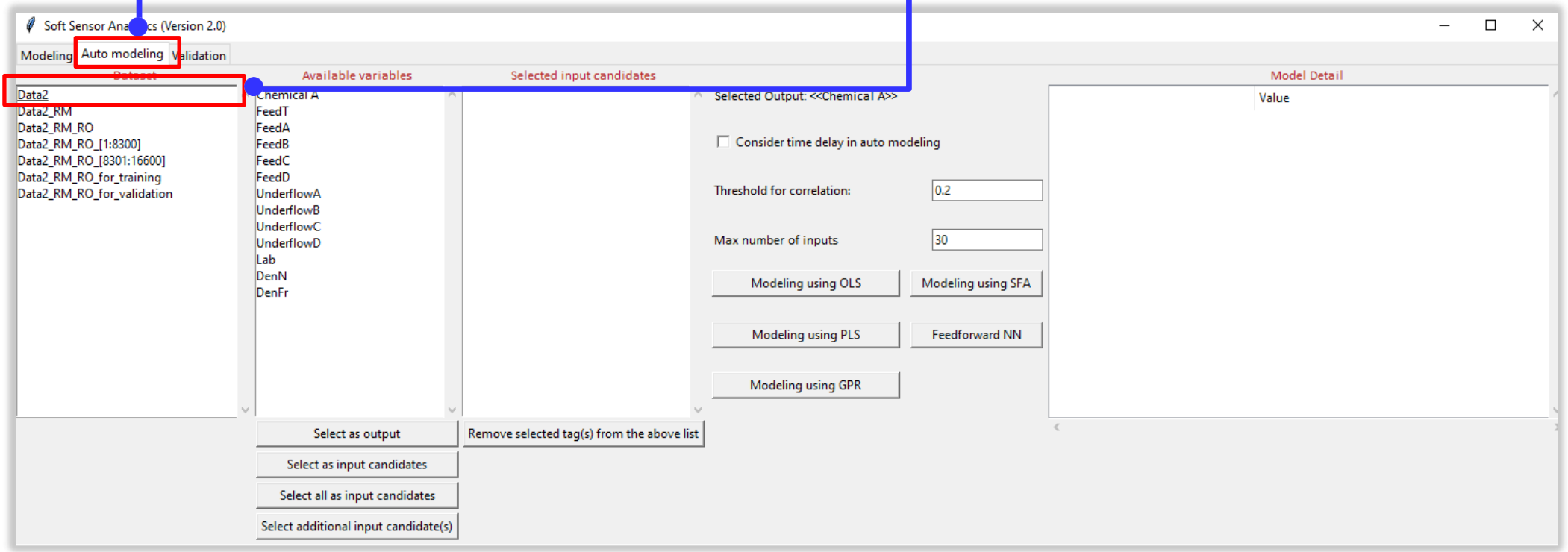
- **Click** the **"Auto modeling"** tab

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

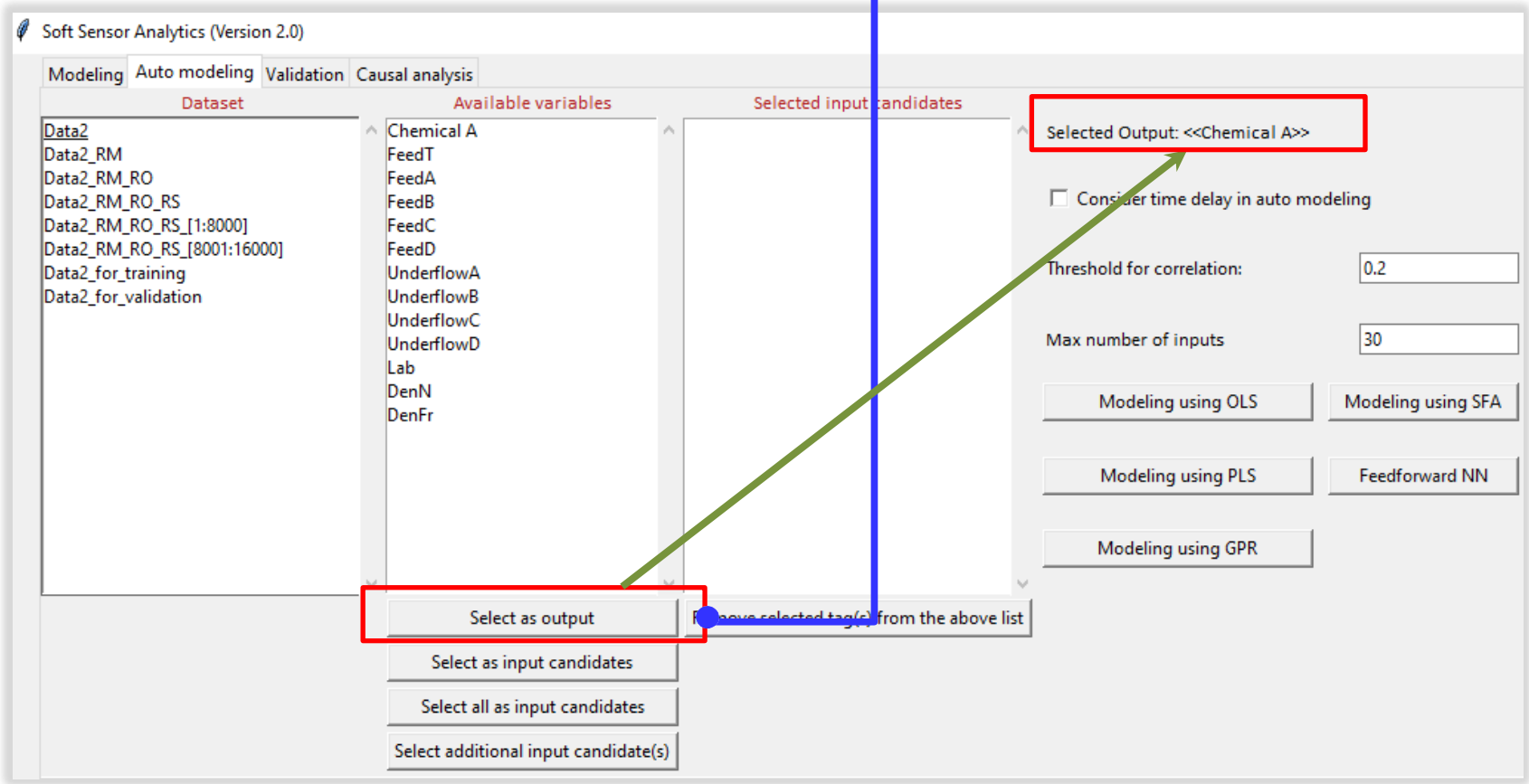


Soft Sensor Analytics

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

Automatic modeling – inputs/output selection

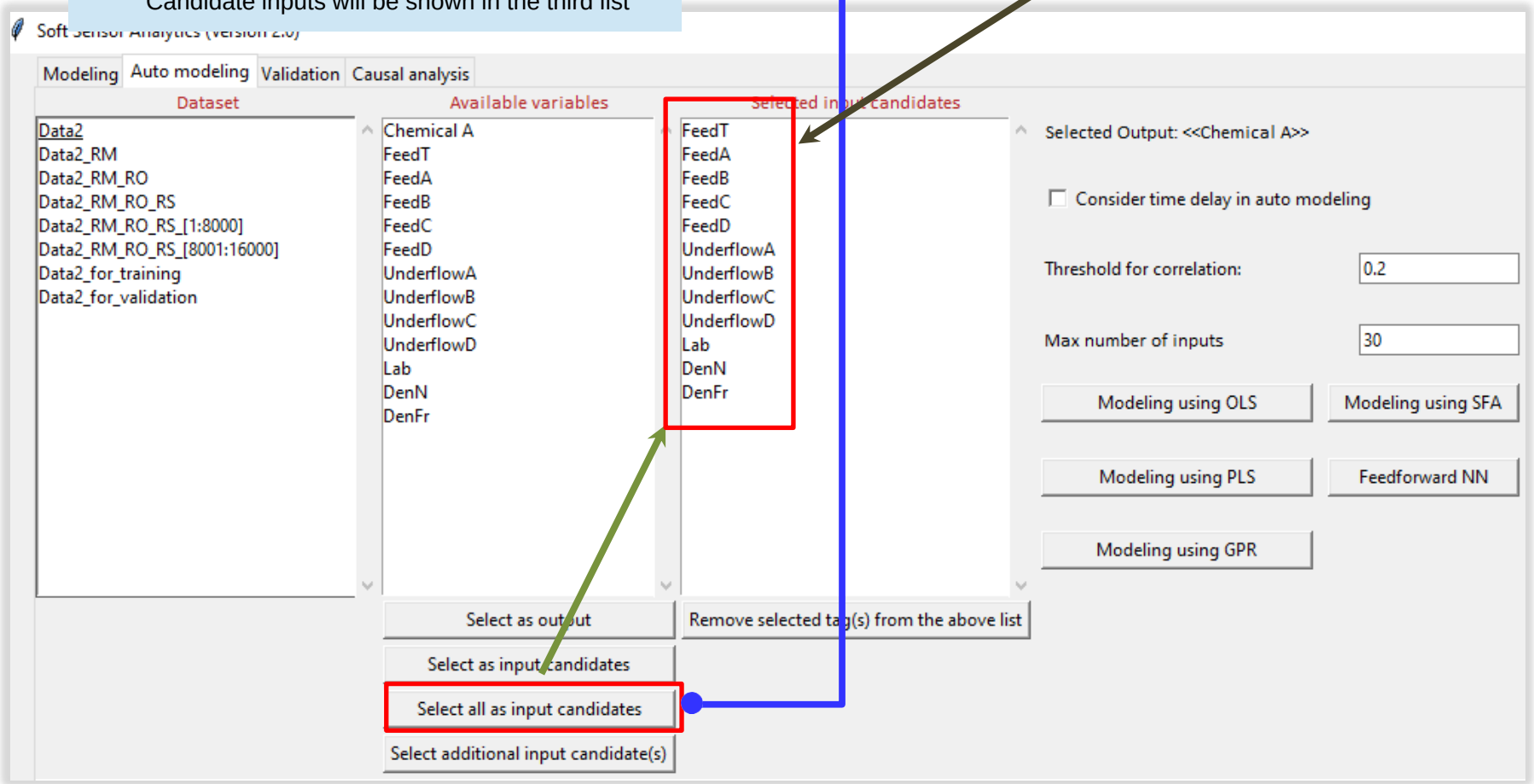
4

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



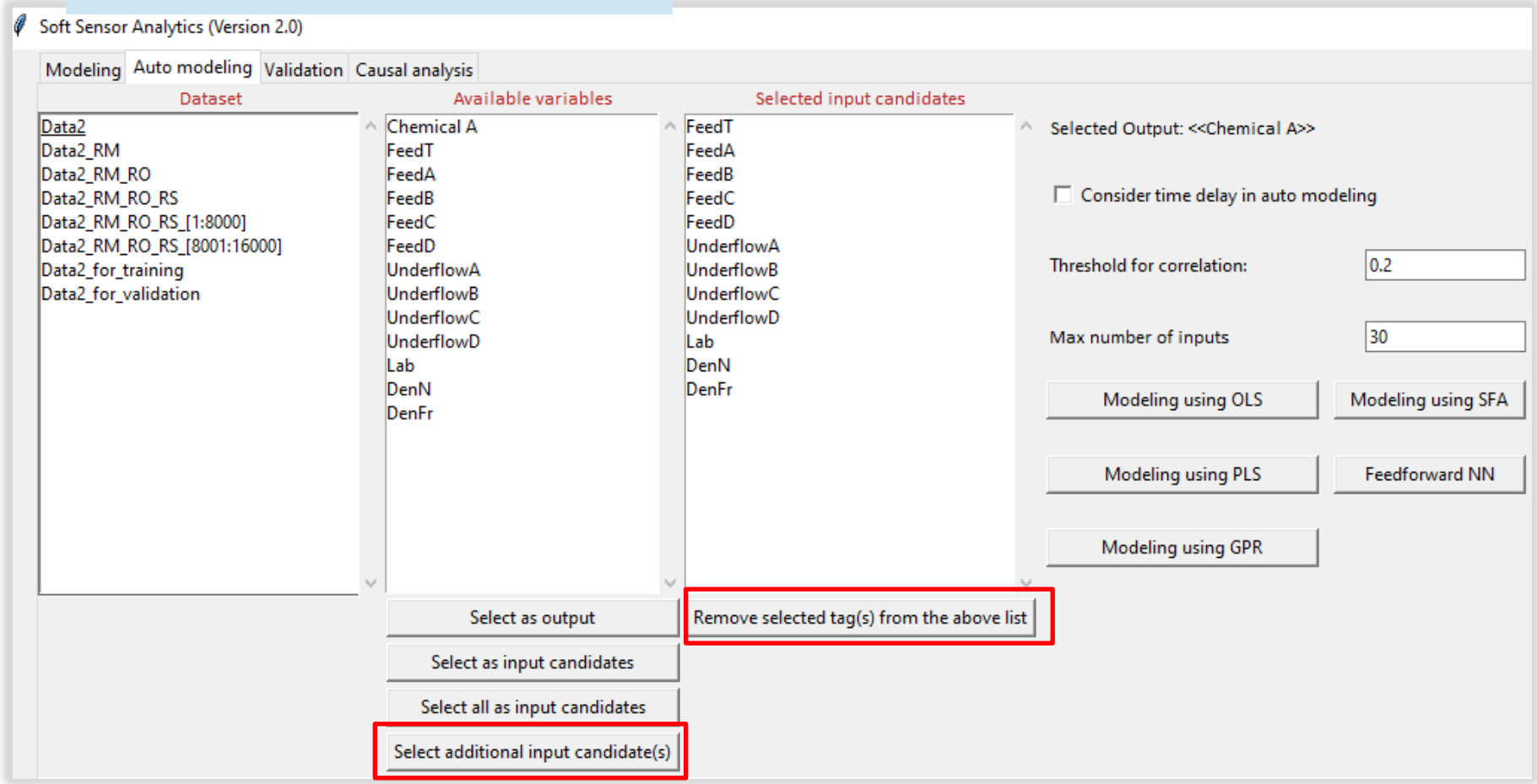
Note: We can also use "Select as input candidates" button to select potential inputs from all variables based on available information/knowledge

Soft Sensor Analytics

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



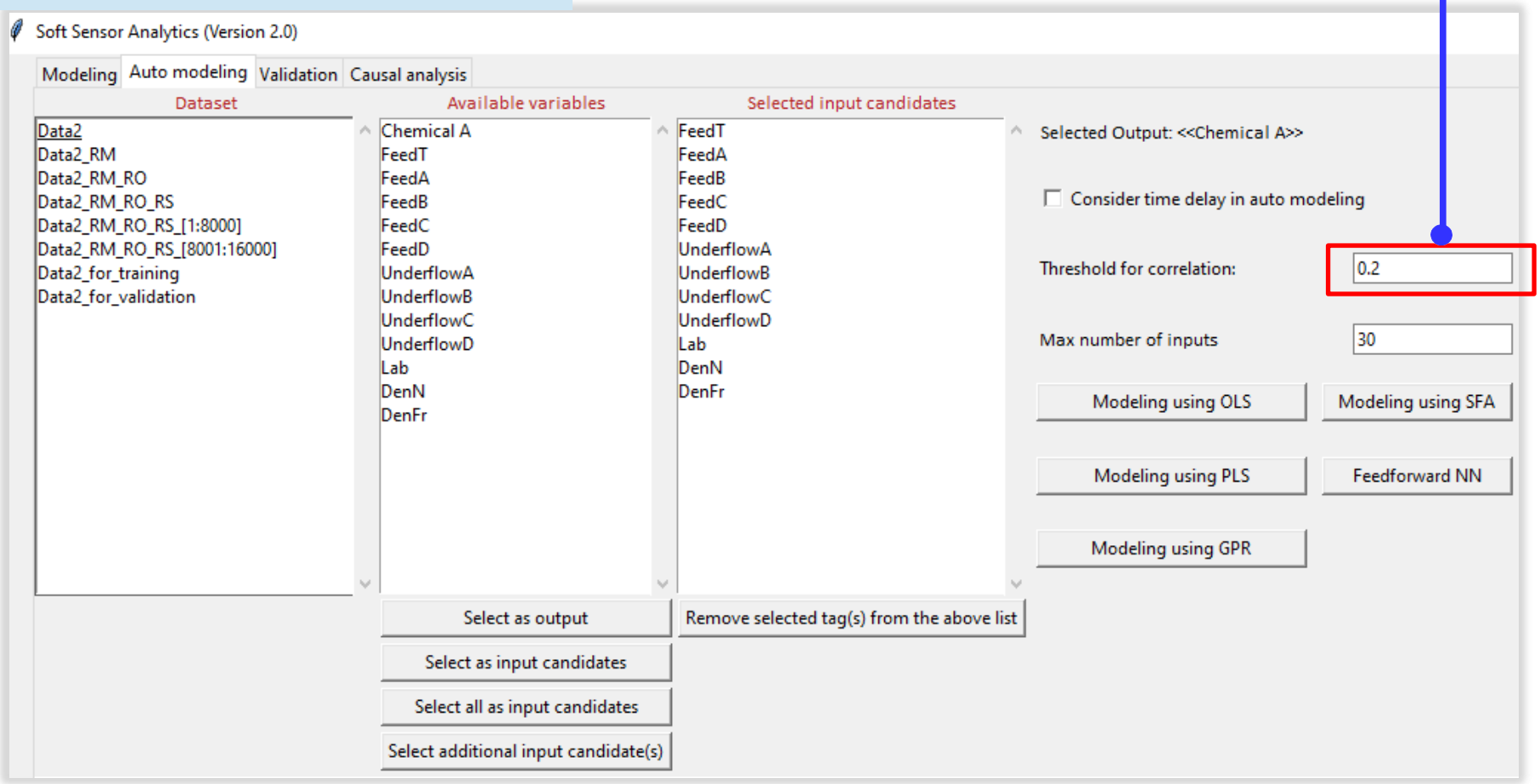
Soft Sensor Analytics

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

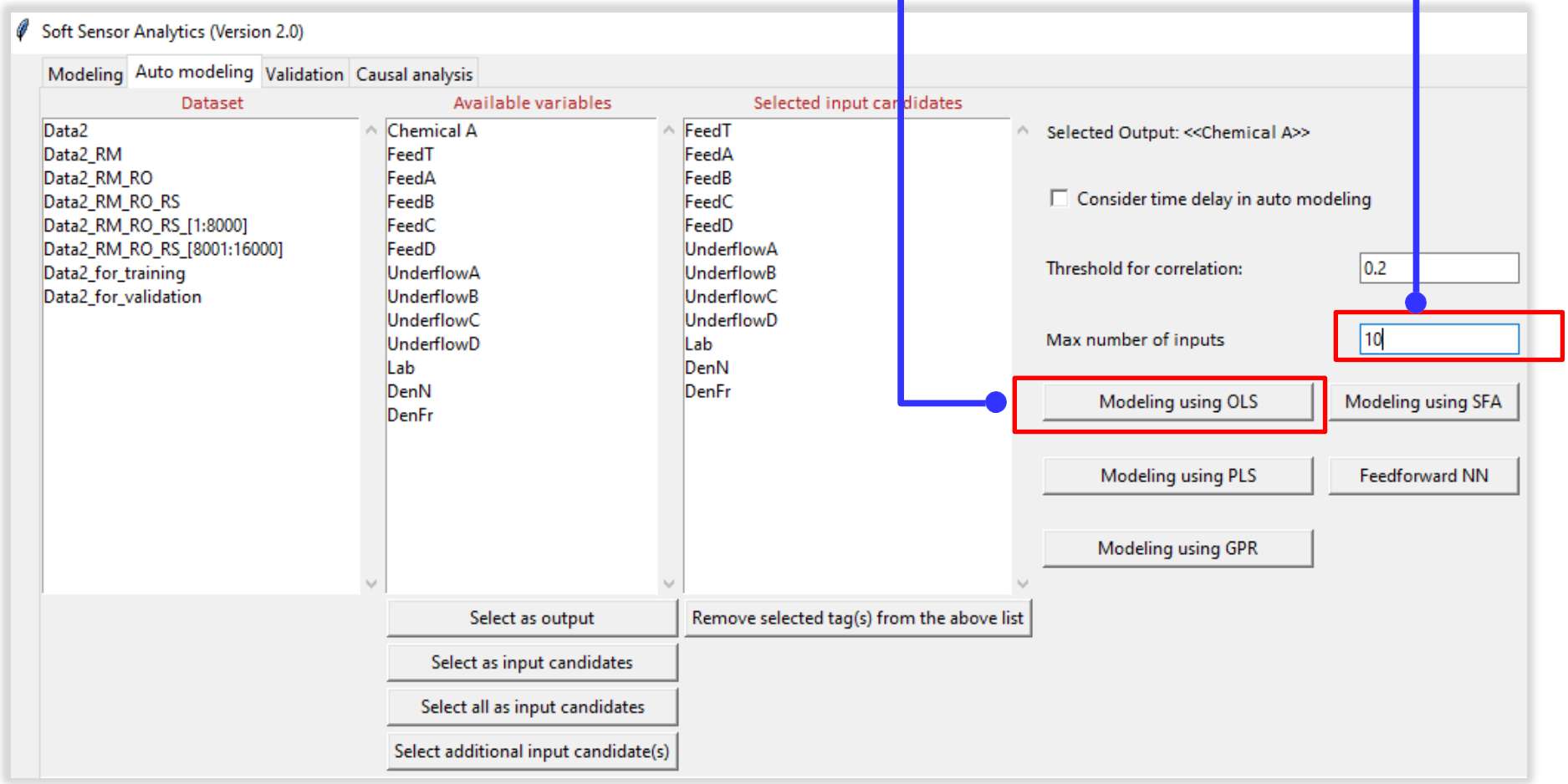


Soft Sensor Analytics

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

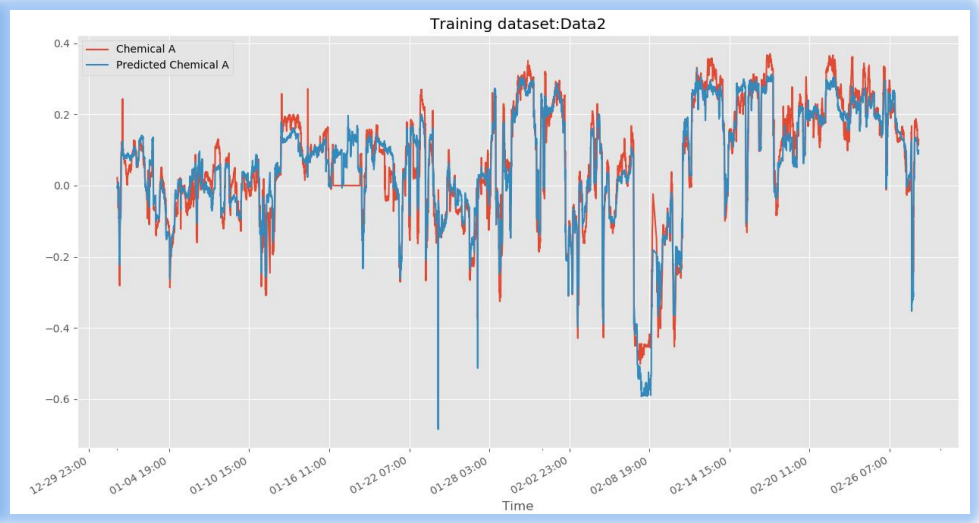


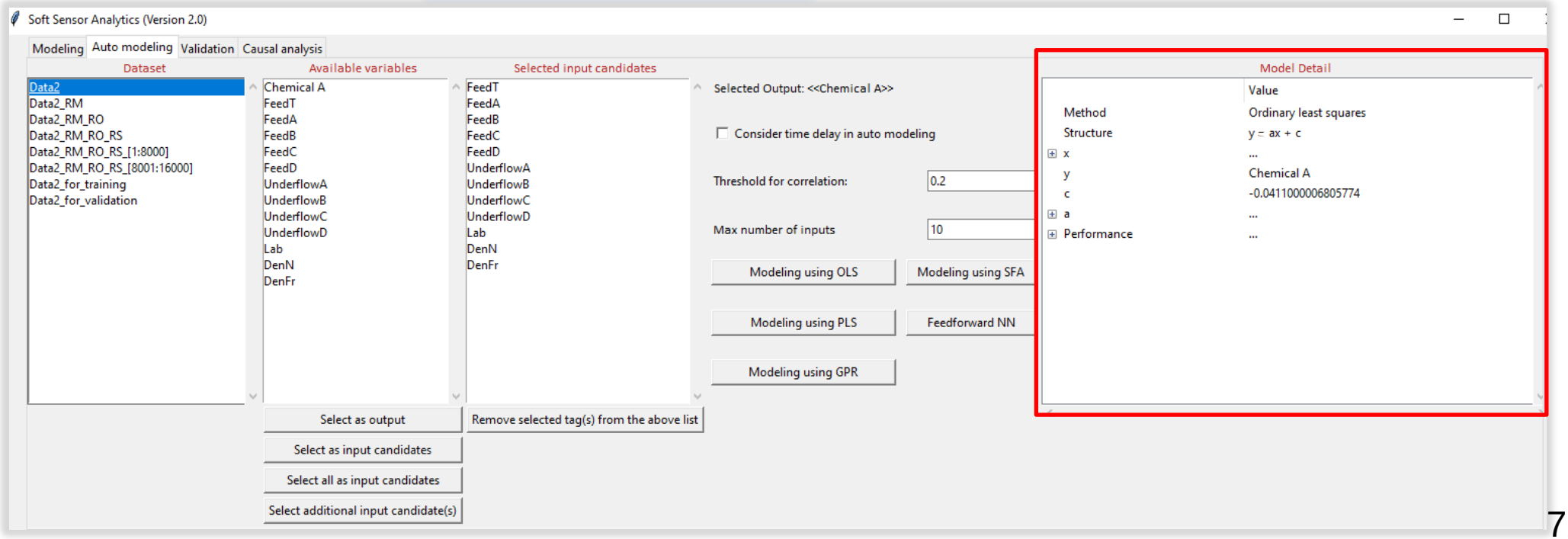
Soft Sensor Analytics

Automatic modeling - Results

Results

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





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

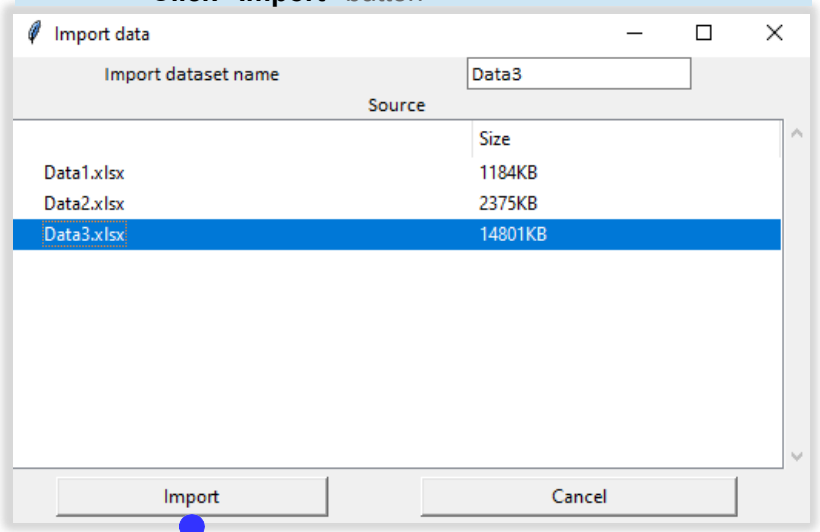
Soft Sensor Analytics

Import data sets

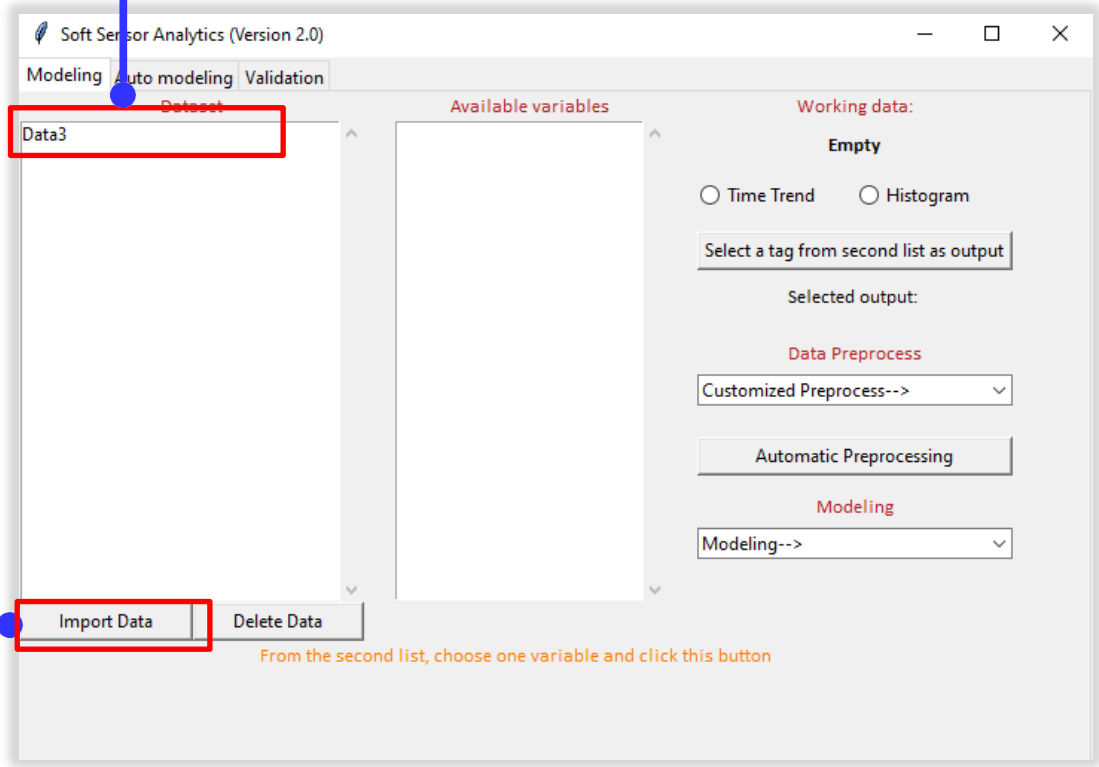
Note: Please save data sets in the ".xlsx" format

1

- **Click "Import Data"** button
- **Select** data file **"Data3.xlsx"**
- **Click "Import"** button



"**Data3**" has been imported

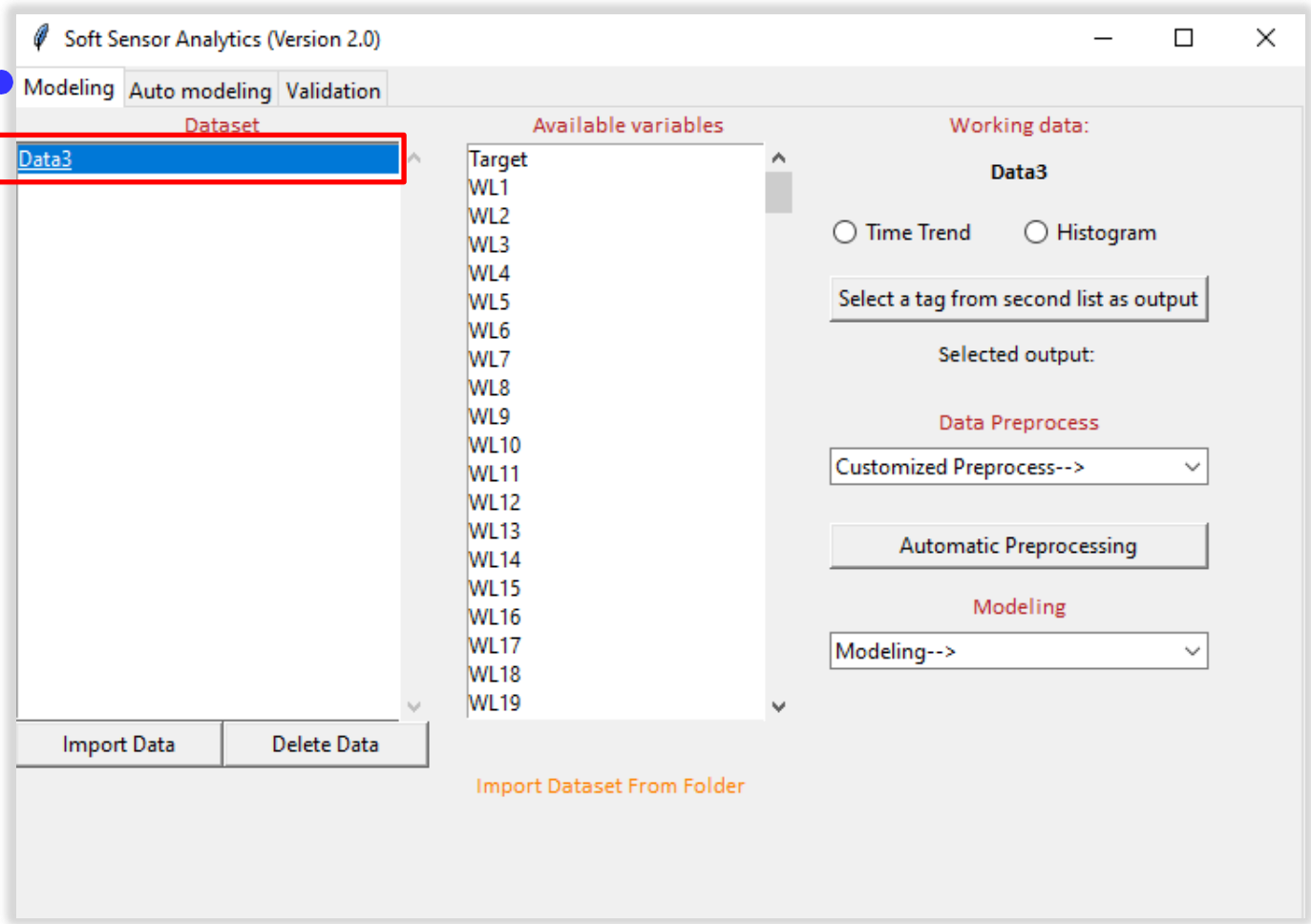


Soft Sensor Analytics

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

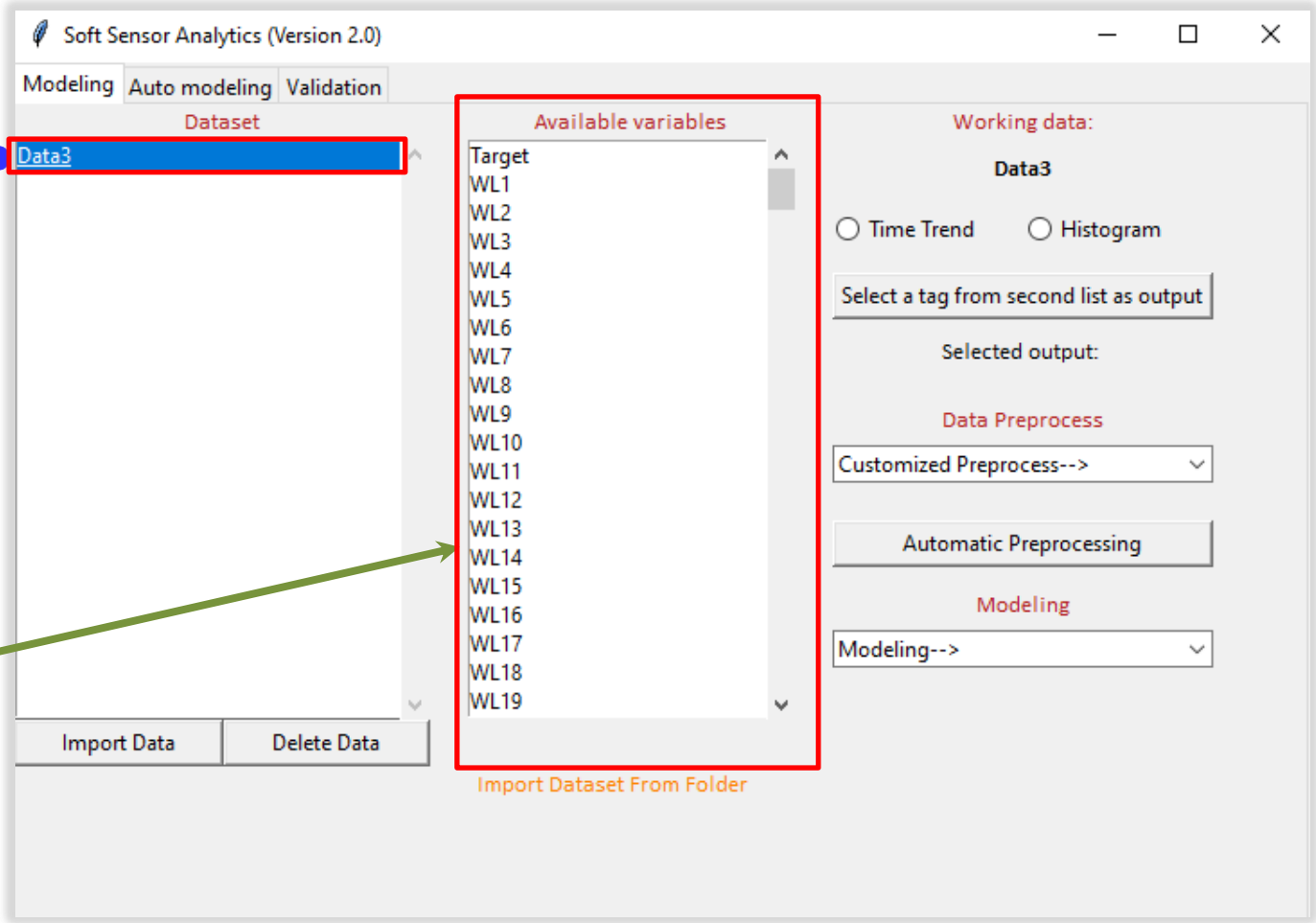
Variables in Data3

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.

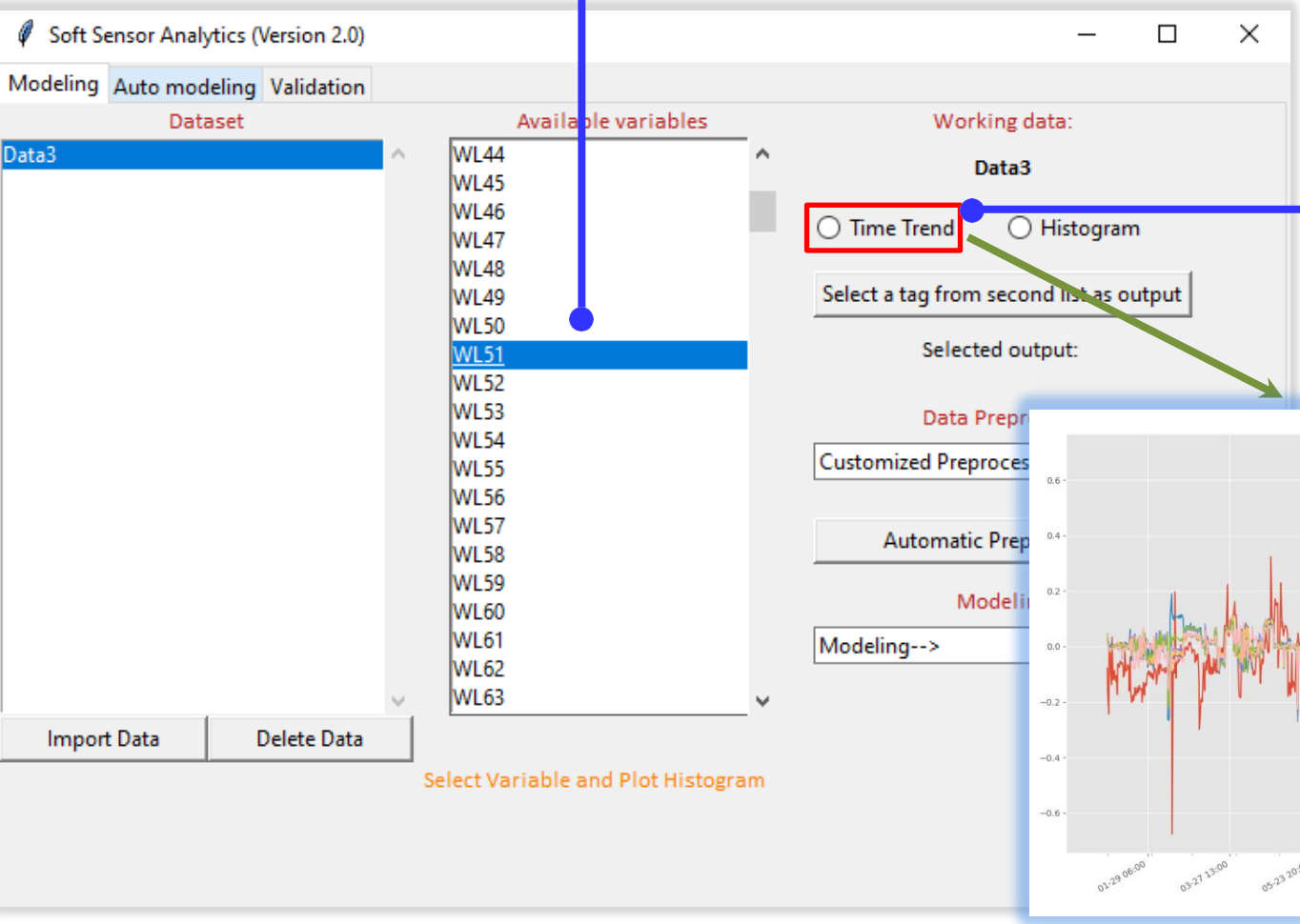


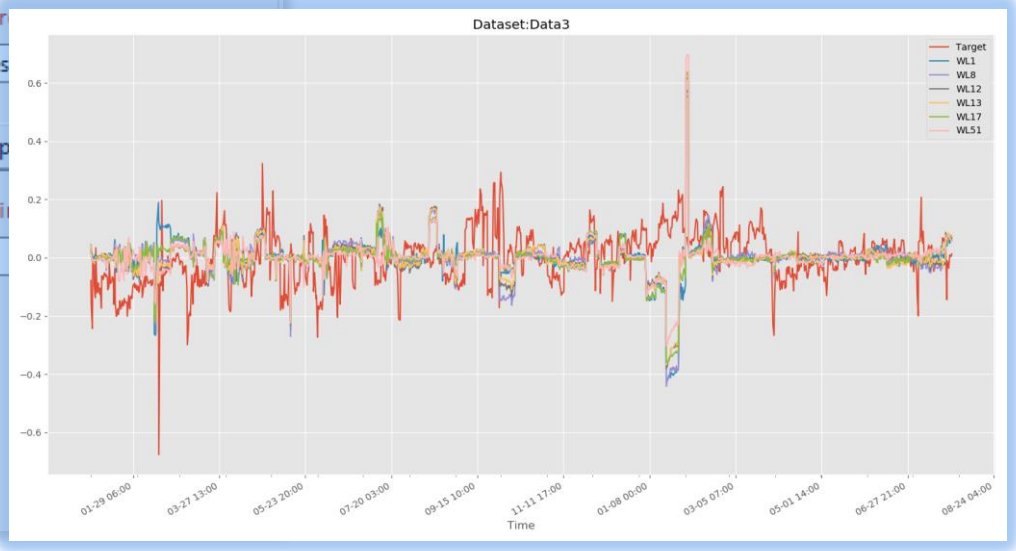
Soft Sensor Analytics

Visualize several variables of Data3

3 **Click** to select variables for visualization

4 **Click "Time Trend"** to visualize the selected variables



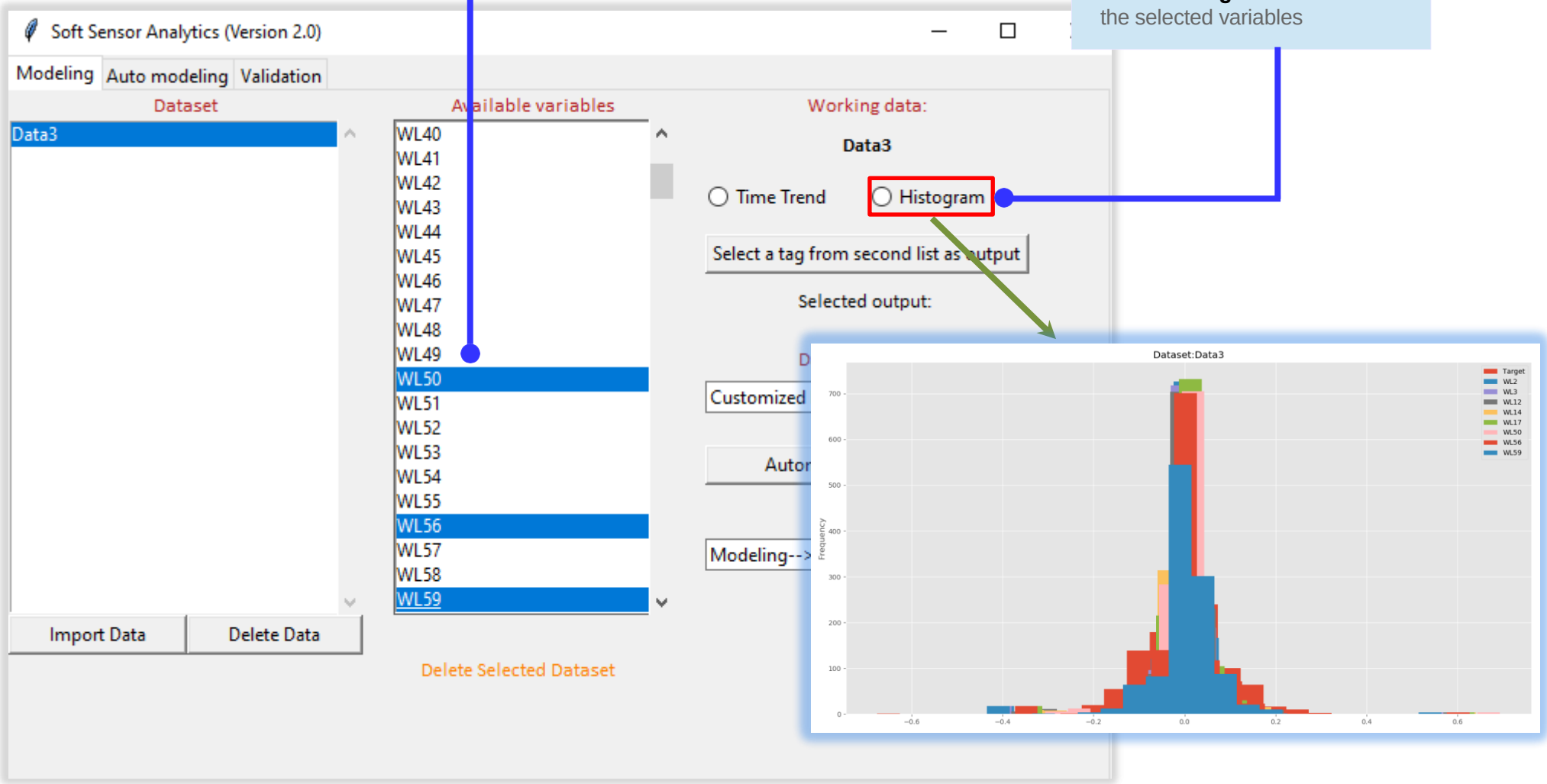


Soft Sensor Analytics

Visualize several variables of Data3

5 **Click** to select variables for visualization

6 **Click "Histogram"** to visualize the selected variables



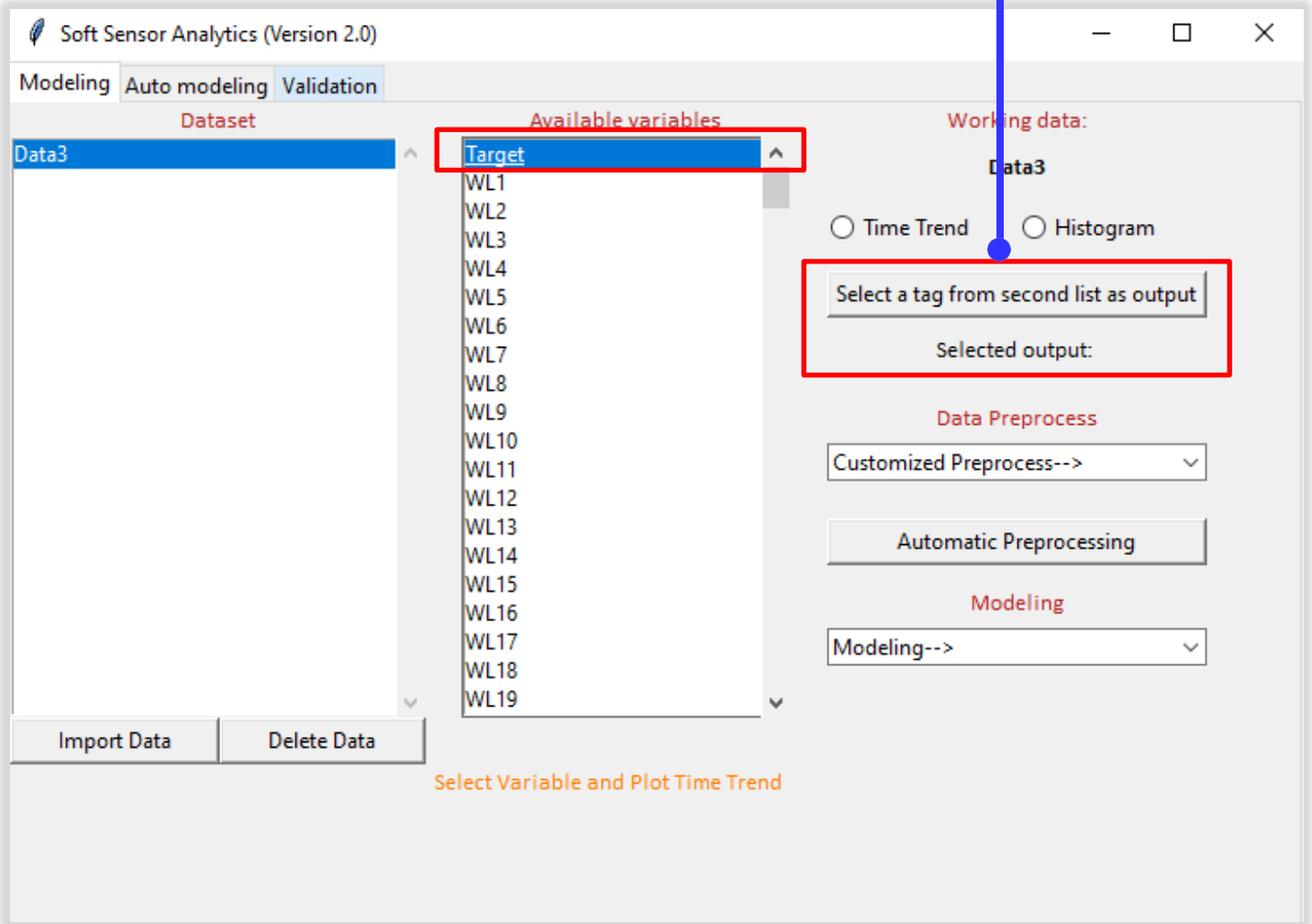
Soft Sensor Analytics

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



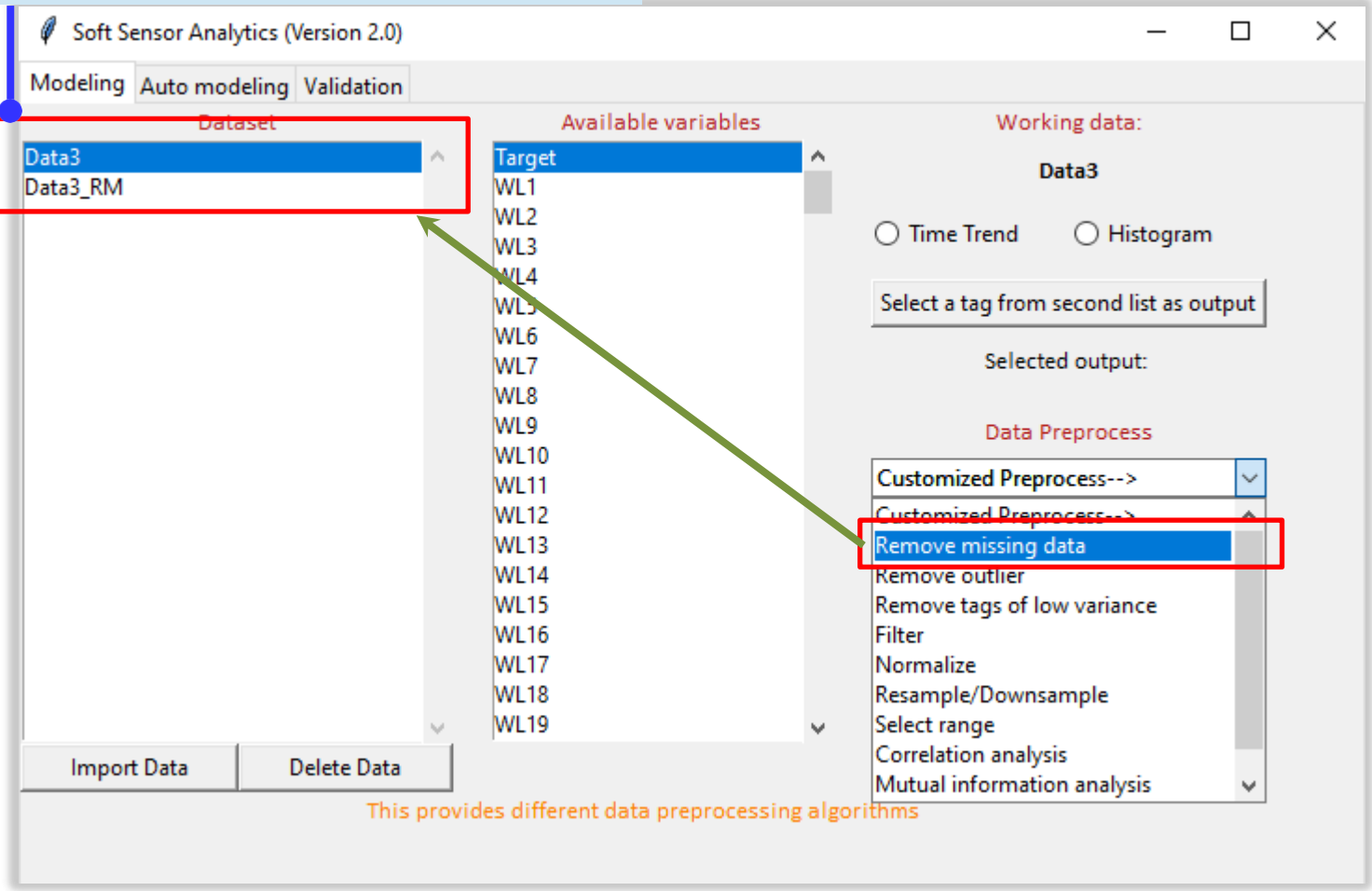
Soft Sensor Analytics

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"



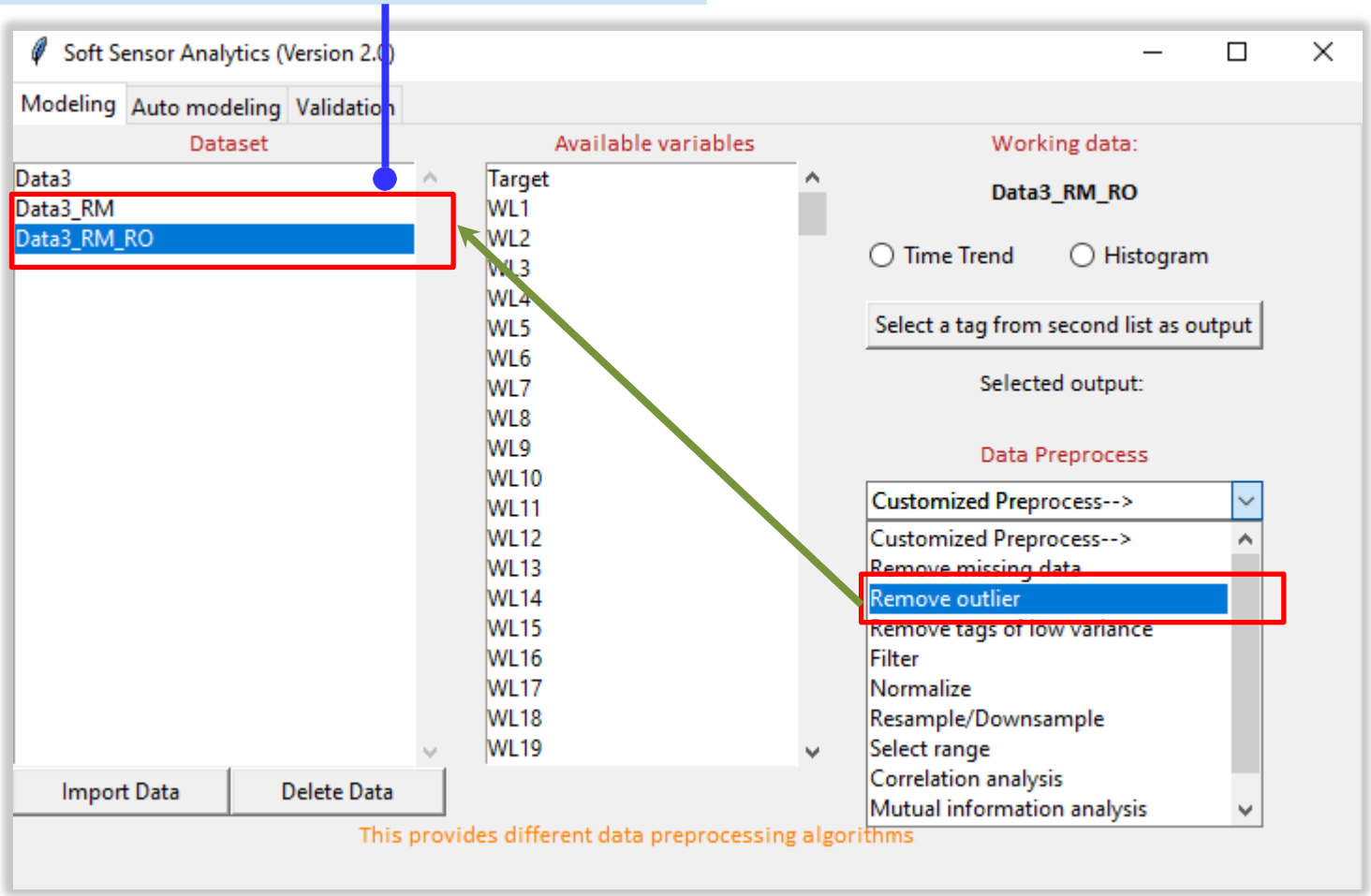
Soft Sensor Analytics

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"



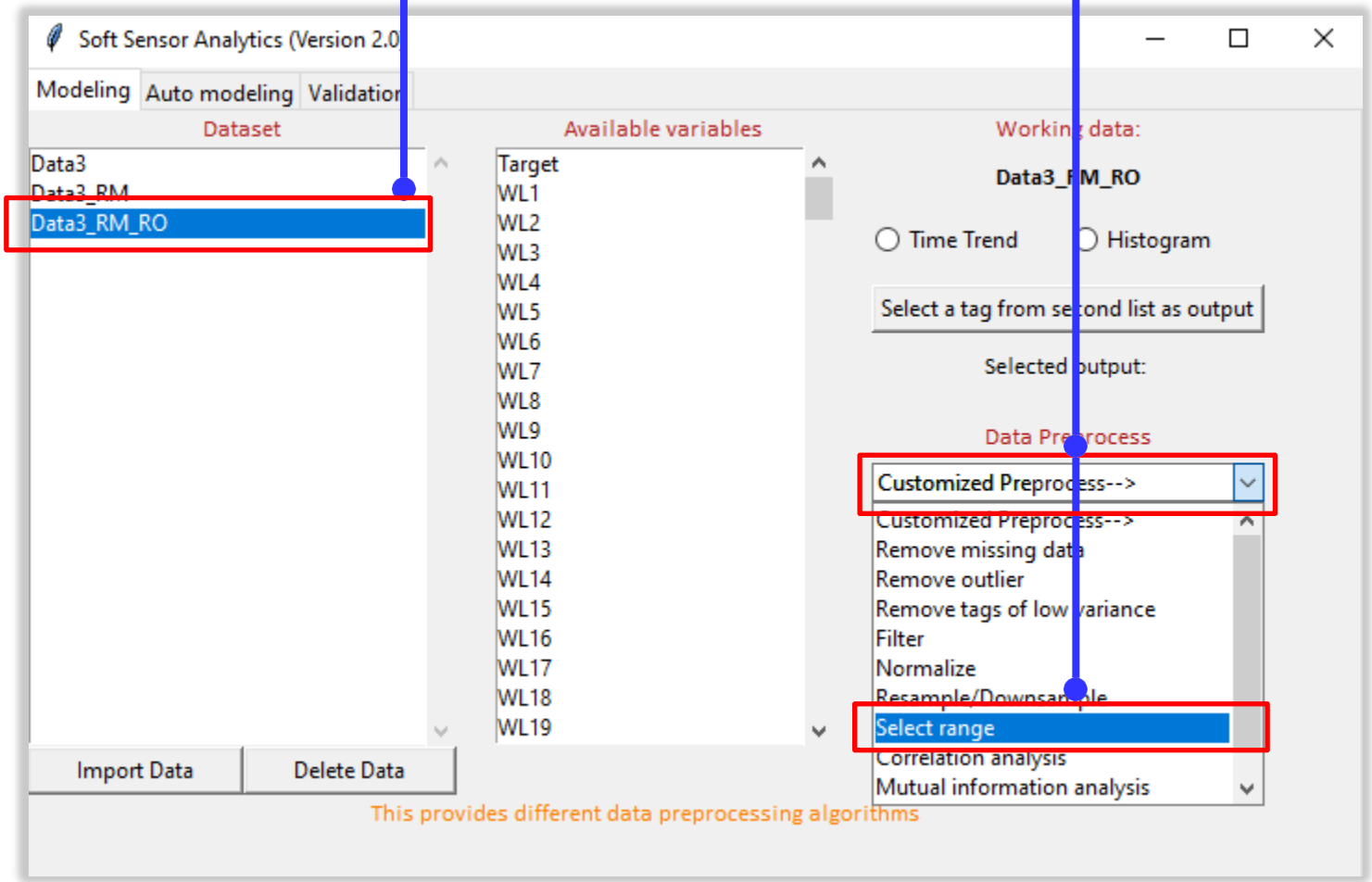
Soft Sensor Analytics

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



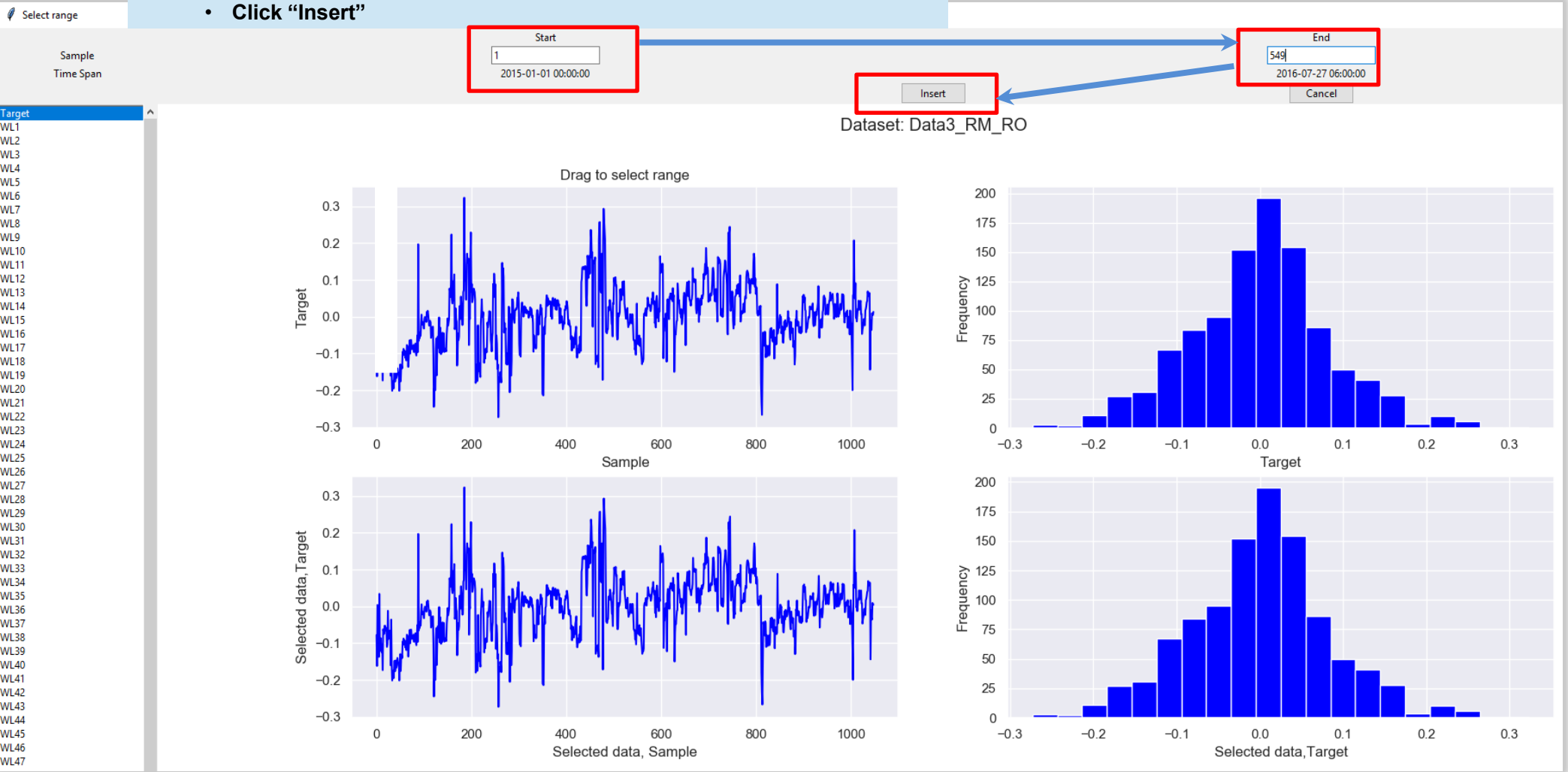
Soft Sensor Analytics

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"**



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

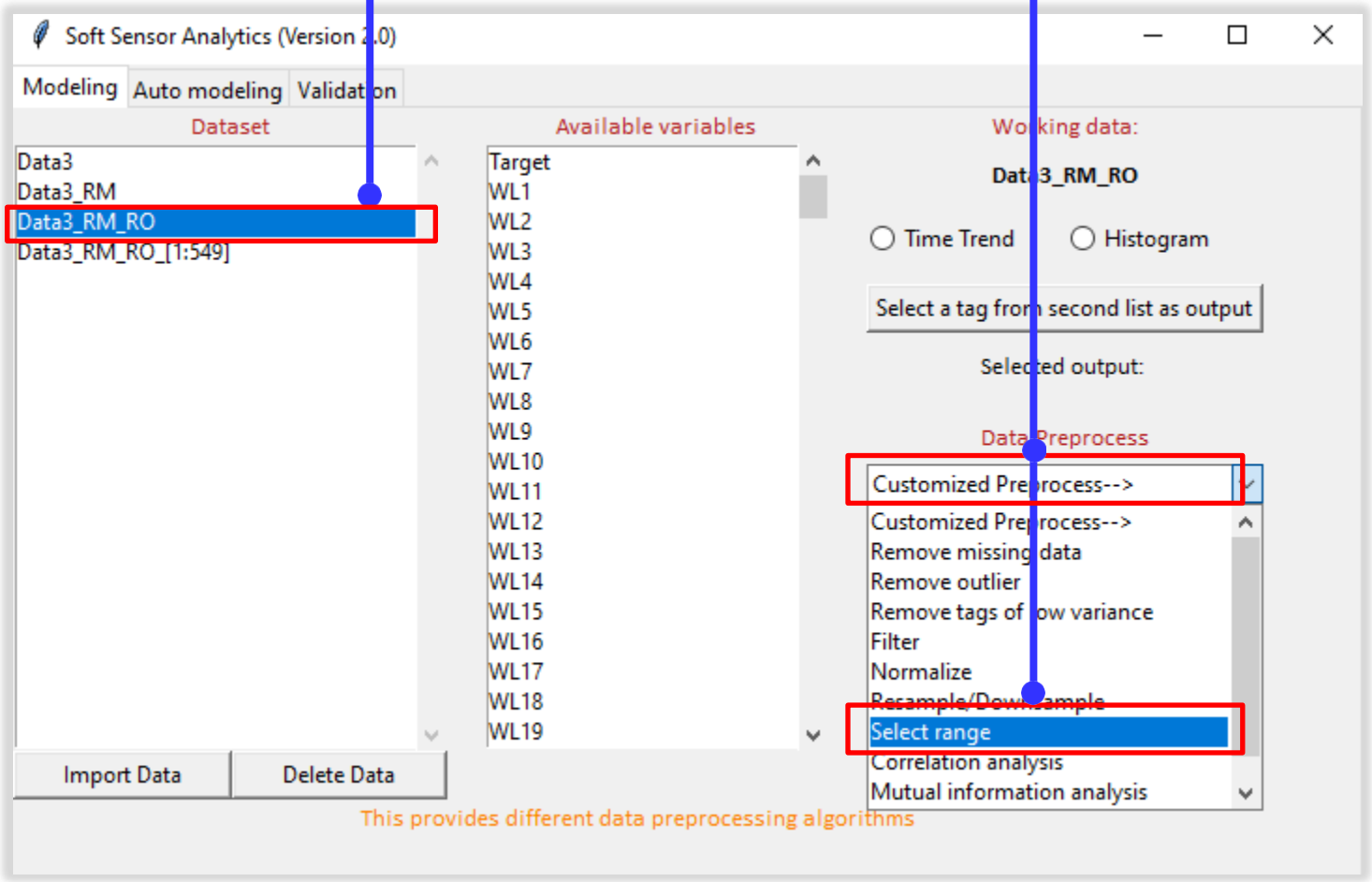
Soft Sensor Analytics

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



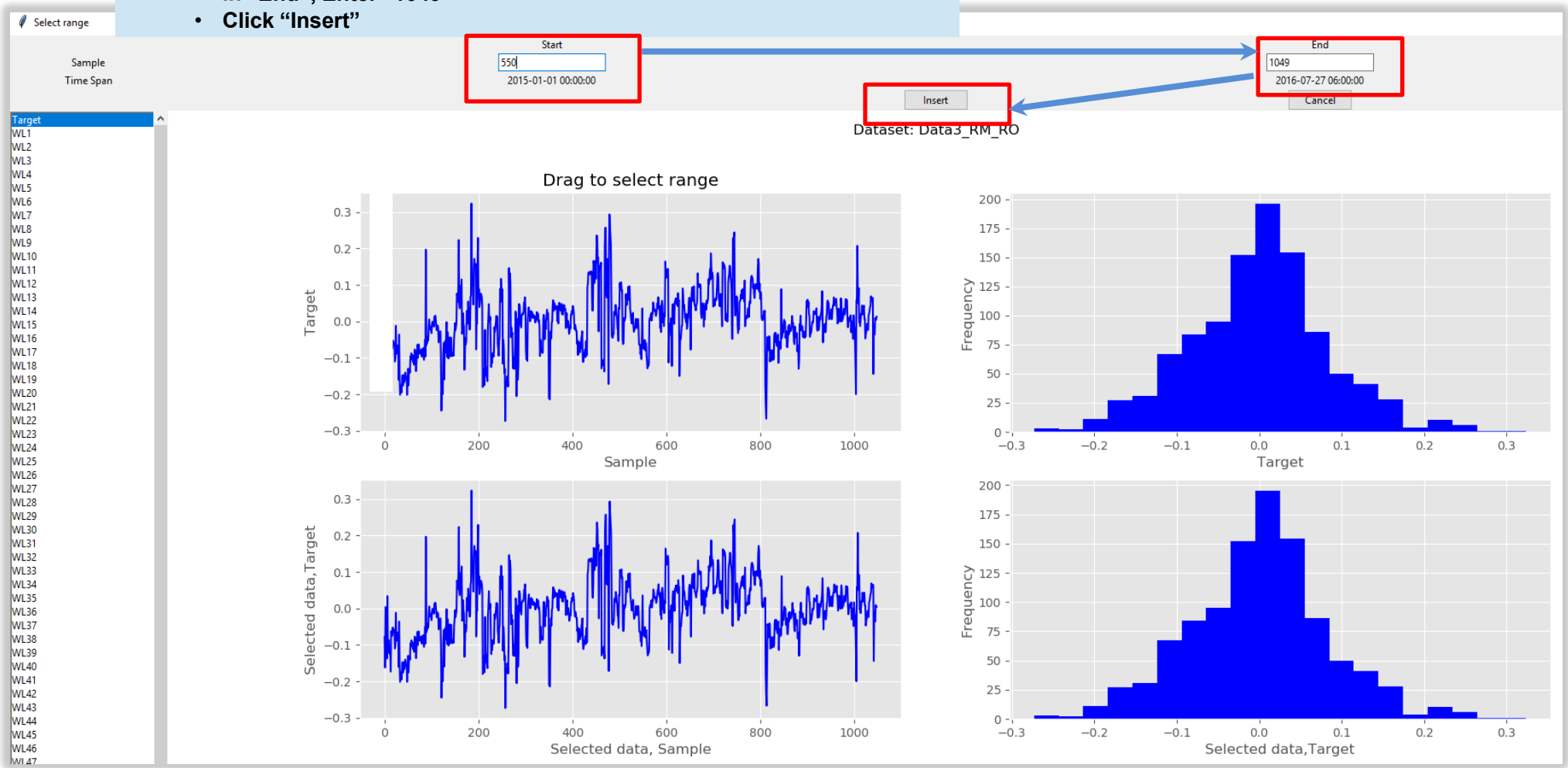
Soft Sensor Analytics

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"**

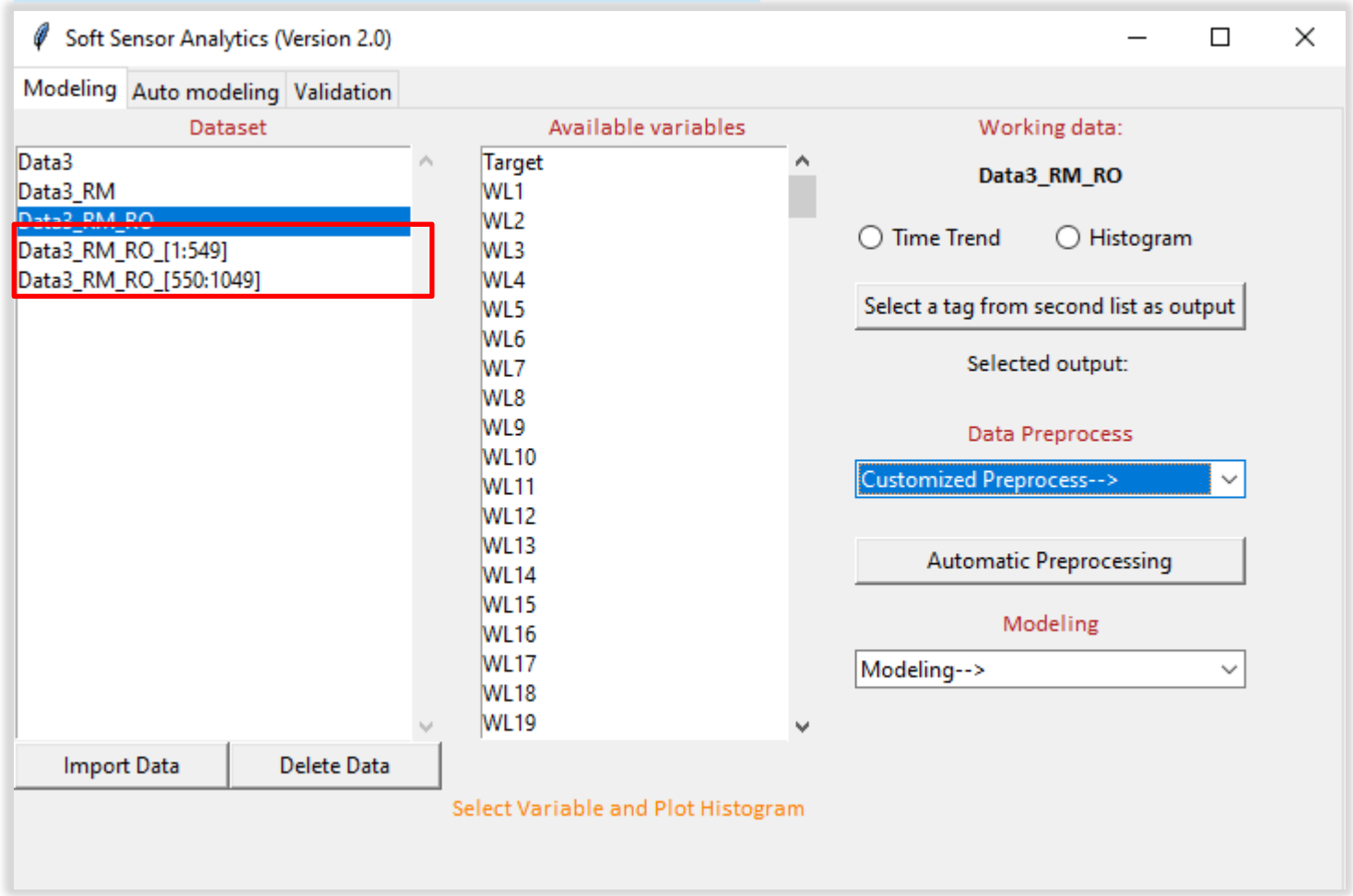


Note: Please close the current popup figures before taking next action

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



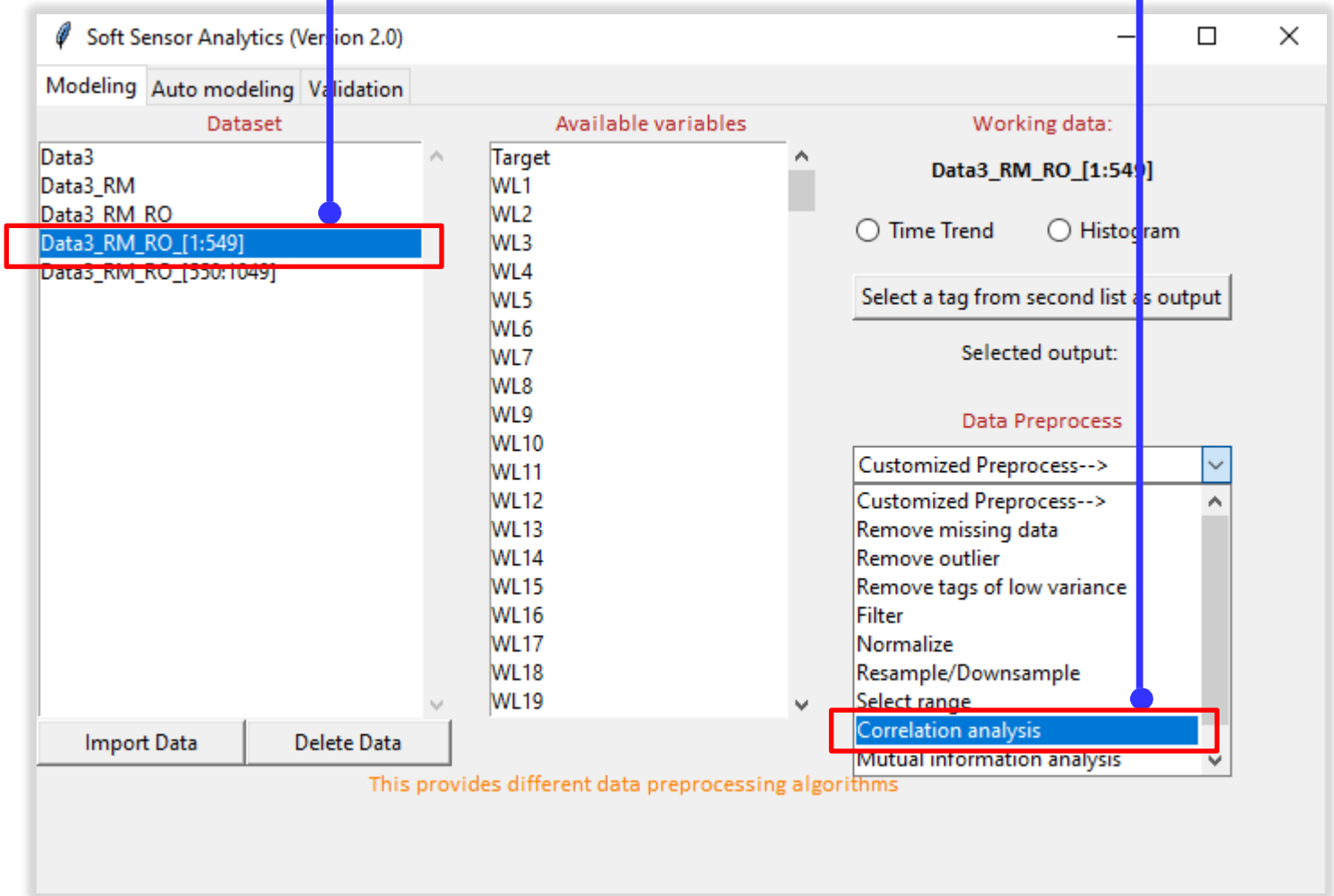
Soft Sensor Analytics

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



Soft Sensor Analytics

Preprocessing - Correlation analysis

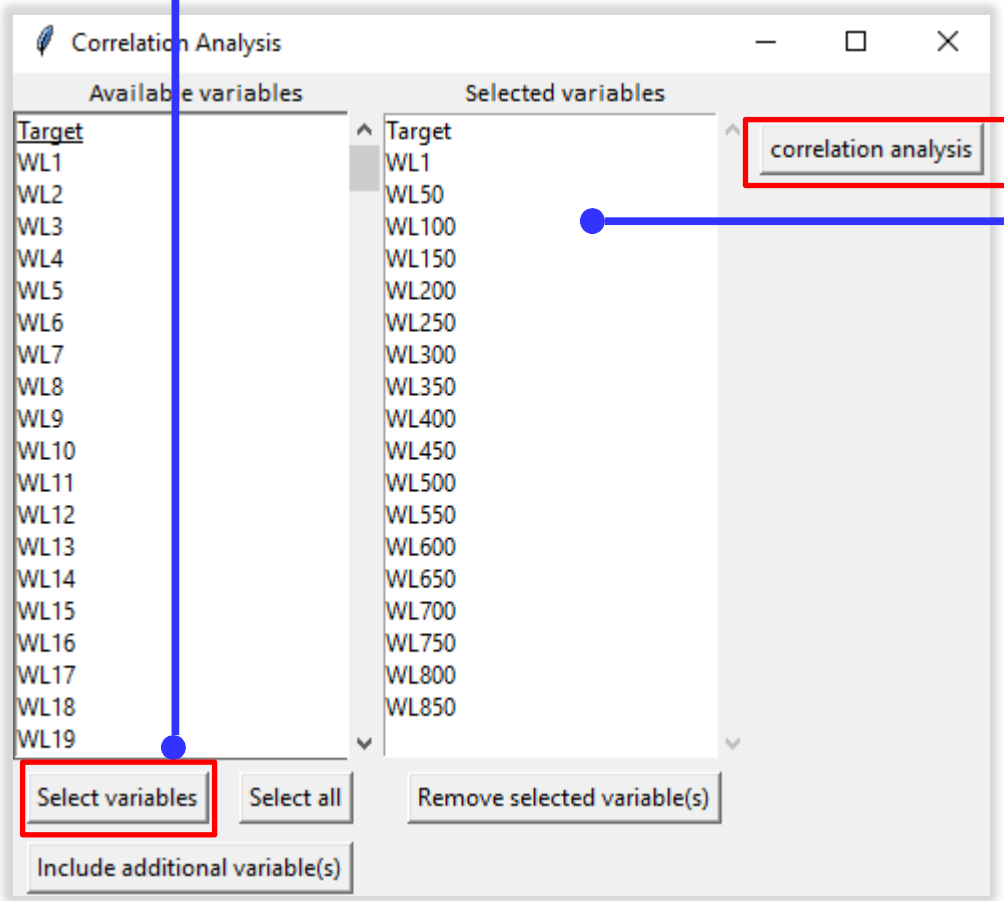
3

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



4

Correlation analysis

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

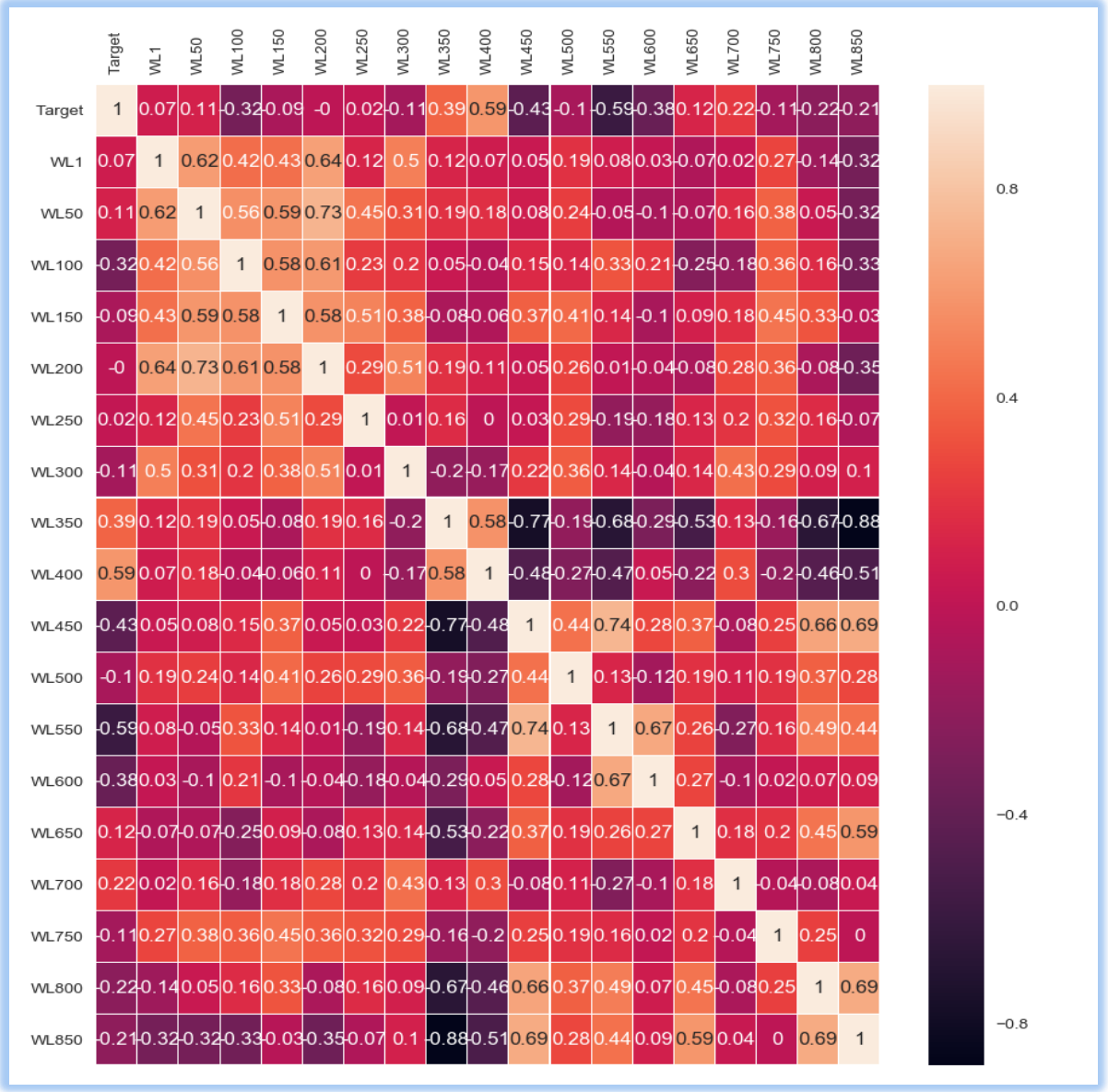
Soft Sensor Analytics

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.



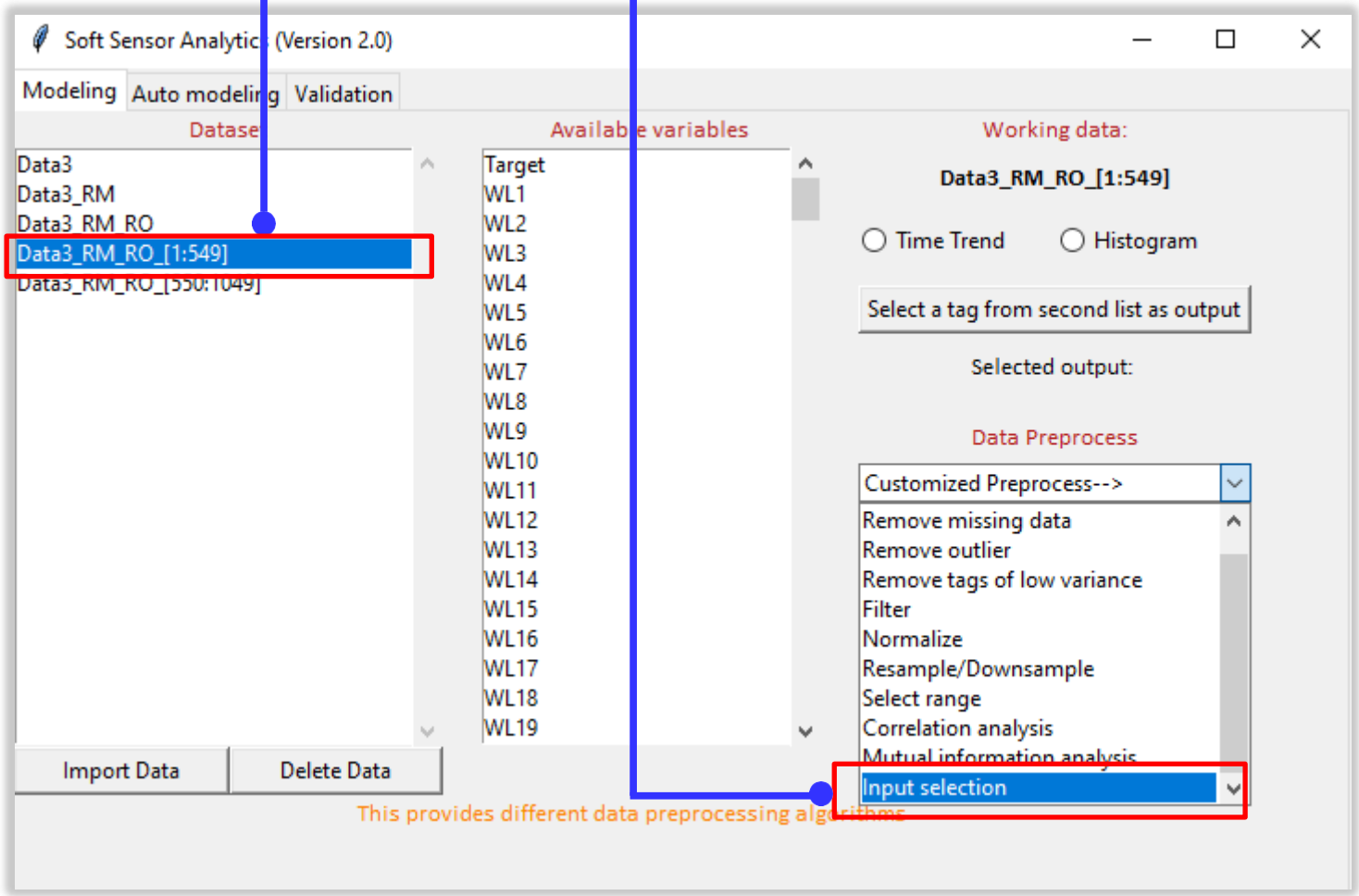
Soft Sensor Analytics

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**

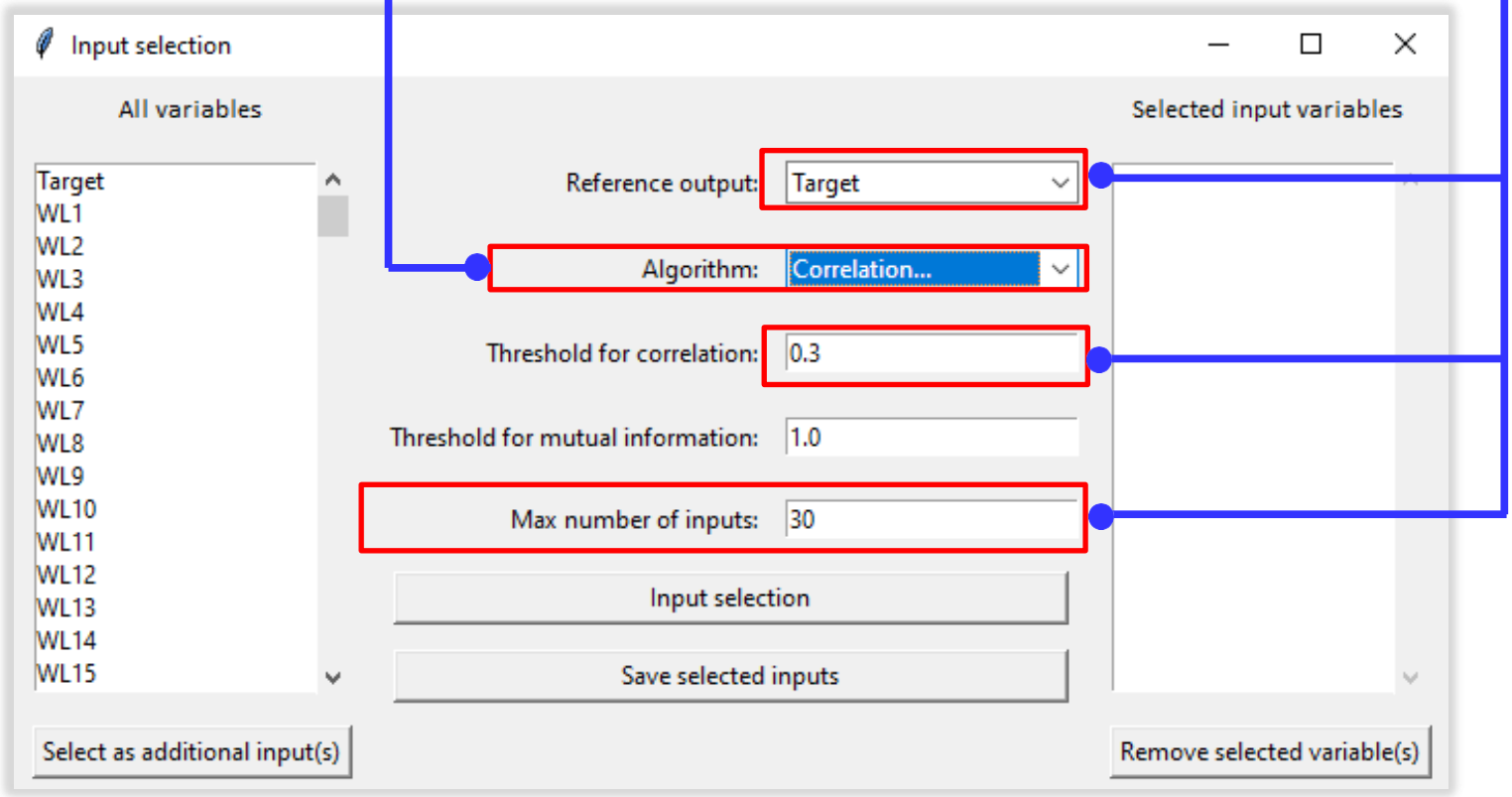


Soft Sensor Analytics

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



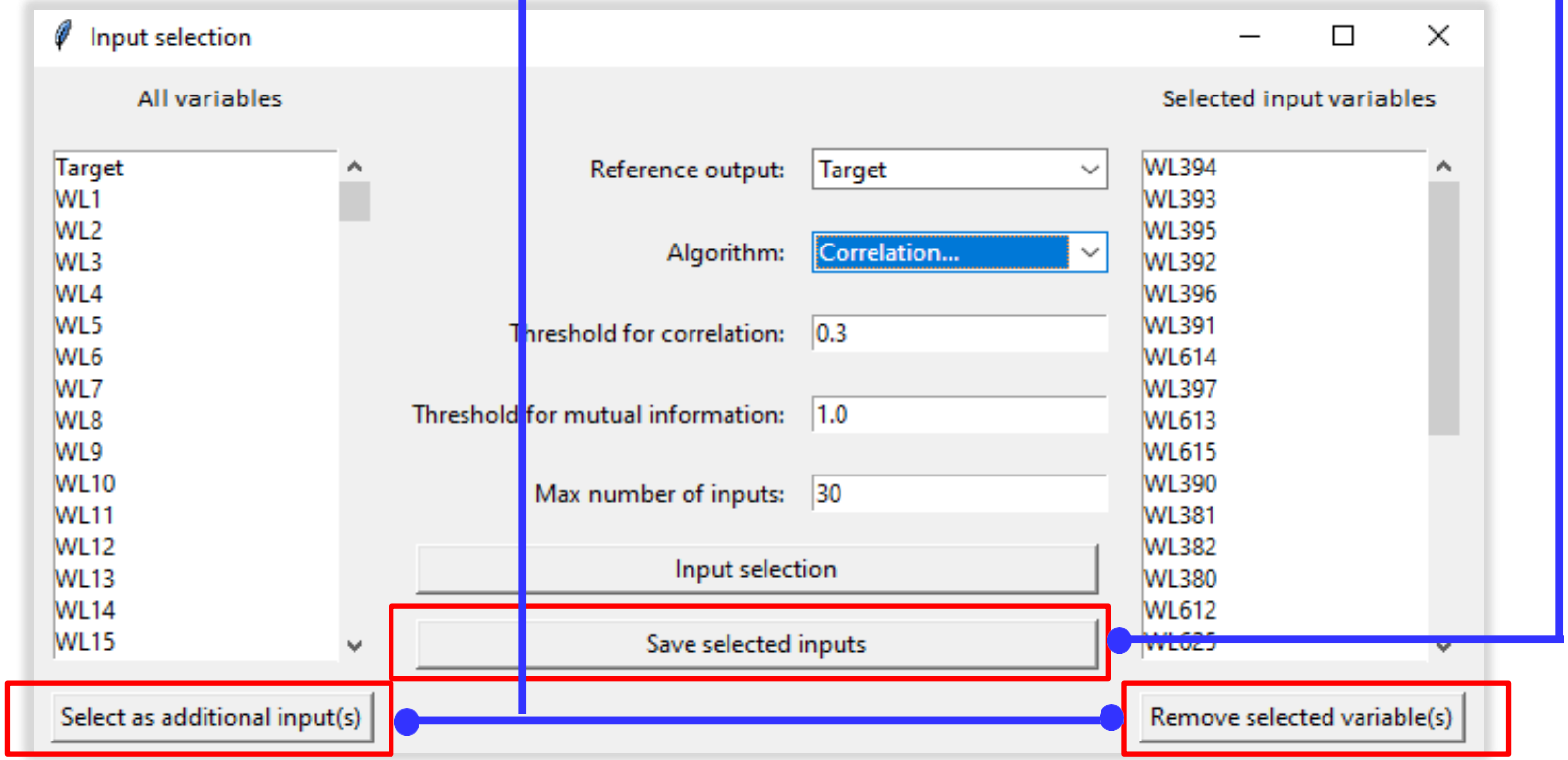
Soft Sensor Analytics

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

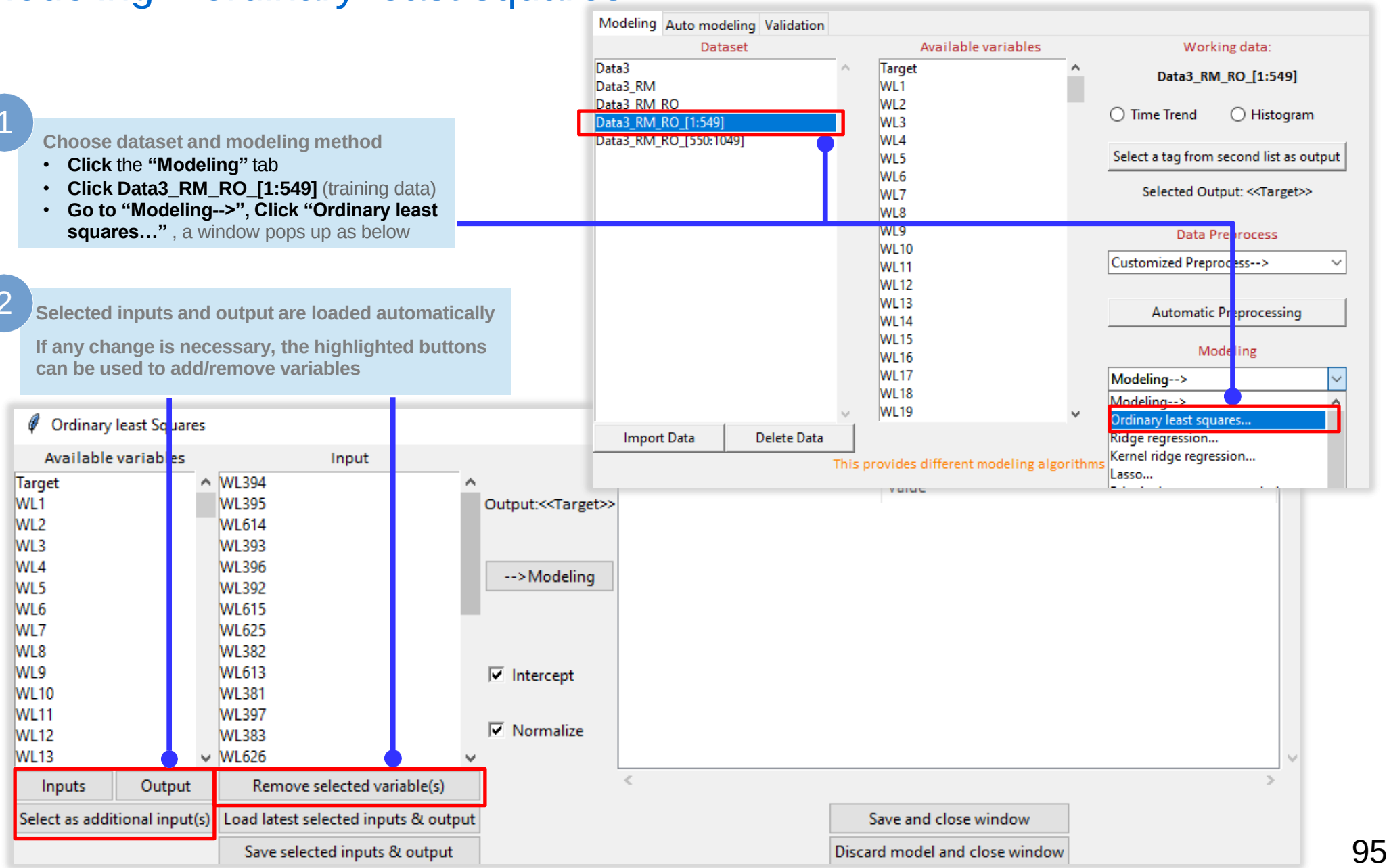
Choose dataset and modeling method

- **Click** the **"Modeling"** tab
- **Click Data3_RM_RO_[1:549]** (training data)
- **Go to "Modeling-->", Click "Ordinary least squares…"** , a window pops up as below

2

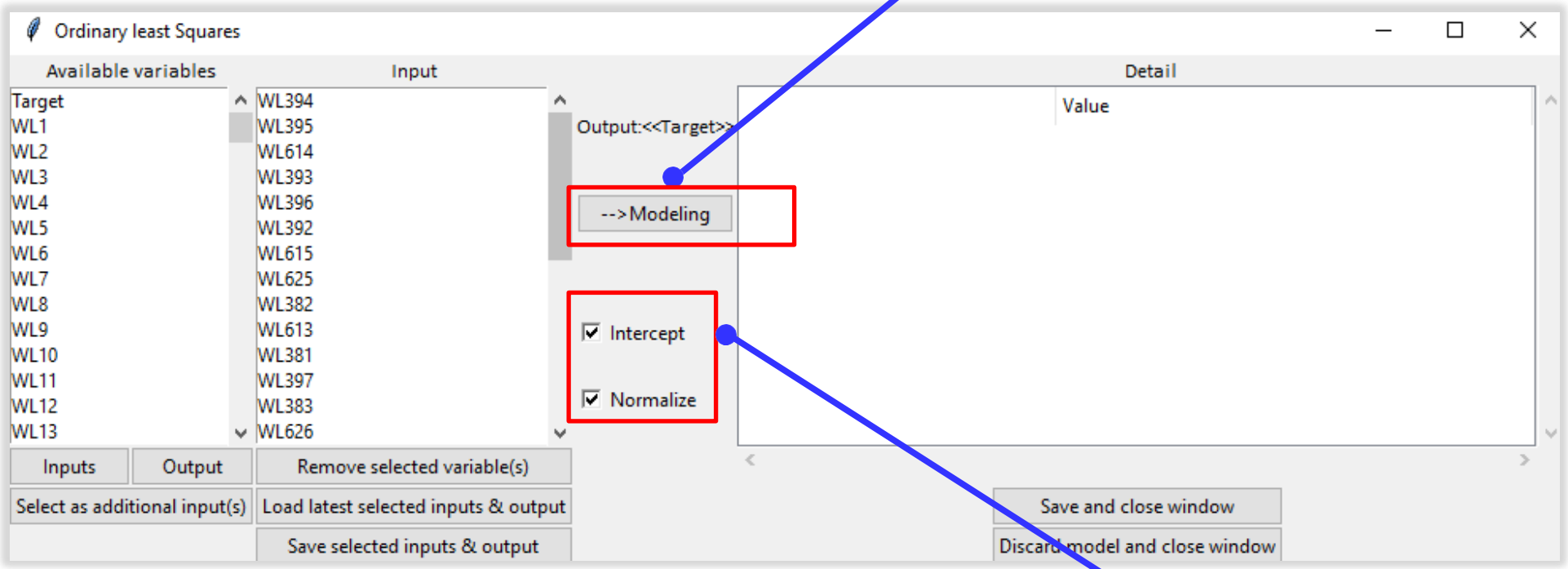
Selected inputs and output are loaded automatically

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



Soft Sensor Analytics

Modeling – ordinary least squares (select inputs manually)



4

Modeling
Click "-->Modeling" to perform regression

3

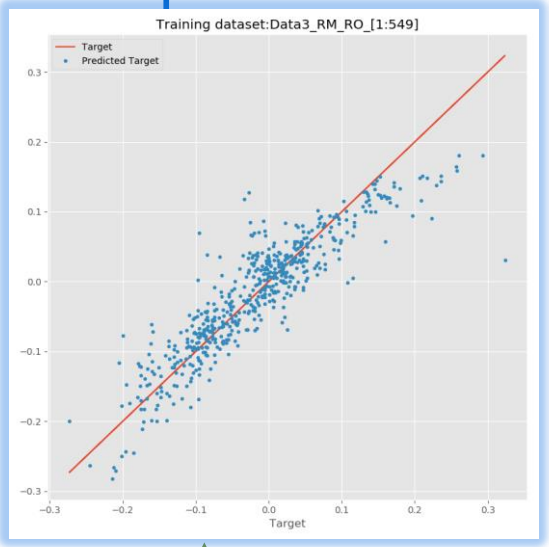
Options
Tick "Intercept" or/and **"Normalize"**.

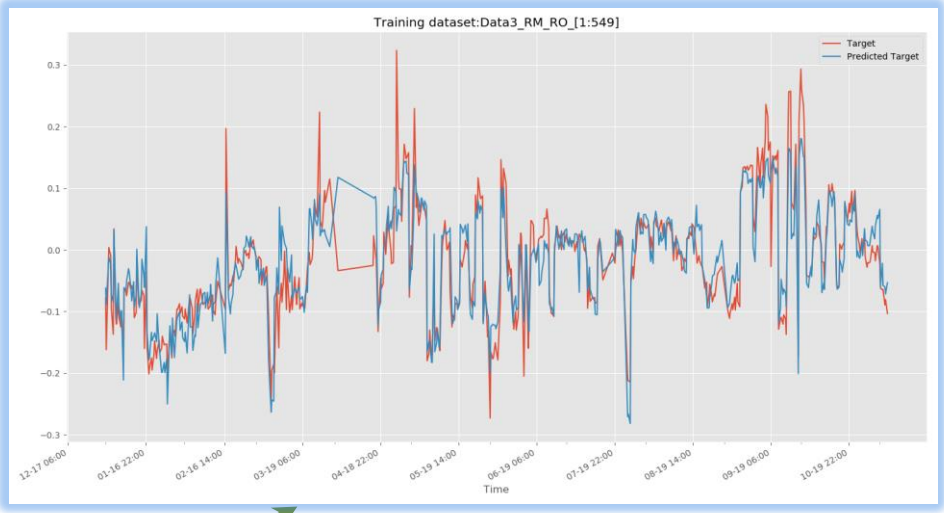
Soft Sensor Analytics

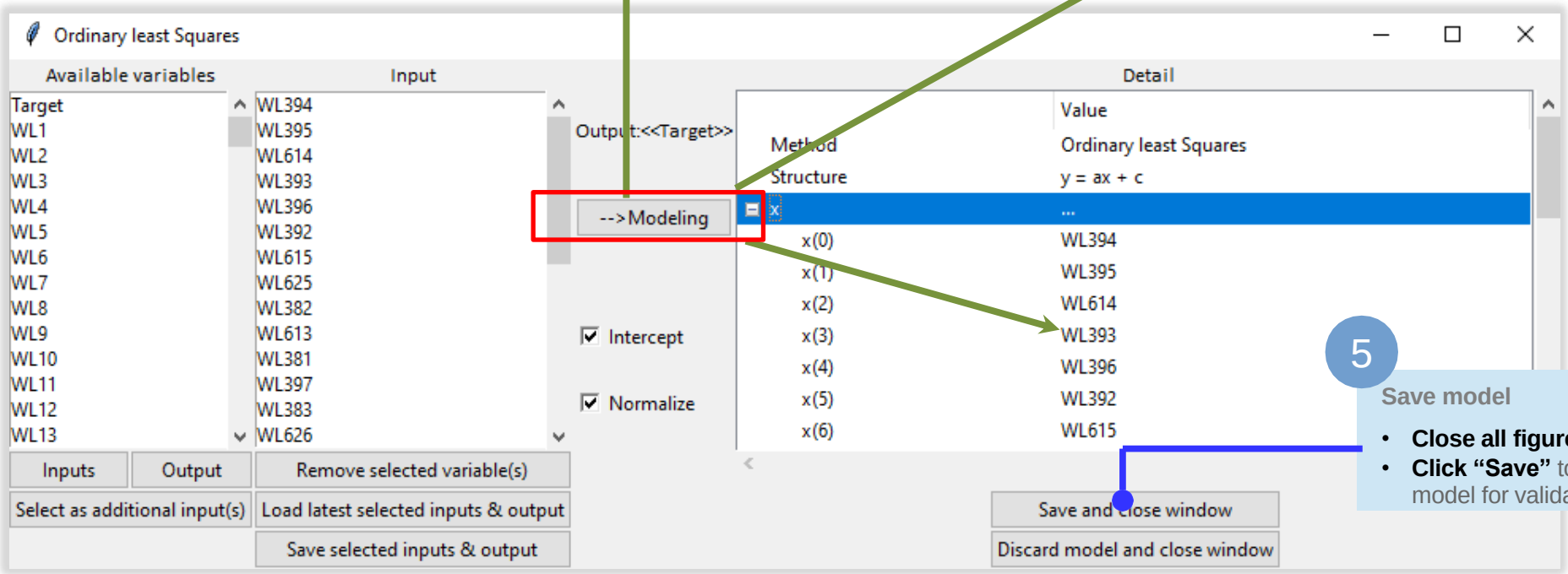
Results – ordinary least squares

Results

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







5

Save model

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

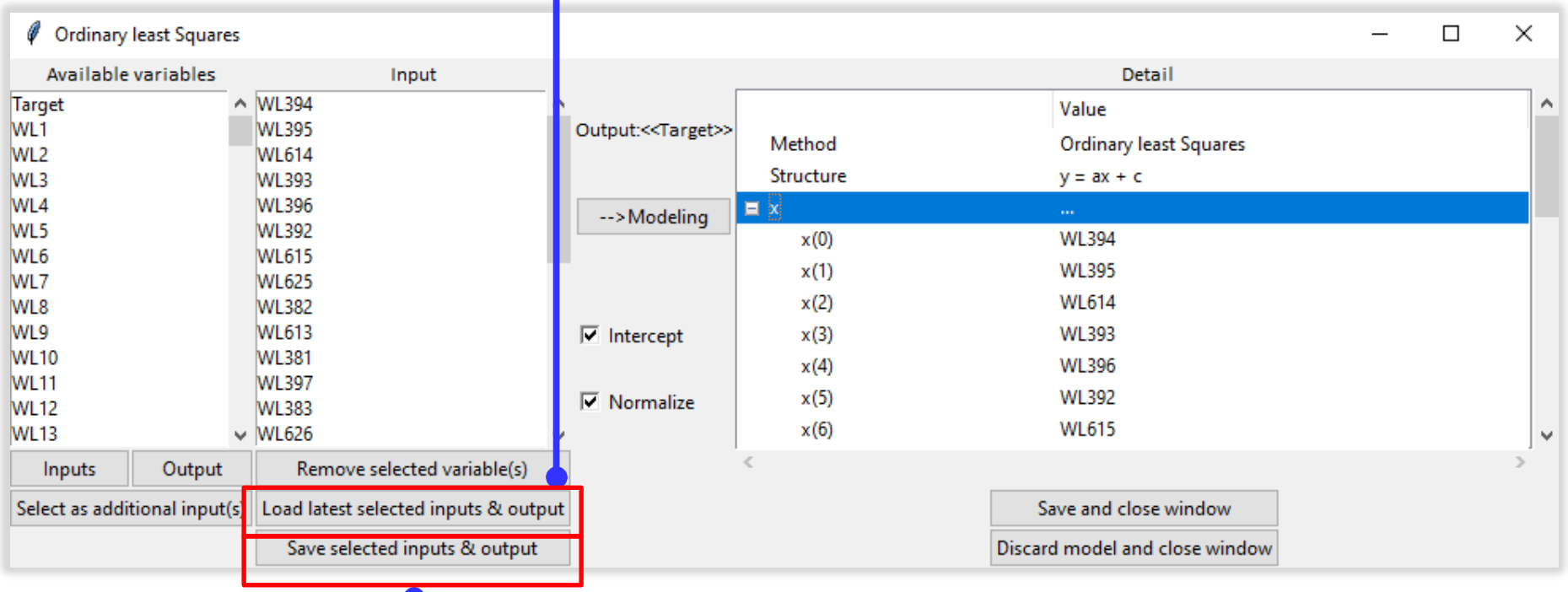
Soft Sensor Analytics

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



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

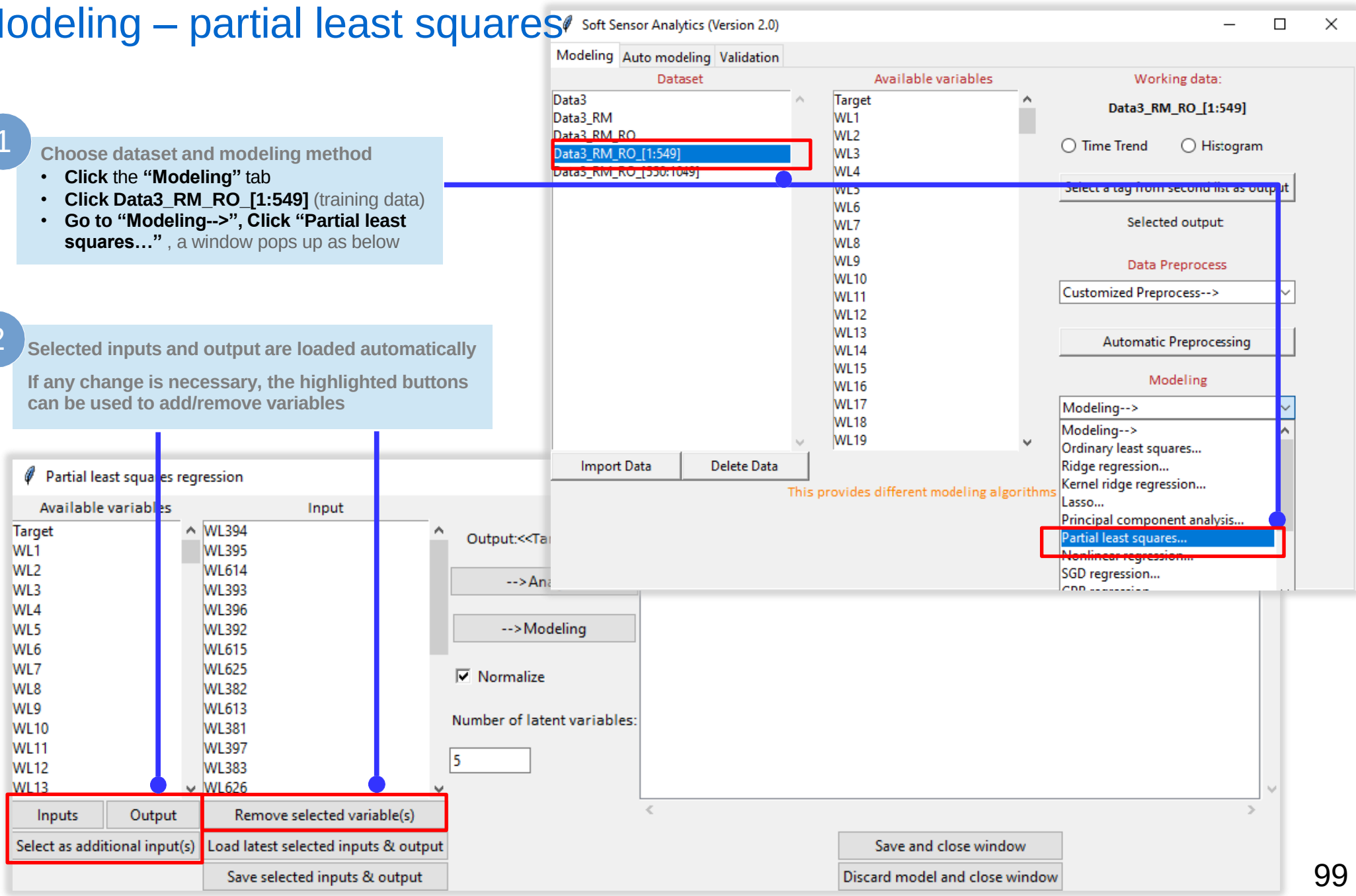
Choose dataset and modeling method

- **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



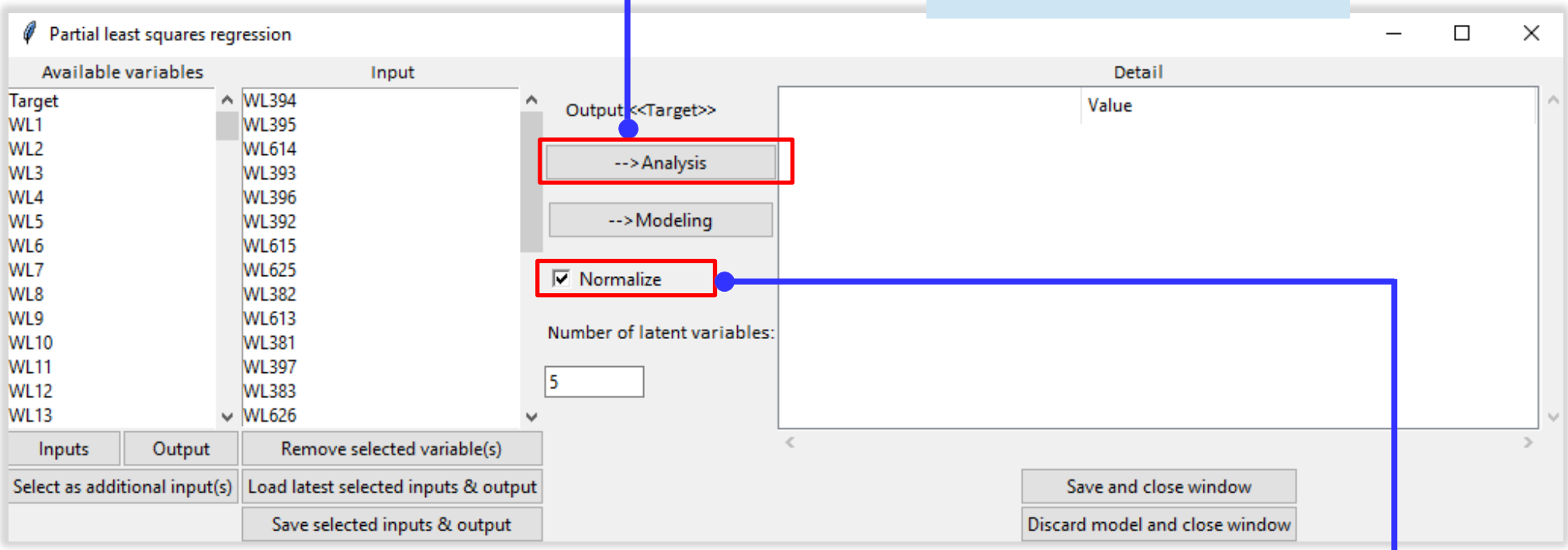
Soft Sensor Analytics

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

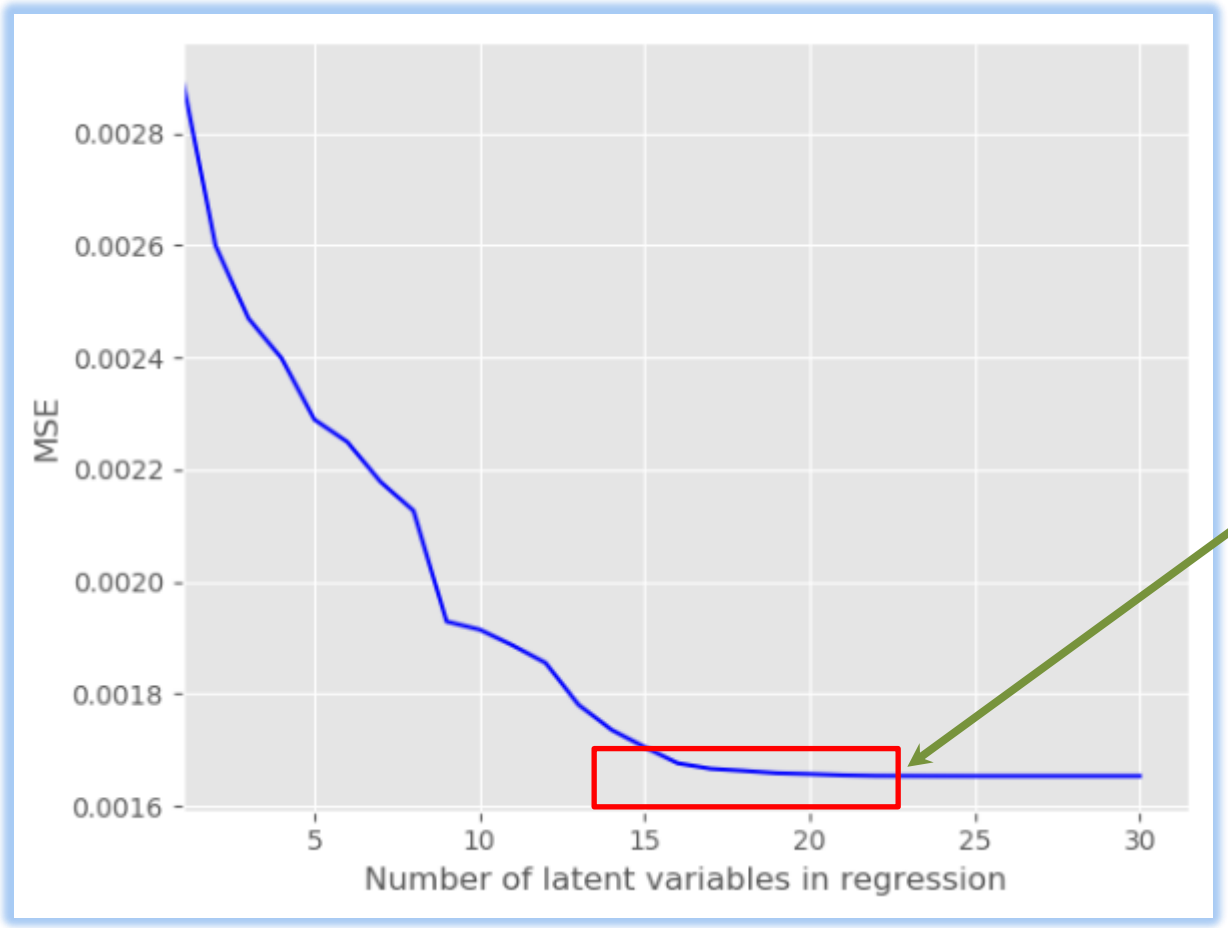


3

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.

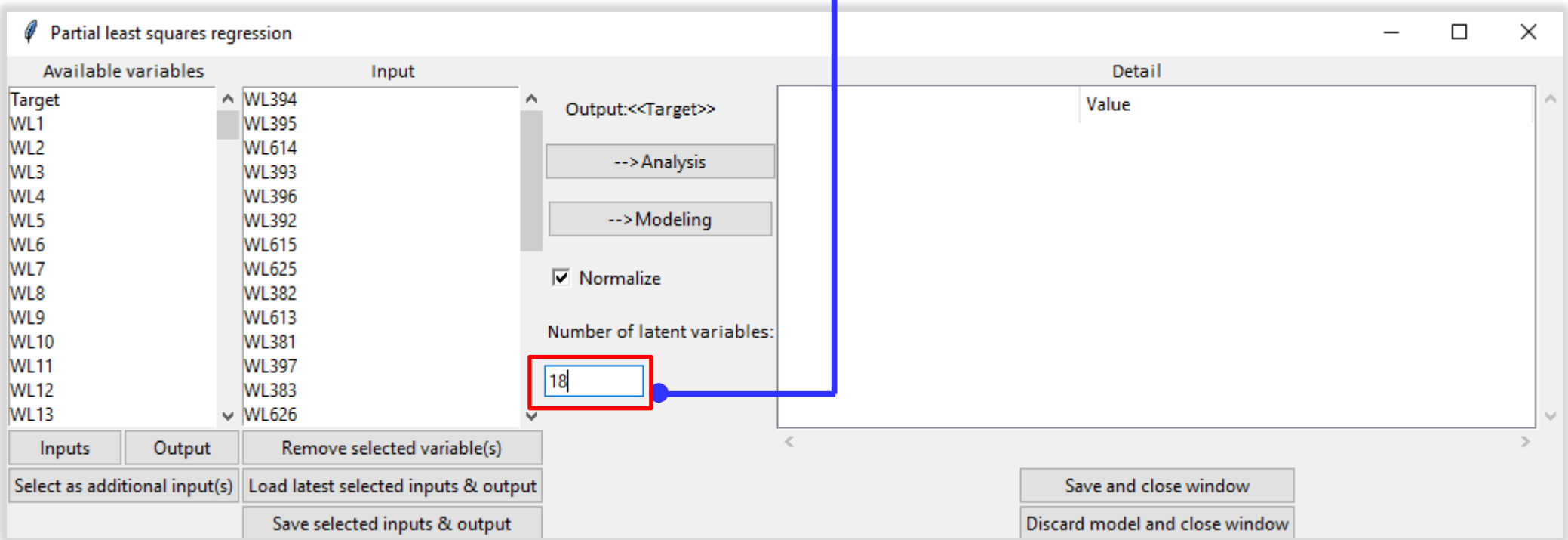
Soft Sensor Analytics

Modeling – PLS (determine # of latent variables)

5

Determine # of latent variables

- **Close all figures**
- **Enter "18"** in the **"number of latent variables"** box

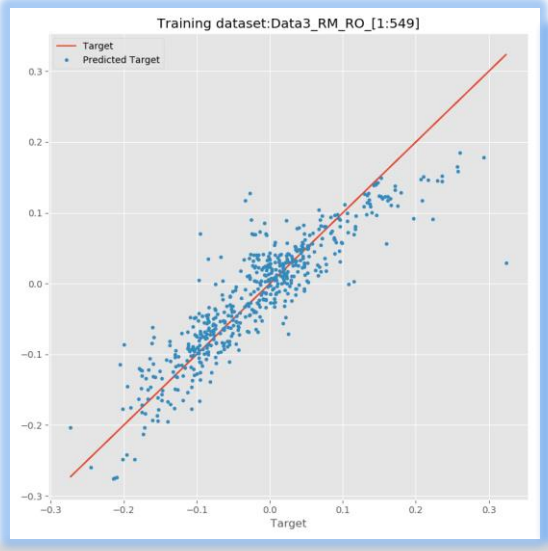


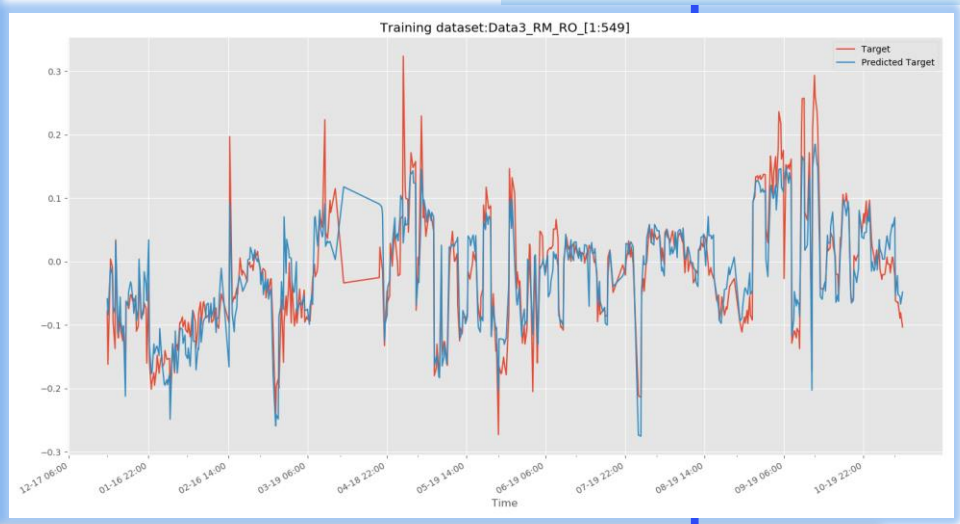
Soft Sensor Analytics

Results – PLS

6

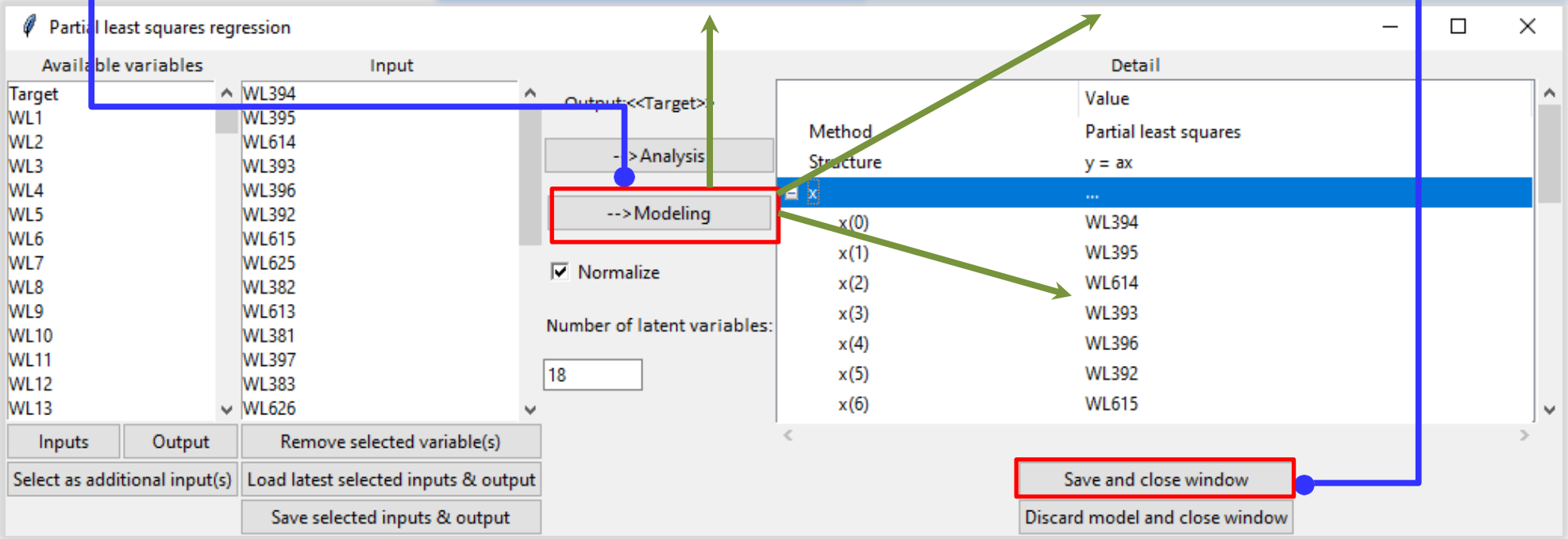
Regression
Click "-->Modeling"
- scatter plot and time trend are generated
- model information is shown in **"Detail"** column





7

- **Close figures**
- **Click "Save"** to save this model



Soft Sensor Analytics

Automatic modeling using Gaussian Process Regression

1

Automatic modeling

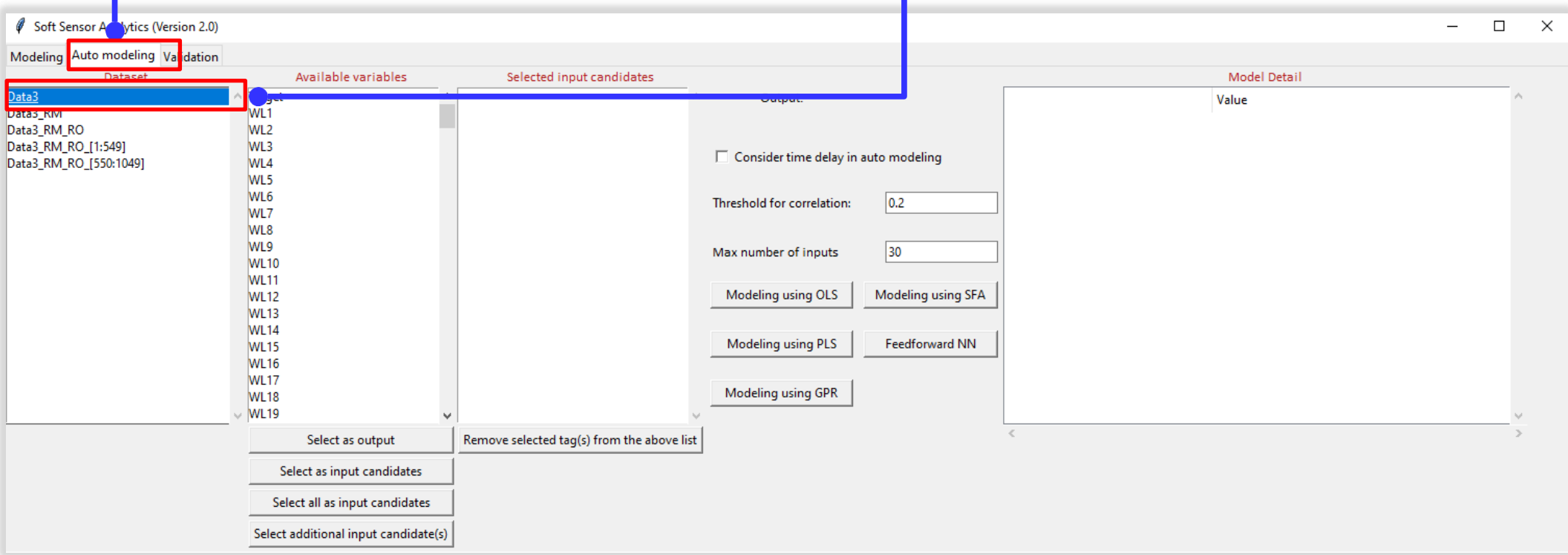
- **Click** the **"Auto modeling"** tab

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

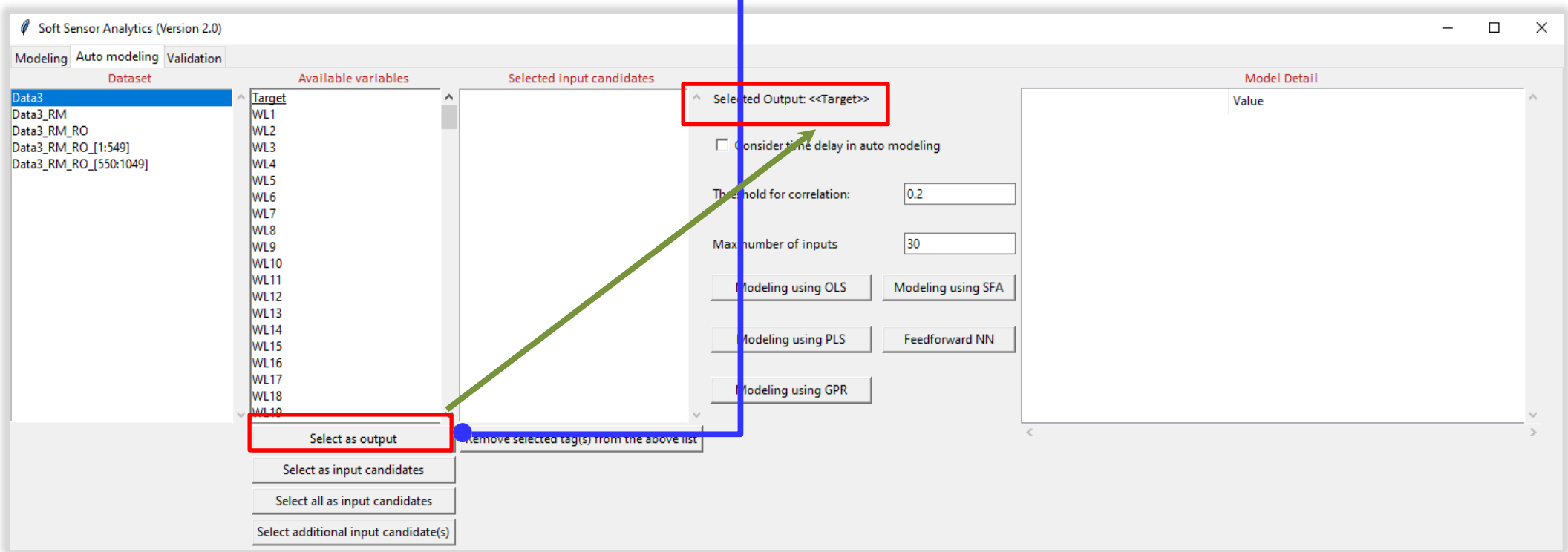


Soft Sensor Analytics

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

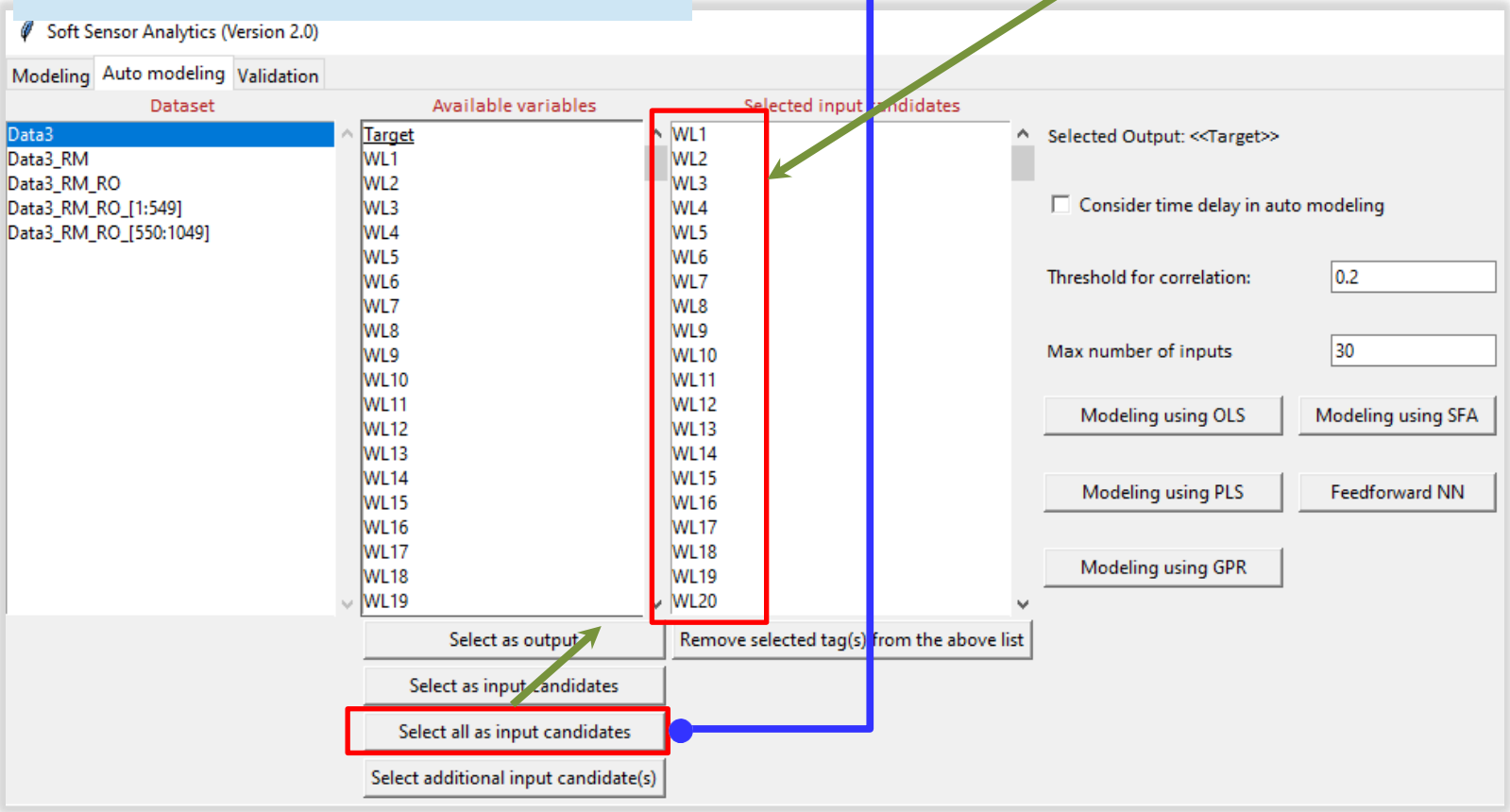
Automatic modeling – inputs/output selection

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



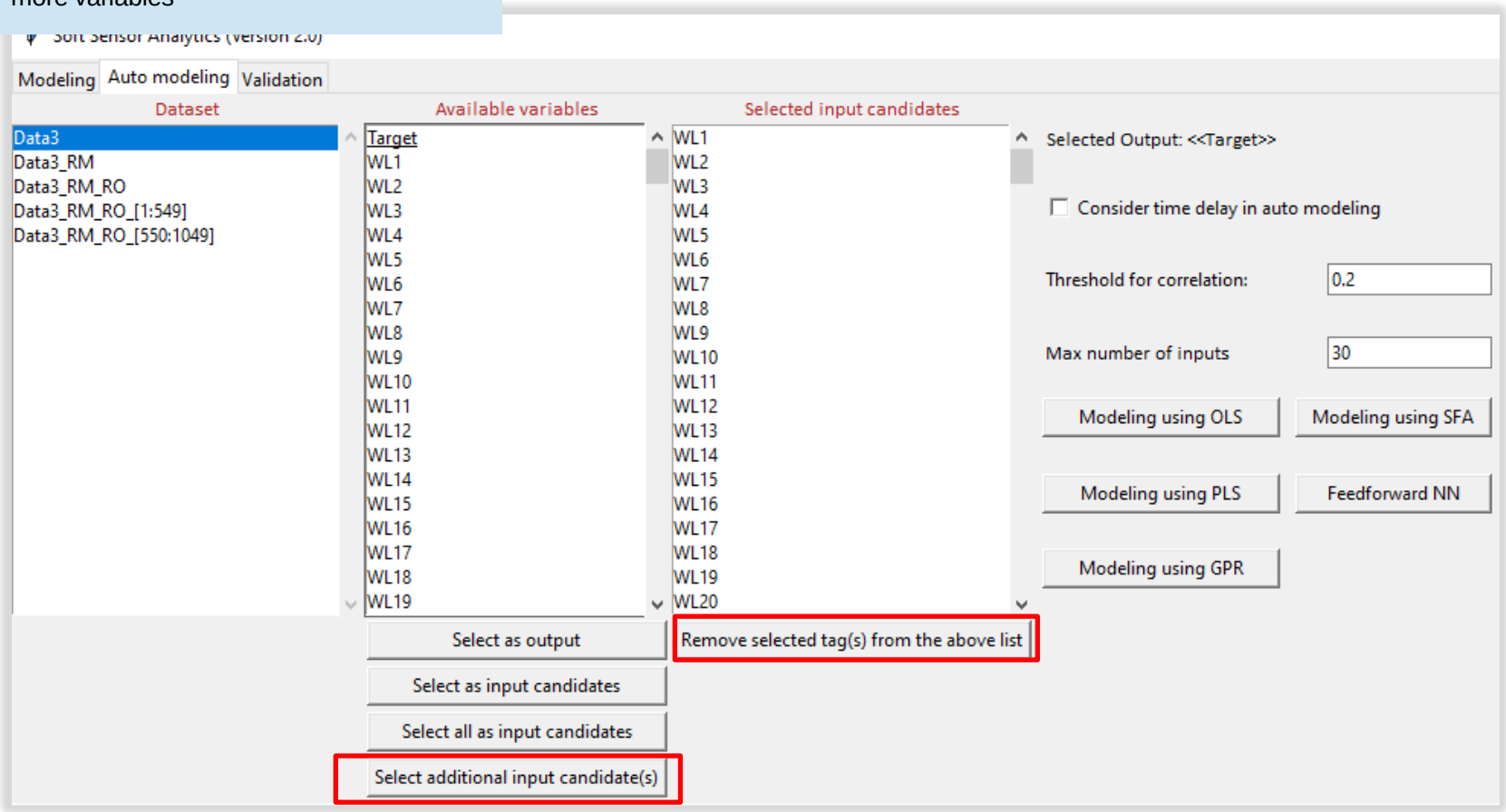
Note: We can also use "Select as input candidates" button to select potential inputs from all variables based on available information/knowledge

Soft Sensor Analytics

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



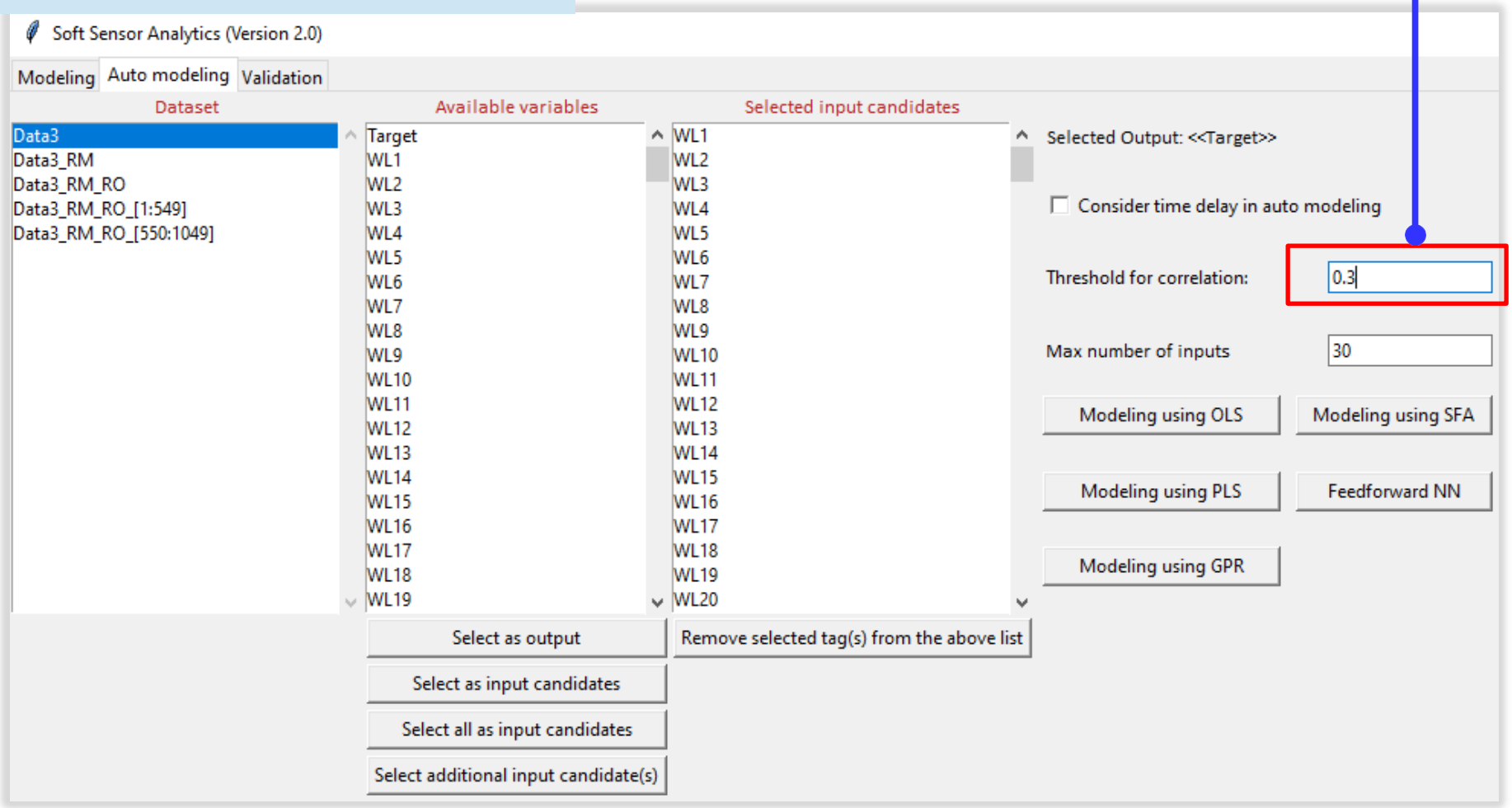
Soft Sensor Analytics

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

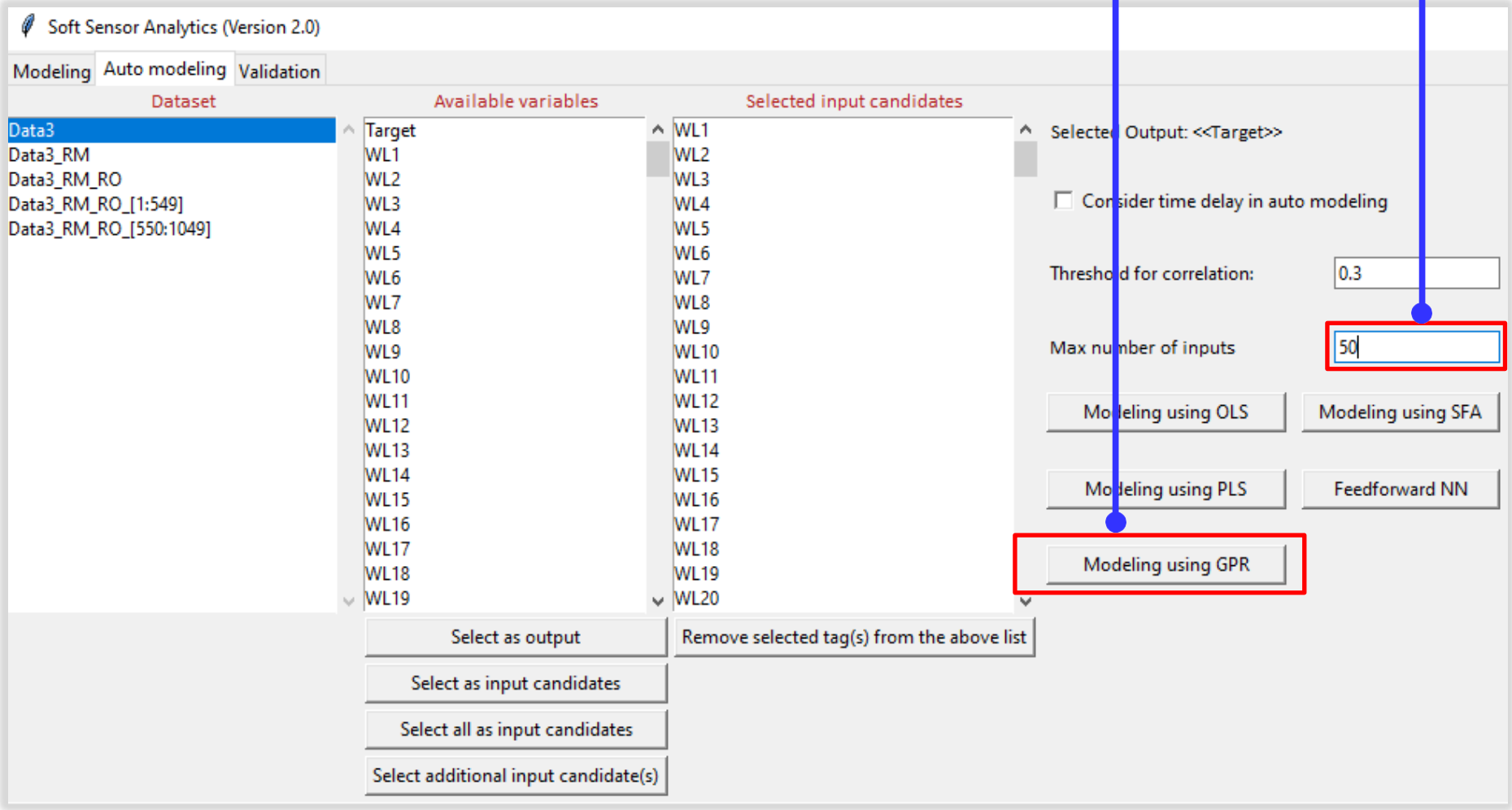


Soft Sensor Analytics

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

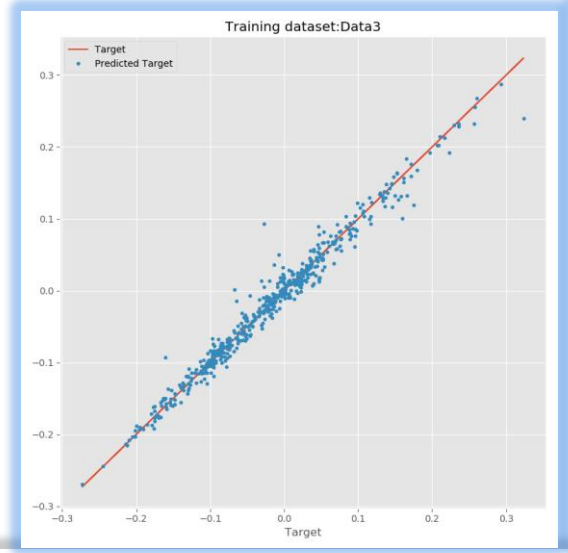


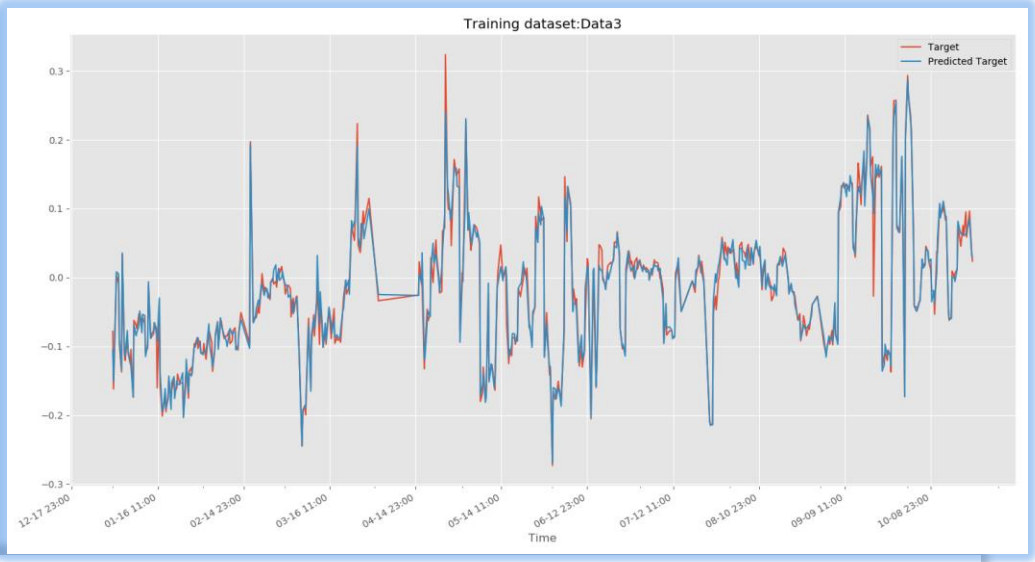
Soft Sensor Analytics

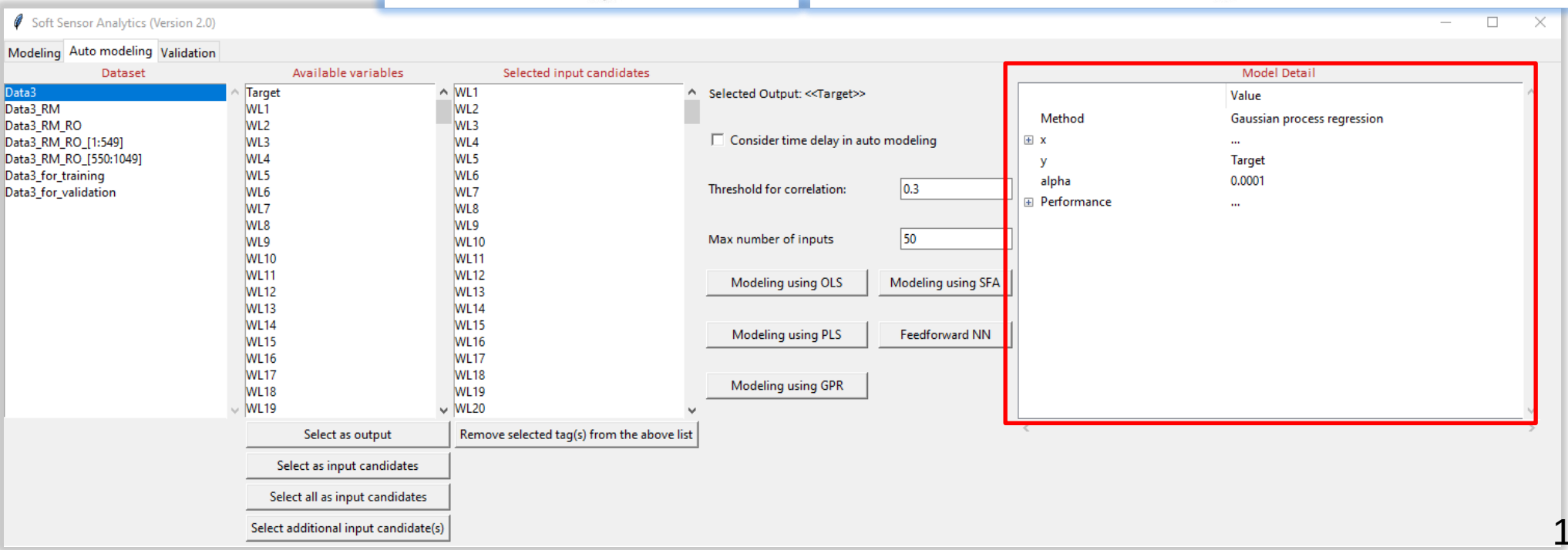
Automatic modeling using GPR - Results

Results

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

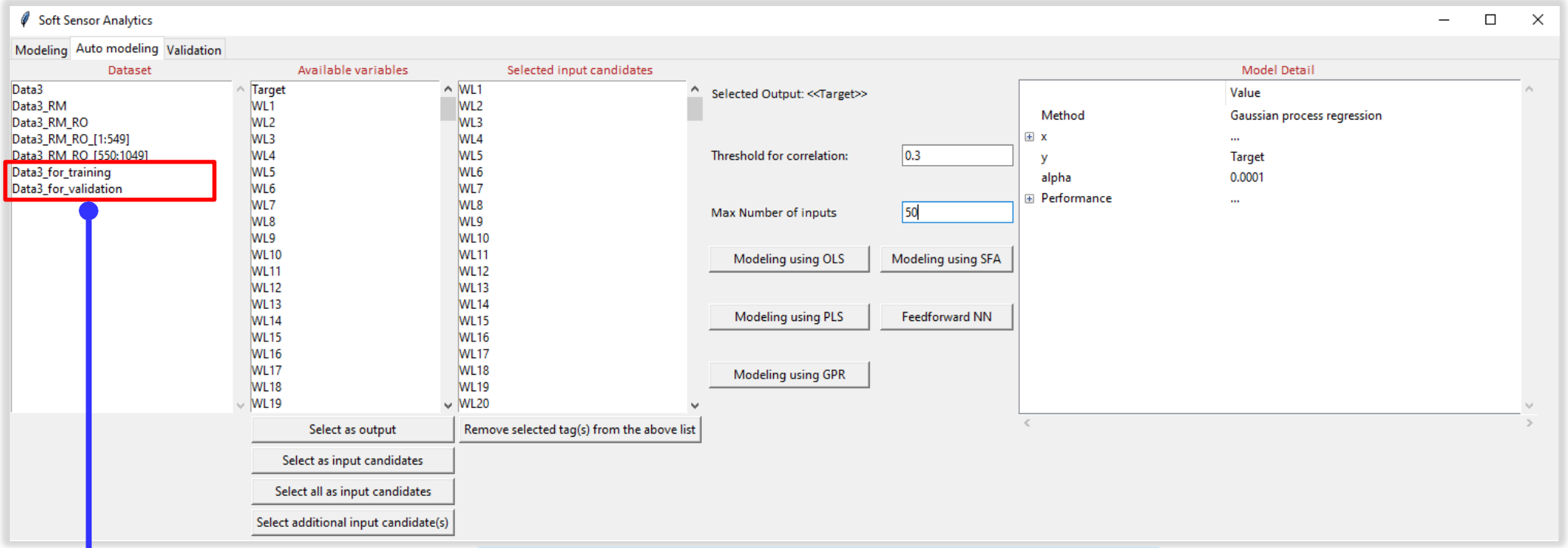






Soft Sensor Analytics

Automatic modeling - Results



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

Soft Sensor Analytics

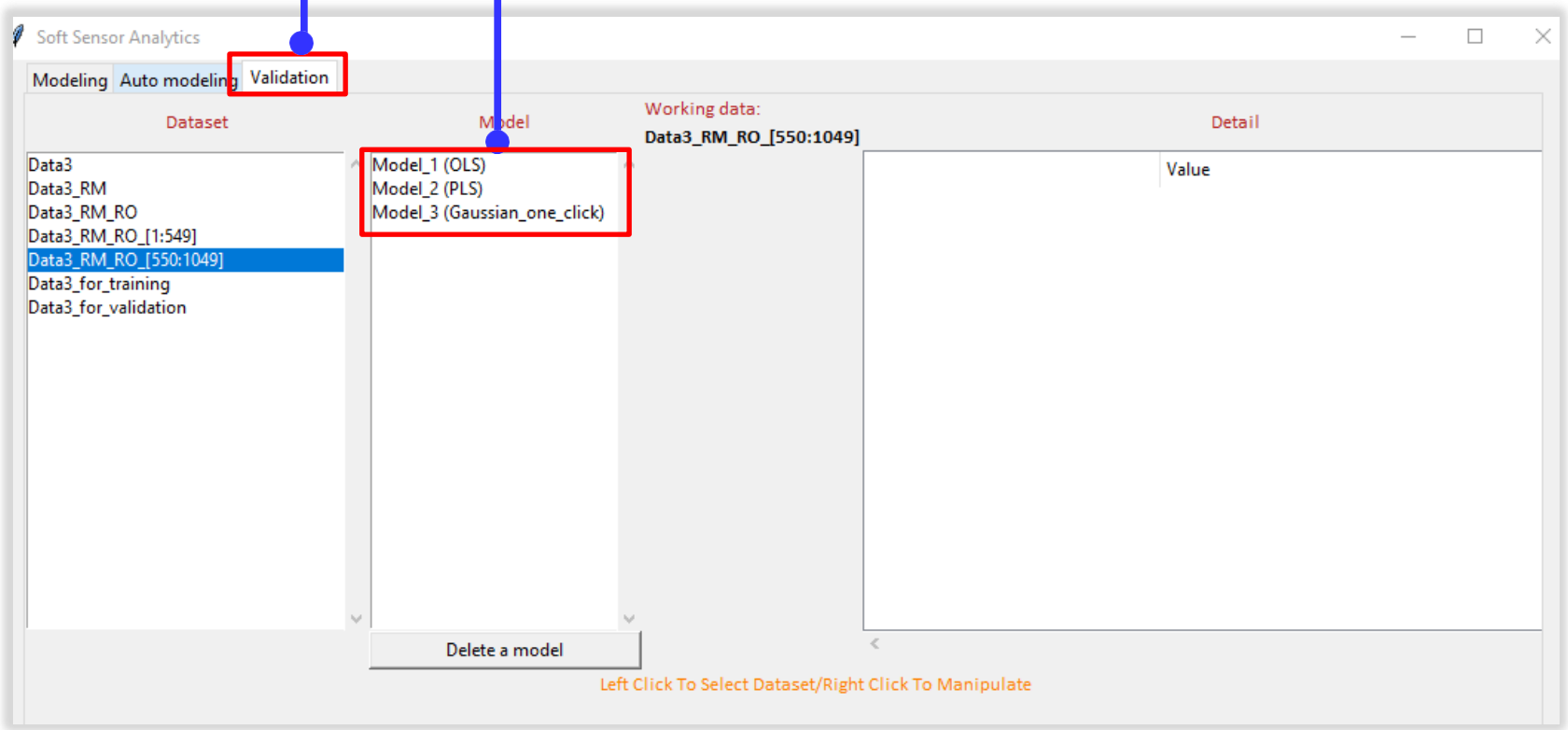
Validation of the models

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

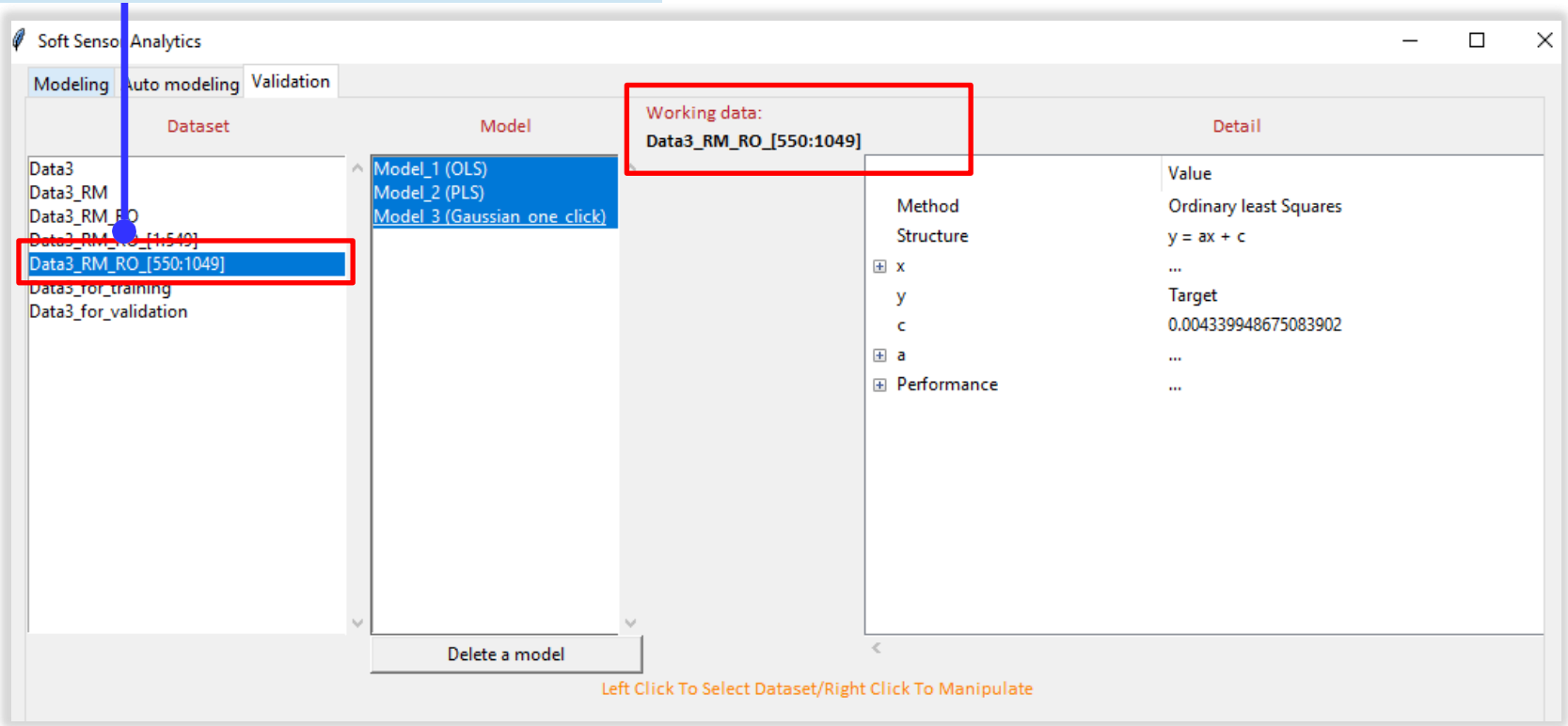
Validation of the models

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



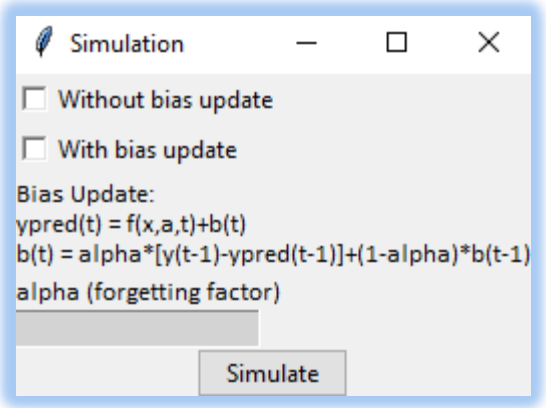
Soft Sensor Analytics

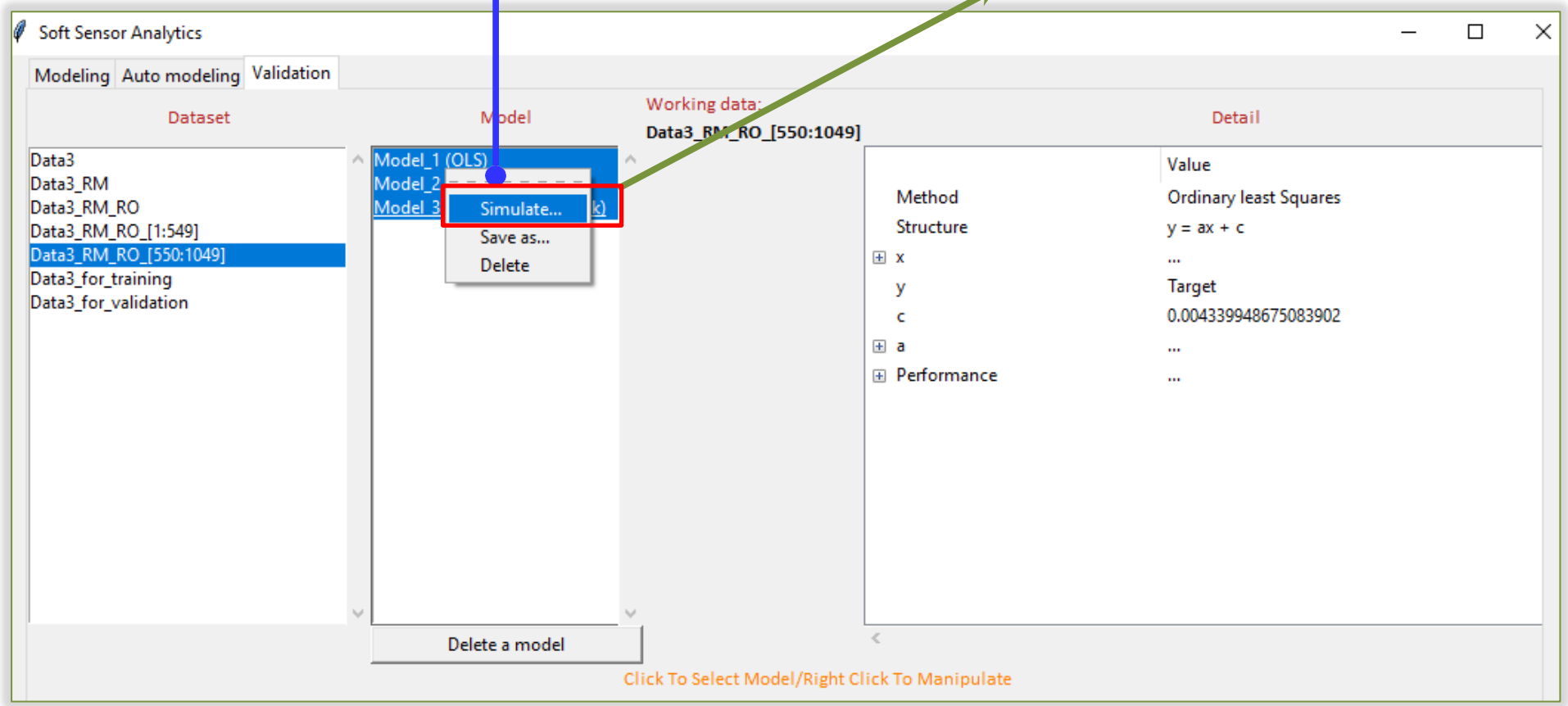
Validation of the models

3

Choose "Simulate" to continue

- **Right-Click** on the two models
- **Click "Simulate…"** button, a window will pop up





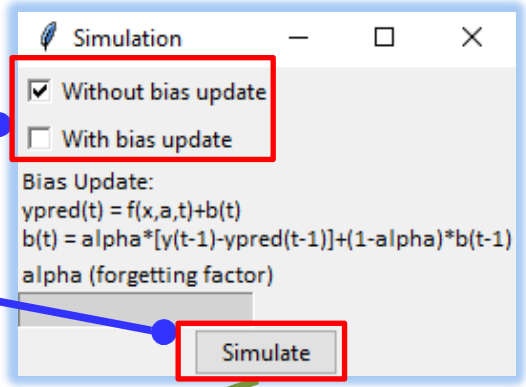
Soft Sensor Analytics

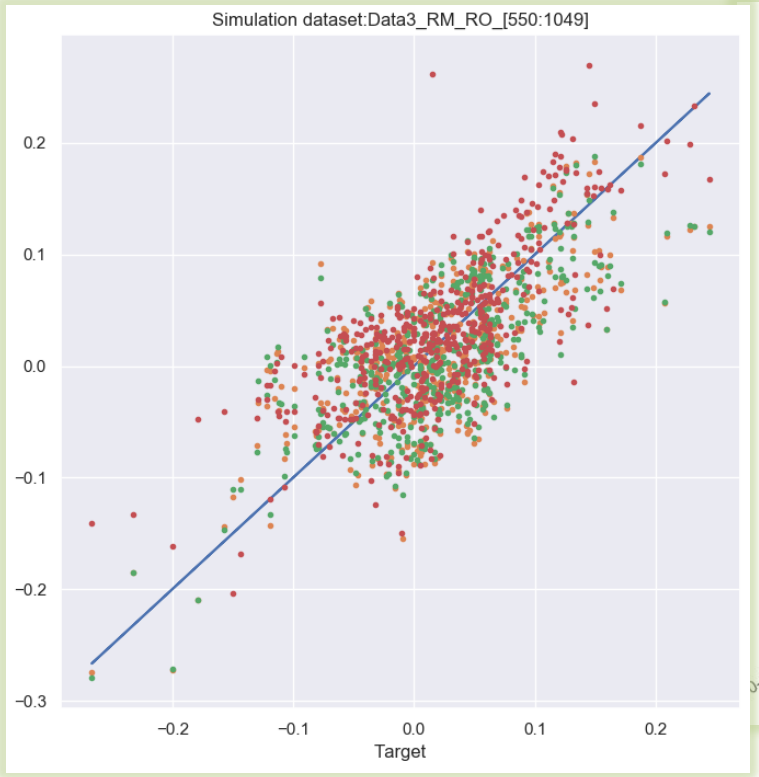
Validation of the models

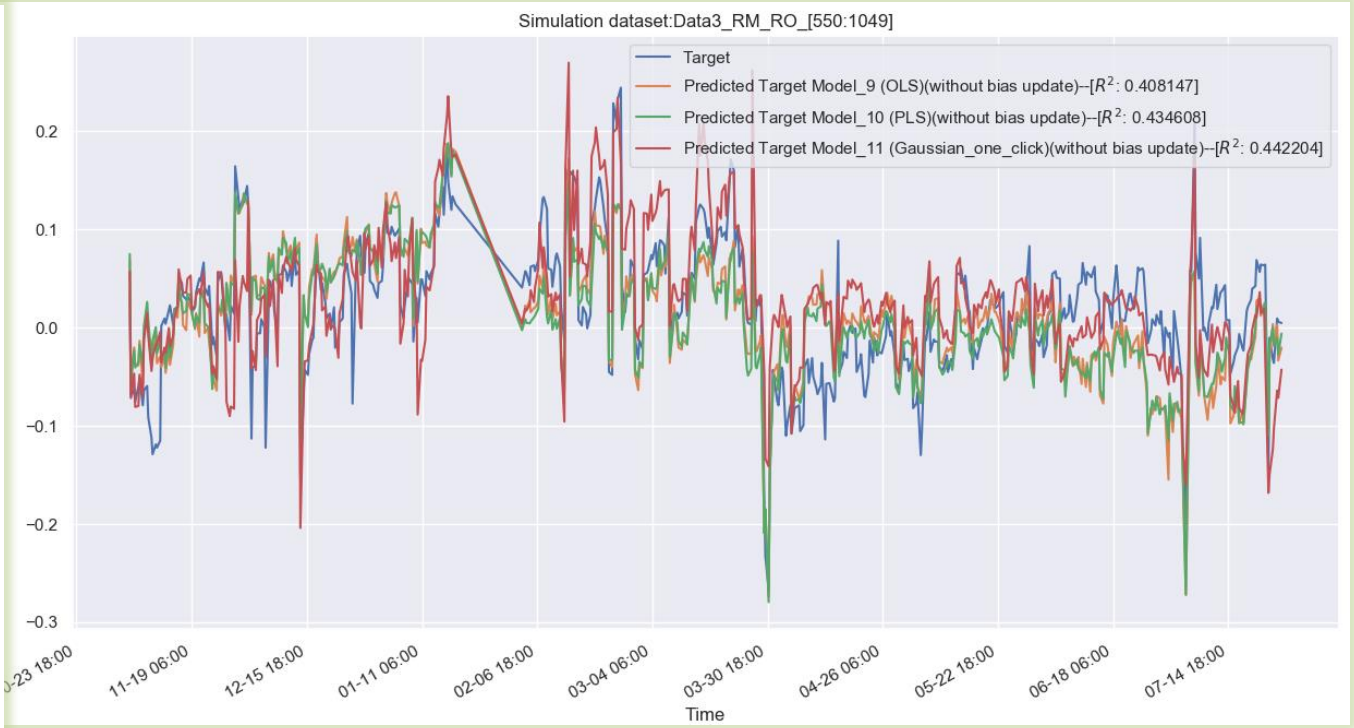
4

Model validation

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







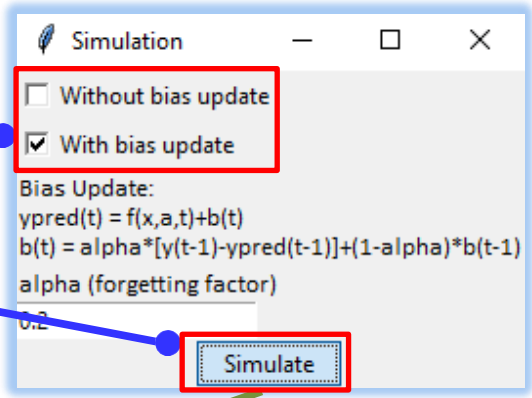
Soft Sensor Analytics

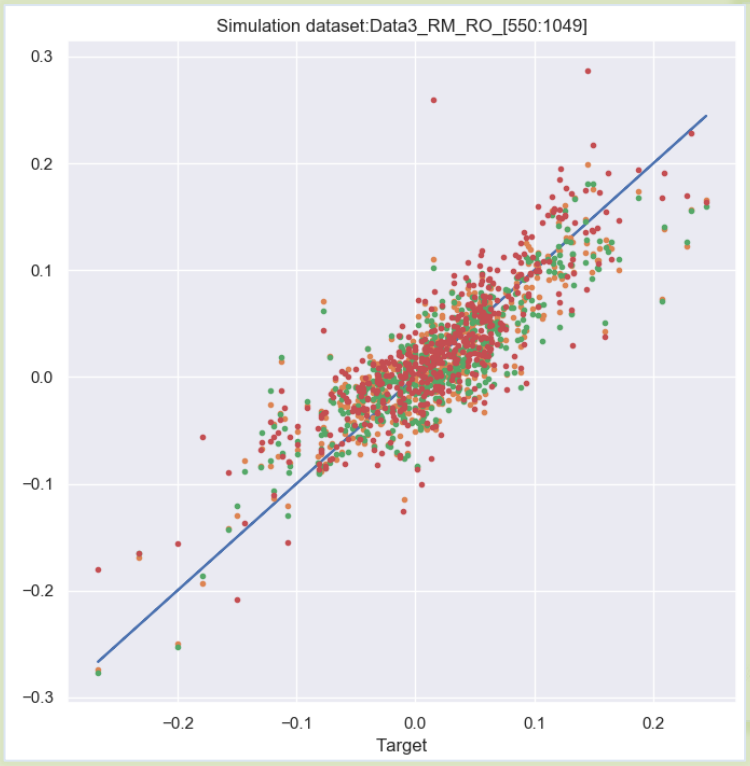
Validation of the models

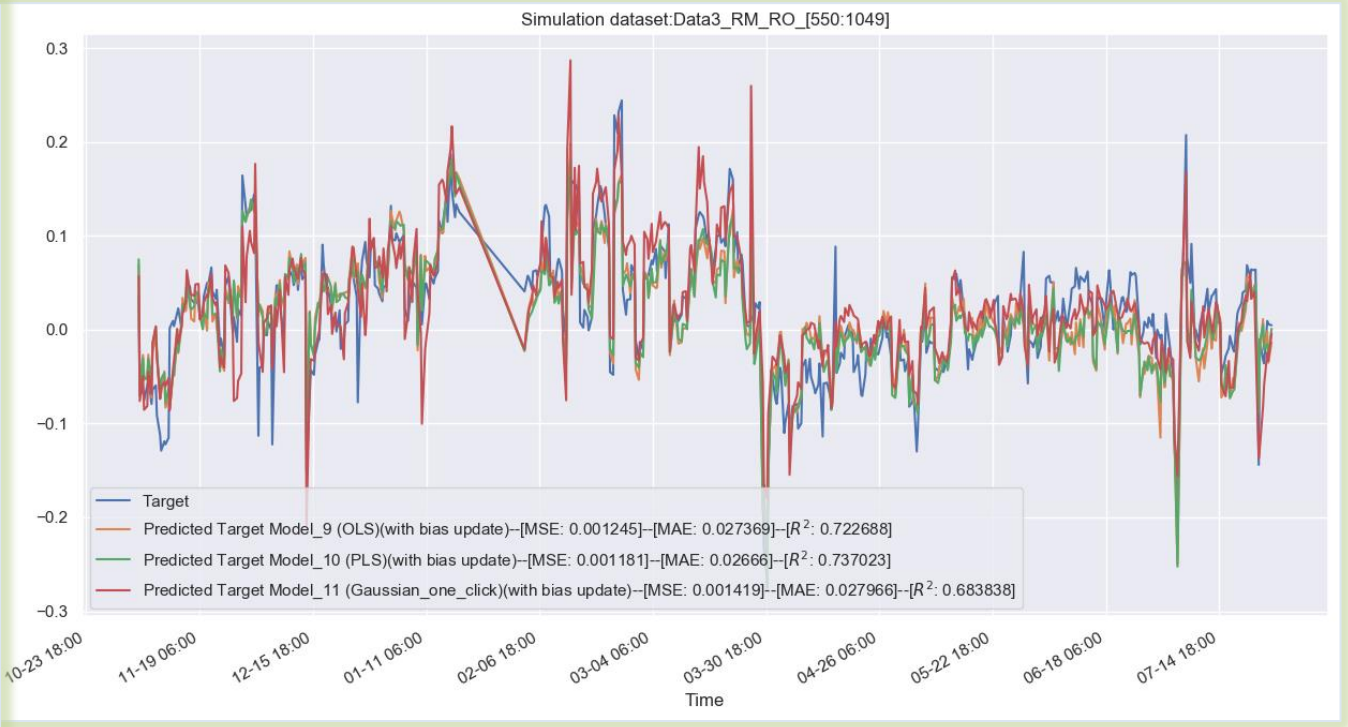
5

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 (SiO2) 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

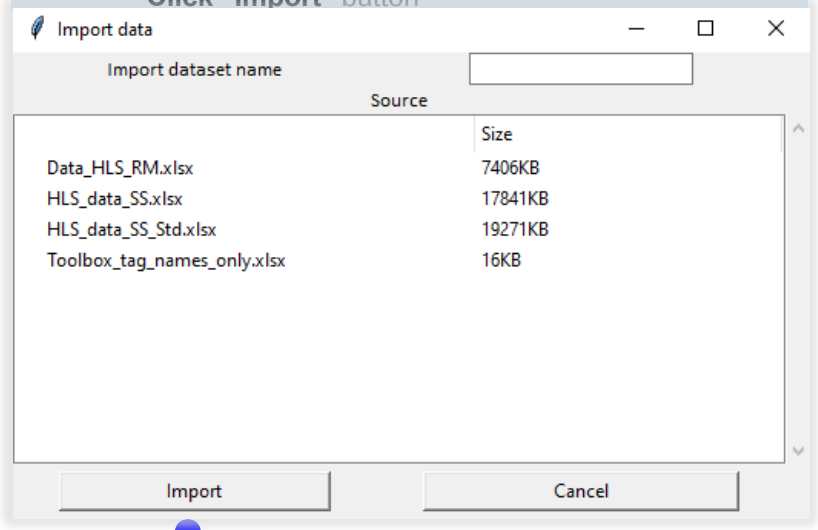
Soft Sensor Analytics

Import data sets

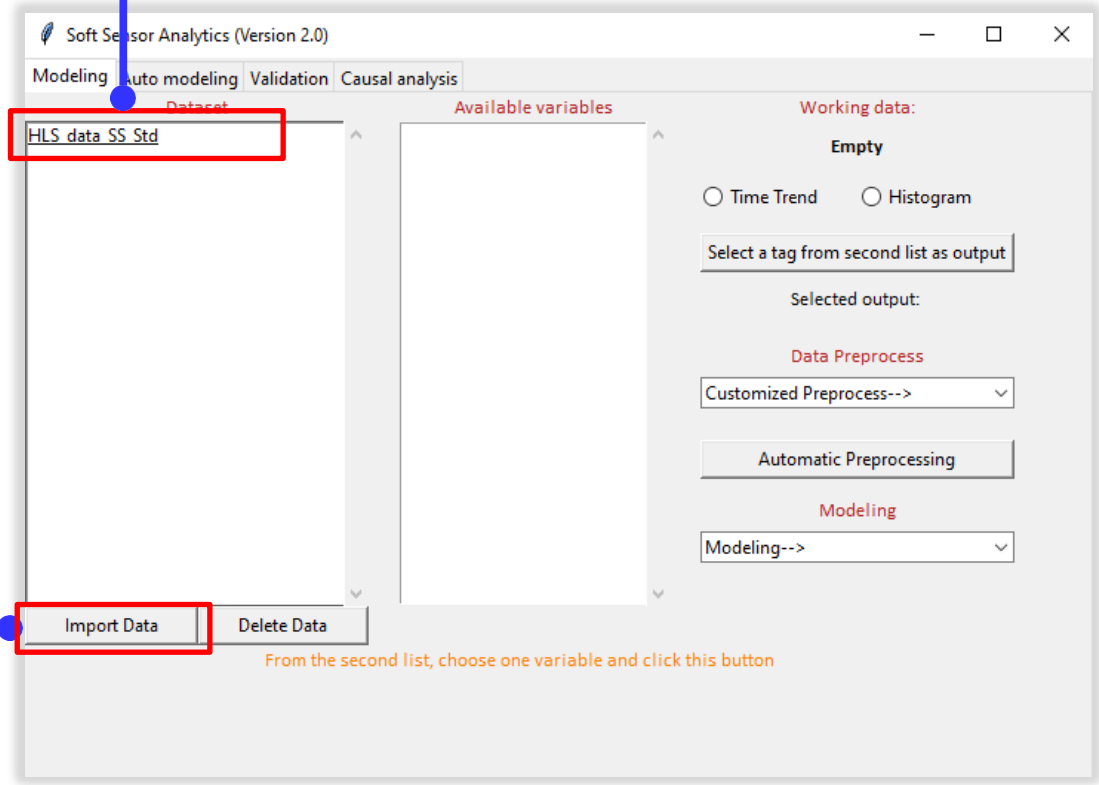
Note: Please save data sets in the ".xlsx" format

1

- **Click "Import Data"** button
- **Select** data file **"HLS_data_SS_Std.xlsx"**
- **Click "Import"** button



"**HLS_data_SS_Std**" has been imported

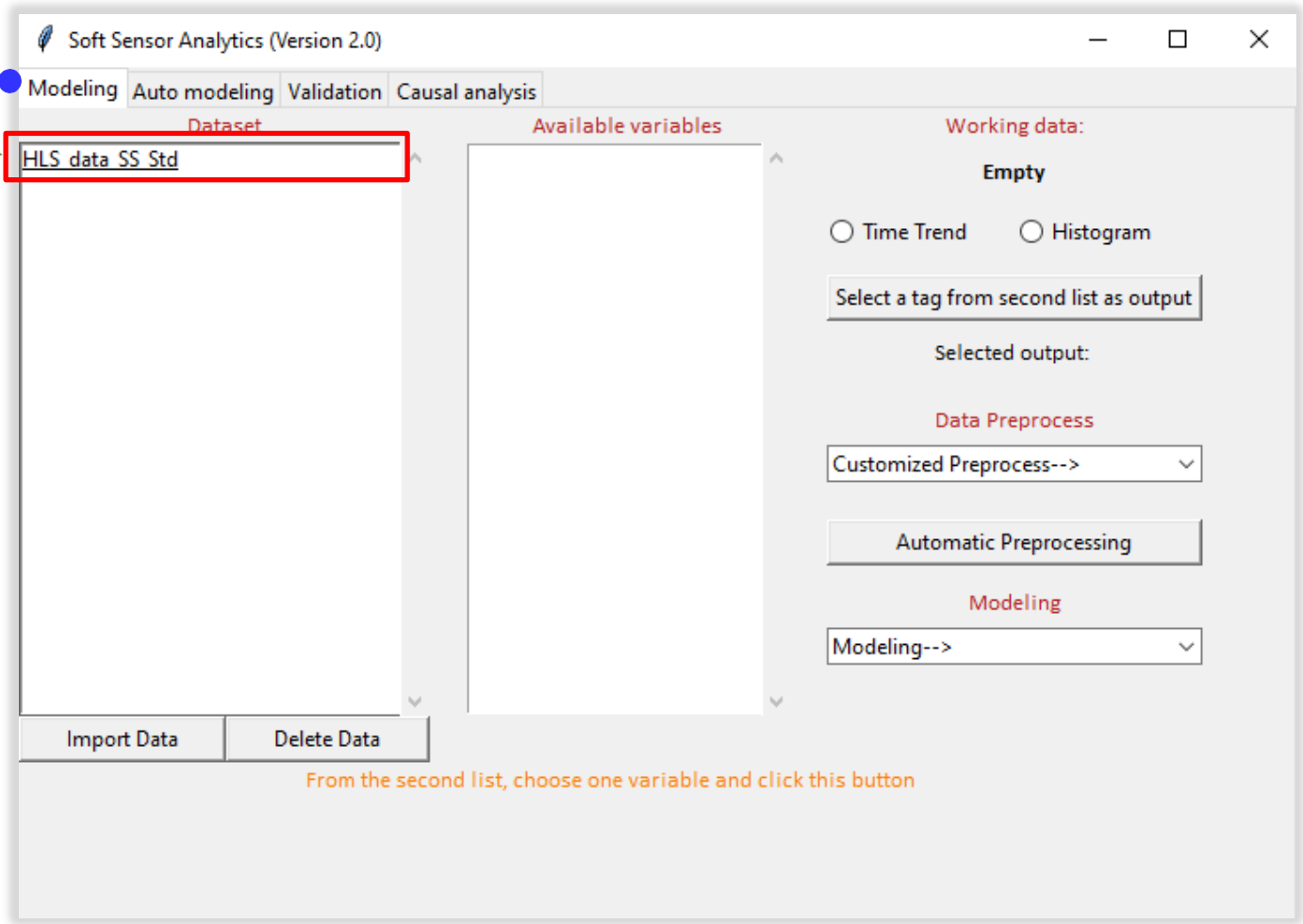


Soft Sensor Analytics

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"



Soft Sensor Analytics

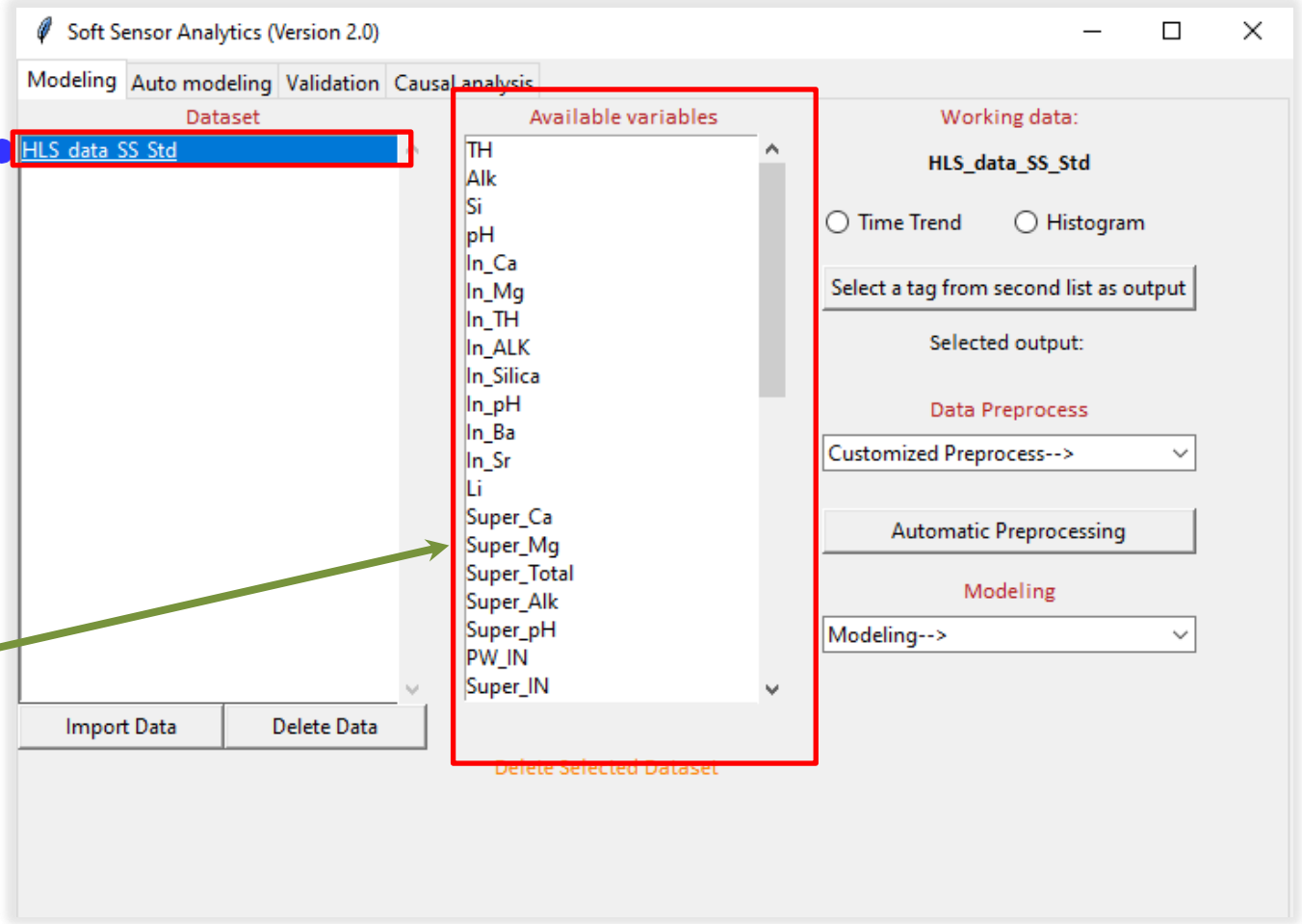
Variables in the dataset

2

Choose Data3 to work on it

- **Click** on **“HLS_data_SS_Std”** from left workspace (i.e., column **“Dataset”**)

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



Soft Sensor Analytics

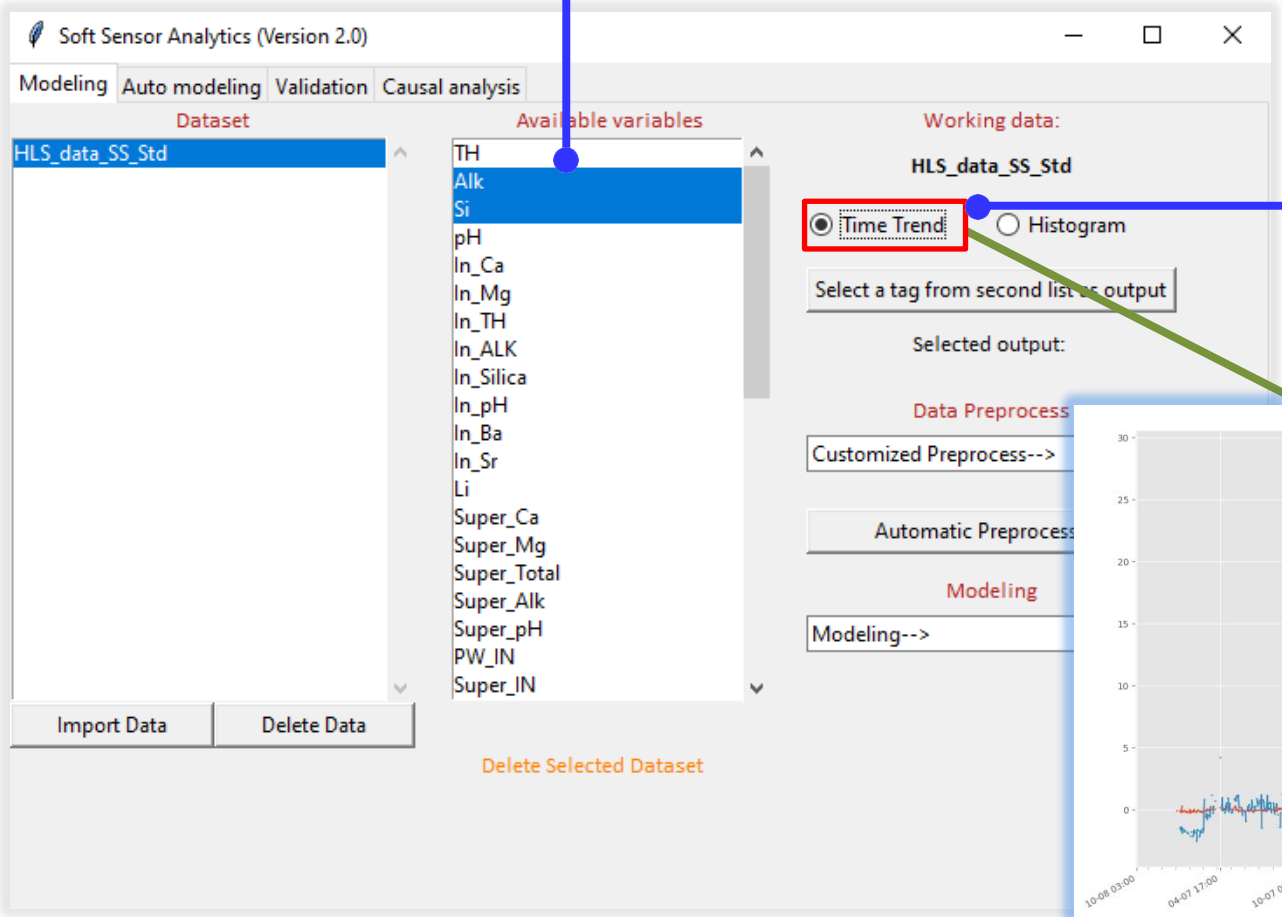
Visualize some of the variables

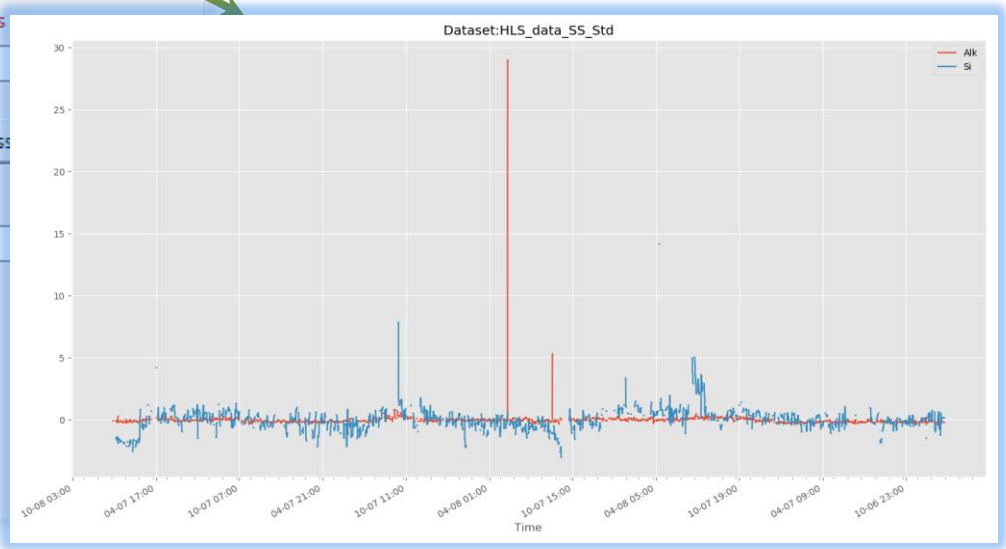
3

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

4

Click "Time Trend" to visualize the selected variables



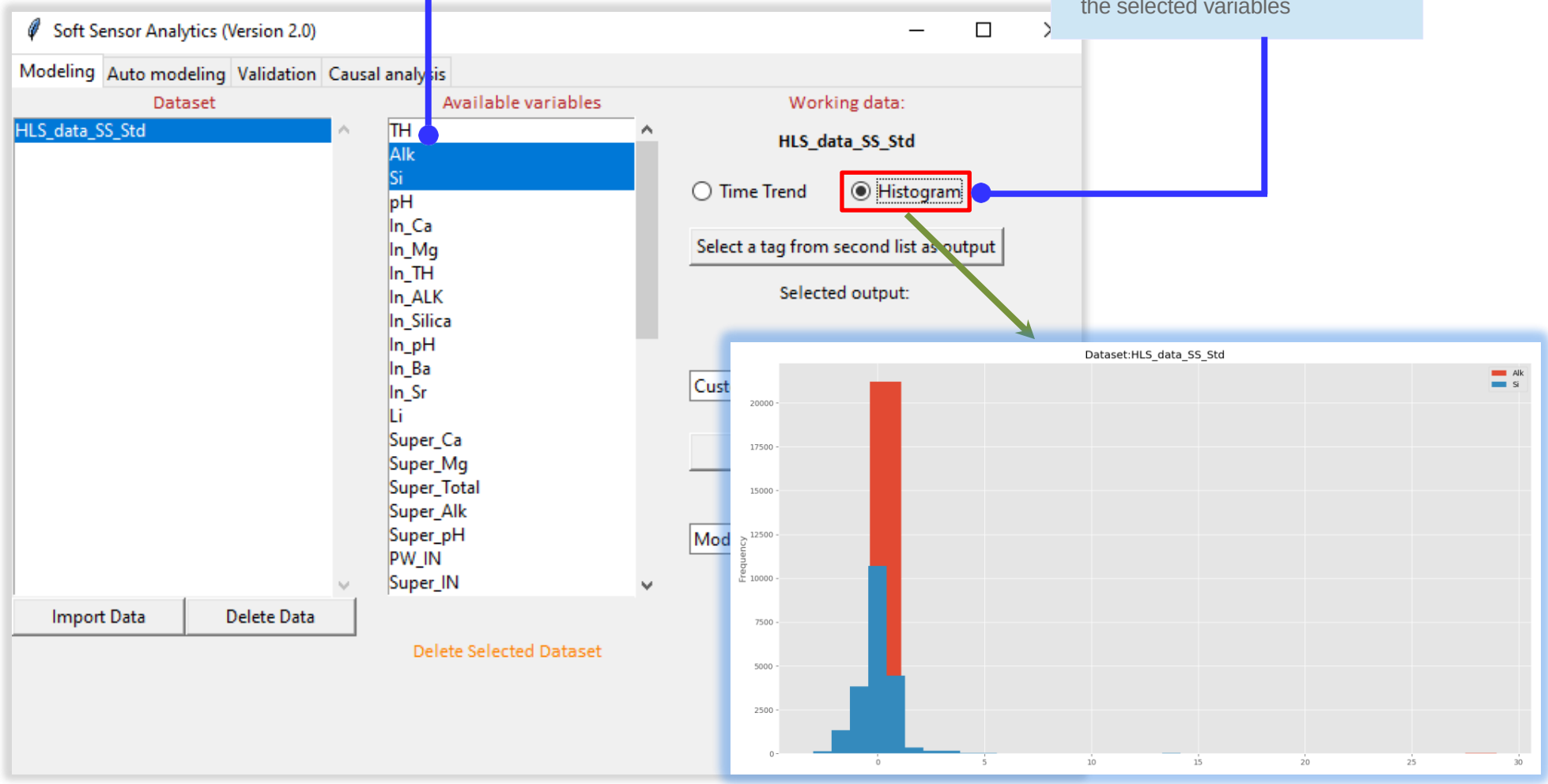


Soft Sensor Analytics

Visualize some of the variables

5 **Click** to select variables for visualization

6 **Click "Time Trend"** to visualize the selected variables



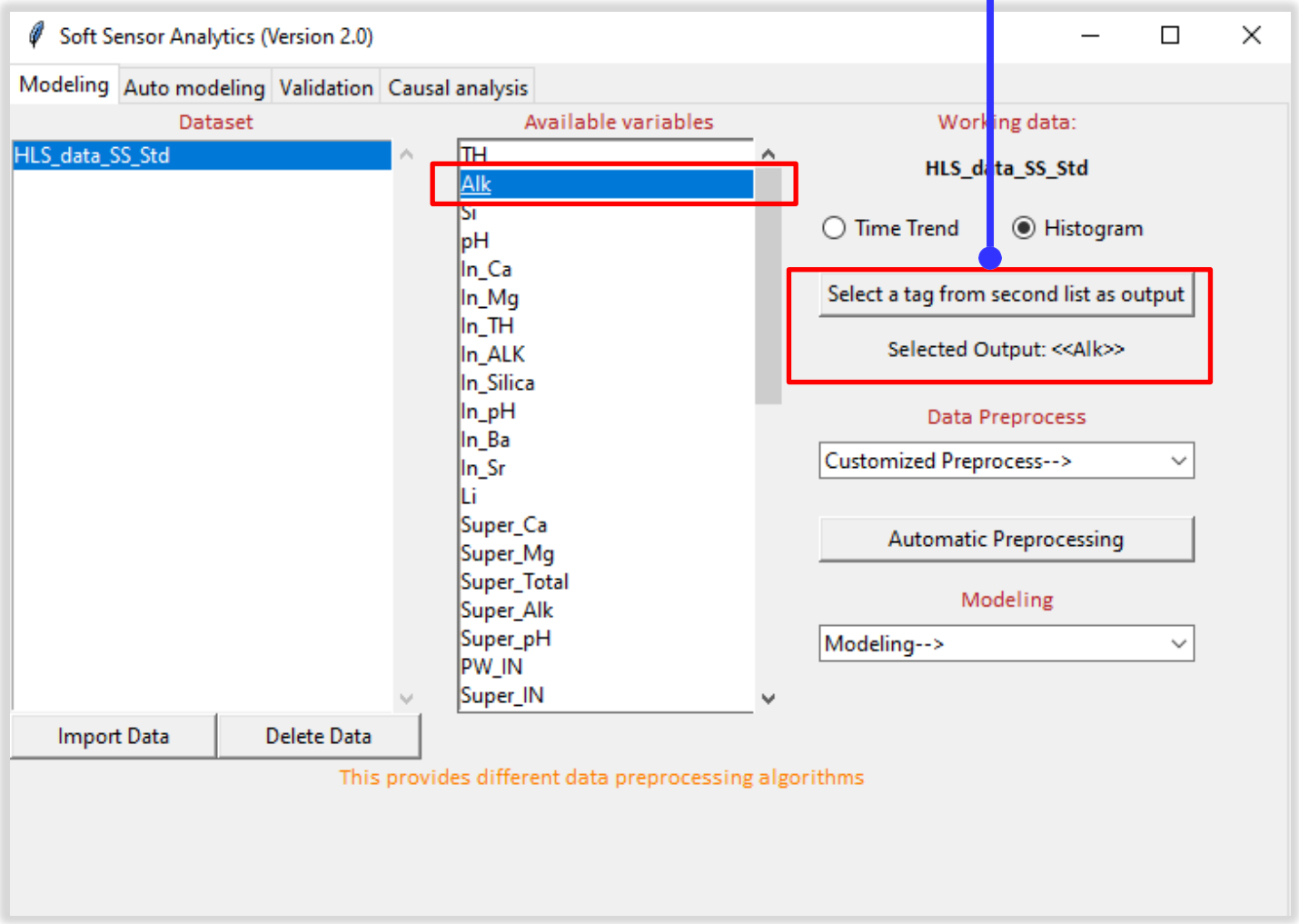
Soft Sensor Analytics

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



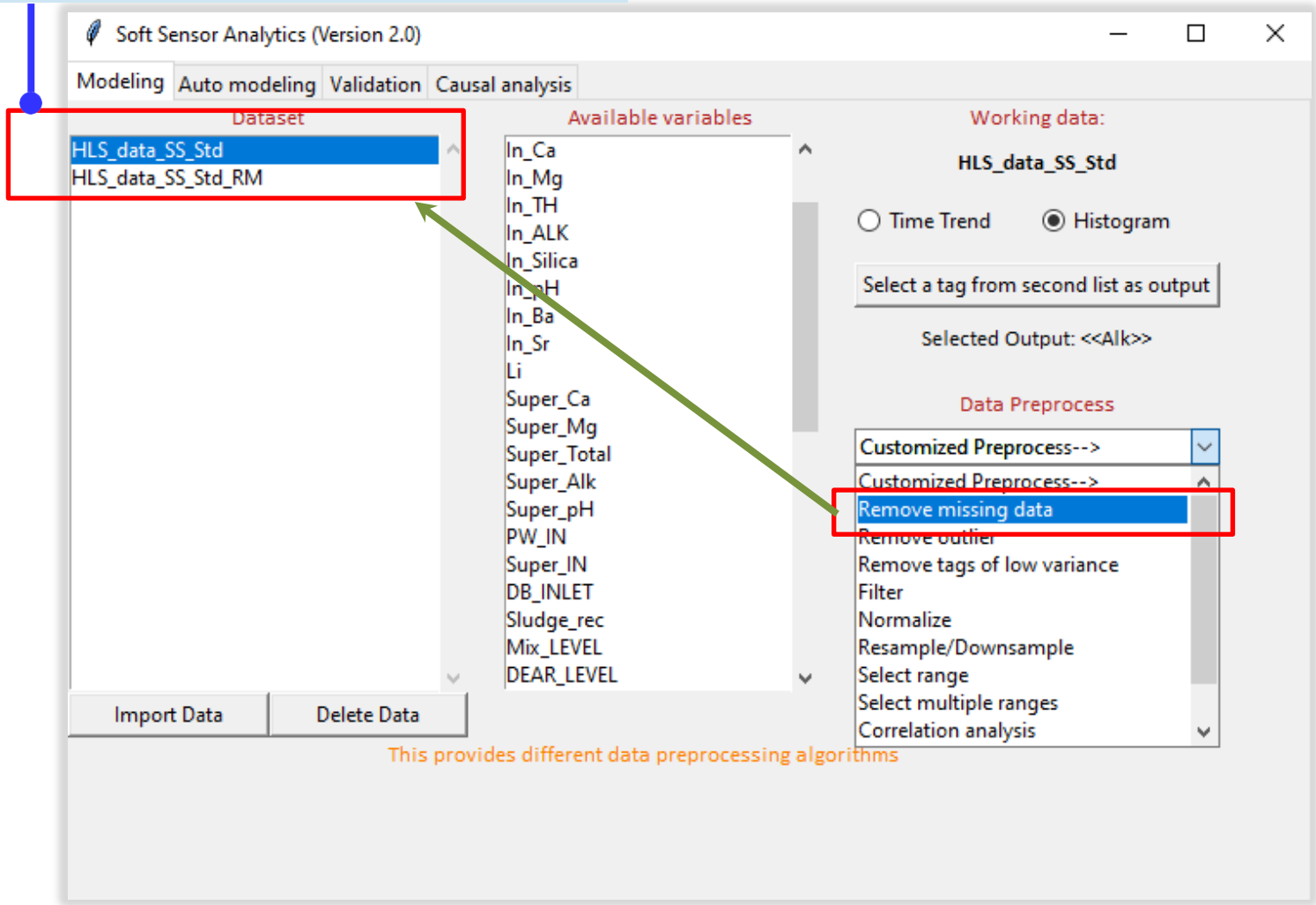
Soft Sensor Analytics

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"



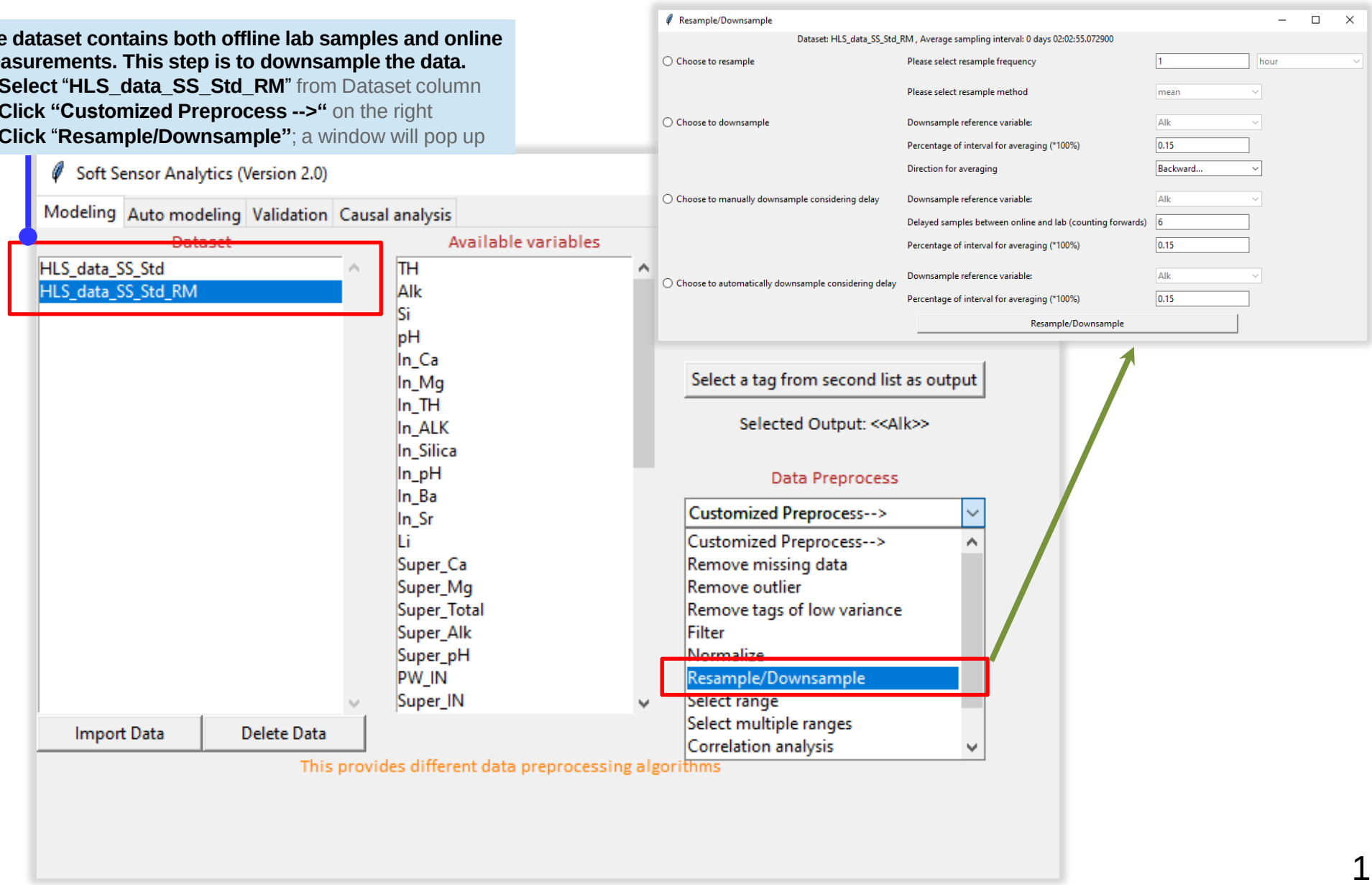
Soft Sensor Analytics

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

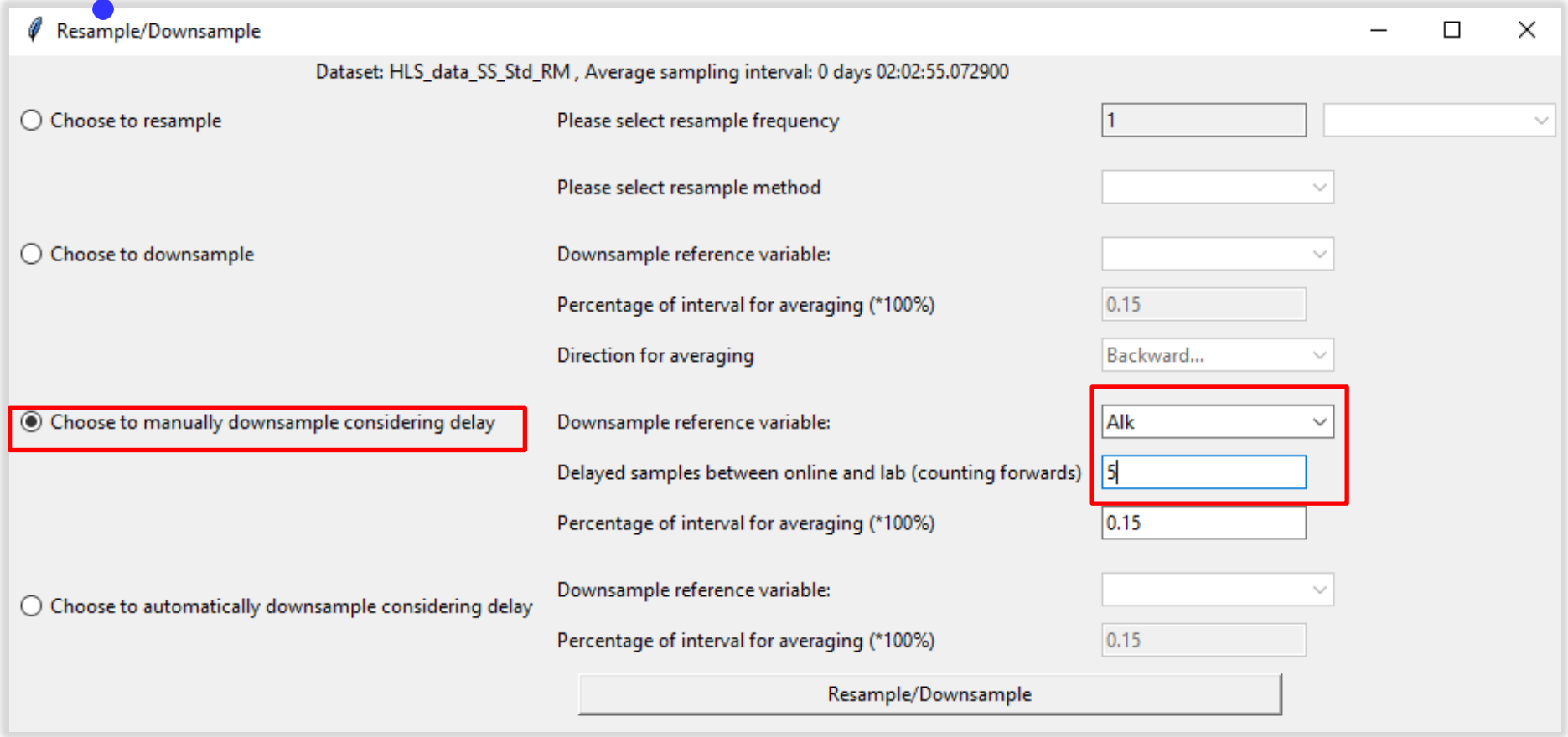


Soft Sensor Analytics

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)

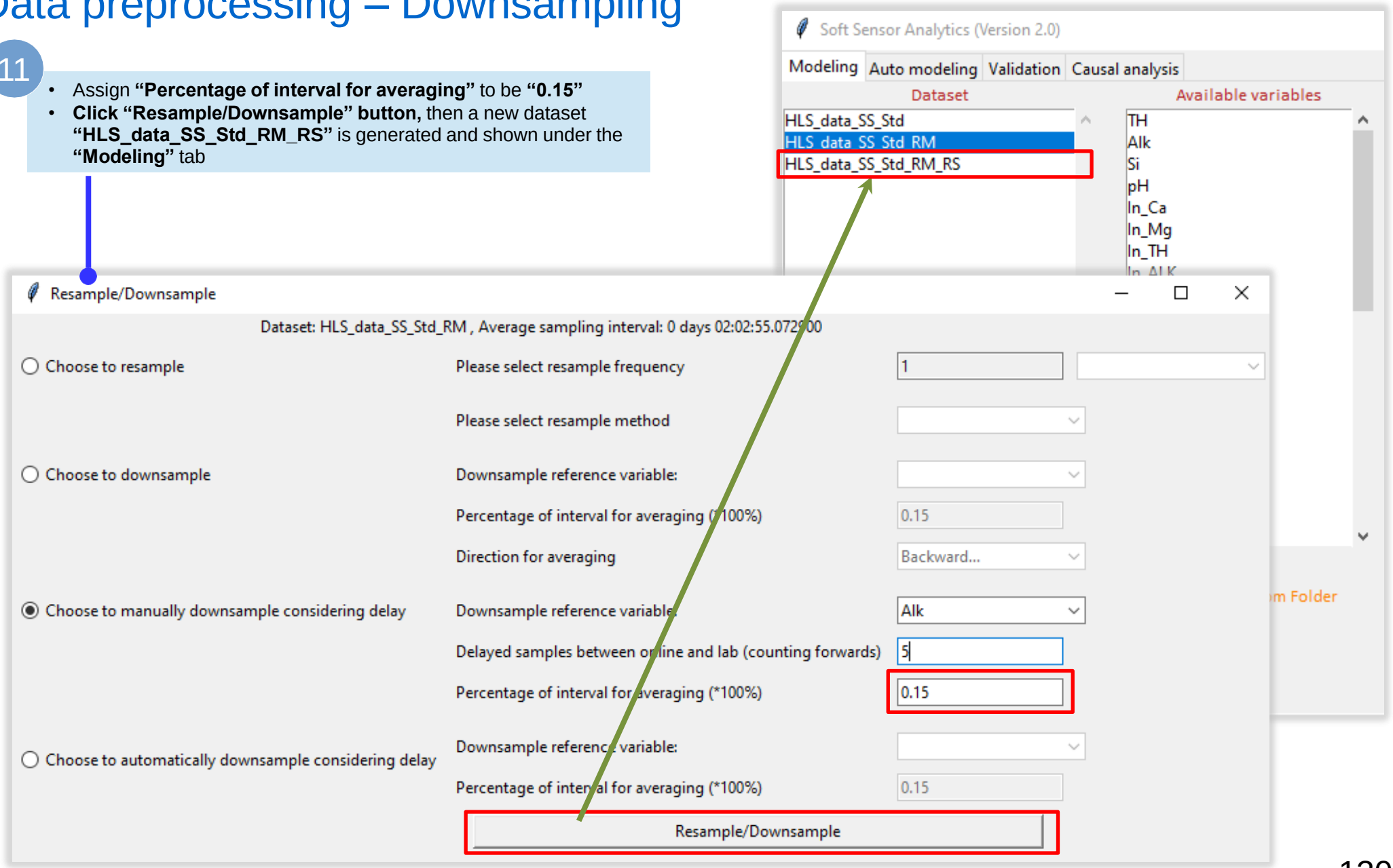


Soft Sensor Analytics

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



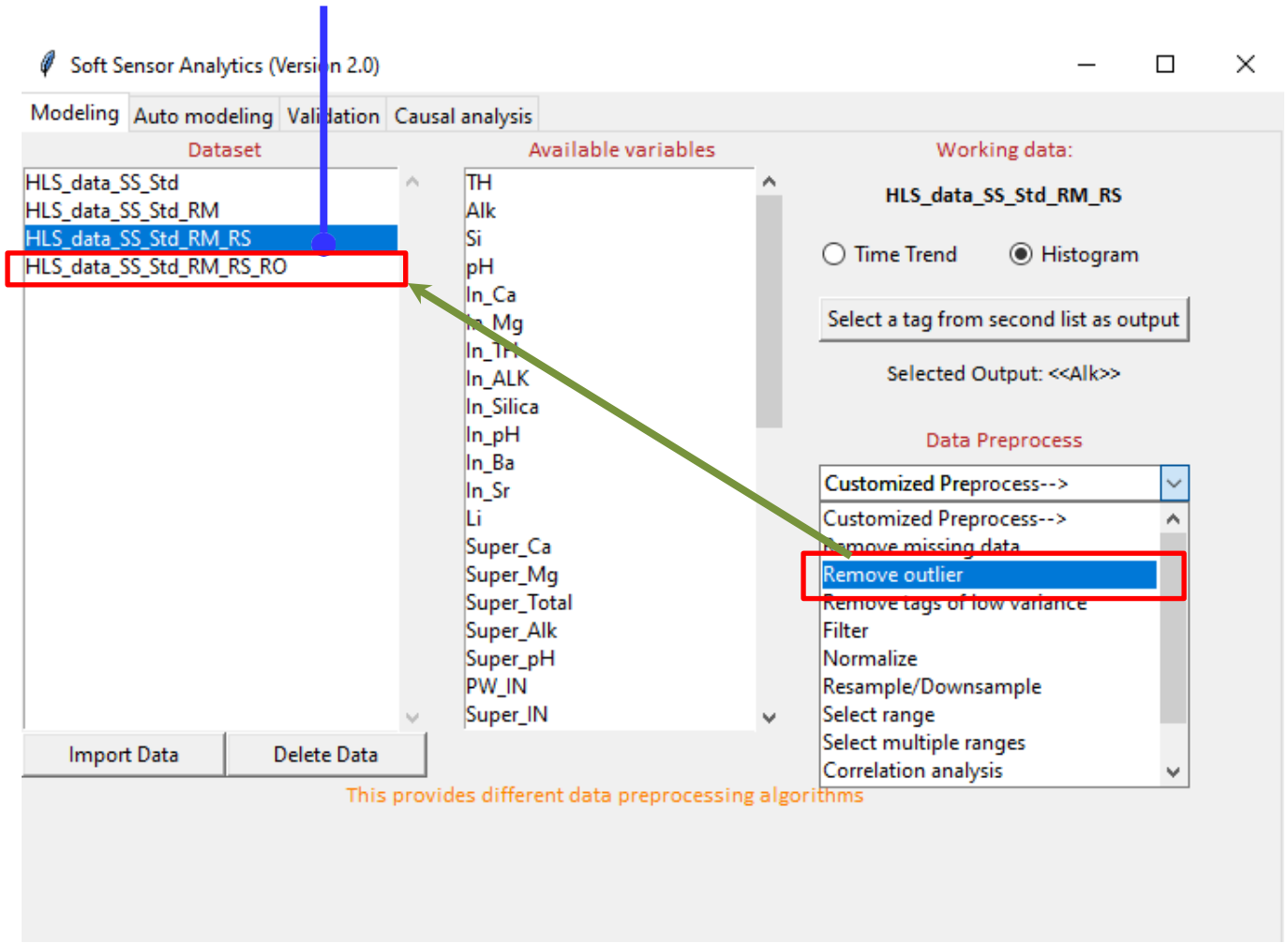
Soft Sensor Analytics

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



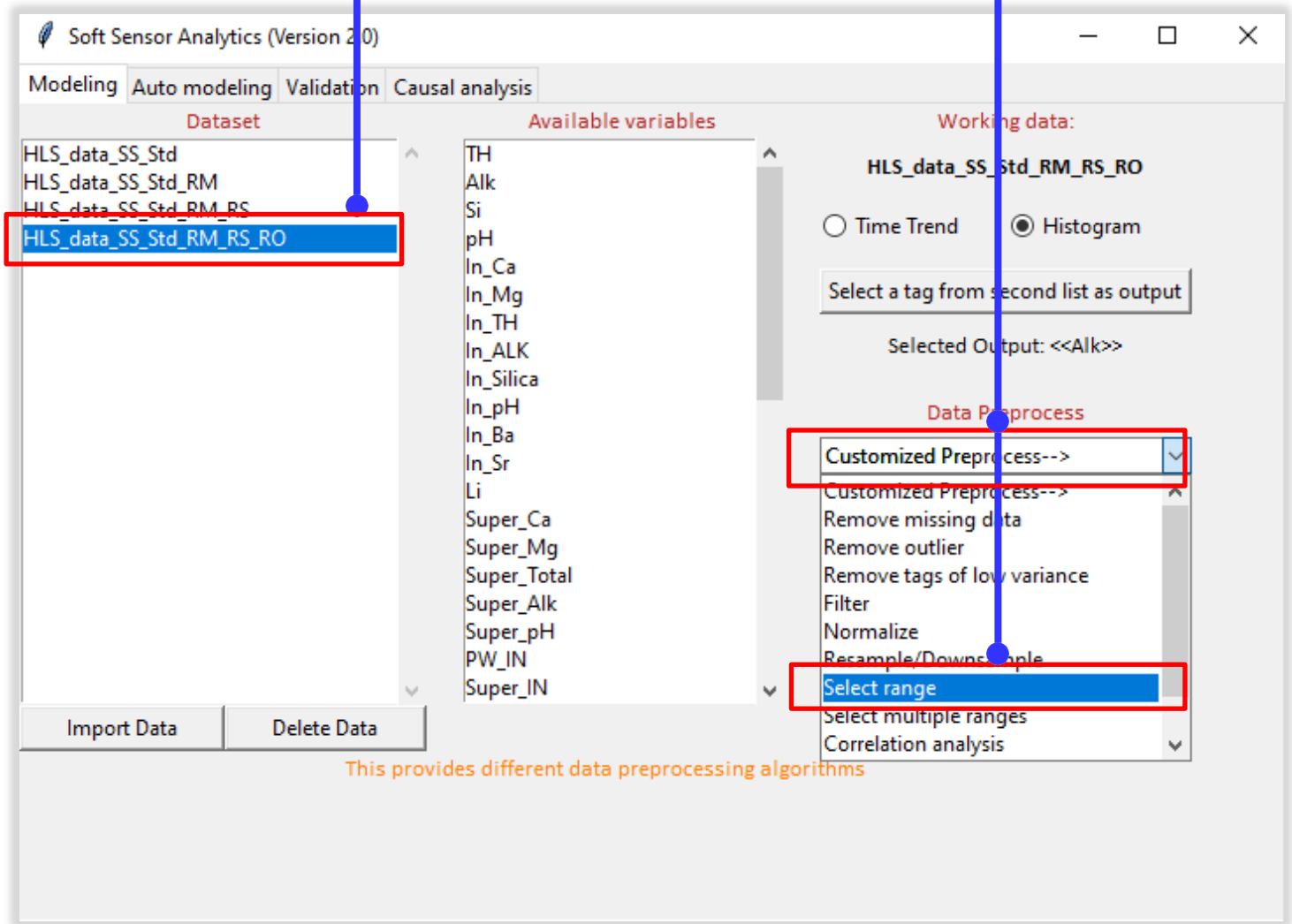
Soft Sensor Analytics

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



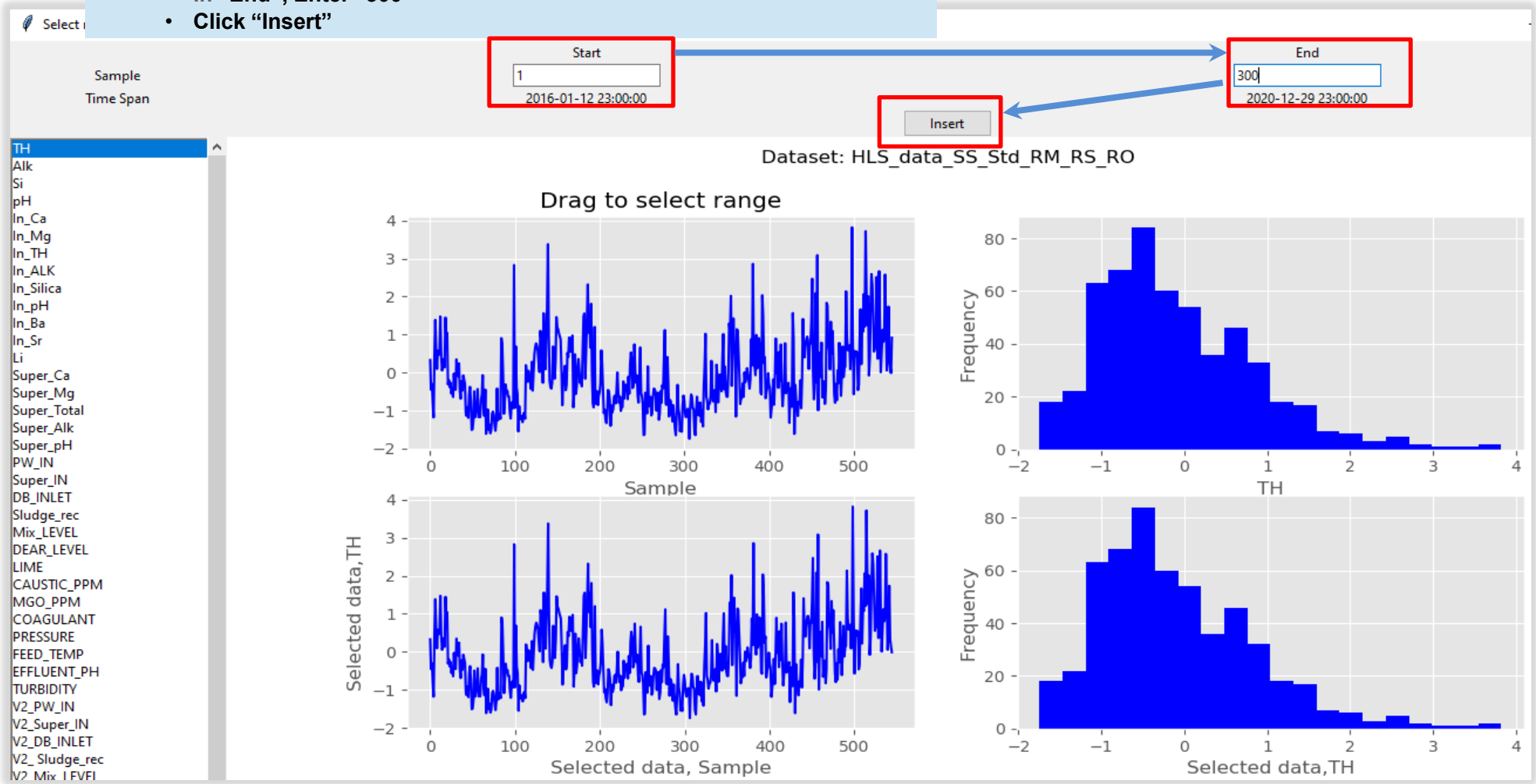
Soft Sensor Analytics

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"**



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

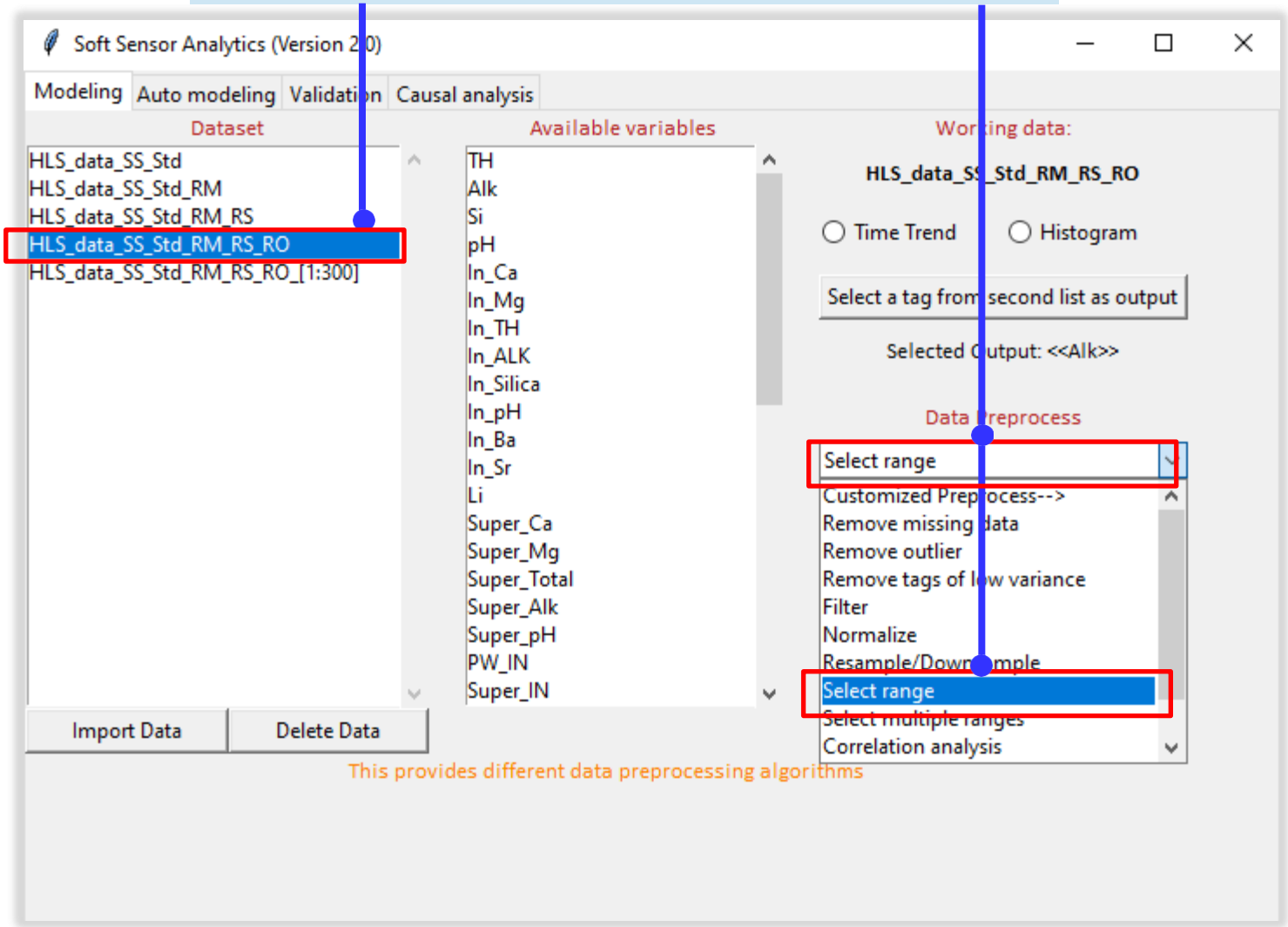
Soft Sensor Analytics

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



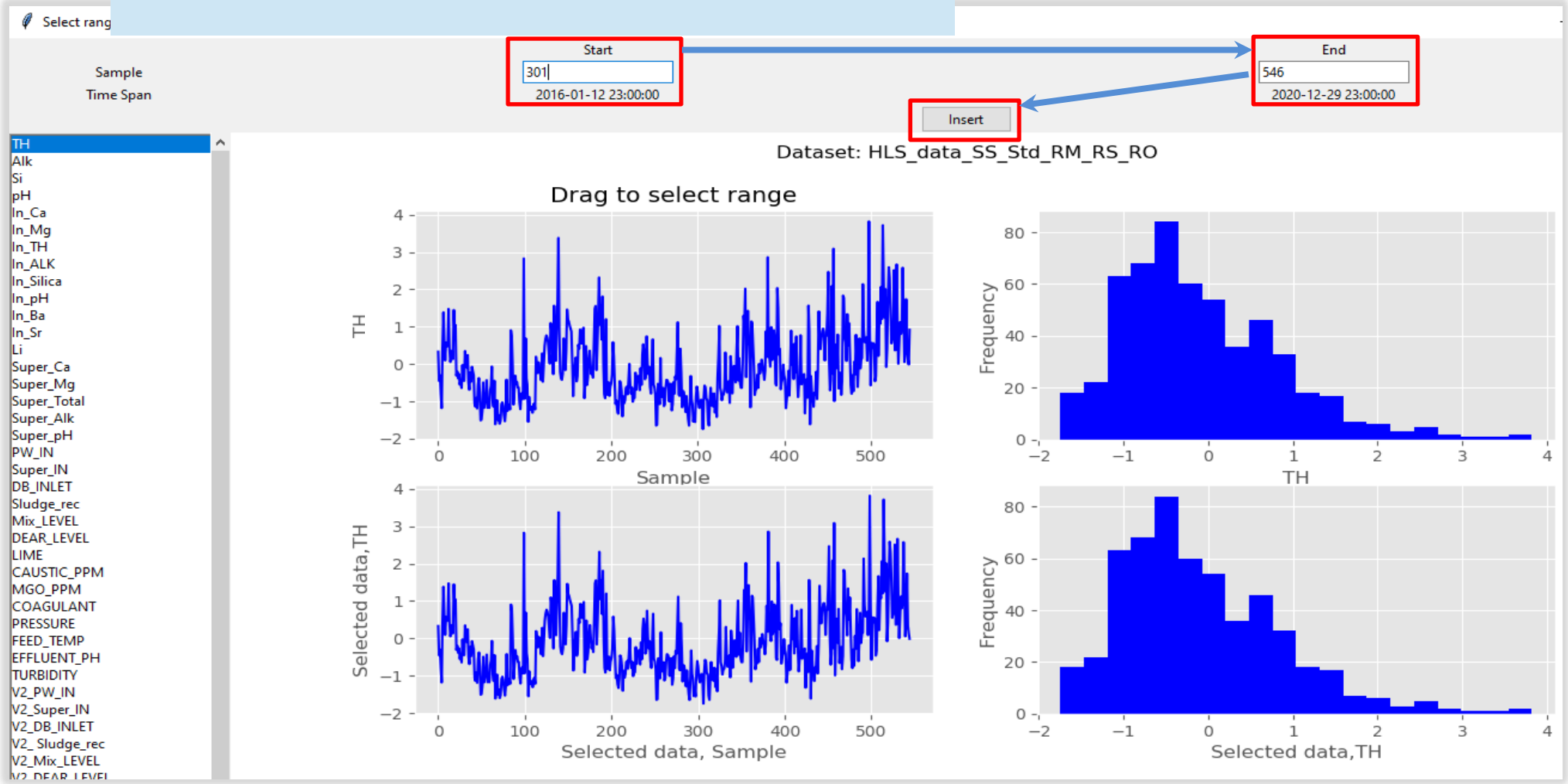
Soft Sensor Analytics

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"**

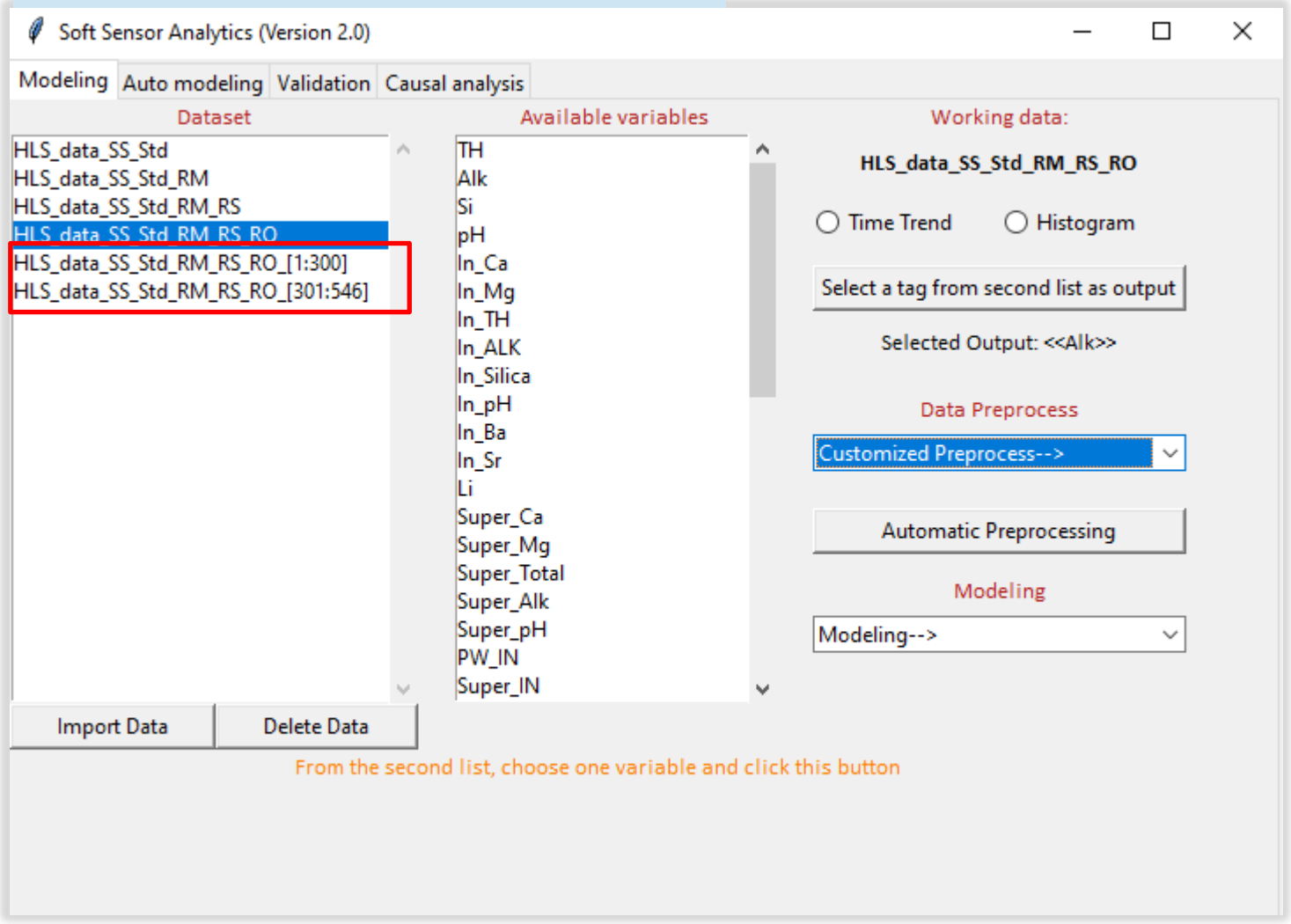


Note: Please close the current popup figures before taking next action

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



Part III-A: Prediction of Alkalinity

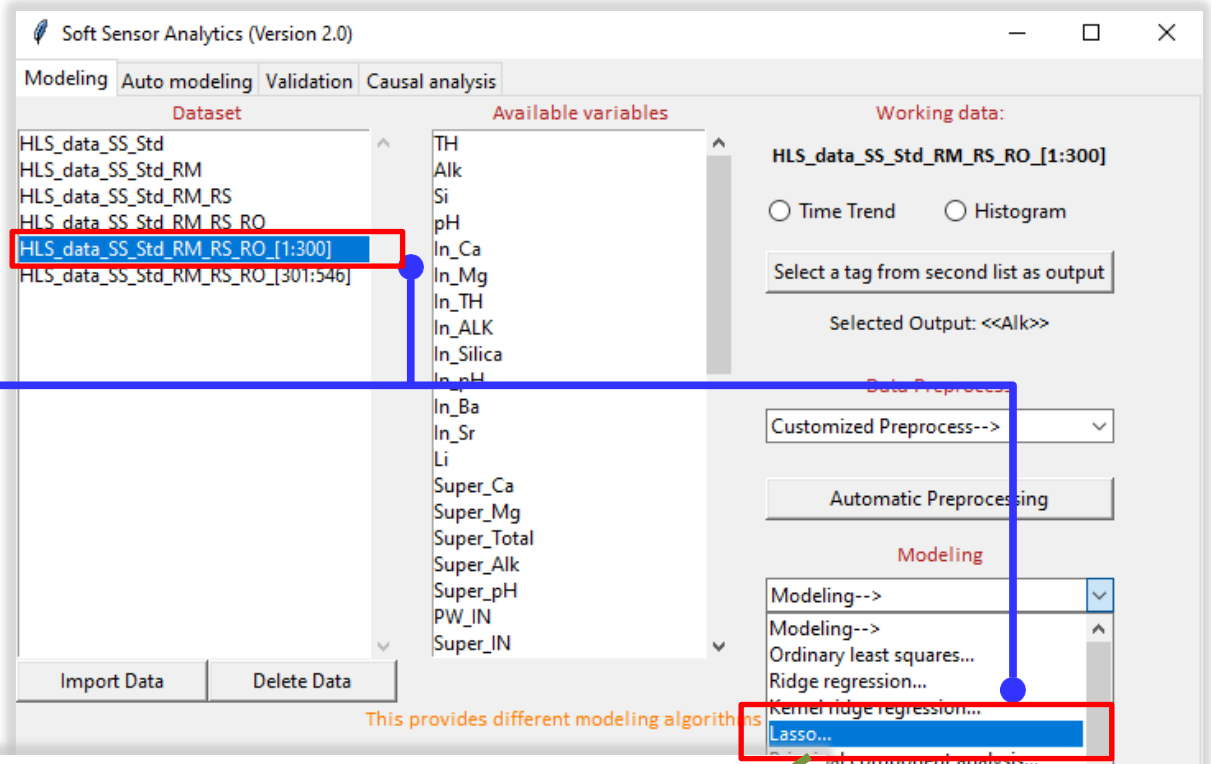
Soft Sensor Analytics

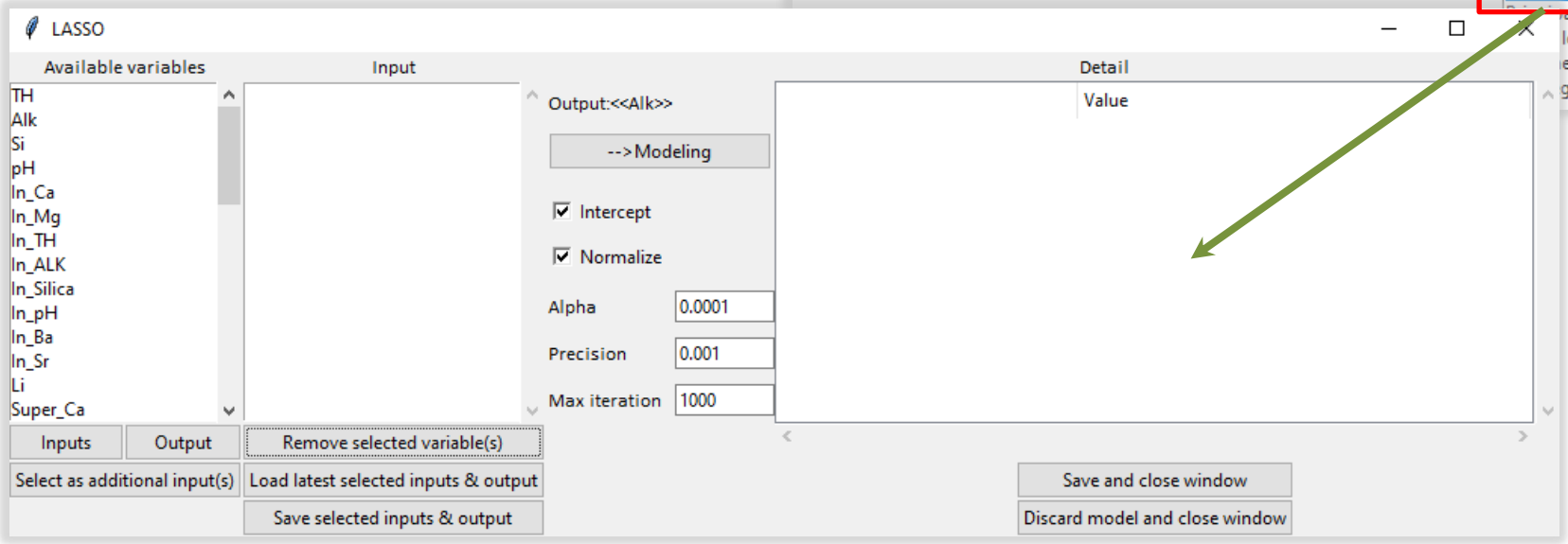
Modeling – Lasso regression

1

Choose dataset and modeling method

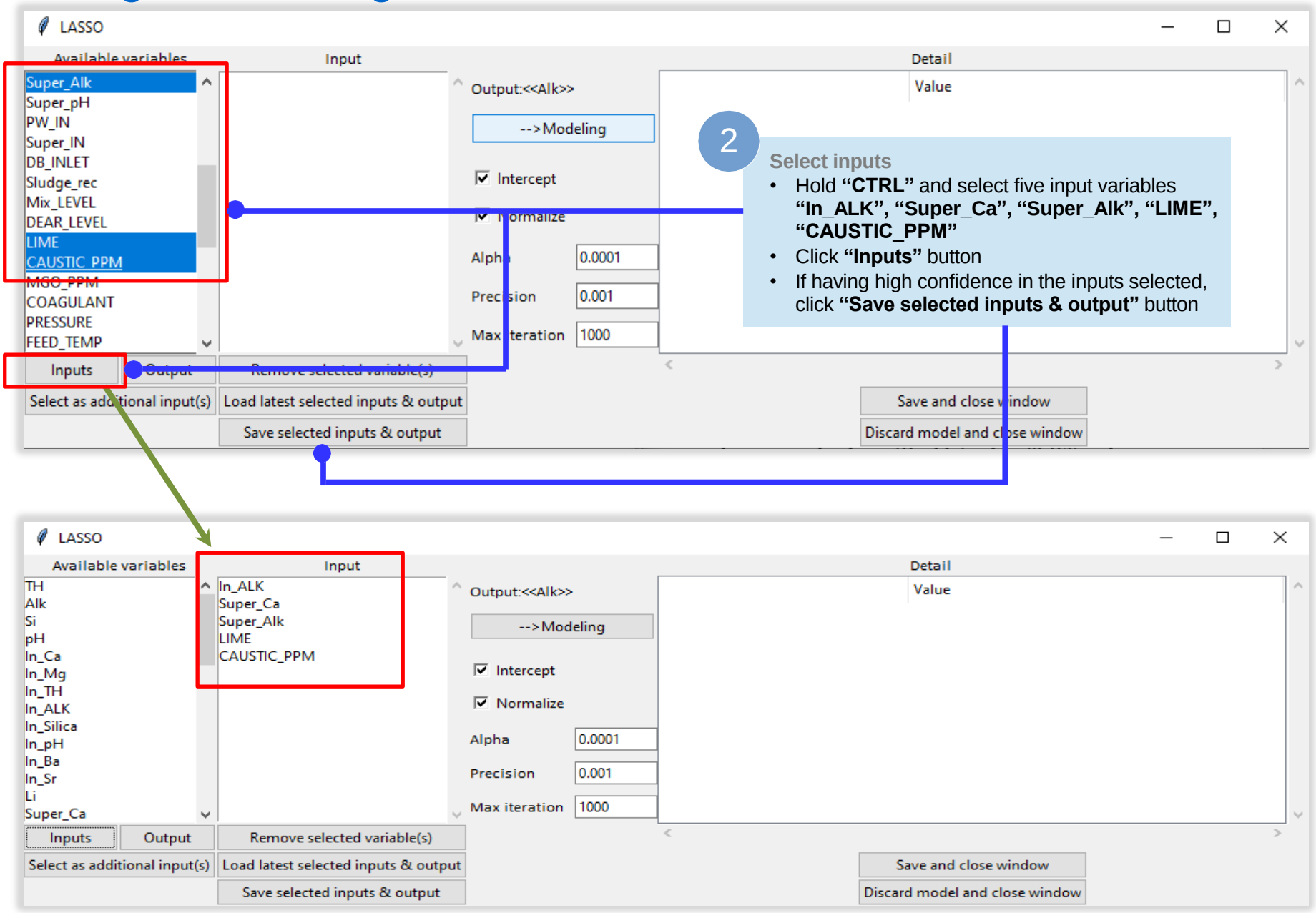
- **Click** the **"Modeling"** tab
- **Click HLS_data_SS_Std_RM_RS_RO _[1:300]** (training data)
- **Go to "Modeling-->", Click "Ordinary least squares…"** , a window pops up as below





Soft Sensor Analytics

Modeling – Lasso regression

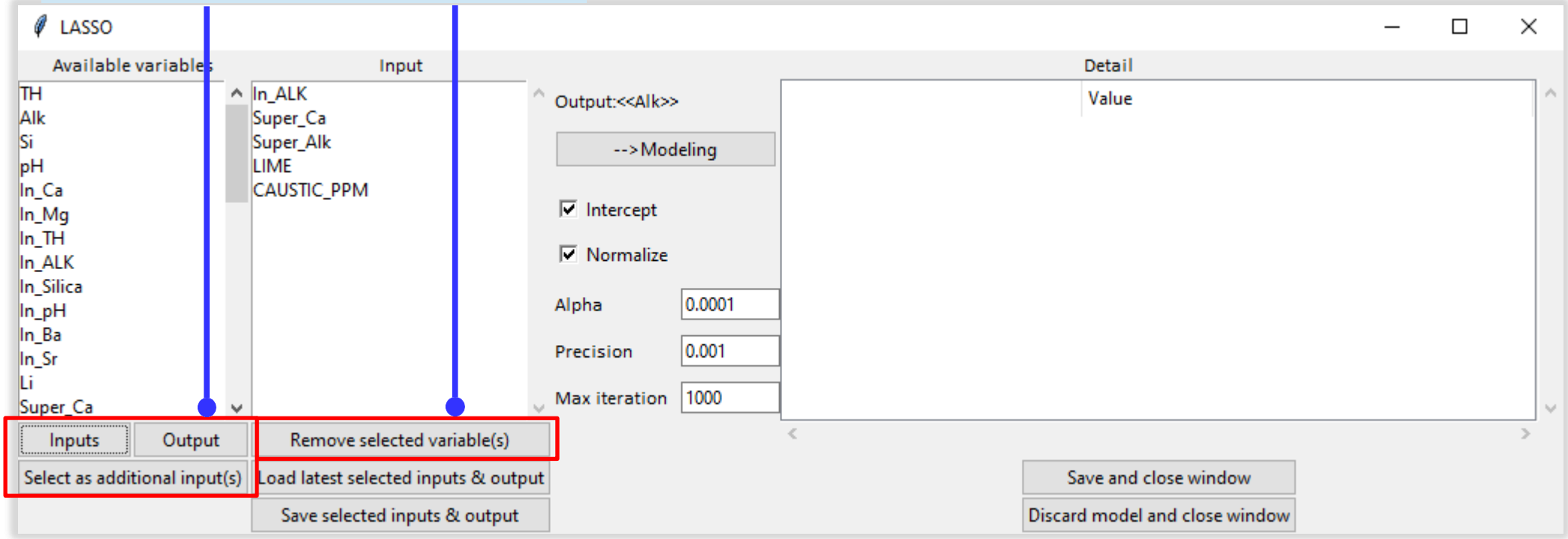


Soft Sensor Analytics

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

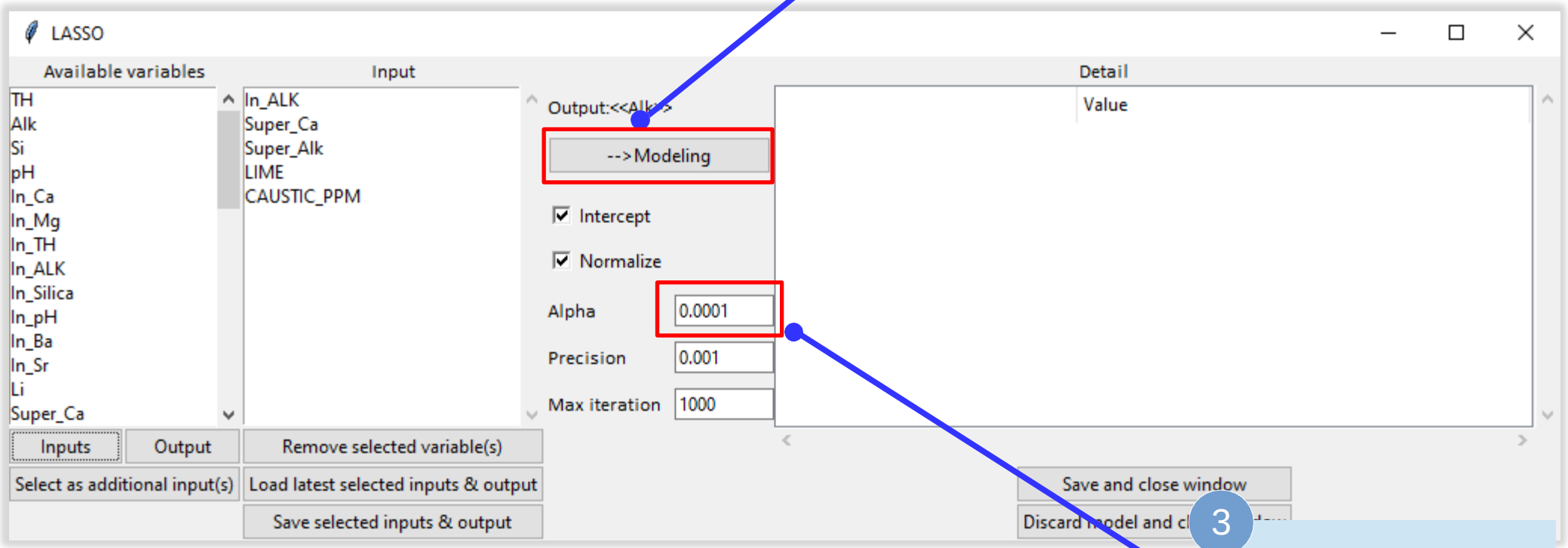


Soft Sensor Analytics

Modeling – Lasso regression

4

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

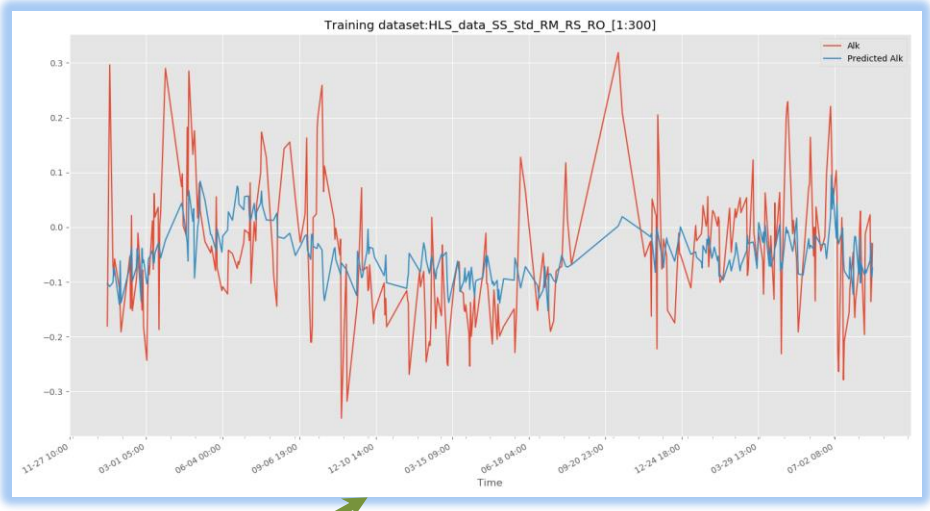


3

Options
Set **"Alpha"** as **"0.0005"**.

Soft Sensor Analytics

Results – Lasso regression for predicting alkalinity



LASSO

Available variables	Input		Detail	
TH	In_ALK	Output <<Alk>>		Value
Alk	Super_Ca	--> Modeling	Window size for averaging	1 sample(s)
Si	Super_Alk		Time delay	0 sample(s)
pH	LIME	☑ Intercept	Method	Lasso
In_Ca	CAUSTIC_PPM	☑ Normalize	Structure	y = ax + c
In_Mg		Alpha 0.0001	⊞ x	...
In_TH		Precision 0.001	y	Alk
In_ALK		Max iteration 1000	c	-0.05203576393196815
In_Silica			⊞ a	...
In_pH			alpha	0.0001
In_Ba			⊞ Performance	...
In_Sr				
Li				
Super_Ca				

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

5

Save model

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

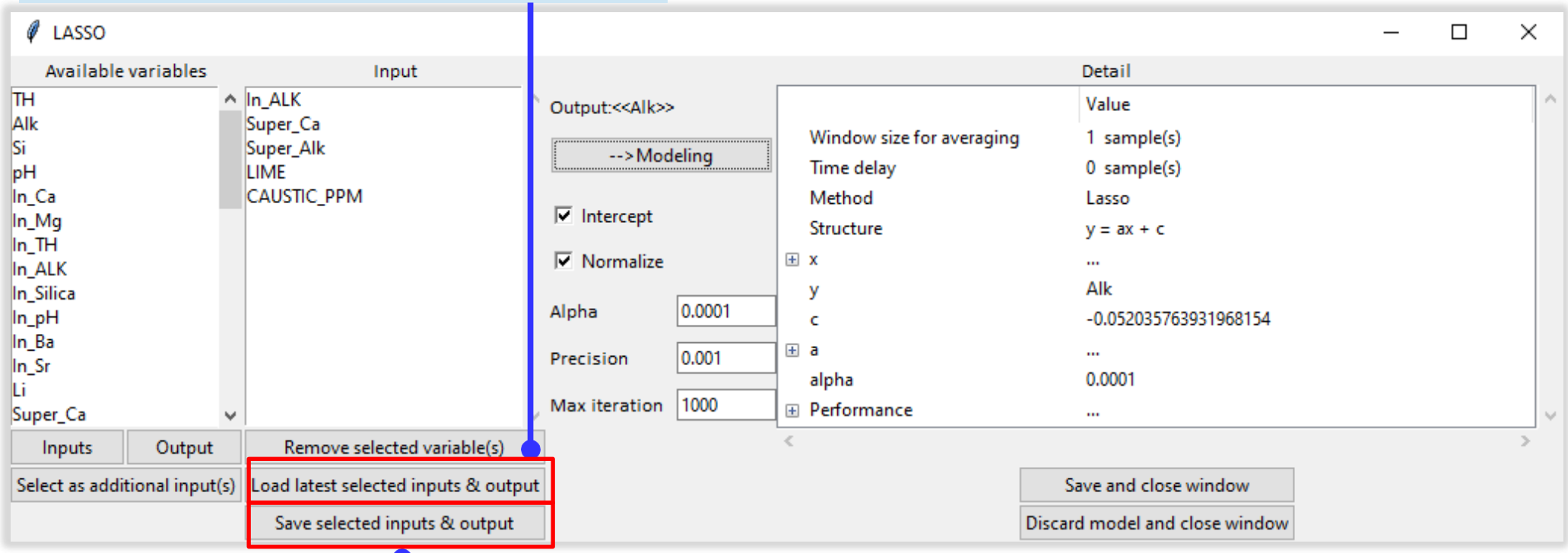
Soft Sensor Analytics

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



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

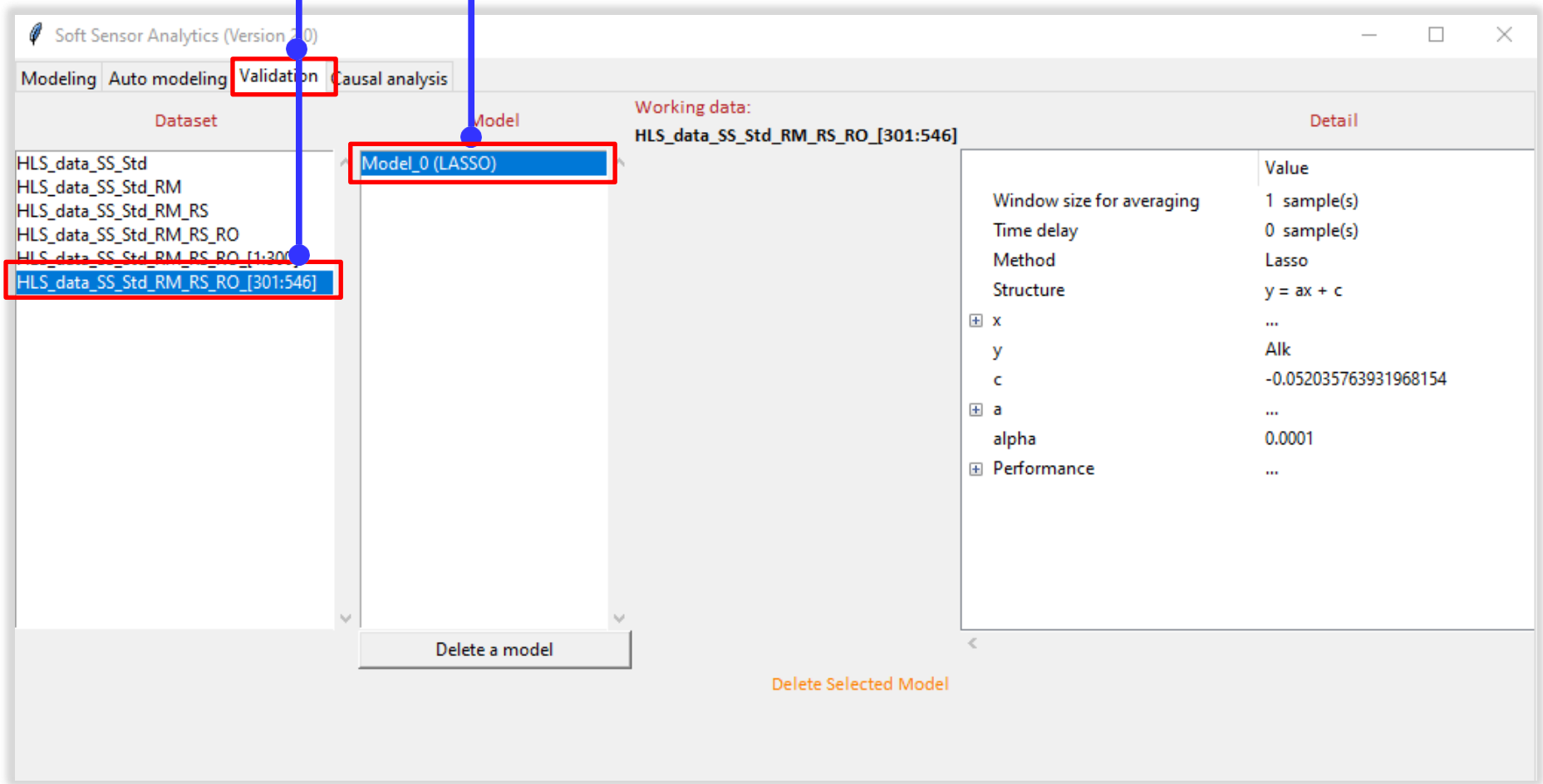
Validation of the models

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



Soft Sensor Analytics

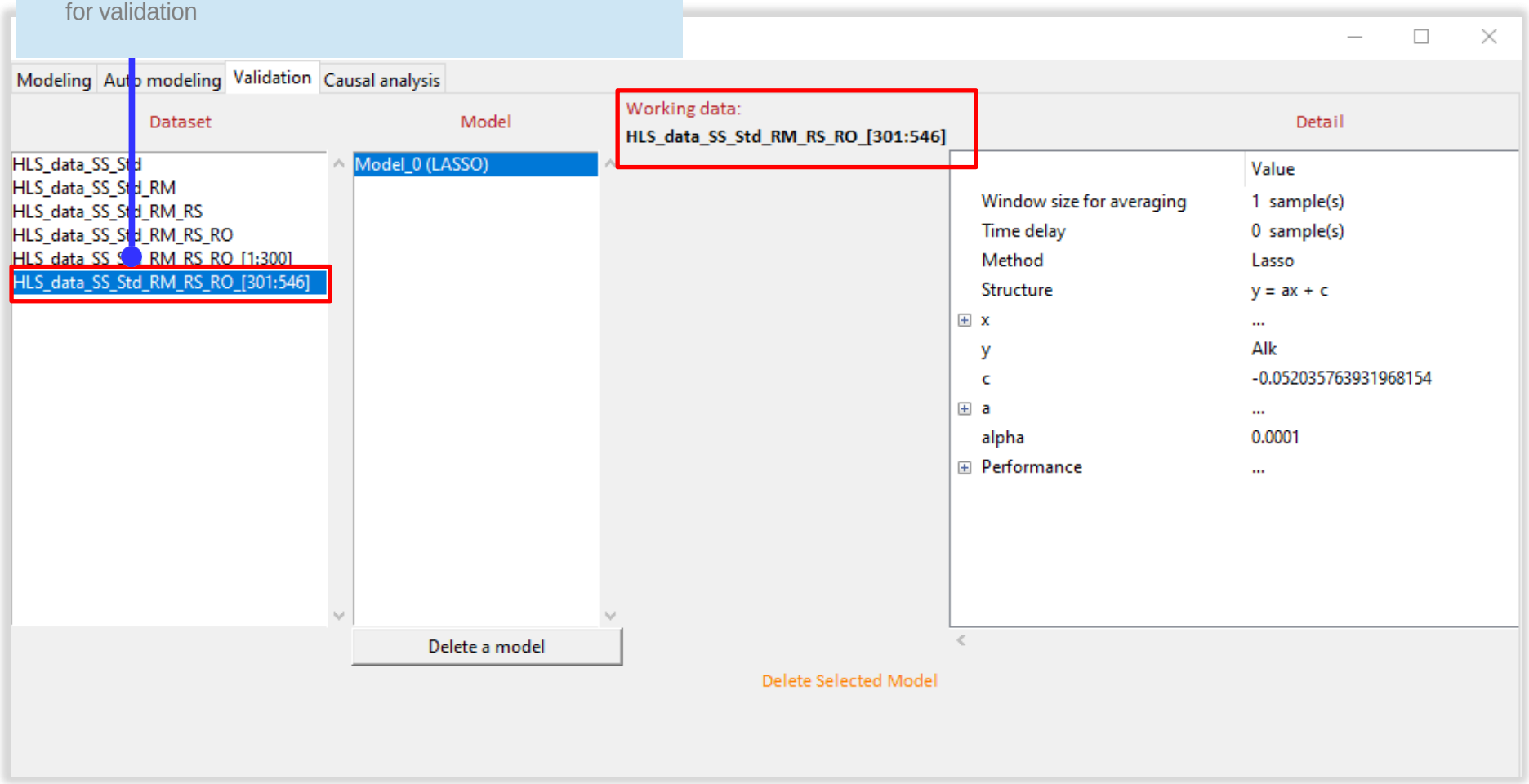
Validation of the model

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



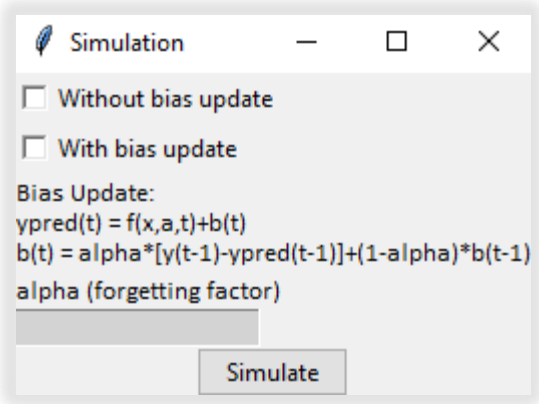
Soft Sensor Analytics

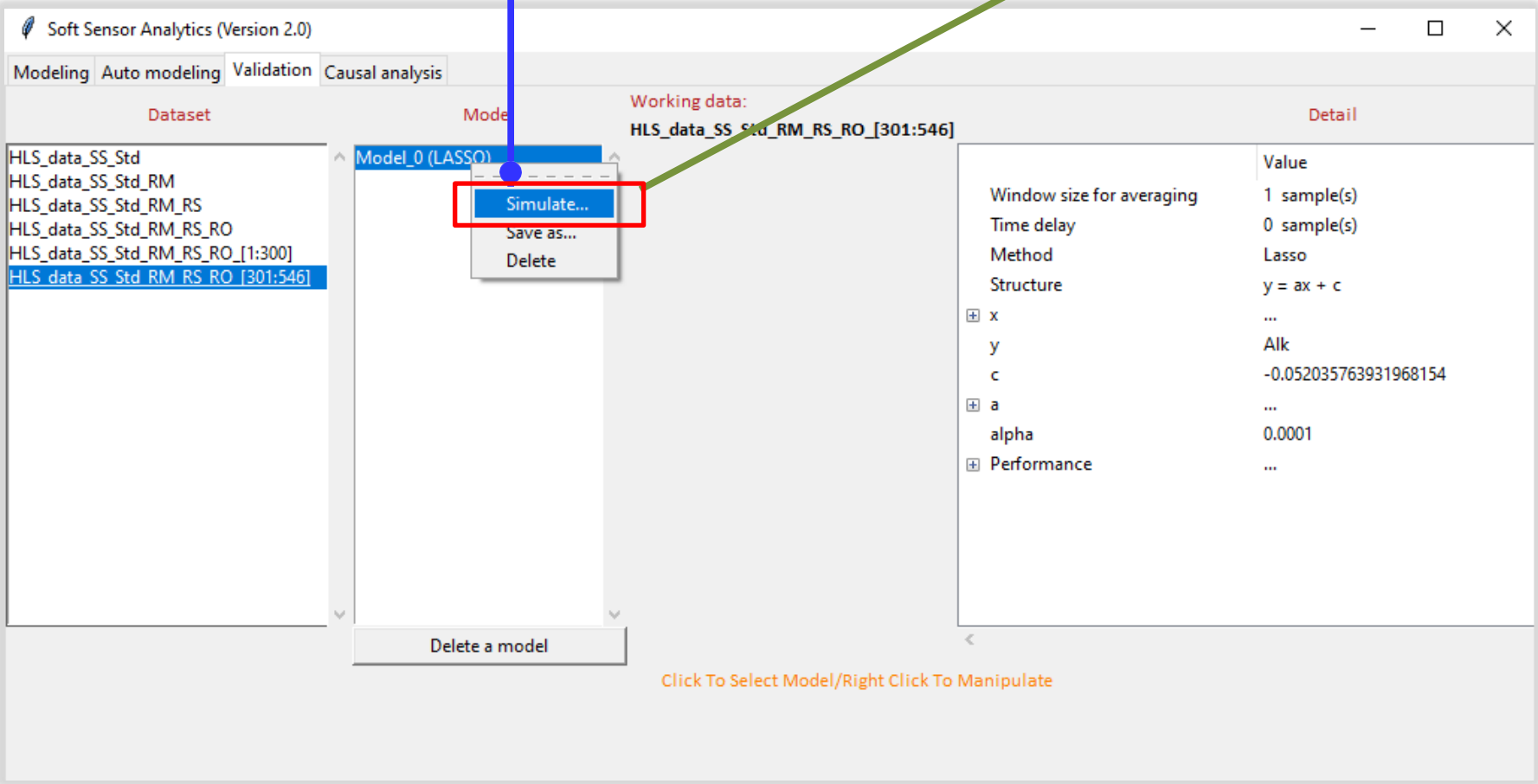
Validation of the model

3

Choose "Simulate" to continue

- **Right-Click** on the two models
- **Click "Simulate…"** button, a window will pop up





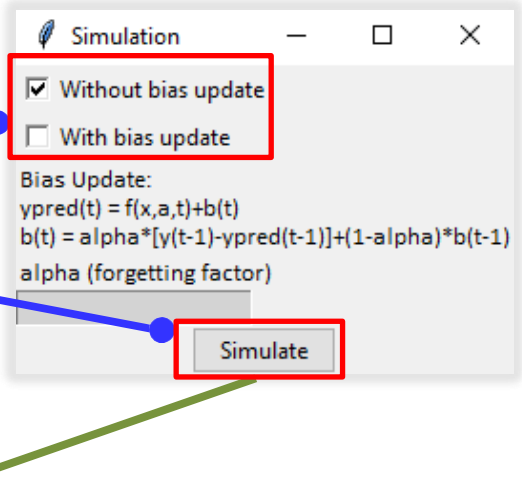
Soft Sensor Analytics

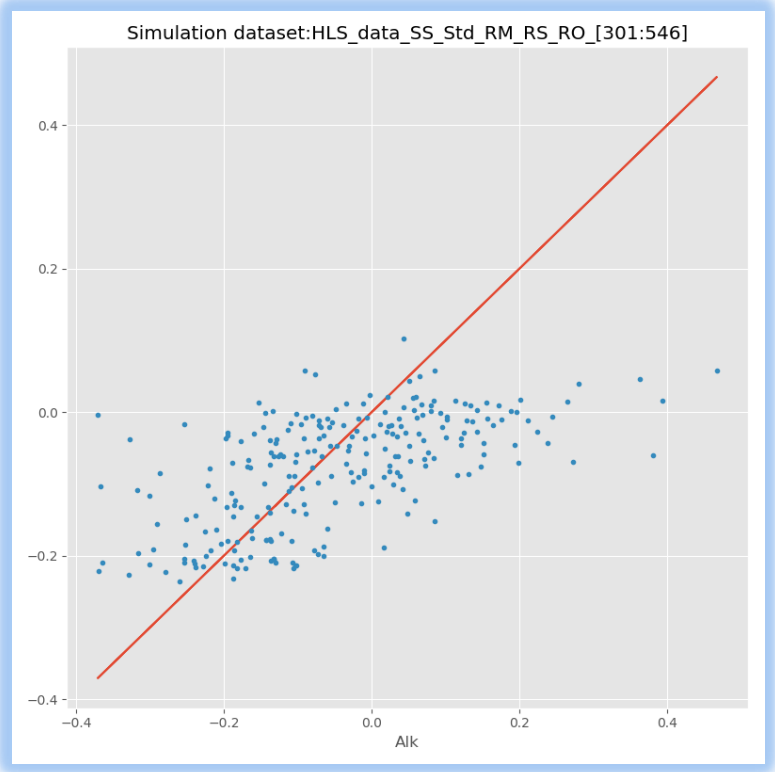
Validation of the model

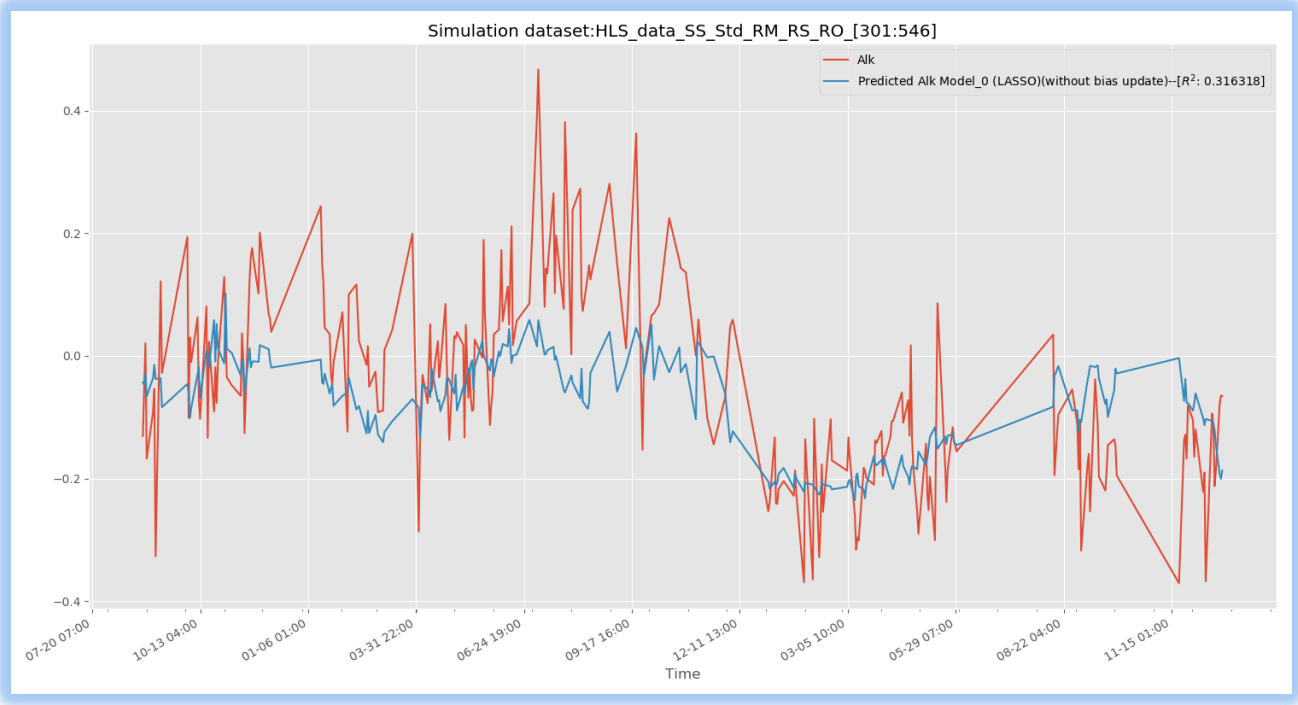
4

Model validation

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







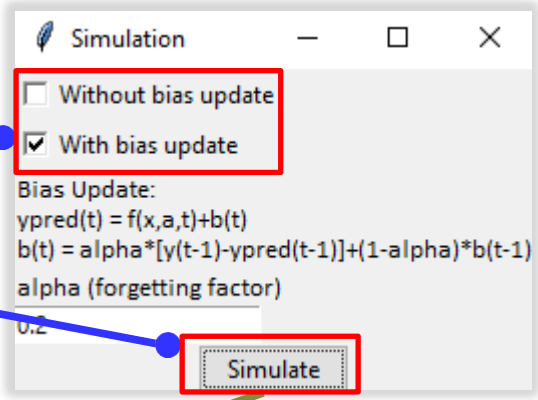
Soft Sensor Analytics

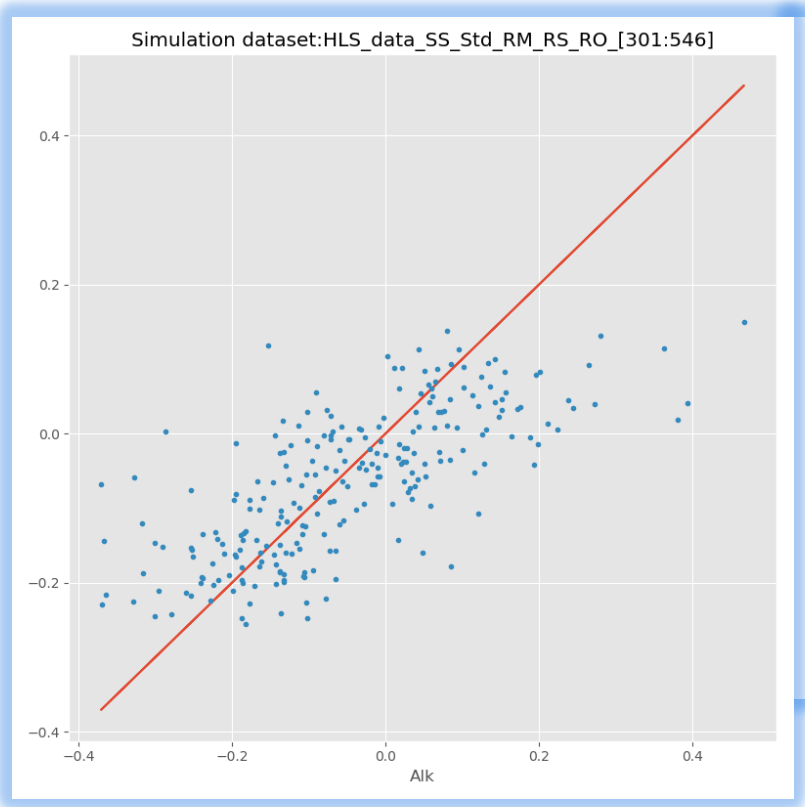
Validation of the model

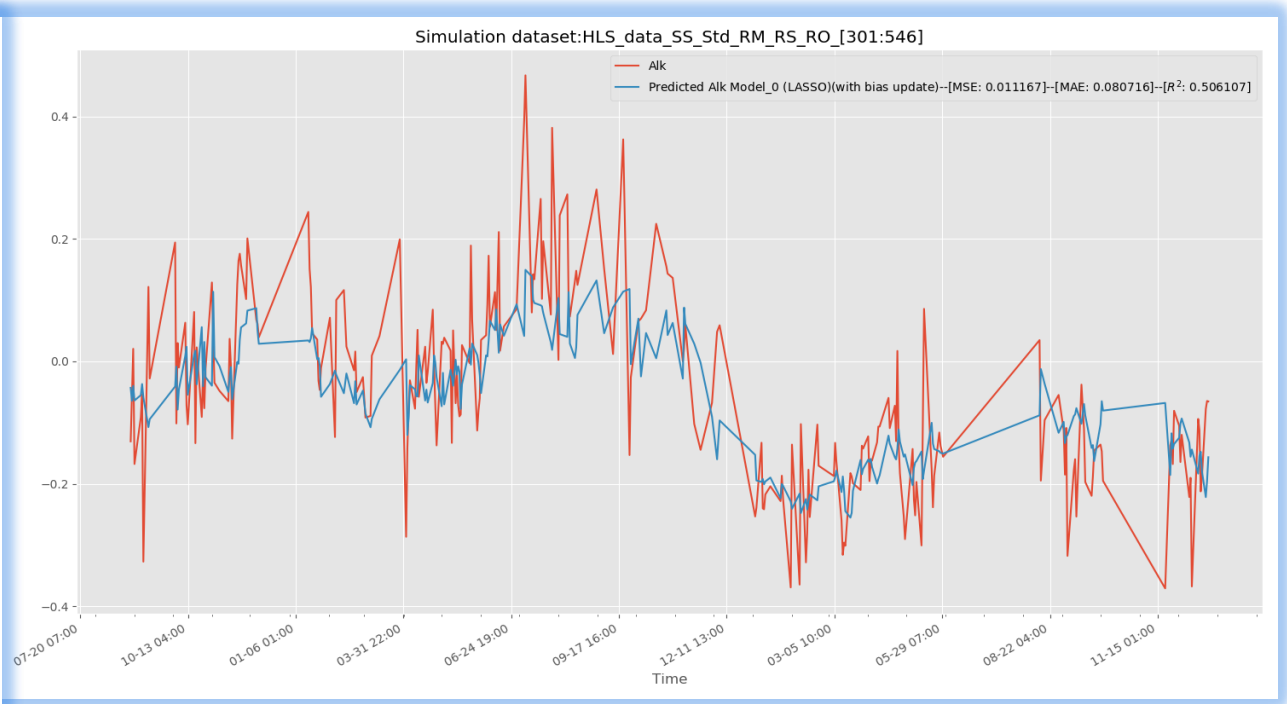
5

Model validation

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







Part III-B: Prediction of Silica

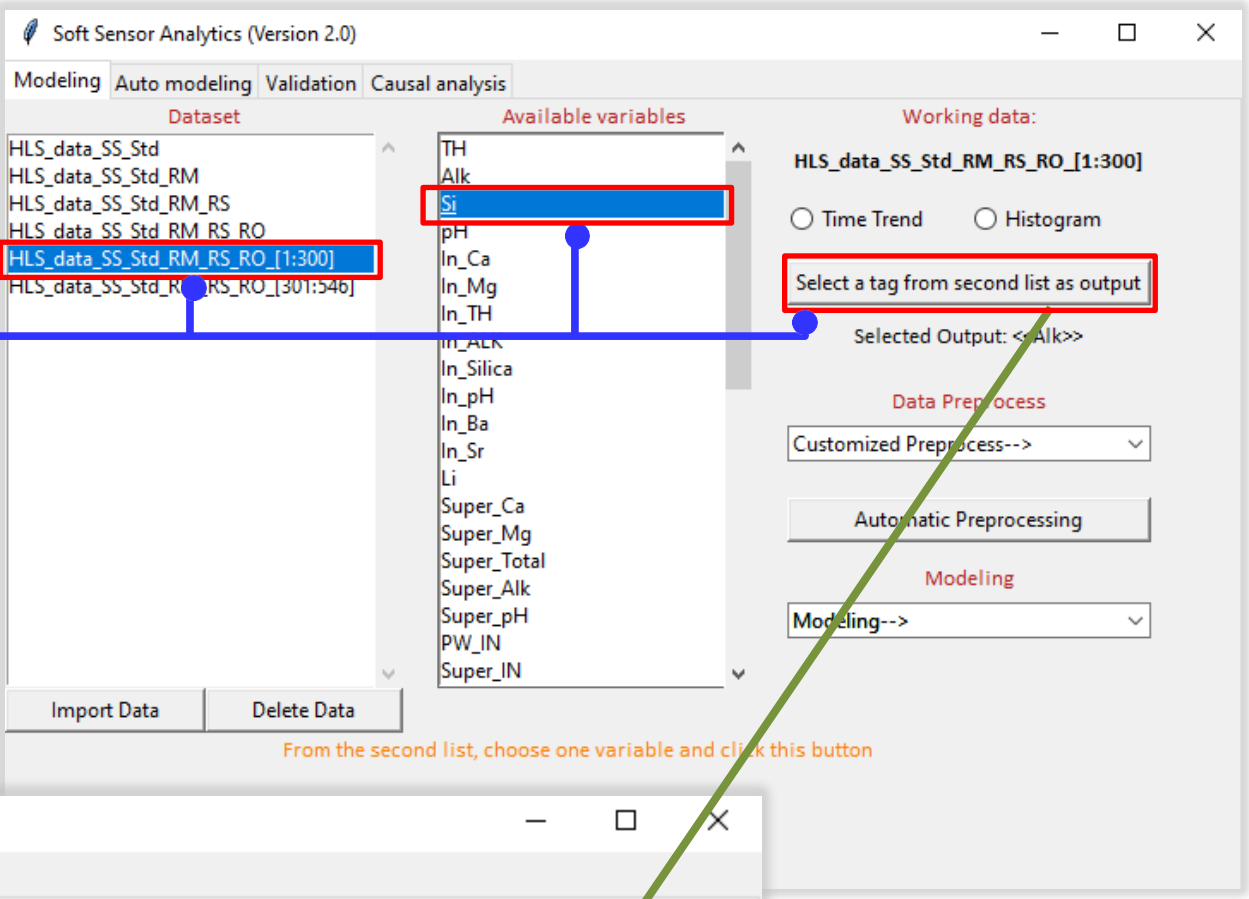
Soft Sensor Analytics

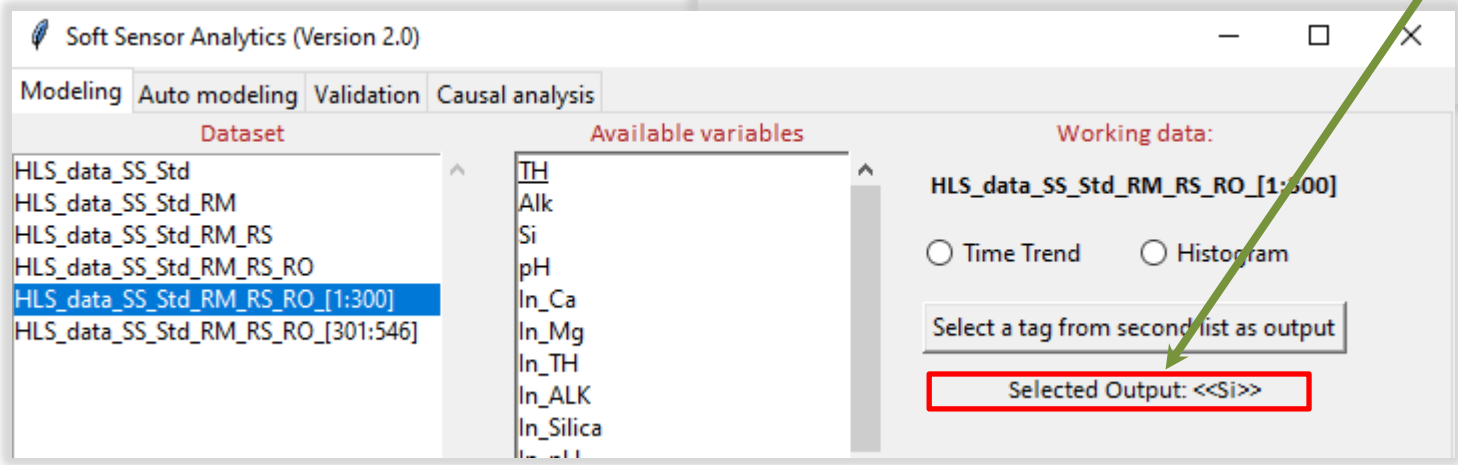
Change output to be predicted

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





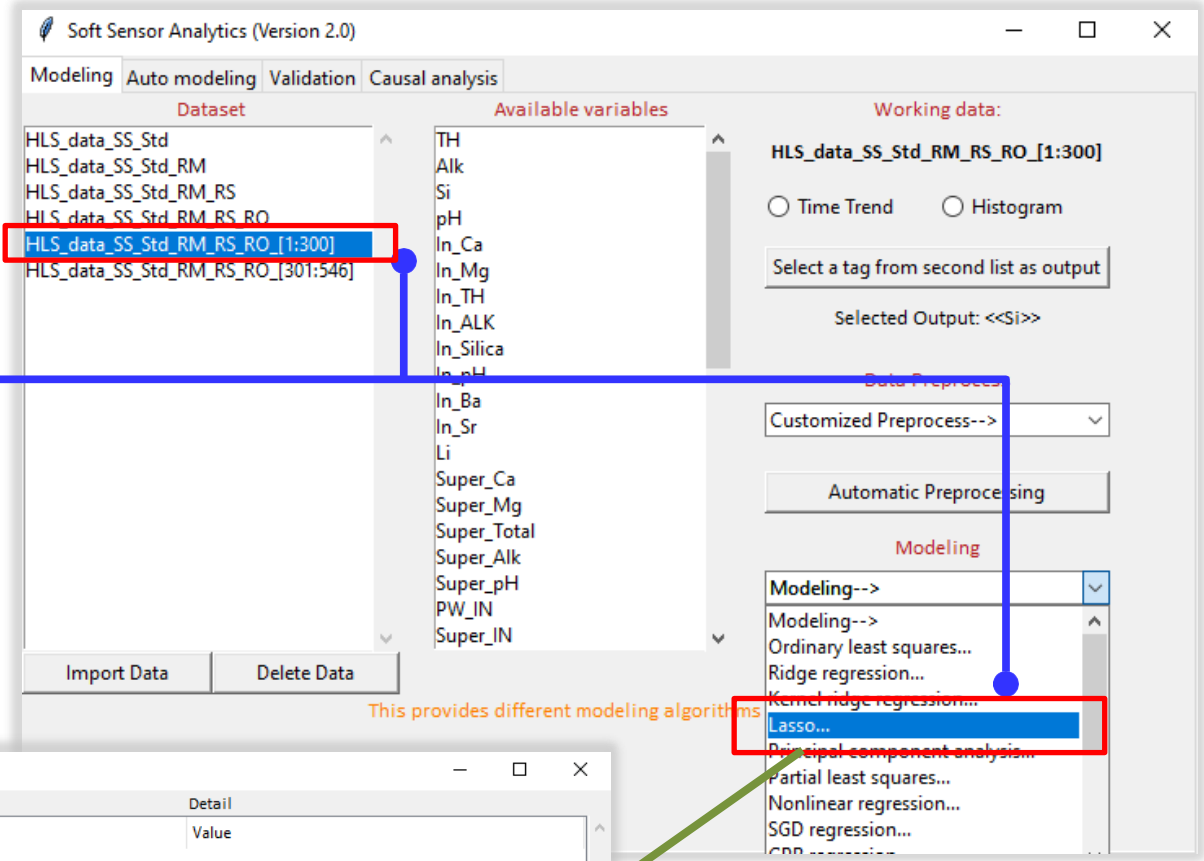
Soft Sensor Analytics

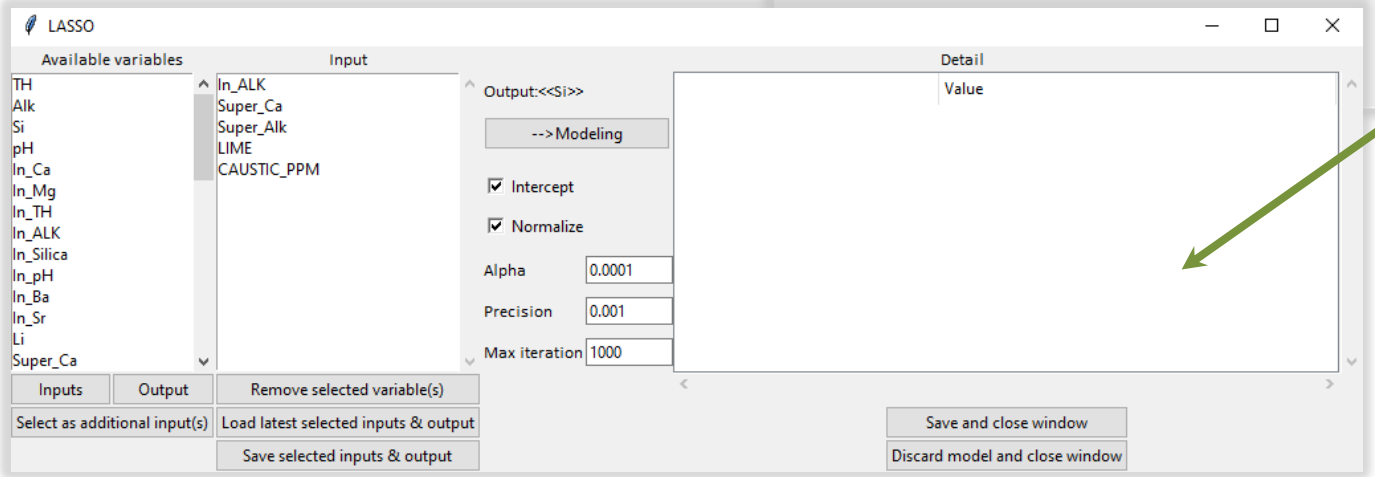
Modeling – Lasso regression

2

Choose dataset and modeling method

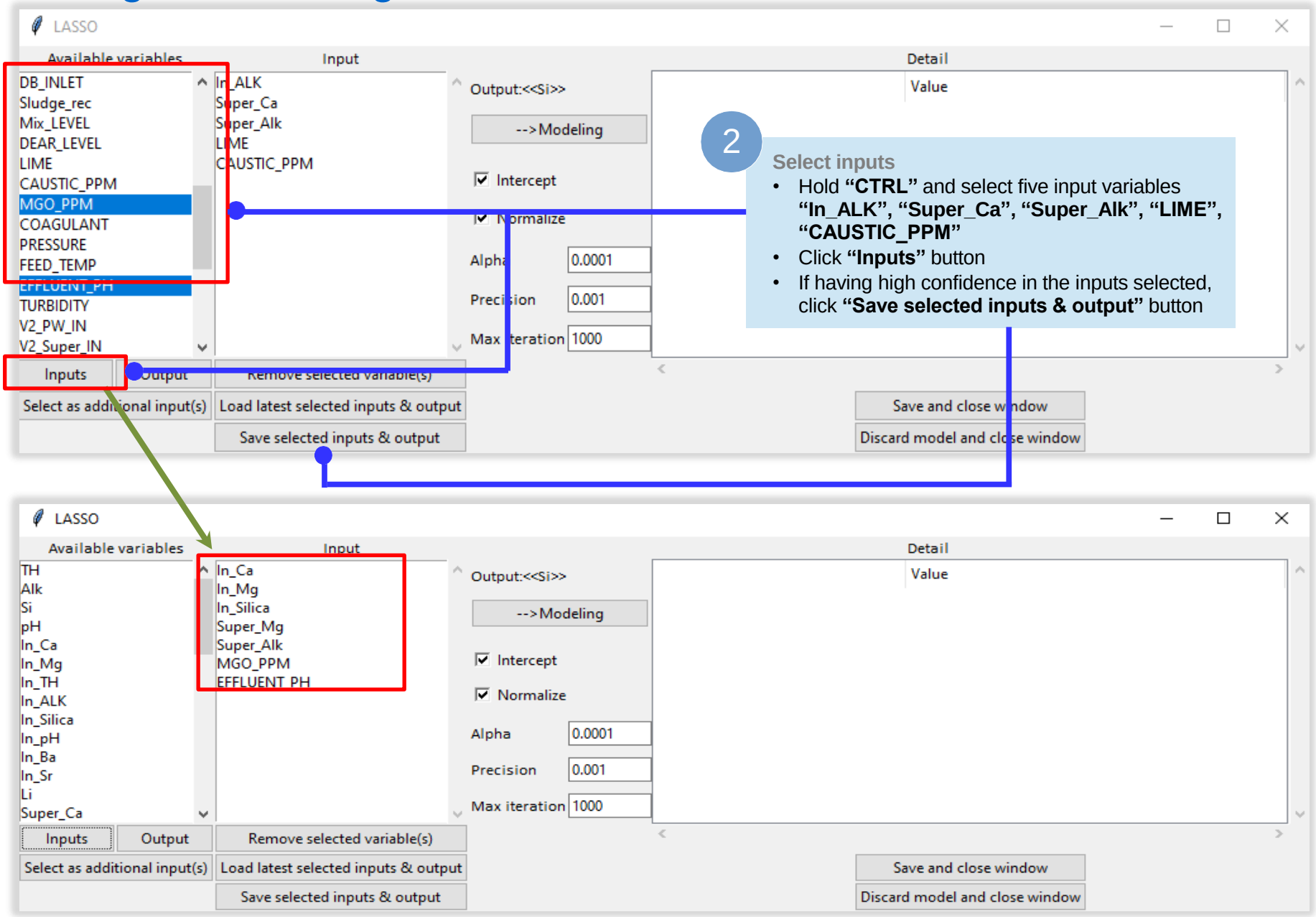
- **Click** the **"Modeling"** tab
- **Click HLS_data_SS_Std_RM_RS_RO _[1:300]** (training data)
- **Go to "Modeling-->", Click "Ordinary least squares…"** , a window pops up as below





Soft Sensor Analytics

Modeling – Lasso regression

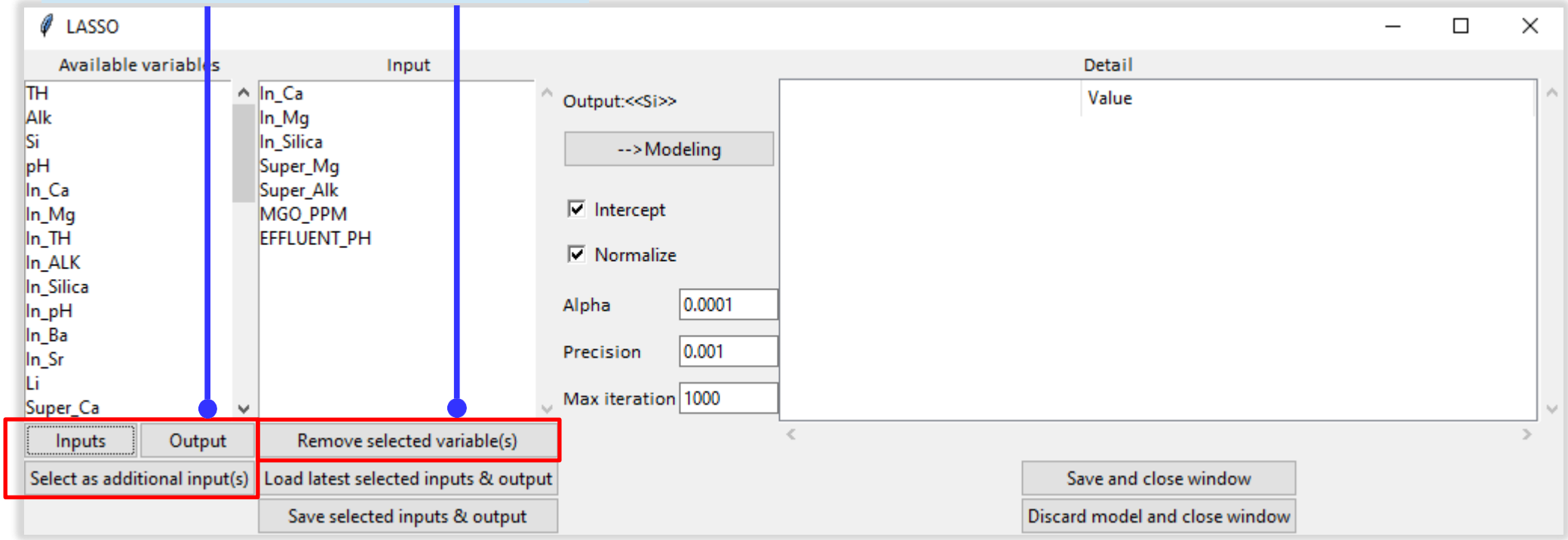


Soft Sensor Analytics

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

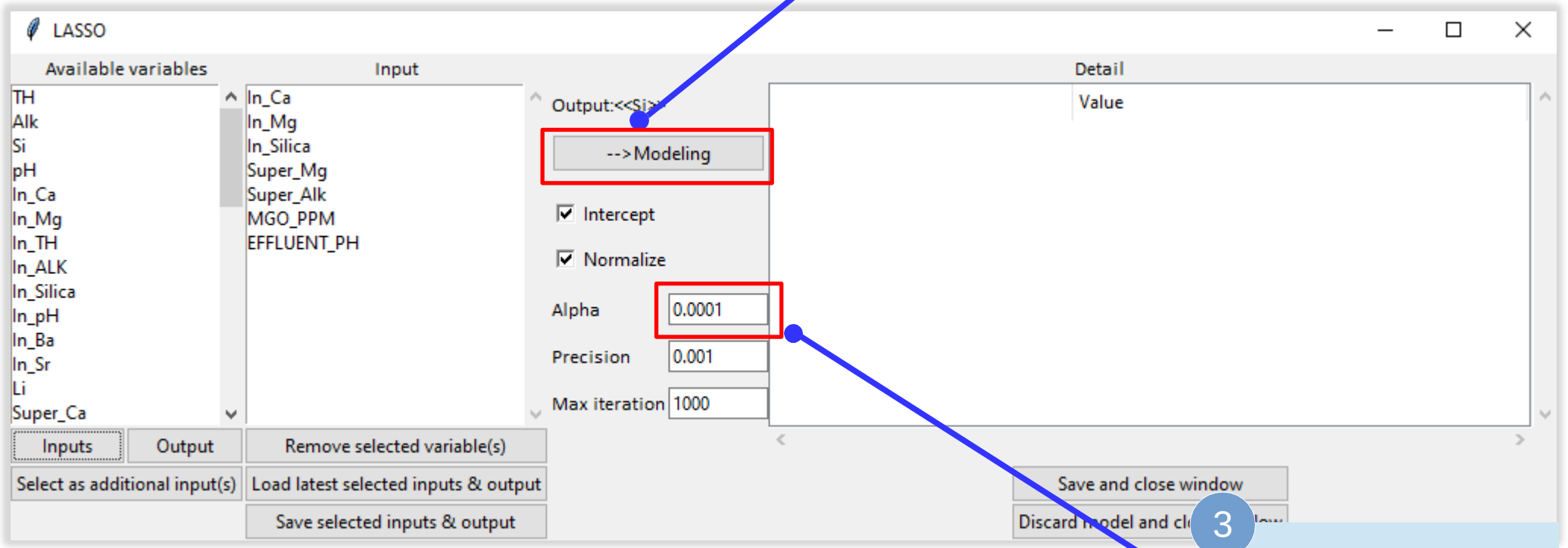


Soft Sensor Analytics

Modeling – Lasso regression

4

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

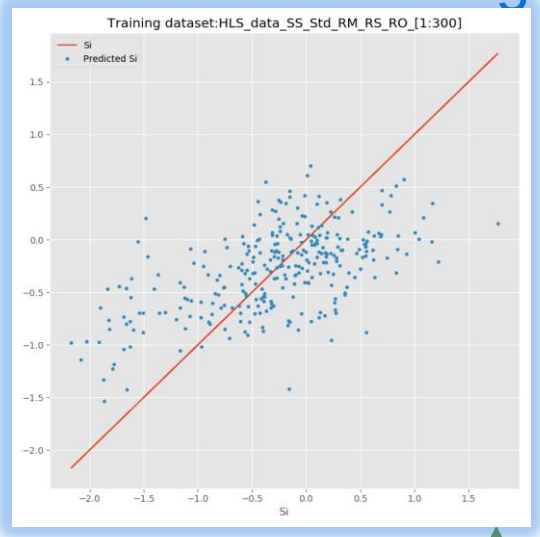


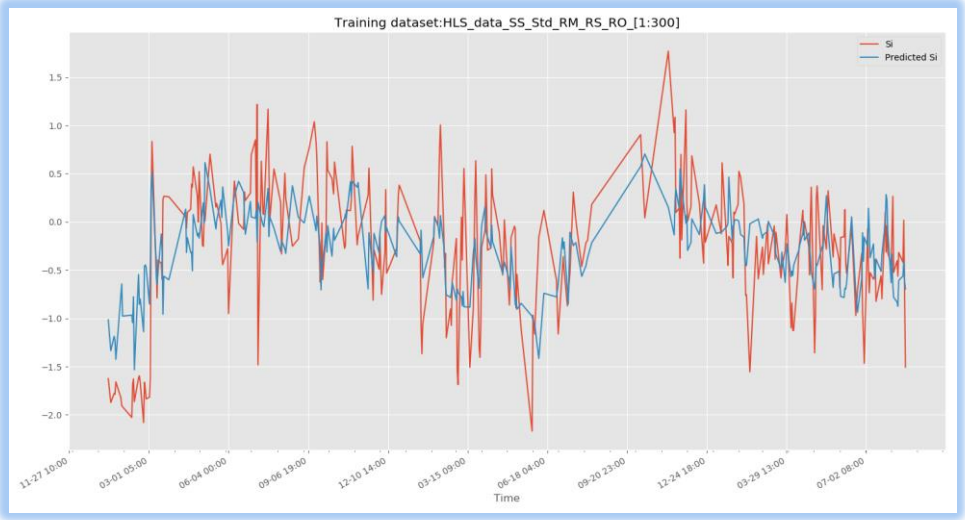
3

Options
Set **"Alpha"** as **"0.0005"**.

Soft Sensor Analytics

Modeling results – Lasso regression for Silica





LASSO

Available variables: TH, Alk, Si, pH, ln_Ca, ln_Mg, ln_TH, ln_ALK, ln_Silica, ln_pH, ln_Ba, ln_Sr, Li, Super_Ca

Input: ln_ALK, Super_Ca, Super_Alk, LIME, CAUSTIC_PPM

Output <<Alk>>

--> Modeling

☑ Intercept

☑ Normalize

Alpha 0.0001

Precision 0.001

Max iteration 1000

Detail

	Value
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	...

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

5

Save model

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

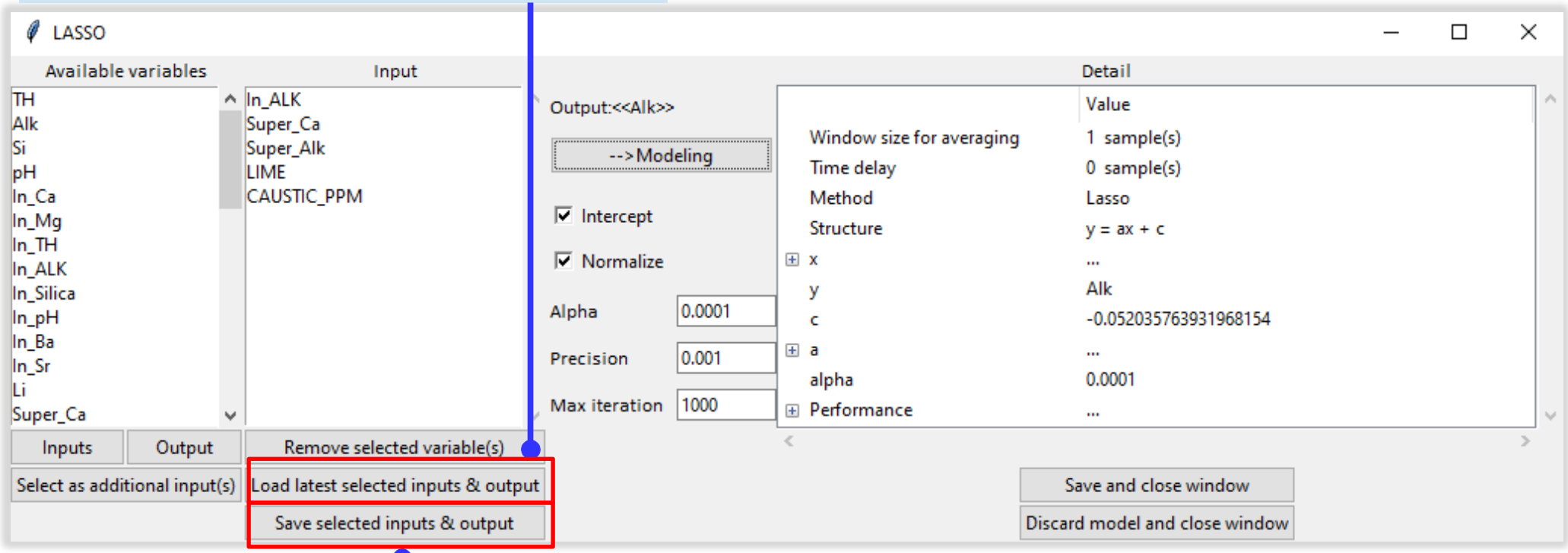
Soft Sensor Analytics

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



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**".

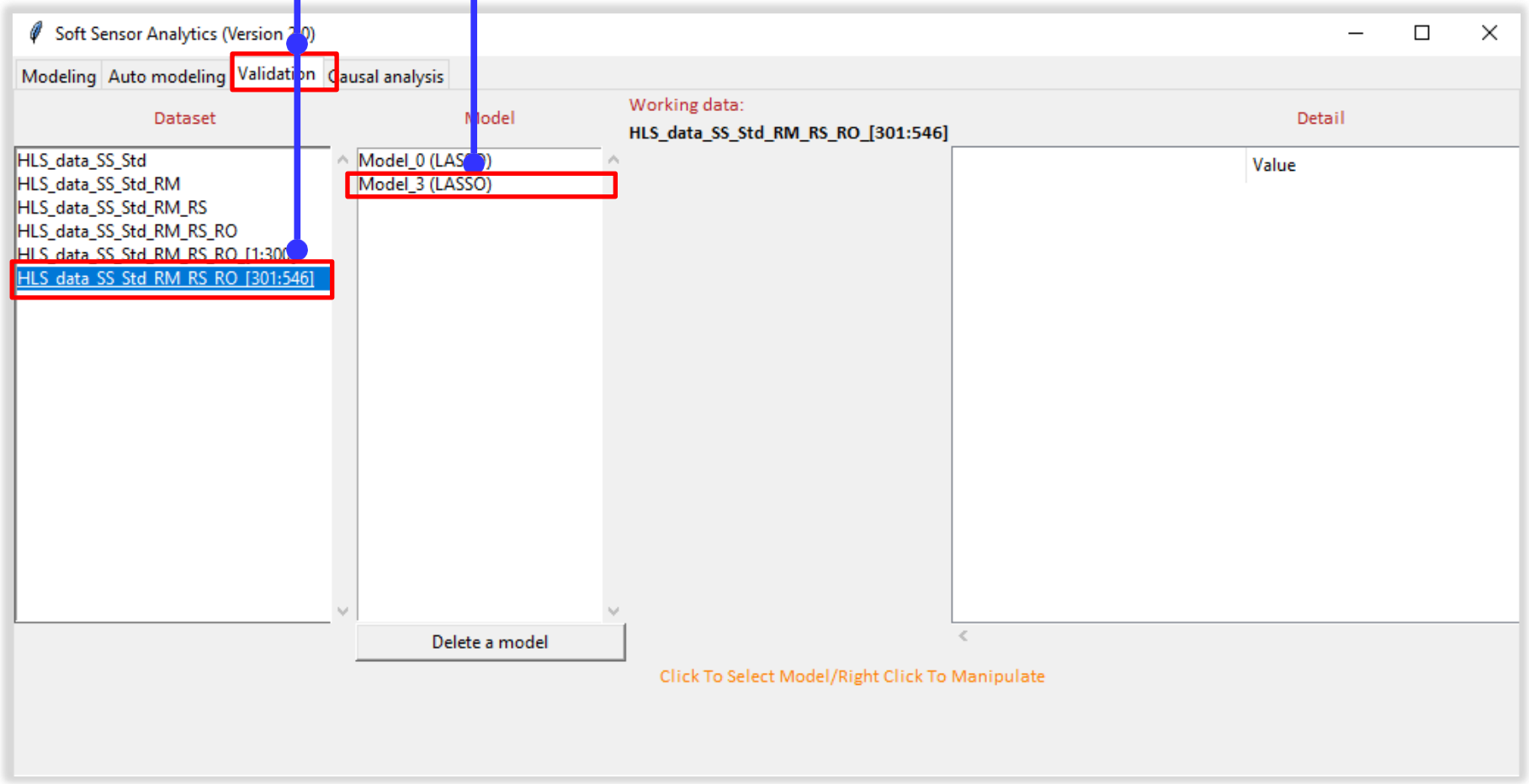
Soft Sensor Analytics

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

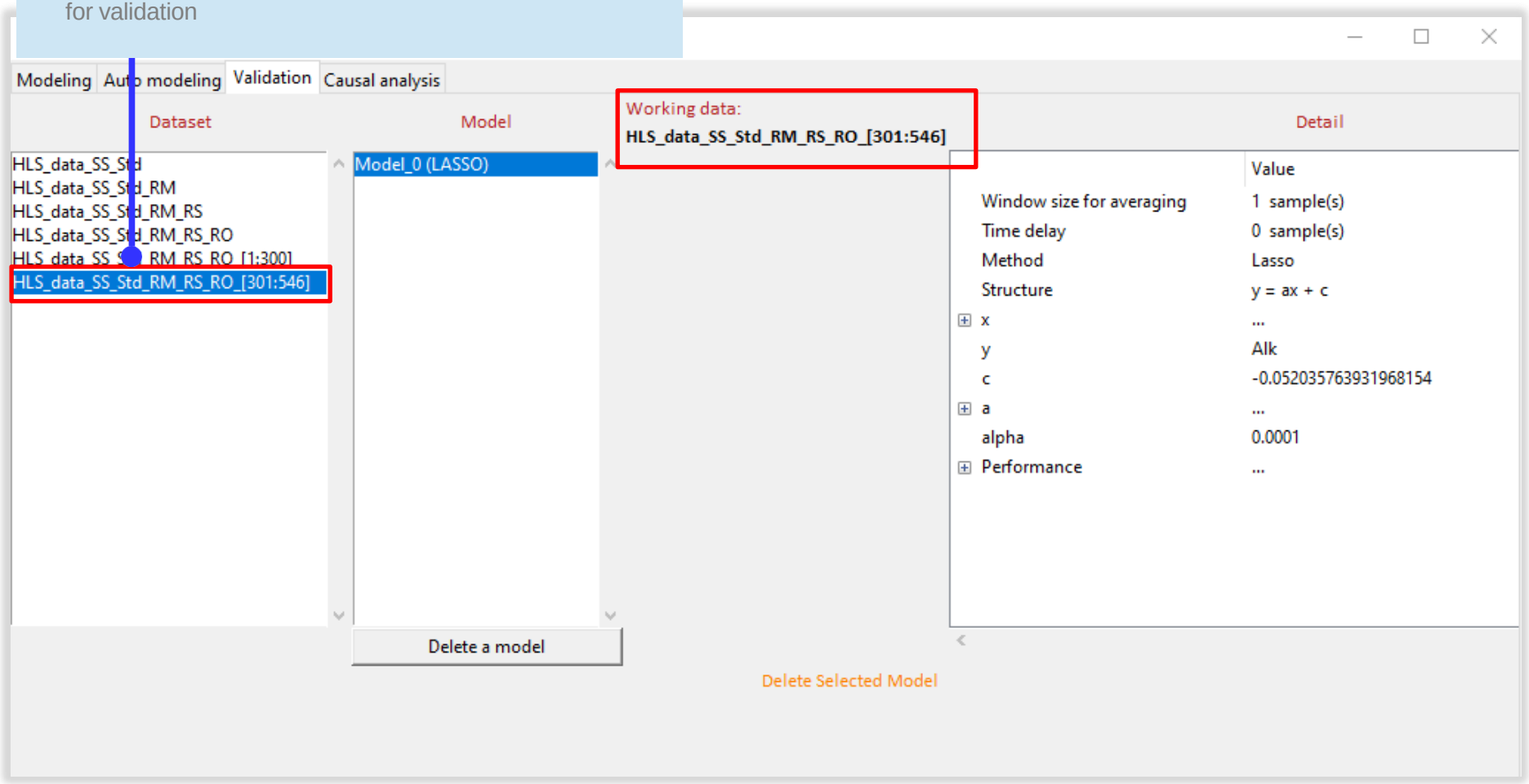
Validation of the model

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



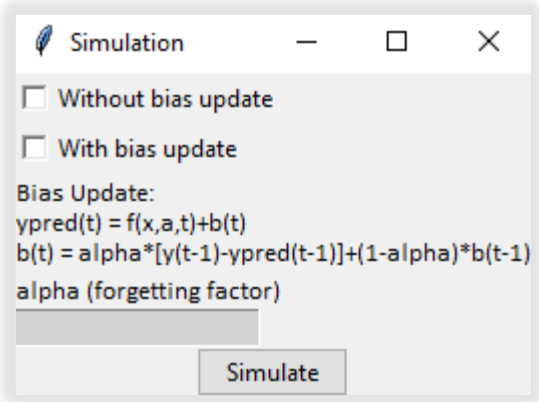
Soft Sensor Analytics

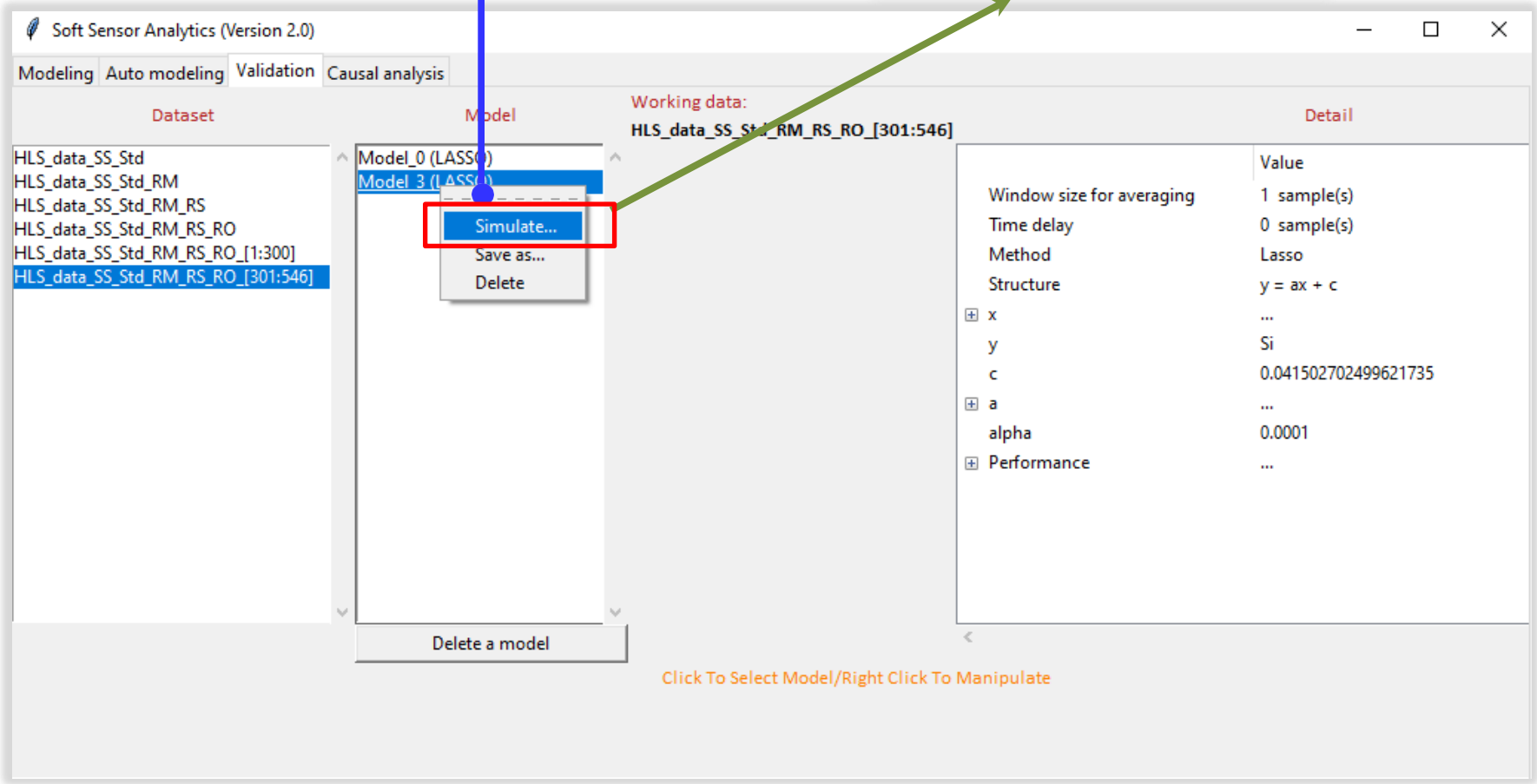
Validation of the model

3

Choose "Simulate" to continue

- **Right-Click** on the two models
- **Click "Simulate…"** button, a window will pop up





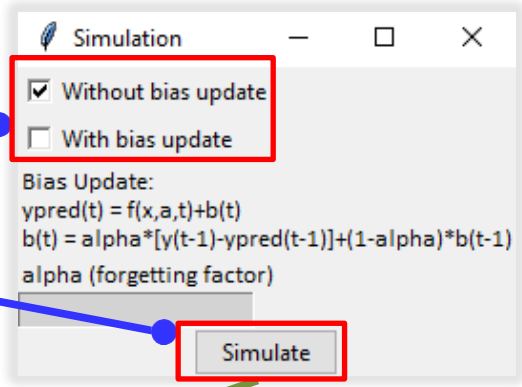
Soft Sensor Analytics

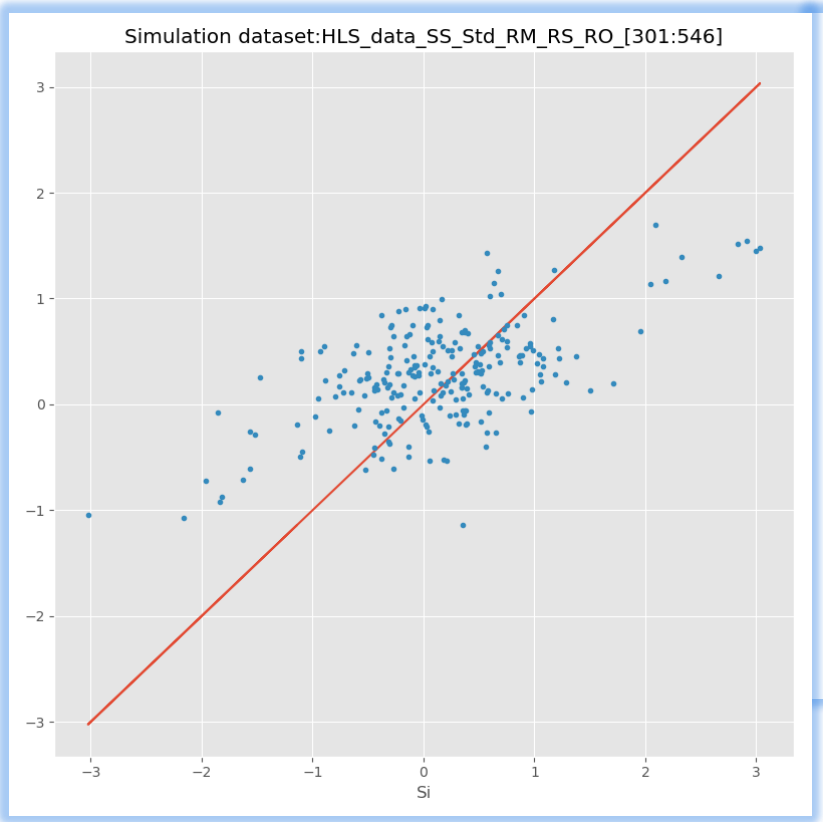
Validation of the models

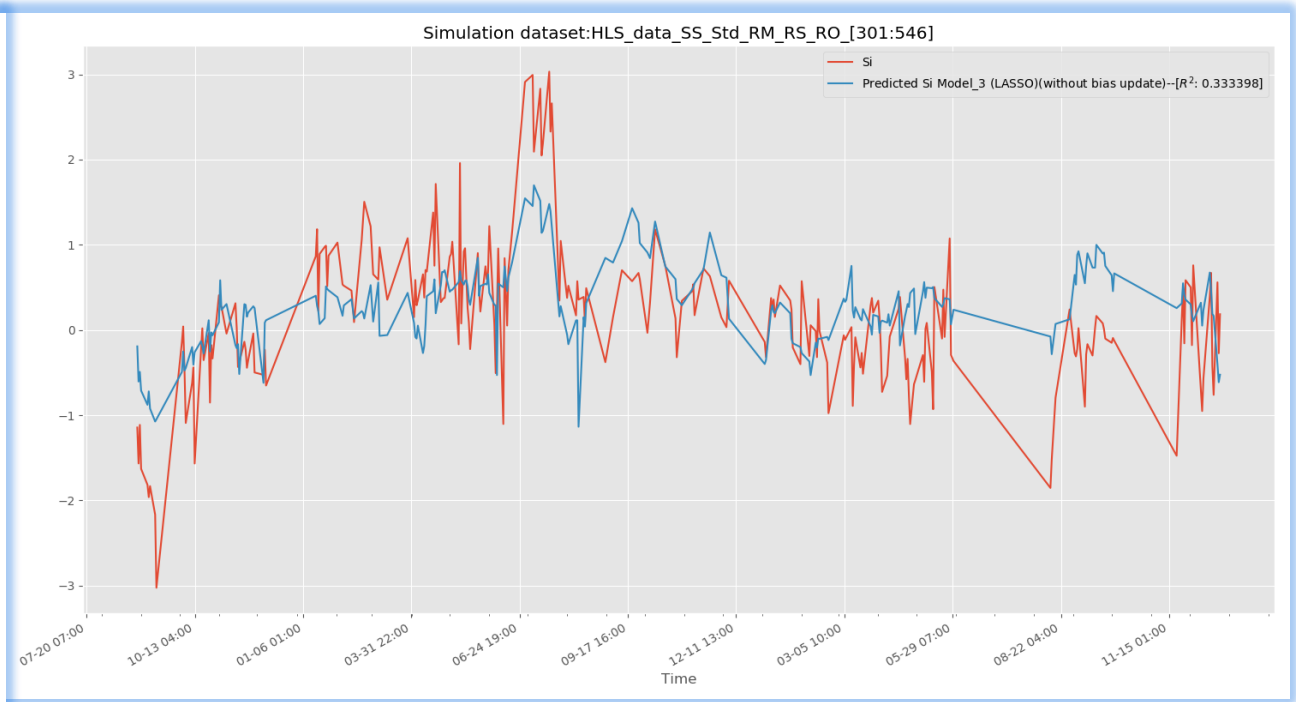
4

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