



RapidMiner – Time Series Functionality

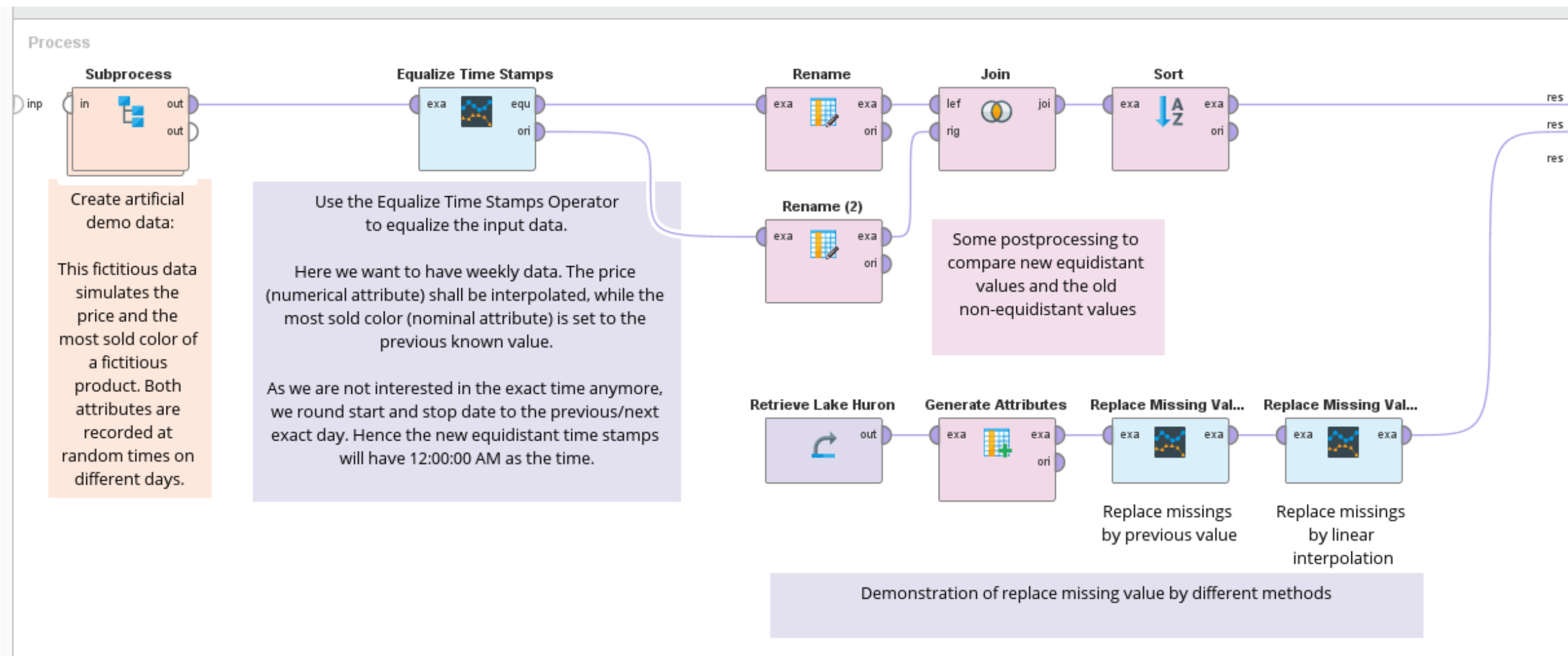


21 & 22 October 2020



Transformation

Equalize and Replace Missings





Transformation

Equalize and Replace Missings

7 Days

7 Days

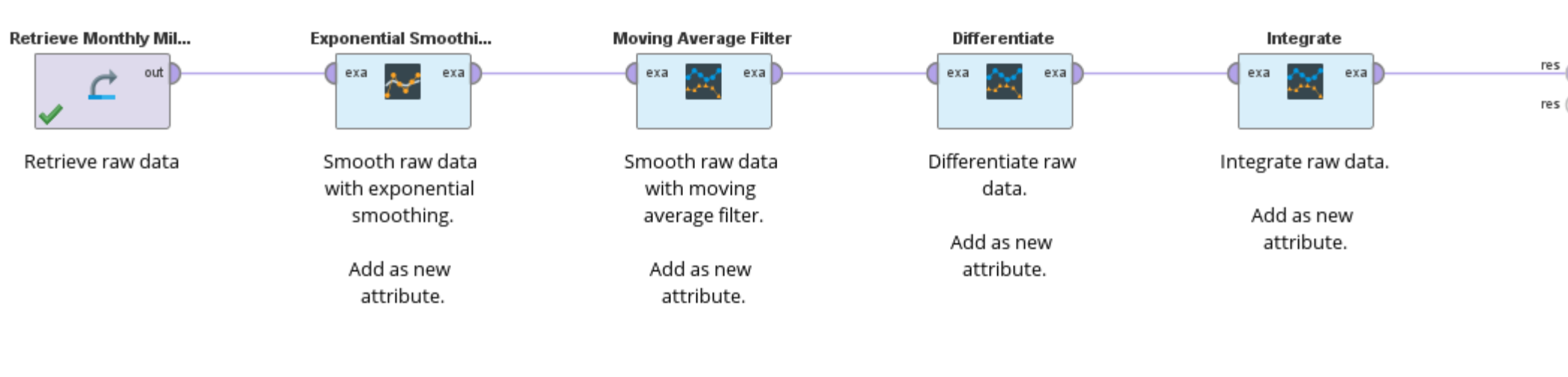
Date	Price (equalized)	Most sold color (equalized)	Price (original)	Most sold color (original)
Aug 2, 2019 12:00:00 AM CEST	108	blue	?	?
Aug 2, 2019 2:46:56 PM CEST	?	?	108	blue
Aug 7, 2019 8:45:56 PM CEST	?	?	139	green
Aug 9, 2019 12:00:00 AM CEST	141.079	green	?	?
Aug 12, 2019 5:34:56 AM CEST	?	?	147	red
Aug 14, 2019 12:22:56 PM CEST	?	?	185	green
Aug 16, 2019 12:00:00 AM CEST	182.909	green	?	?

Date	Lake surface level / feet	data with missing values	data with missing values replaced by previous value	data with missing values replaced by linear interpolation
Jul 1, 1875 2:00:00 AM CET	580.380	580.380	580.380	580.380
Jul 1, 1876 2:00:00 AM CET	581.860	581.860	581.860	581.860
Jul 1, 1877 2:00:00 AM CET	580.970	580.970	580.970	580.970
Jul 1, 1878 2:00:00 AM CET	580.800	?	580.970	580.380
Jul 1, 1879 2:00:00 AM CET	579.790	579.790	579.790	579.790



Transformation

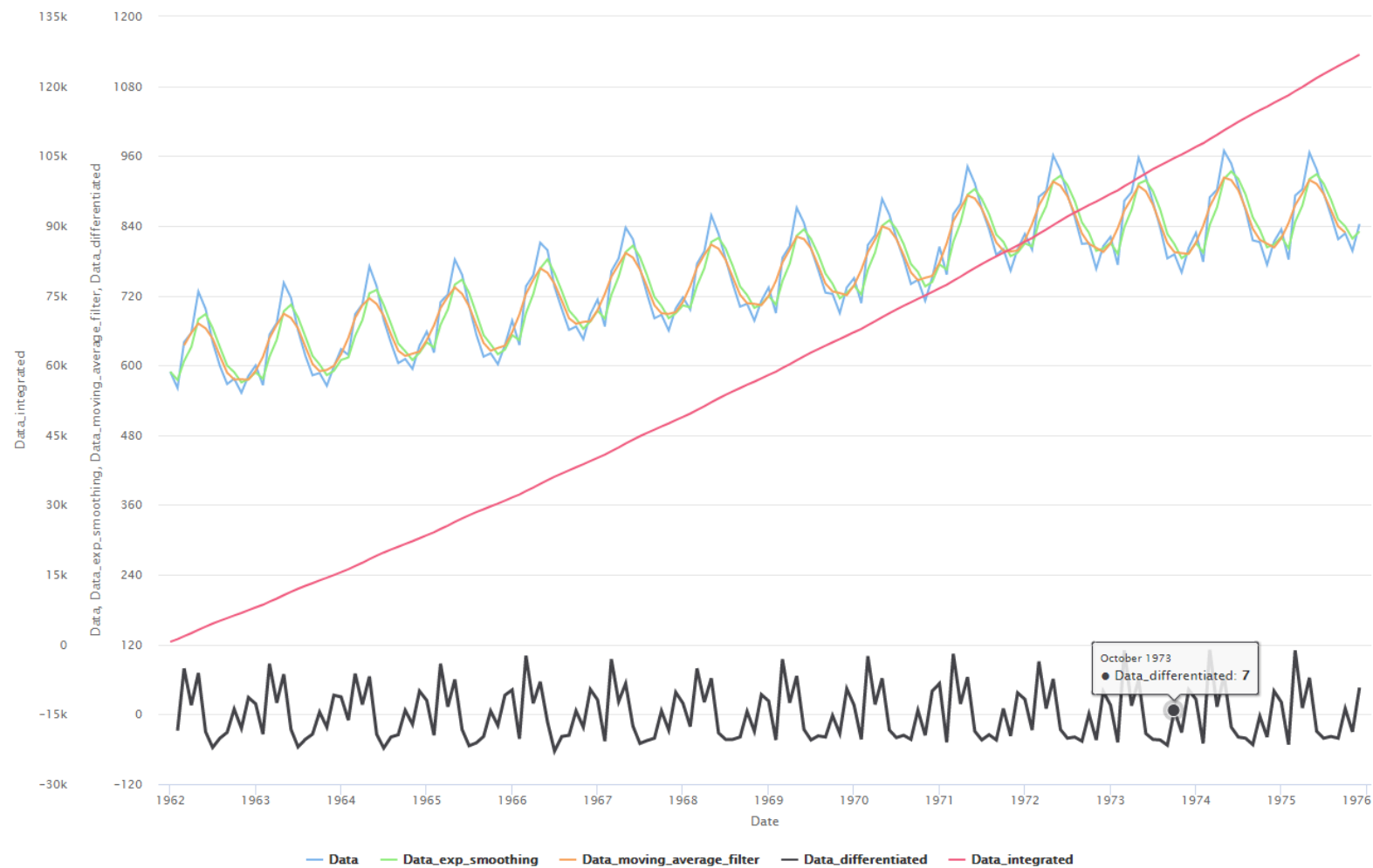
Differentiate, Integrate and Smooth





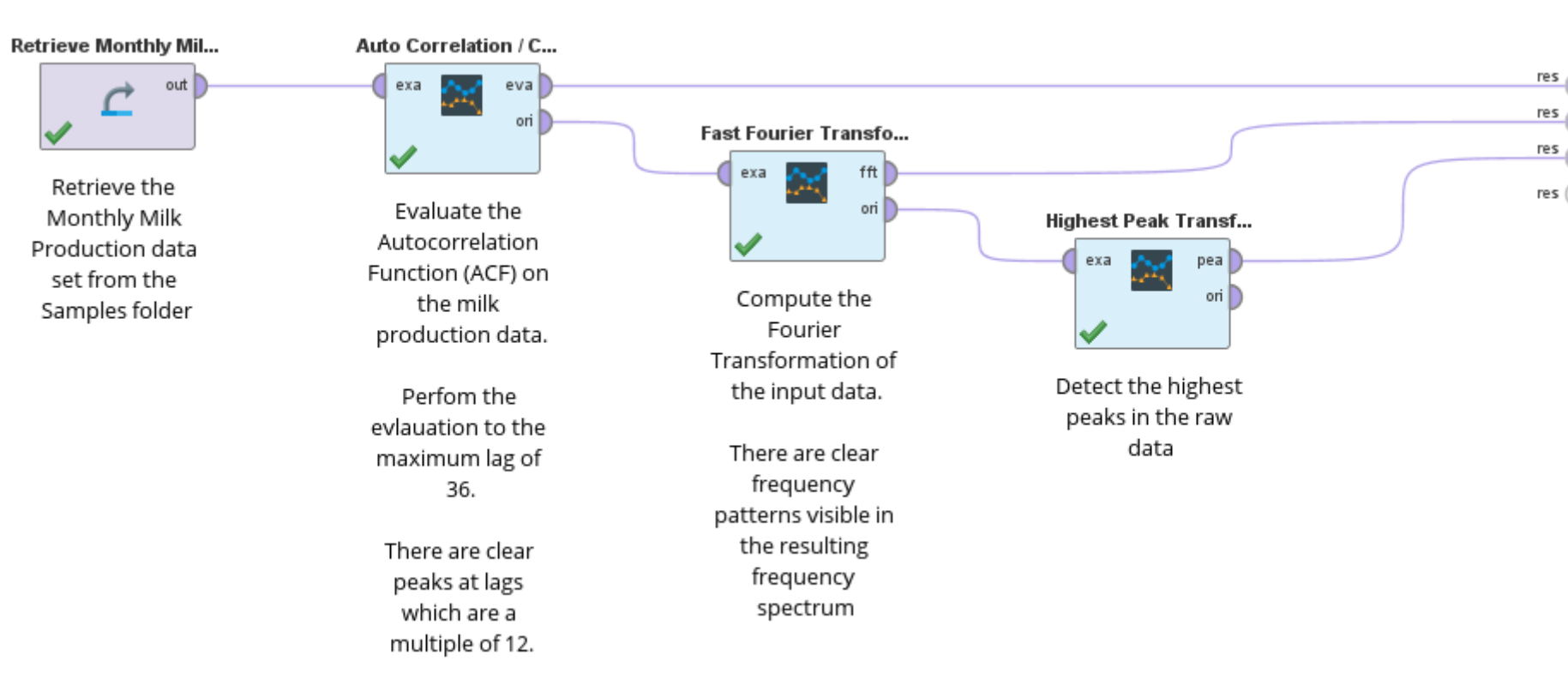
Transformation

Differentiate, Integrate and Smooth



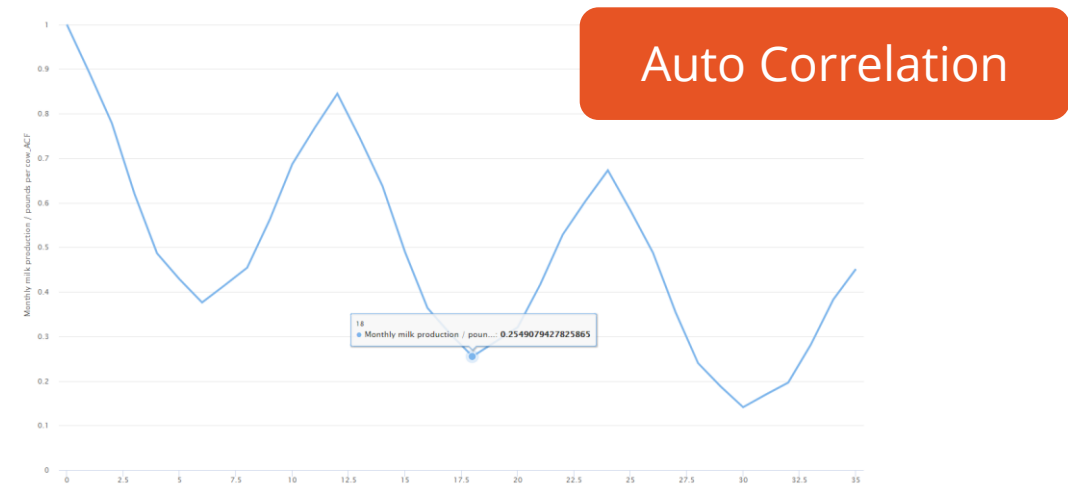
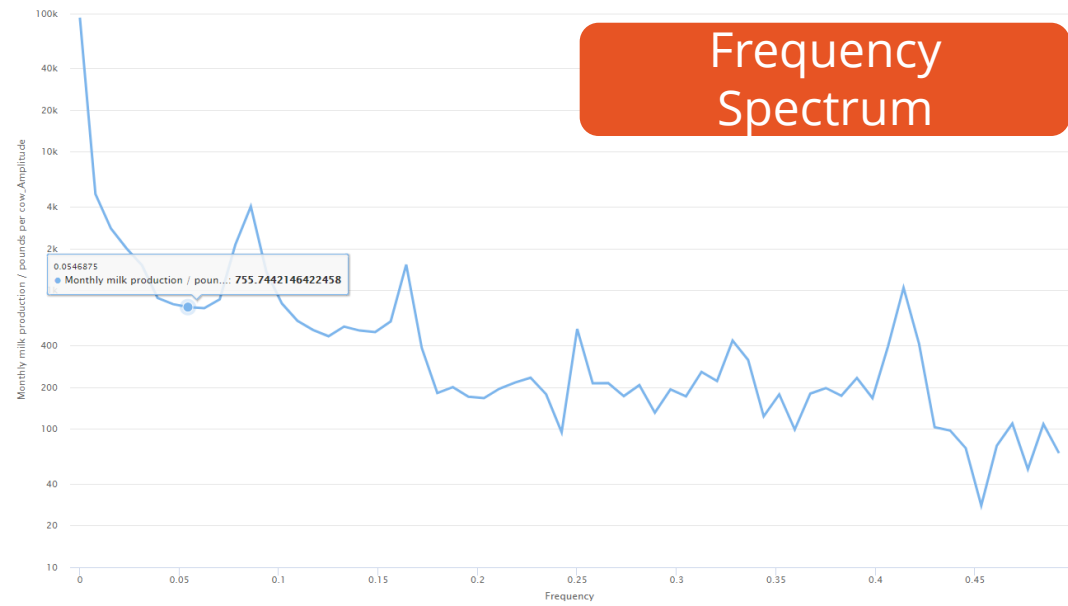
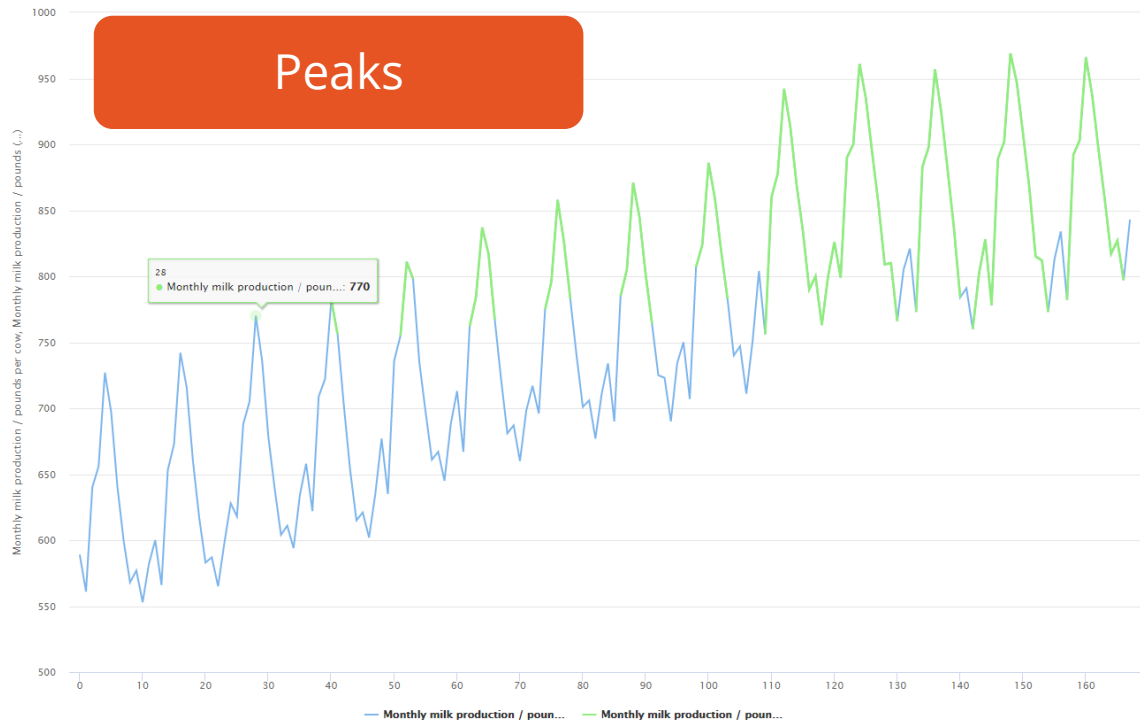


Transformation - Auto Correlation, Fourier Transformation & Peak Detection





Transformation - Auto Correlation, Fourier Transformation & Peak Detection

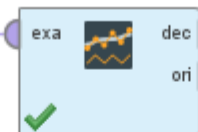


Decomposition Classic and STL

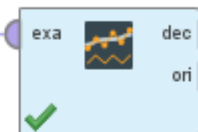
Retrieve Monthly Mil...



Classic Decompositi...



STL Decomposition



The Classic Decomposition operator is used to split the monthly milk production data into trend, seasonal and remainder component.

The seasonal and remainder component does not depend on the trend level so the 'additive' decomposition mode is used.

The seasonal pattern occurs every year and we have monthly data, so the seasonality is set to 12.

The STL Decomposition operator is used to split the monthly milk production data into trend, seasonal and remainder component.

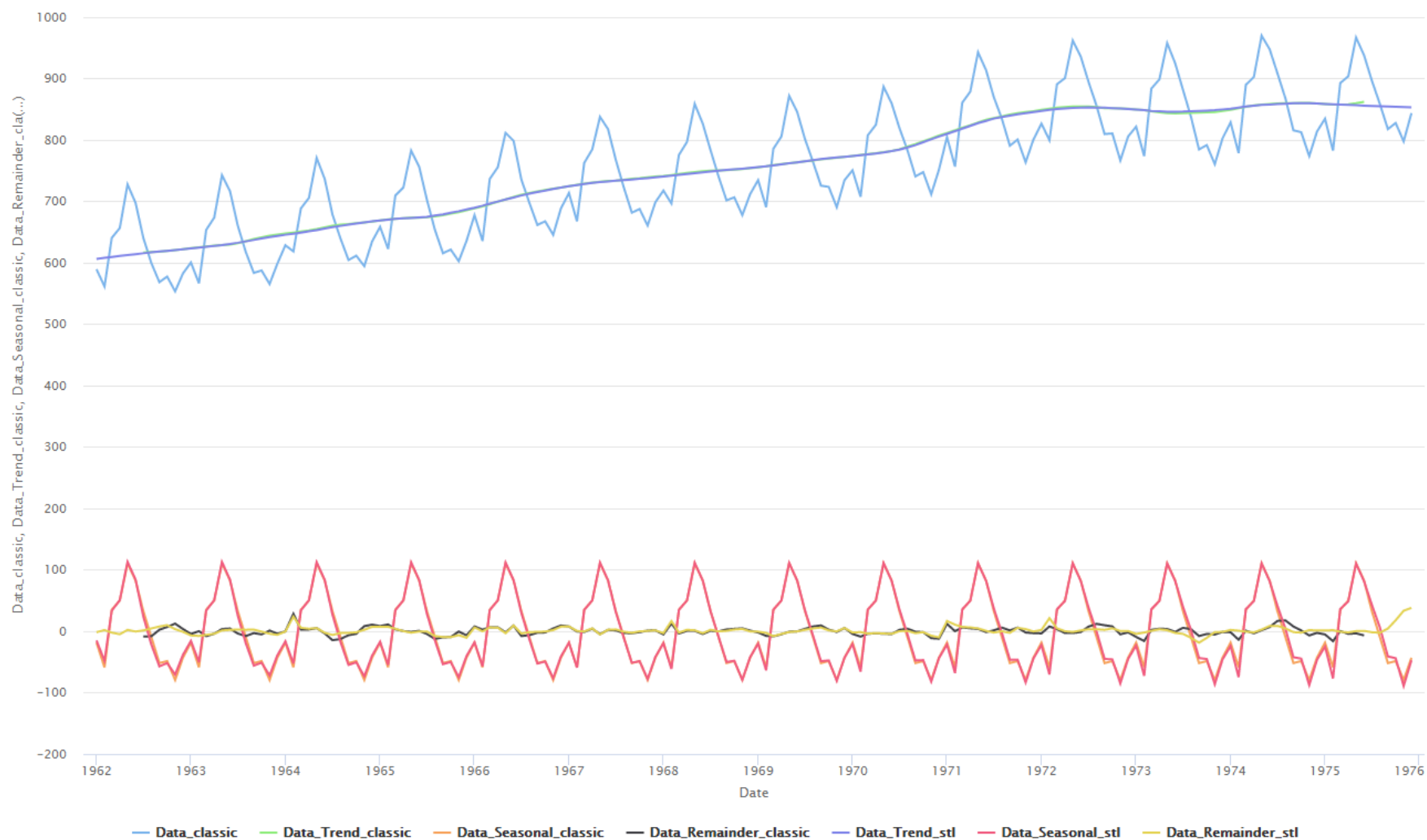
The seasonal pattern occurs every year and we have monthly data, so the seasonality is set to 12.

The seasonal width is set to 35 to allow for a only slowly changing seasonal components.

All other settings are set to the default values of the operator, which results in a appropriate decomposition of the time series

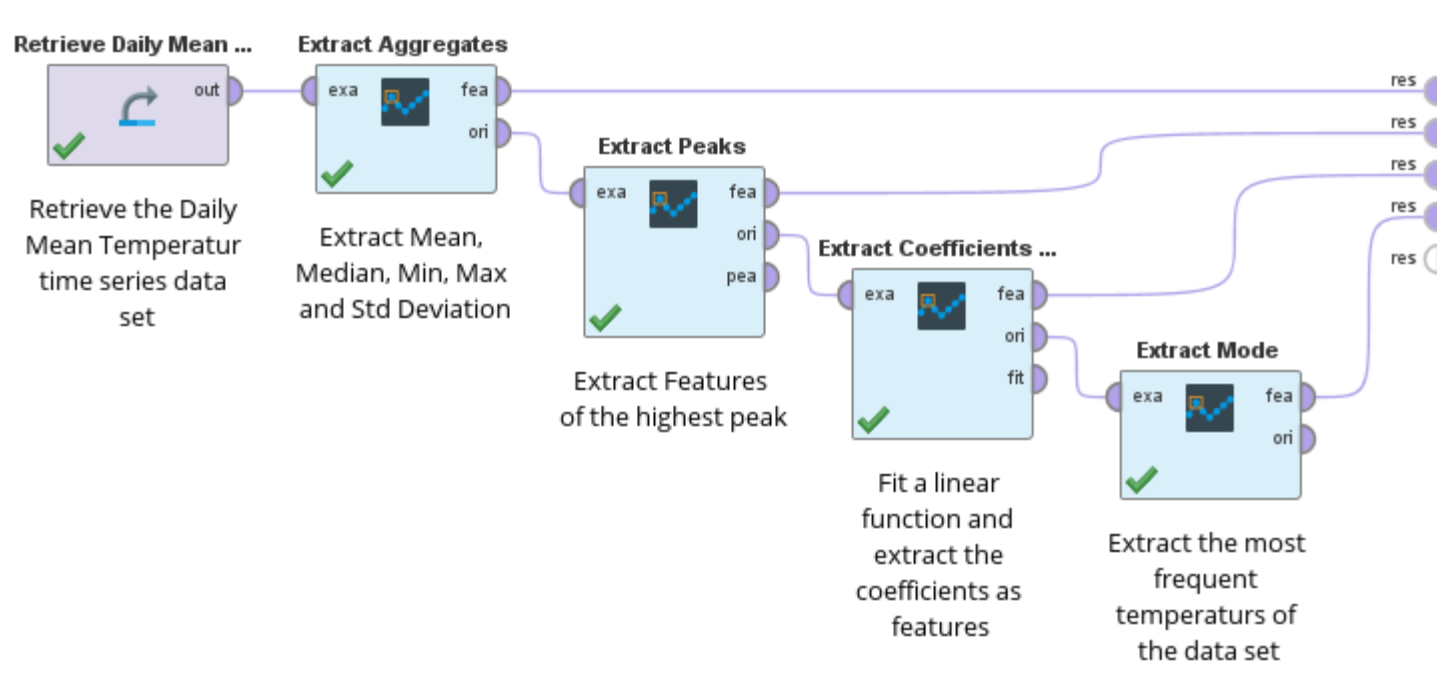


Decomposition Classic and STL





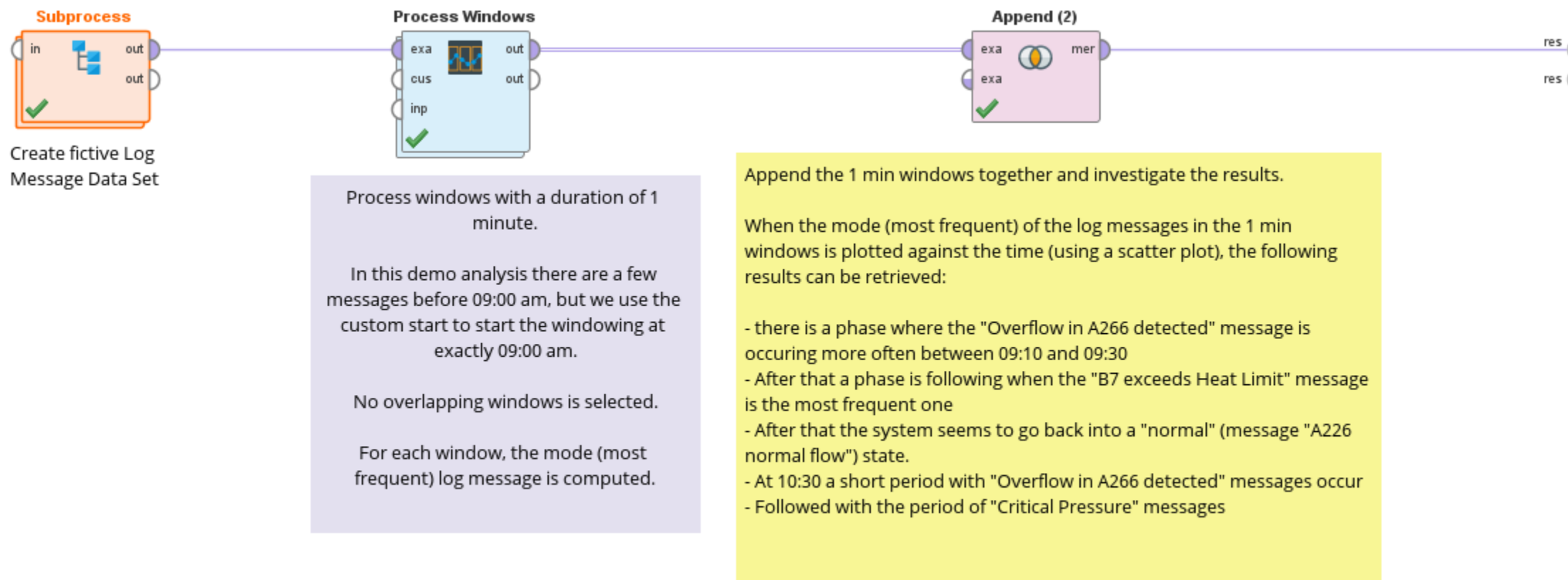
Feature Extraction



id	Daily Mean Temperature / C
mean	12.956433949349758
median	13.5
min	0.0
max	35.0
std deviation	7.8945899856608115
peak_1_extremum_position	398.0
peak_1_high_low_amplitude	30.0
peak_1_type	maximum
peak_1_value	32.5
peak_1_width	5.0
polynomial_fit_coeff_0	13.754624243088173
polynomial_fit_coeff_1	-0.0010934113612855022
polynomial_fit_discr	62.06923204629619
mode_order_1	13.5
mode_order_2	16.5
mode_order_3	14.0
mode_order_4_k_0	18.0
mode_order_4_k_1	5.0
mode_order_4_k_2	16.0



Windowing



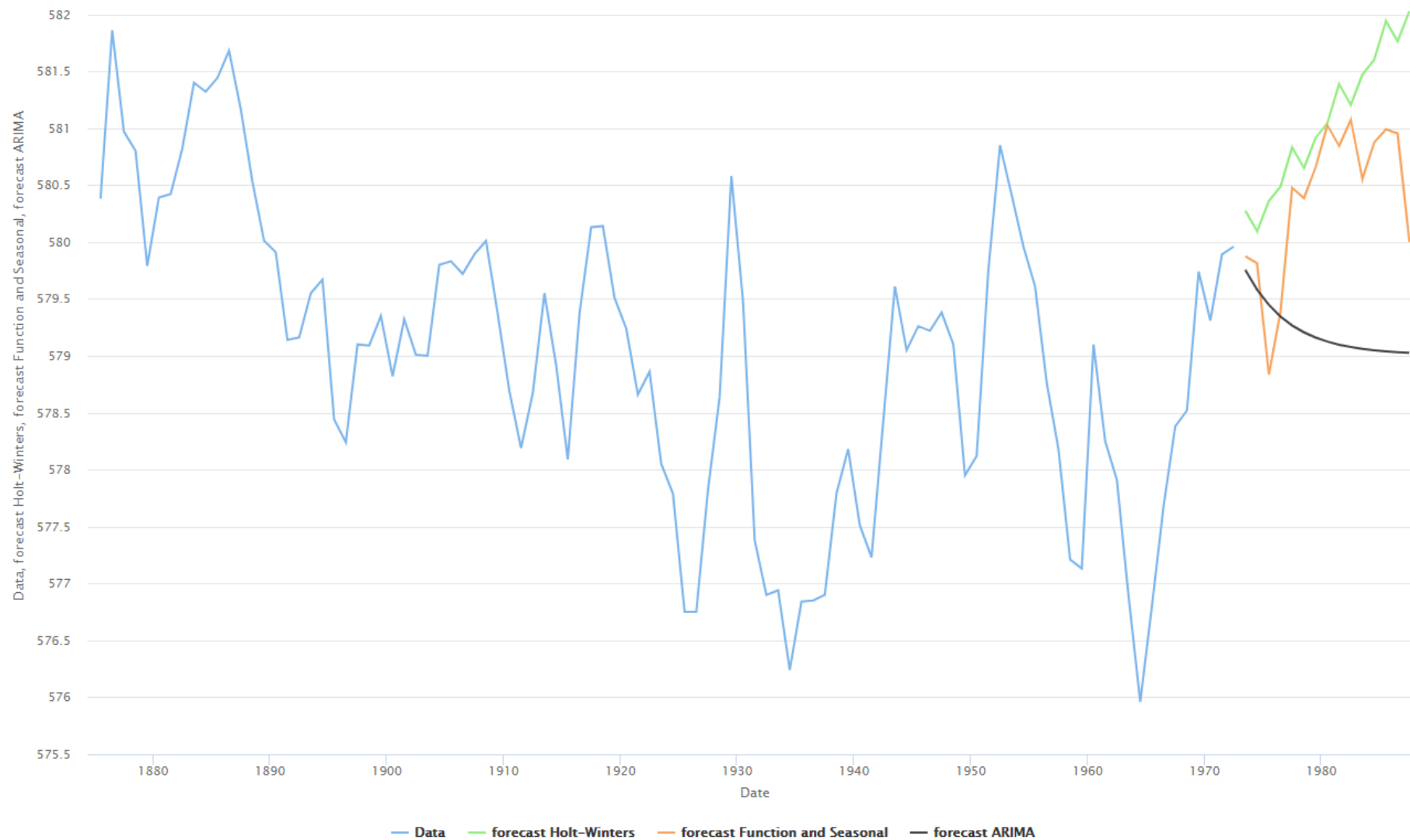


Windowing



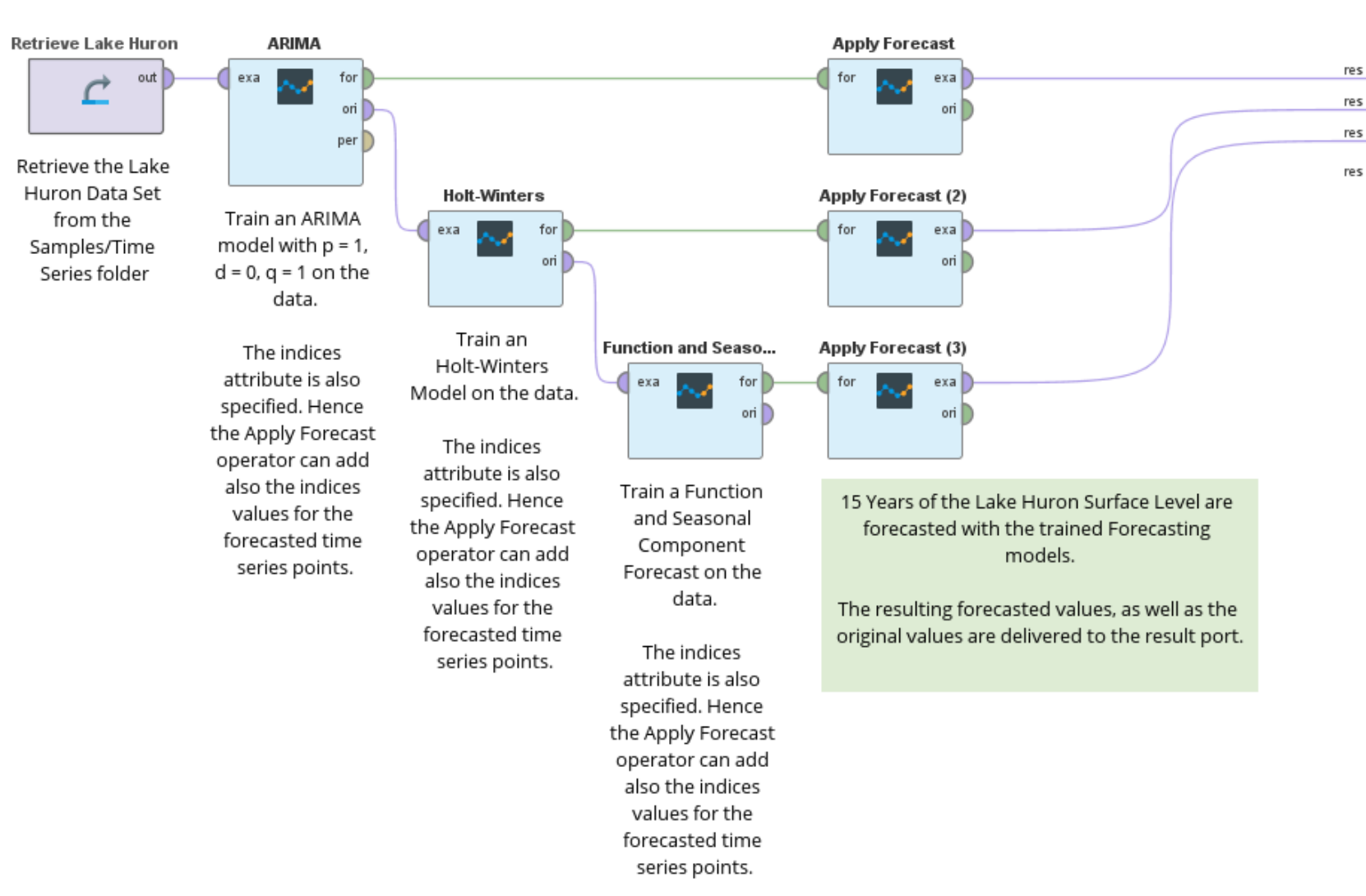


Forecasting





Forecasting





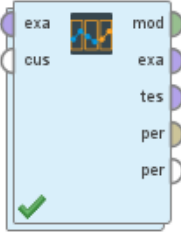
Validation

Retrieve Lake Huron



Retrieve the Lake Huron data set from the Samples/Time Series folder

Forecast Validation

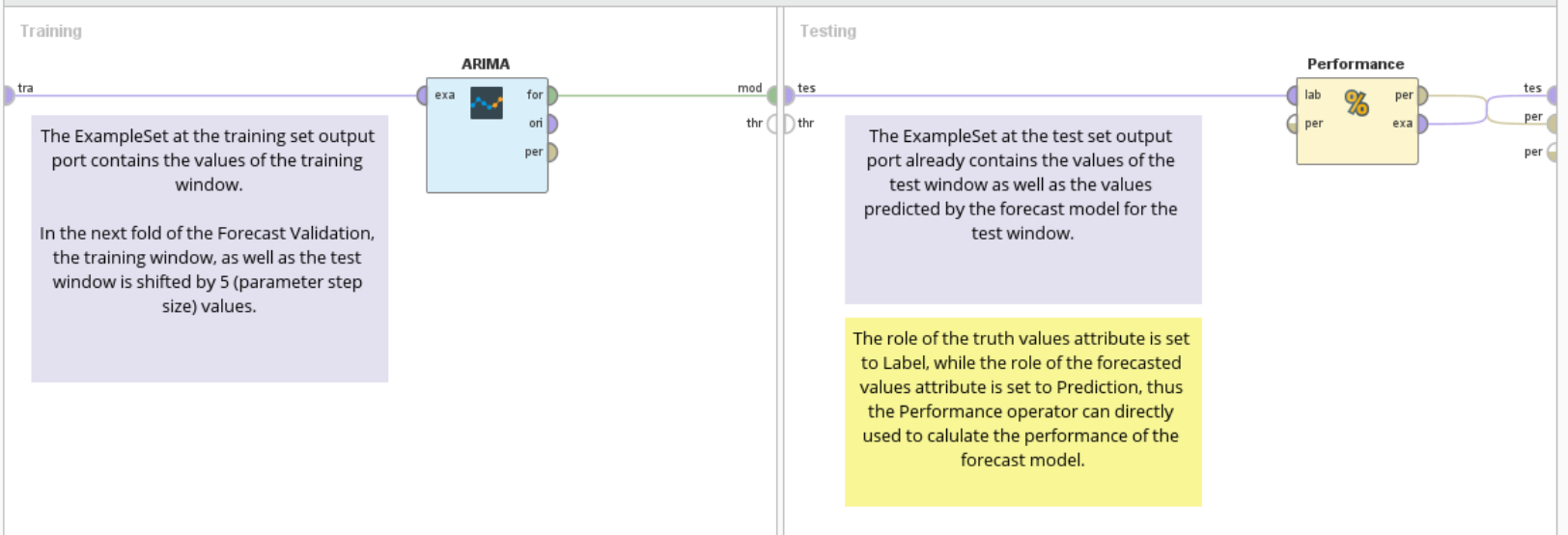


Perform a Forecast Validation to train an ARIMA model and validate its forecast performance on the consecutive values in the series (called the horizon).

The window size is set to 20, so the training window on which the ARIMA model is trained has 20 values in each validation step.

The horizon size is set to 5, so the next 5 values are predicted by the ARIMA model and compared to the true values in the Testing phase of each validation step.

For more details about the use of the operator, please have a look at the comments inside the Forecast Validation operator.





Validation

