

Scripts [#9]

This is a description of the content in the “Scripts” folder.

The scripts can perform the cleaning procedure which we applied on the dataset. The scripts are provided in Python language (Python v. 3.8.6) and the python-packages applied are numpy, pandas, bokeh, pickle, datetime and scipy.signal.

The following describes the aim of the scripts, the input data, and the output is and finally the methods applied.

Aim:

Make a flag where data is erroneous and make a timeseries without flagged errors.

Input:

Timeseries from Scada system or Danova in UTC0-time. Datetime does not need to be equidistant.

Two columns: ‘time’: YYYY-MM-DD HH:MM:SS

‘value’: depth [m] or flow [m³/s]

Output:

Two files containing:

A pickle file:

‘time’: YYYY-MM-DD HH:MM:SS (equidistant one-minute values)

‘level’: Corrected values with non-values

And a .csv file:

‘time’: YYYY-MM-DD HH:MM:SS (equidistant one-minute values)

‘raw_value’: mean “raw”-value of the aggregated one-minute values

‘value_no_errors’: only values without flagged error, else nan values

‘ffill’: Indication of forward filling values needed, because of iFix’ settings of data-saving functions

‘stamp’: Flag if value is stamped error (from iFix)

‘outbound’: Flag if data is out of bounds (according to config-file)

‘frozen’: Flag if data is frozen.

‘outlier’: Flag if there is a single outlier

‘frozen_high’: Flag if there is a frozen sensor above a given height.

‘depth_s’: Depth of sensors with no errors. Interpolated values

‘level’: Level of the corresponding depth_s.

Method:

Raise flags:

Stamped error: If the SCADA system gives a poor-quality stamp to the data value.

Man remove id: Manual remove data by indicating in .csv file the period of manually delete data.

Outbound id: If data is outside accepted area. Criteria set by standard values and manually edited data.

Frozen id: If values are the same in a given period, it is flagged as a frozen sensor. The value is given individually. A window size determine if the row of equally values are frozen

Outlier id: Removes if there is a spike in values with threshold (height) = 0.2 m and 1 value in width (meaning the level cannot increase and drop suddenly).

The methods are applied like this:

- Scada
 - Forward fill data savings point (get one-minute values) (only ifix)
 - Stamped error (only ifix)
 - Manual remove
 - Out of bound
 - Frozen
- Danova
 - Out of bound
 - Manual remove
 - Frozen
 - Outlier

When every potential error is identified and flagged, a new timeseries is calculated, where data is aggregated as mean value for one-minute values. The value_no_errors timeseries is forward filled when necessary and interpolated for empty timestamps with a max interpolation of 5 timesteps (5 minutes).

The usages of these data are on own responsibility. VCS Denmark and DTU cannot be held responsible.