Tutorial: Using GAQ-qc to inspect recent trace gas time series

This document provides a step-by-step guide on how to use the GAW-qc dashboard for inspecting recent time series (up to 12 months) from a GAW station. Additional information about GAW-qc can be found here.

The beta version of GAW-qc is currently available at https://app-503-gaw-dev.azurewebsites.net/.

Contents

1. Prepare the data

GAW-qc requires **hourly averages** of **CH_4**, **CO**, **CO_2**, or **O_3** mole fractions. Units must be **ppm** for CO_2 and **ppb** for the other species. To calculate hourly averages you can follow the same procedure adopted for submitting data to the GAW World Data Centres.

The **file format** for the input data can be either **csv** (comma- or semicolon-separated values) or **Excel** (.xls or .xlsx). Only one species (at one height) can be analyzed at a time, meaning that the file will only have **two columns**: time (beginning of the 1-hour interval) and value (hourly average) for one species.

The **time format** should ideally be *yyyy-mm-dd hh:mm:ss*. Other similar date formats (e.g. yyyy/mm/dd or dd-mm-yyyy) are also supported. If the time format is not supported, or if it is not consistent throughout the file, you will get a specific error message. **Local time** can be used if preferred. The time should indicate the beginning of the interval (e.g., 12:00:00 indicates the interval between 12:00 and 13:00).

NB: If you use Excel for file preparation (not recommended!): Excel will automatically change the format of your time column according to your system settings. To avoid this, select the entire column and set the cell format to text (right-click -> Format Cells -> Category: Text)

Missing values can be omitted or indicated by a negative number or any non-numeric code.

Here is an example:

	A	B	C	D	E	F
1	Time (UTC)	CH4 (ppb)				
2	2024-01-01 00:00:00	2013.01				
3	2024-01-01 01:00:00	2013.19				
4	2024-01-01 02:00:00	2014.06				
5	2024-01-01 03:00:00	2013.61				
6	2024-01-01 04:00:00	2014.31				
7	2024-01-01 05:00:00	2015.18				
8	2024-01-01 06:00:00	2017.16				
9	2024-01-01 07:00:00	-999.99				
10	2024-01-01 08:00:00	2014.98				

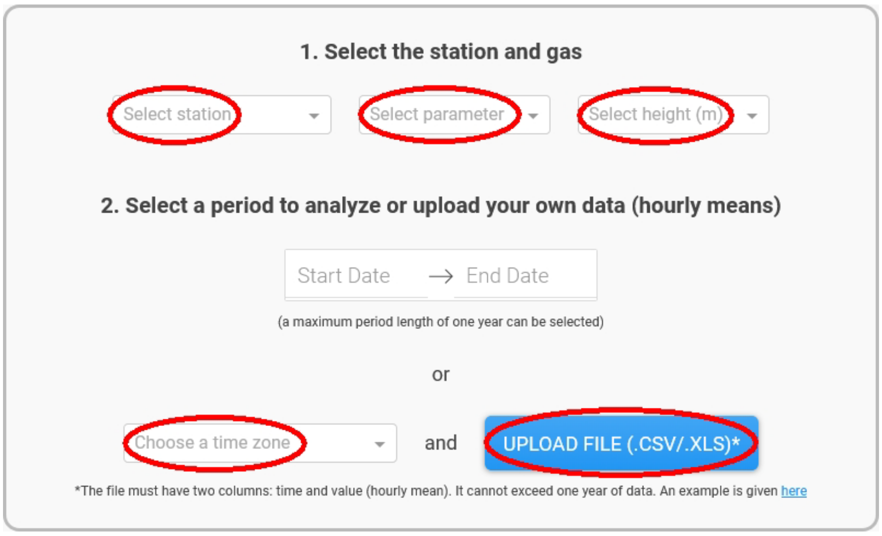
And the same example as csv file:

```
Time (UTC),CH4 (ppb)
2024-01-01 00:00:00,2013.01
2024-01-01 01:00:00,2013.19
2024-01-01 02:00:00,2014.06
2024-01-01 03:00:00,2013.61
2024-01-01 04:00:00,2014.31
2024-01-01 05:00:00,2015.18
2024-01-01 06:00:00,2017.16
2024-01-01 07:00:00,-999.99
2024-01-01 08:00:00,2014.98
```

NB: The first row (header) is ignored by GAW-qc. Nevertheless, it is always good practice to use informative column names. Additional columns (other than the first two) are also ignored.

2. Upload the data

Open the GAW-qc webpage (see above). At the bottom of the page, you will find the input form:



1. Select your station from the list
2. Select the gas that you want to analyze
3. Select the height of the inlet
4. Choose the time zone that you use (if left empty, UTC is assumed)
5. Click on UPLOAD FILE and select the file that you prepared

If you have not prepared a file yet, you can fill in *Start Date* and *End Date* instead. These are used to inspect historical data that are already archived in the World Data Centres and allow you to familiarize with the dashboard before you use it on new data.

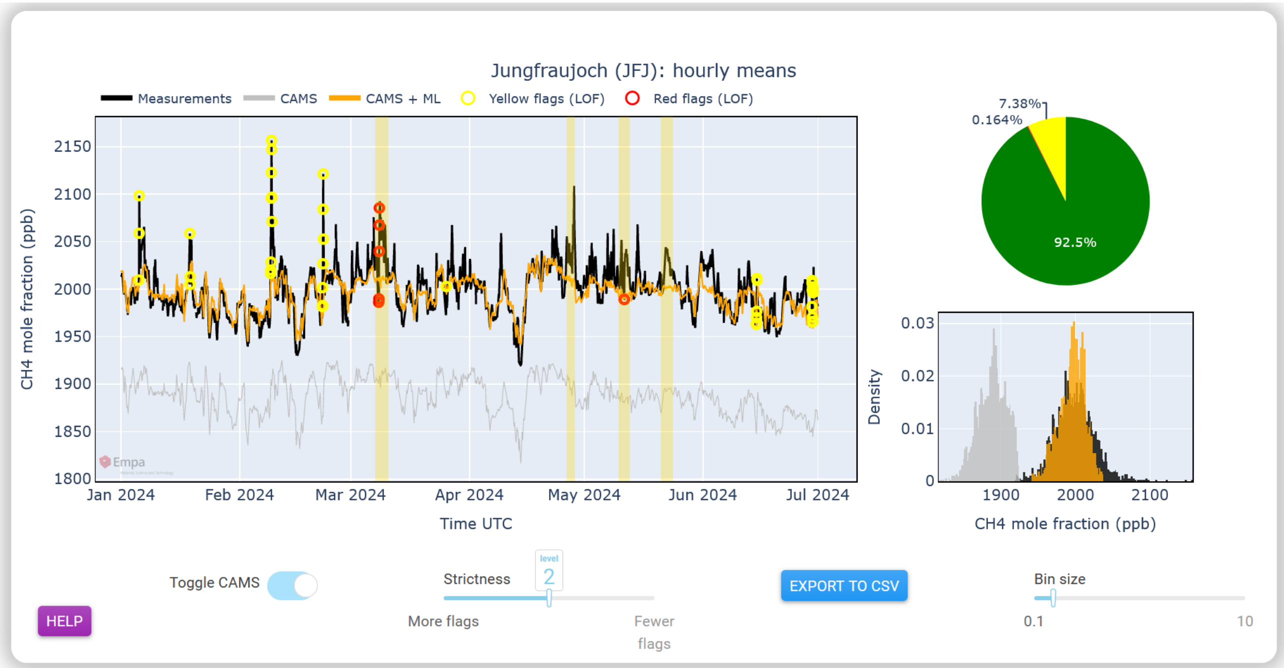
The dashboard will usually take between 5-30 seconds to load, depending on the station and parameter. Try again if nothing happens after one minute or so.

> **NB**: The dashboard will appear at the bottom of the page. Depending on the resolution of your screen it may be necessary to scroll down to see it!

3. Analyze the data

3.1 Hourly plot

In the following we analyze data for CH_4 at Jungfraujoch (JFJ) for the period from January to June 2024. The first panel of the dashboard shows the time series of hourly averages:



The **measurements** are shown in black, while the orange curve gives the **expected values** (based on numerical forecasts by Copernicus). The expected values are optimized CAMS forecasts using comparisons of historic observations, CAMS original forecasts and machine learning. This allows a downscaling from the gridded CAMS output to the station location, including correction for systematic biases. The original CAMS forecast (gray line) is only shown for the sake of completeness. Different time series can be toggled by clicking on the legend on the top. Large deviations of the black curve from the orange one are highlighted by the **yellow areas**. These are the periods that deserve closer attention (for example, through an inspection of the station logbook).

In addition, individual measurements are flagged by the **yellow** and **red circles**. These are **statistical outliers** that should also be investigated. Red flags are generally of higher concern, either because they have a larger magnitude or (as in this case) because they overlap with a yellow area.

> **NB:** You can increase or decrease the number of flags and yellow areas by acting on the *Strictness* slide. This can be useful if you feel that the algorithm is flagging too much (or not enough).

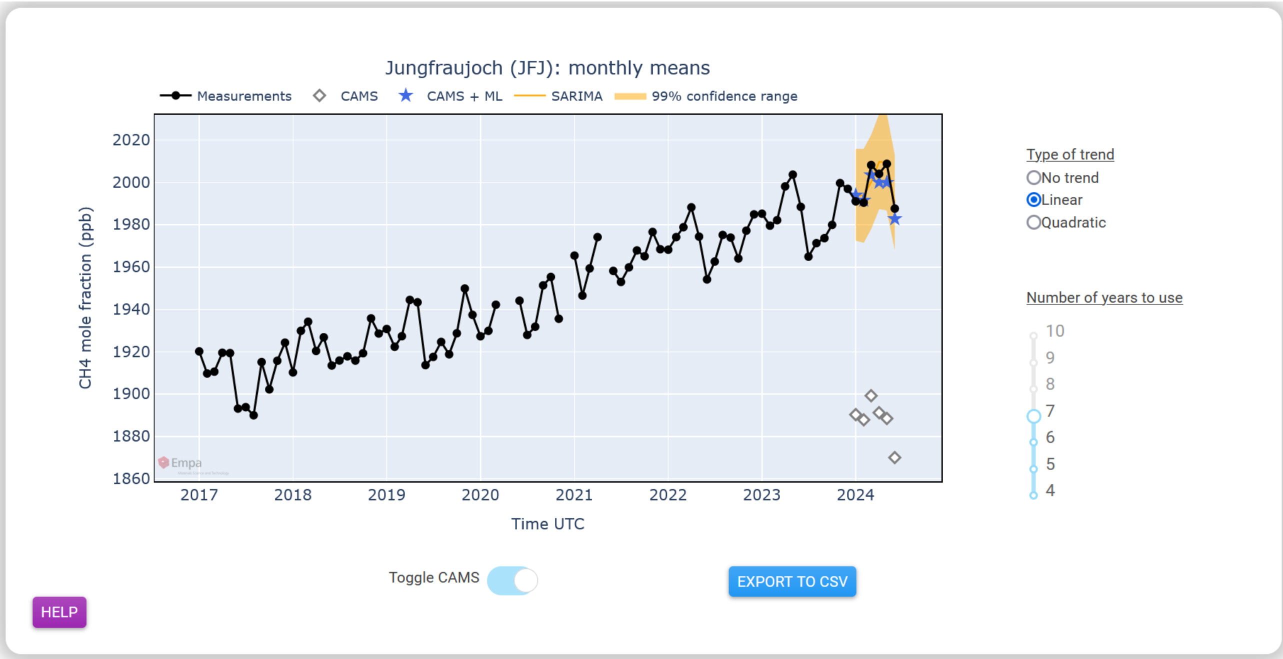
The **pie chart** on the top right corner shows the fraction of measurement that are either in a yellow area or have a yellow flag (7.38%) and those that have a red flag (0.164%). In this example, 92.5% of measurements are considered not suspicious (green).

The **histogram** on the bottom right corner shows the distribution of the three time series (same color code as in the main plot). The expected values (orange) have typically the tendency to underestimate extremes, as it is also clear in this example.

> **NB:** The accuracy of the expected values can vary considerably from station to station, depending on the performance of the CAMS forecasts and of the machine learning model. The worse the performance, the less meaningful the yellow areas will be.
> **The expected values are currently not available for CO_2.**

3.2 Monthly plot

The second panel of the dashboard shows the time series of monthly averages:



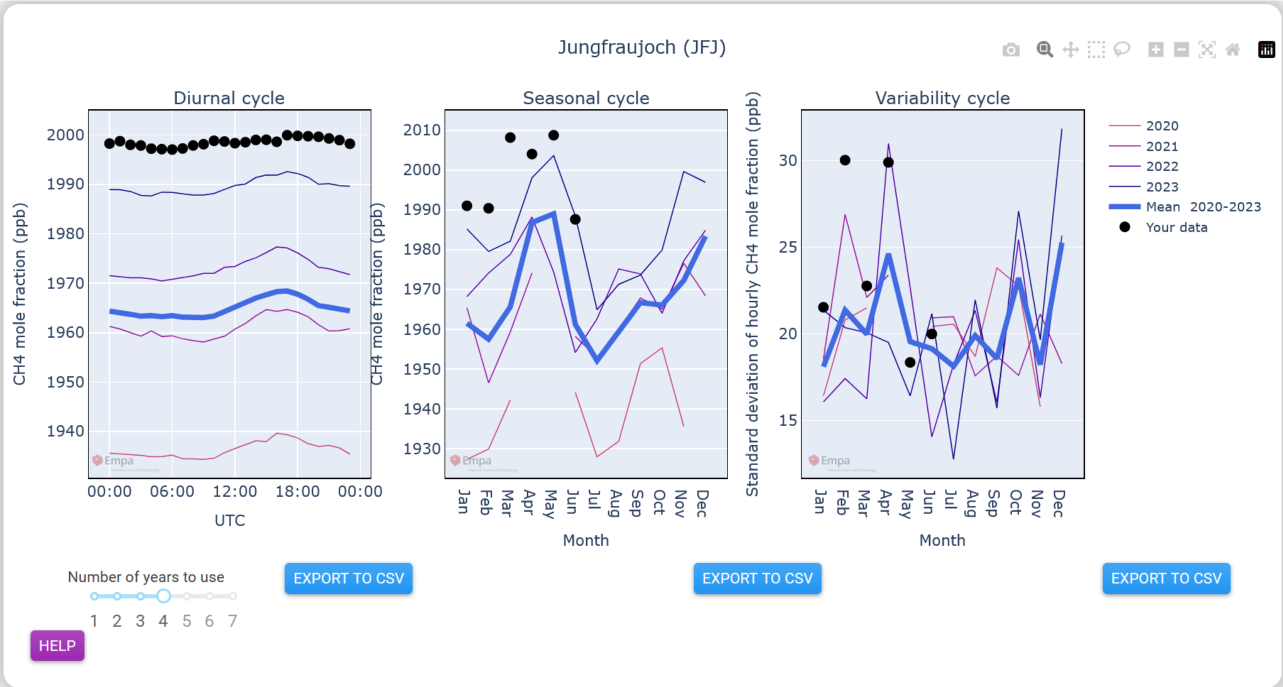
Here we also see historical data up to 10 years in the past (set by the vertical slide on the bottom-right corner). Those data are used to make the SARIMA-based statistical prediction shown in **orange**. The debiased numerical (CAMS-based) prediction is indicated by the **blue stars**.

In this example, the measurements (black) are within the orange shading and very close to the blue stars, meaning that data at monthly scale do not raise any concern. If one month would have been outside the confidence range of the statistical prediction, it would have been flagged with a **red circle**.

NB: If the trend in the historical data is clearly different from linear, or if there is no clear trend, you may get a better statistical prediction by selecting a different type of trend on the top-right corner.

3.3 Visual comparisons

The third and last panel of the dashboard shows three plots where the data for the analyzed period are compared with previous years (up to seven):



The **diurnal cycle** shows the mean variability of the gas concentration during a day in the uploaded months (January to June). In the example, the maximum in the afternoon is less pronounced in 2024 (black dots) than in previous years, something that might be worth investigating. The curves are shifted vertically because of the underlying trend.

The **seasonal cycle** shows the monthly averages of each year. In the example, the maximum in spring appears consistently in all years. Again, the curves are shifted because of the trend.

The **variability cycle** shows the standard deviation of the hourly averages for each month and each year. It can be useful to spot periods where data have too little variability or are too noisy. In the example, the variability in February 2024 is higher than in previous years, arguably because of the peaks in the hourly plot.

NB: You can export the data (and the flags) from each plot of the dashboard to a csv file by using the respective *EXPORT TO CSV* button. You can also save the plots as png by clicking on the camera icon appearing on the top-right corner of each panel.