

Empa.QATool



End-User Documentation

Version 25, February 2021

Quality Assurance tool

Render as: Stack plot Scatter plot
 Use separate Y-Axis per substance Render precision and accuracy plot

Select timeseries: **+** **-**

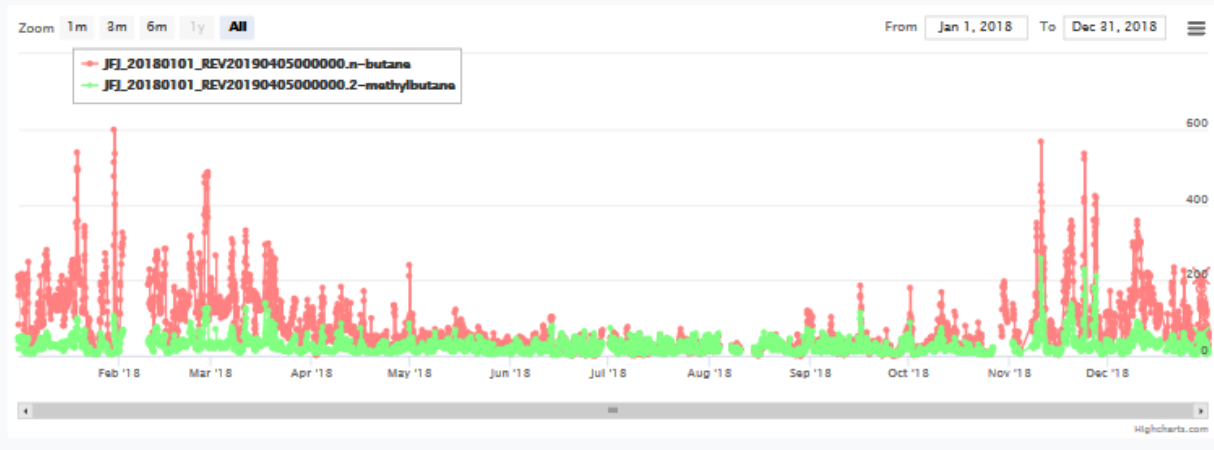
Timeseries	Station	Dataset	Revision	Substance	Color
#1	JFJ	1.1.2018	5.4.2019	n-butane	
#2	JFJ	1.1.2018	5.4.2019	2-methylbutane	

Render Plots

Actions

Flag selected points

Flag date range



Contents

Introduction	3
What is Empa QATool?	3
What can Empa QATool do for me?	3
Sounds great, how can I access it?.....	3
User login	4
Start page.....	5
QATool	6
Plotting area.....	6
Miscellaneous symbols	10
Zooming	10
Tooltips.....	11
Selecting data.....	11
Actions area	12
Flag selected points.....	12
Flag data range.....	14
Set comment on selected point.....	15
Quick start.....	16
Import Data.....	16
CSV Format.....	17
Mandatory part.....	17
Optional part.....	17
Examples	18
1. CO ₂ and NO _x data without end-time.....	18
2. data with only mandatory fields, containing an end time	18
File name.....	18
Compare live data	19
Import of the first dataset	19
Adding a new dataset	20
Export Data	23

Introduction

What is Empa QATool?

Empa QATool is a stack of software that is used for quality assurance (mainly in the ACTRIS, EBAS field) of data. The software stack is made up of different applications. With the "QA Tool" you can compare different substances from the same or from different stations in one plot. In the "Quick Start" you find the most important predefined substance pairs that can be plotted with one click. Under "Compare live data" it is possible to add new data to an existing file of the running year. But it is also possible to export proceeded files or to import other useful data sets for your station like wind speed or carbon monoxide.

What can Empa QATool do for me?

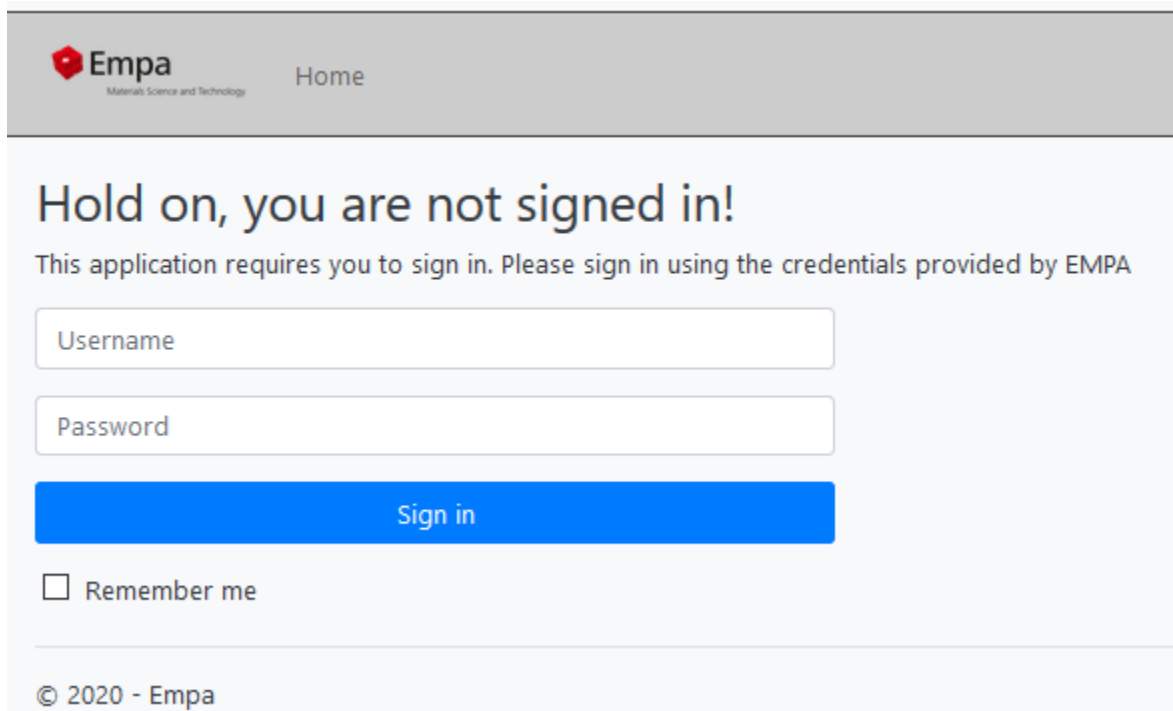
The tool is used to render data in your web-browser interactively and allows you to interact with the data in an efficient way. For example, it allows you to flag data by simply clicking on the point in a scatter or stack plot. It can export flagged data again in different formats, can import custom csv datasets...

Sounds great, how can I access it?

The web application is hosted under the URL <https://voc-qc.nilu.no>

User login

When you open the web-application in your browser and are not logged in, you will automatically be greeted with this view:



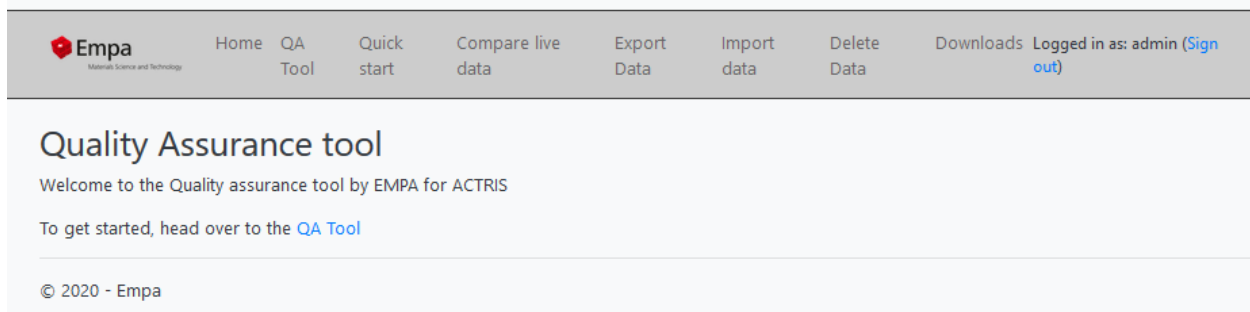
The screenshot shows a web application login page. At the top left is the Empa logo (Materials Science and Technology) and a 'Home' link. The main heading is 'Hold on, you are not signed in!'. Below this is a message: 'This application requires you to sign in. Please sign in using the credentials provided by EMPA'. There are two input fields: 'Username' and 'Password'. A blue 'Sign in' button is positioned below the password field. A 'Remember me' checkbox is located below the button. At the bottom left, the copyright notice '© 2020 - Empa' is visible.

For security reasons, the tool is protected behind a user login. Empa (Stefan Reimann, Matz Hill) should have provided you with a user logon, if not, write a short e-mail to stefan.reimann@empa.ch or matthias.hill@empa.ch, they will gladly provide you with a user login.

Depending on your login / user, you will have a different permission set.

Start page

Once logged in you will be redirected to the start page:



The screenshot shows the top navigation bar of the Empa Quality Assurance tool. The navigation bar includes the Empa logo, a 'Home' link, a 'QA Tool' link, a 'Quick start' link, a 'Compare live data' link, an 'Export Data' link, an 'Import data' link, a 'Delete Data' link, and a 'Downloads' link. On the right side of the navigation bar, it shows 'Logged in as: admin (Sign out)'. Below the navigation bar, the main content area displays the title 'Quality Assurance tool', a welcome message 'Welcome to the Quality assurance tool by EMPA for ACTRIS', and a link to 'QA Tool'. At the bottom of the page, there is a copyright notice '© 2020 - Empa'.

The actions available in the top bar depend on your permission set, also the content of the start page might be different (it is configurable by the administrator).

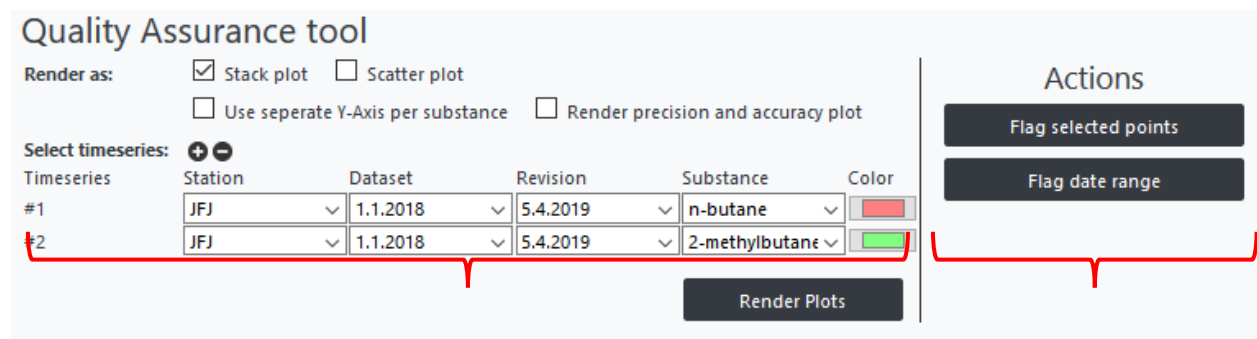
We will go through every action available in this document and explain them in detail.

For a short overview:

- **Home:** Displays the start page
- **QATool:** The QATool is the heart of the whole application. Here data can be plotted and flagged.
- **Quick start:** Preconfigured graphs for the major VOCs, redirects to the QATool
- **Compare live data:** This module is used for quality check of near real-time data.
- **Export Data:** Tool that allows one to export modified data as well as a flag report from the internal SQL database again, in NASA AMES, and CSV formats.
- **Import data:** Can be used to import custom datasets that don't go through the validation process of EBAS, like NOX data for example. Can then be used to plot against other data in the QATool.
- **Delete Datasets:** This action can be used to delete old datasets from the database. This action is only available to administrators.

QATool

This is the heart of the application. It allows users to plot data against each other in scatter and stack plot, and allows to flag data. The view is divided into two sections:



The screenshot shows the 'Quality Assurance tool' interface. On the left is the 'Plotting area' which includes a 'Render as' section with checkboxes for 'Stack plot' (checked), 'Scatter plot', 'Use separate Y-Axis per substance', and 'Render precision and accuracy plot'. Below this is a 'Select timeseries' section with a table of two time series. A red bracket highlights the table. On the right is the 'Actions area' with two buttons: 'Flag selected points' and 'Flag date range'. A 'Render Plots' button is located at the bottom center of the plotting area.

Timeseries	Station	Dataset	Revision	Substance	Color
#1	JFJ	1.1.2018	5.4.2019	n-butane	Red
#2	JFJ	1.1.2018	5.4.2019	2-methylbutane	Green

Plotting area

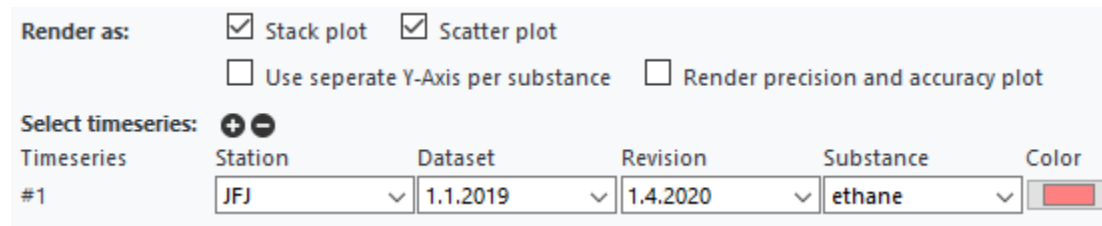
Actions area

We will first focus on the Plotting area

Plotting area

This area is entirely dedicated to plotting data.

Let us take a closer look:



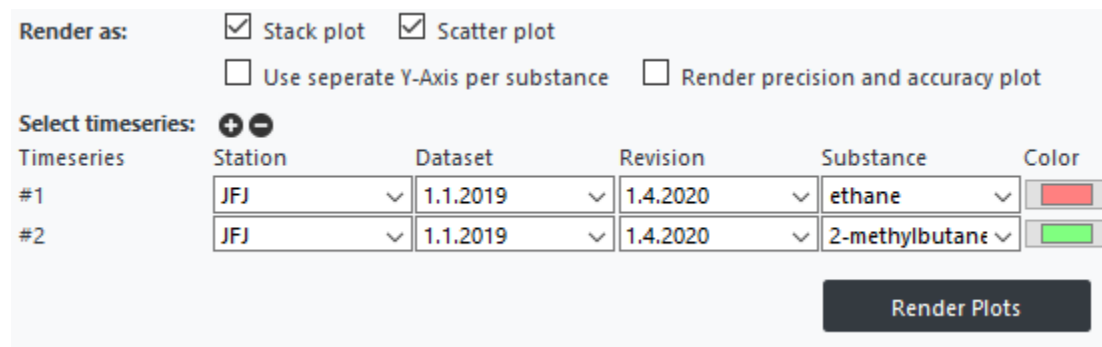
This close-up shows the 'Render as' section with 'Stack plot' and 'Scatter plot' both checked. The 'Select timeseries' section shows a table with one time series selected.

Timeseries	Station	Dataset	Revision	Substance	Color
#1	JFJ	1.1.2019	1.4.2020	ethane	Red

In the “Render as” section you can define which plot types should be displayed.

The black **plus and minus buttons** are used to add or remove time series from the selection. Generally, performance of the tool should be decent for up to four time series rendered simultaneously. Rendering large numbers of time series at the same time will have a negative impact of your pc / browser.

Press the black plus button once, to get two time series:



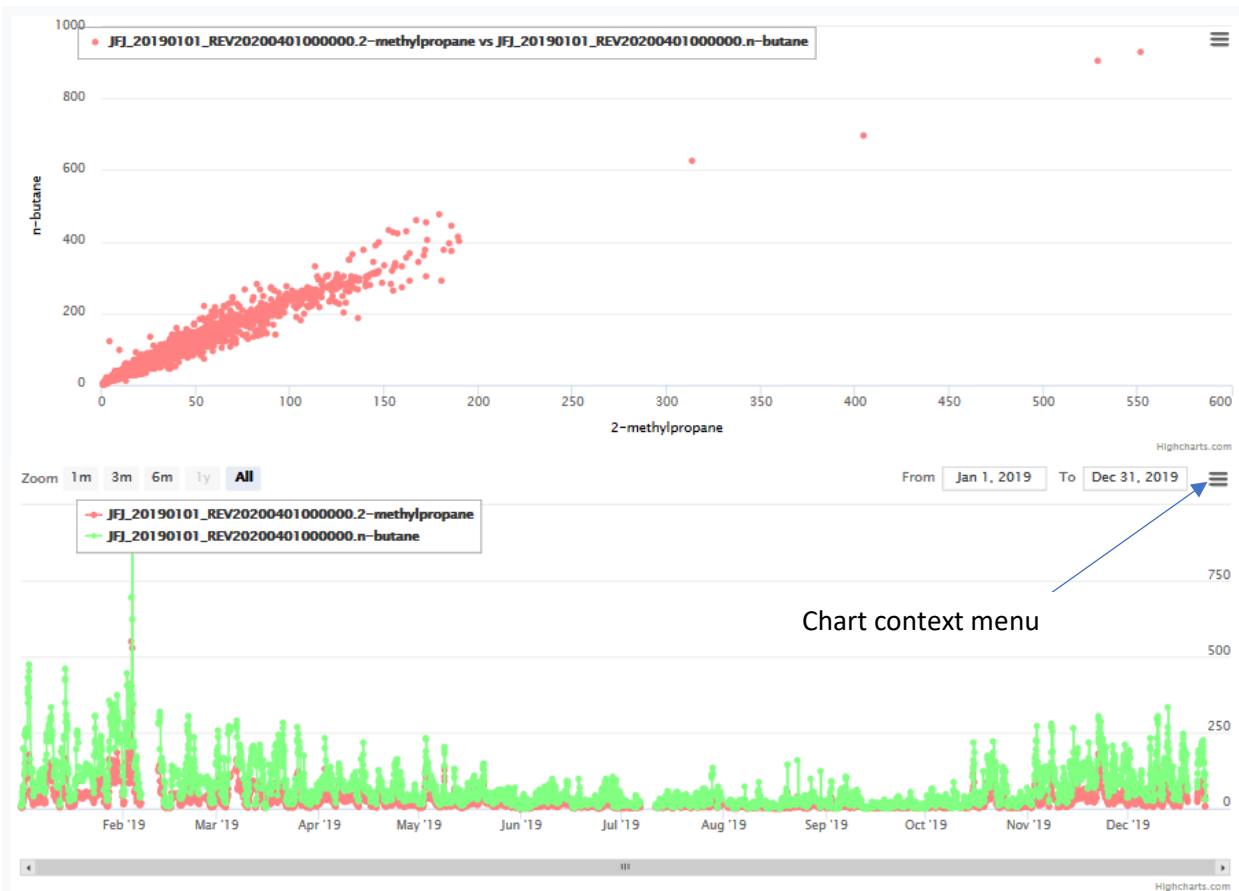
This close-up shows the 'Render as' section with both 'Stack plot' and 'Scatter plot' checked. The 'Select timeseries' section shows a table with two time series selected.

Timeseries	Station	Dataset	Revision	Substance	Color
#1	JFJ	1.1.2019	1.4.2020	ethane	Red
#2	JFJ	1.1.2019	1.4.2020	2-methylbutane	Green

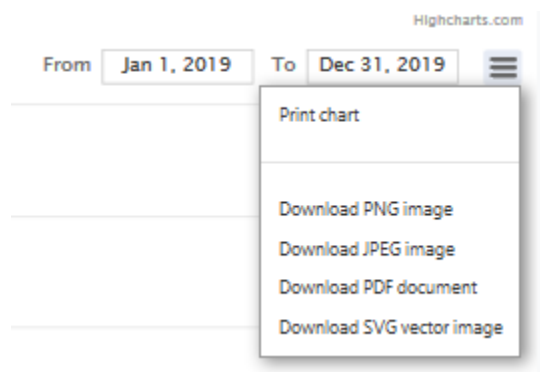
Now select the specific dataset that you want to render / plot and choose what substance(s) you are interested in. The “Color” column sets the color that should be used to render the individual time series.

Once the time series is ready to be rendered, simply press "Render Plots". A loading animation will show up and once everything is loaded and rendered the loading animation will disappear.

In this case, both scatter and stack plot were rendered and displayed.

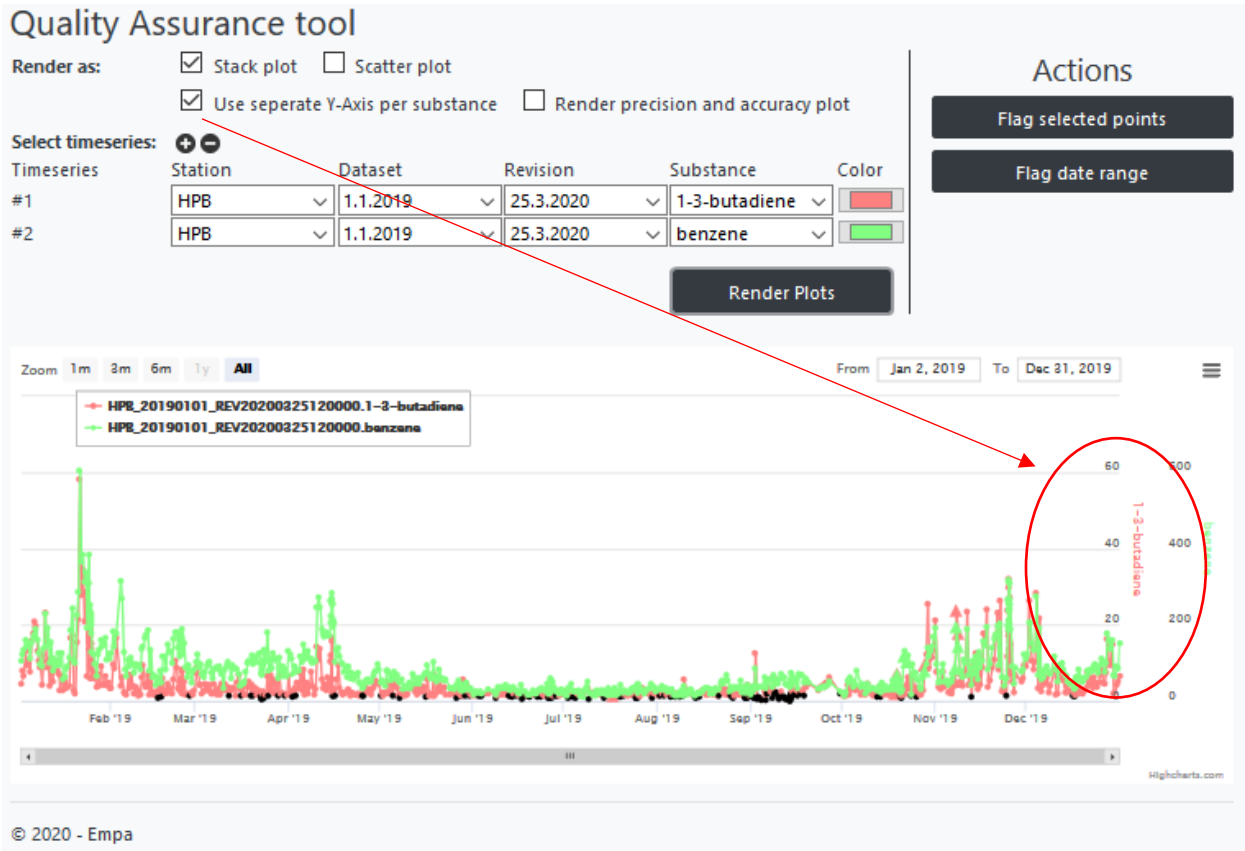


By clicking on the 'Chart context menu' the image can be downloaded in different formats

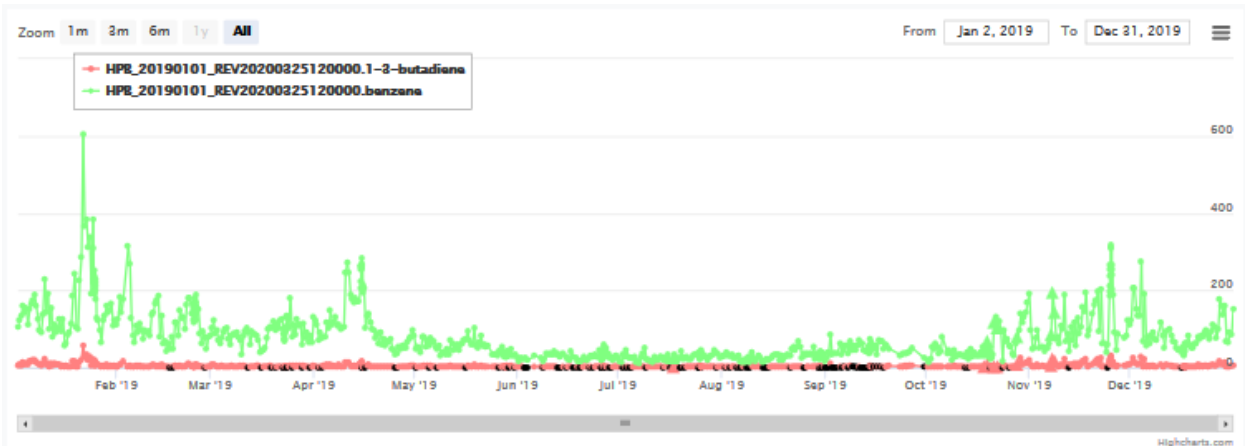


Note: The scatter plot has a few limitations. For example, it can only render data when both time series have the same time stamp. Further, it will always pick two time series together to be rendered (meaning it requires an even number of time series selected). If one has selected four time series, it will plot #1 vs #2 and #3 vs #4 and produce two scatter plots. The color is taken from the first time series rendered (#1, #3, ...).

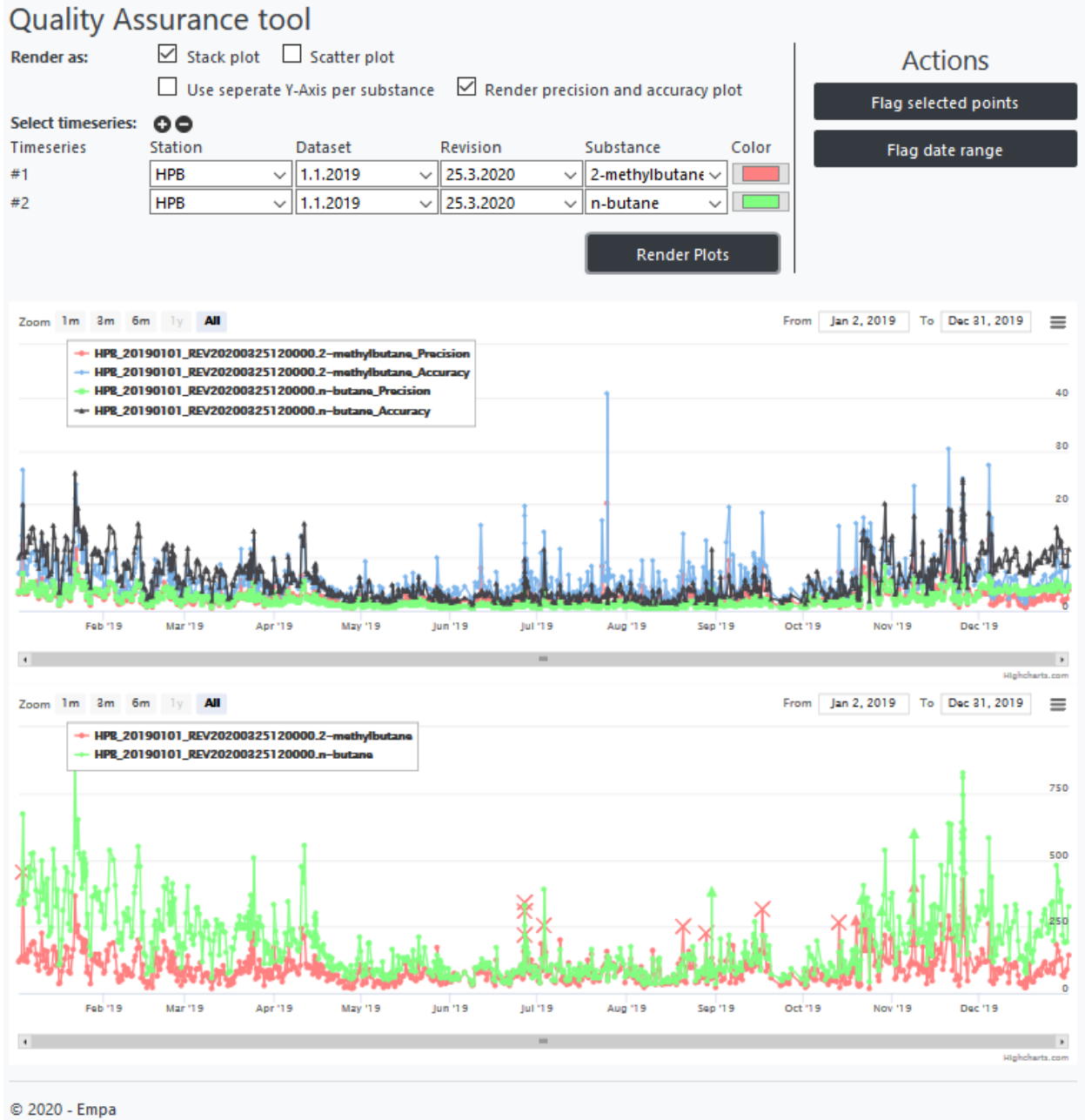
For a better visualization, it is sometimes helpful to have separate y-axis. If you click on 'Use separate Y-axis per substance', and then on render plots, two different Y-axes are displayed.



Same plot with only one Y-axis:



Precision and accuracy (if you have them in your data set) can be visualized in a separate plot if you tick the checkbox 'Render precision and accuracy plot'.



Miscellaneous symbols

Colored dot: Valid data with the flag "0".

Big colored dot: Selected data point

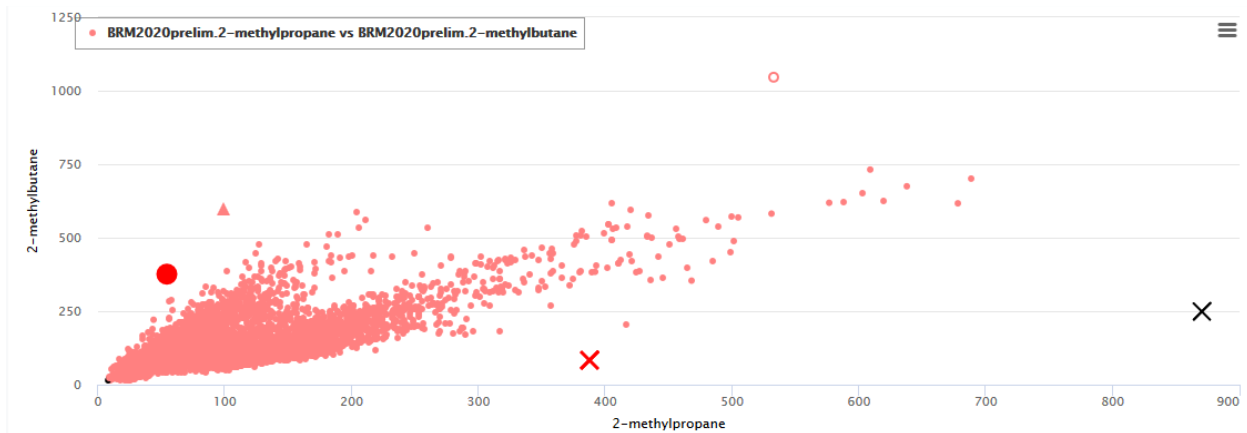
Black dot: Value below detection limit (valid) with the flag "0.147".

Colored triangle: Valid value with a flag (except the 0.147 flag).

Colored cross: data point has to be checked (flagged by the user as 2.000, see "Flag selected points").

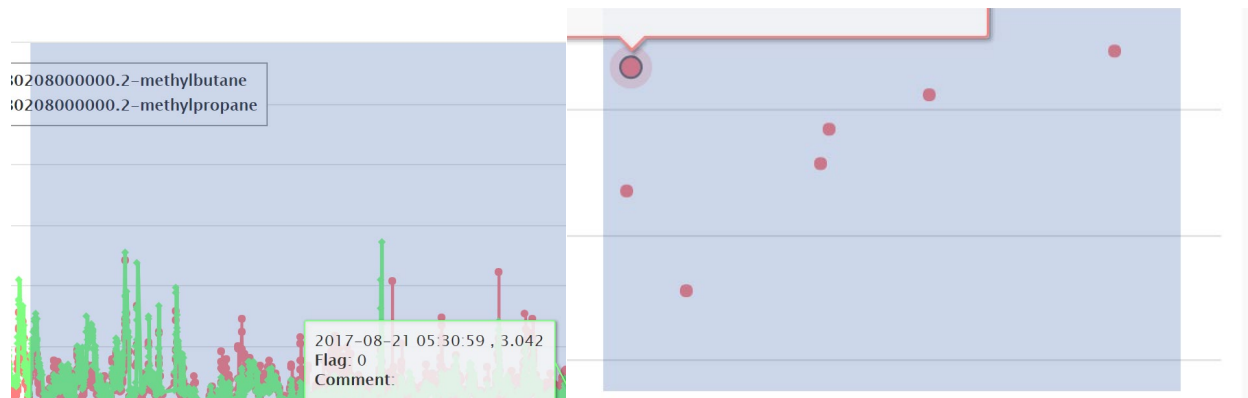
Black cross: Invalid value (flagged by the user as 1.999 see "Flag selected points").

Colored circle: Flagged by the user as 1.000 see "Flag selected points".



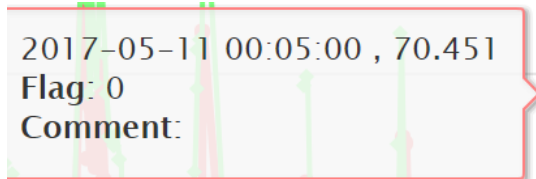
Zooming

Inside the plot, one can zoom in by holding the left mouse button and selecting the area that he wants to zoom in:



Tooltips

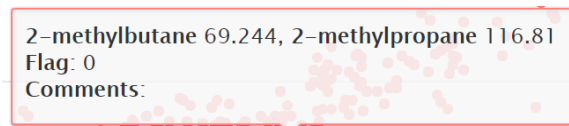
In the stack plot, tooltips are displayed as follows:



In the first line the date and value are displayed, then the flag (if any) and the comment. Flagged data will be rendered as an X:



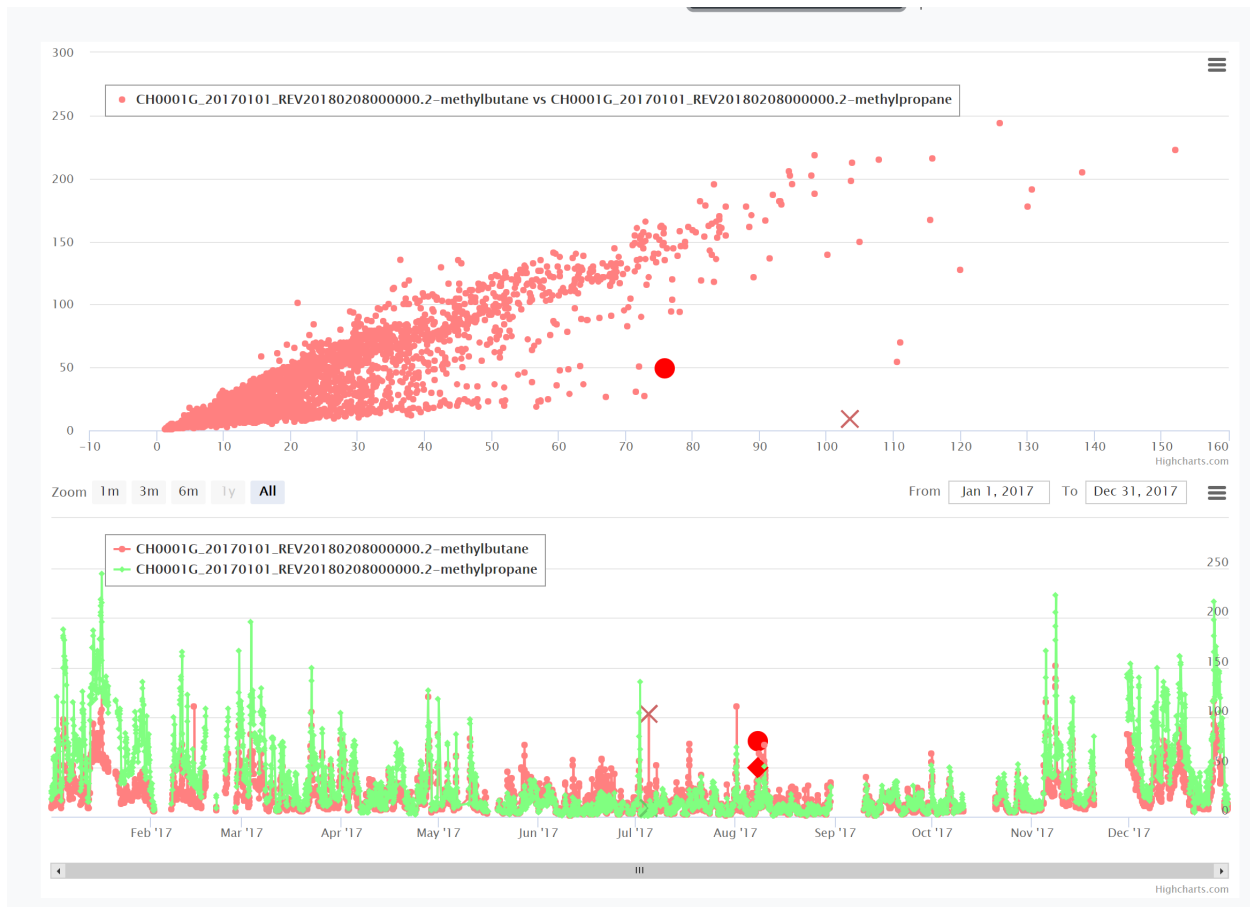
On the scatter plot the tooltip looks a bit different:



It shows us the value of both time series, the flag (if one of the time series is flagged at this point it will display that flag), and the comments (will show comments of both time series).

Selecting data

Data points can be selected when clicking on them. Multiple data points can be selected when pressing on them while holding shift key. Selected points are synchronized between the plots, which makes it easy to spot for example a certain point in the scatter plot on the stack plot:



Actions area

The actions area contains two actions:

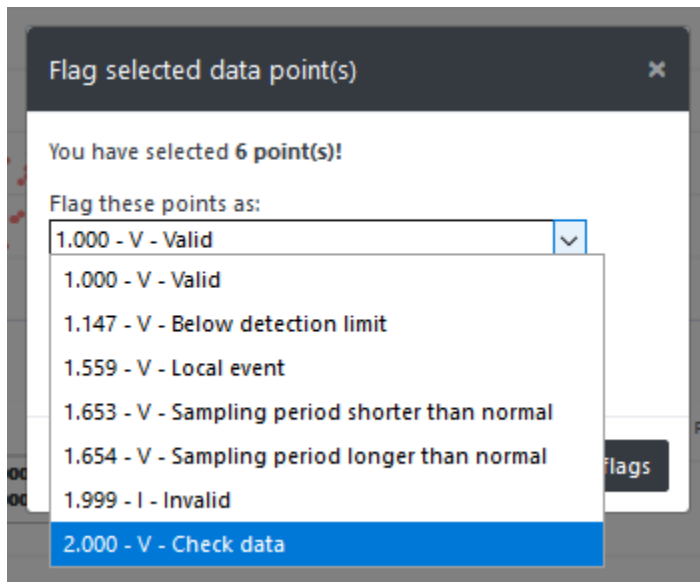
Actions

Flag selected points

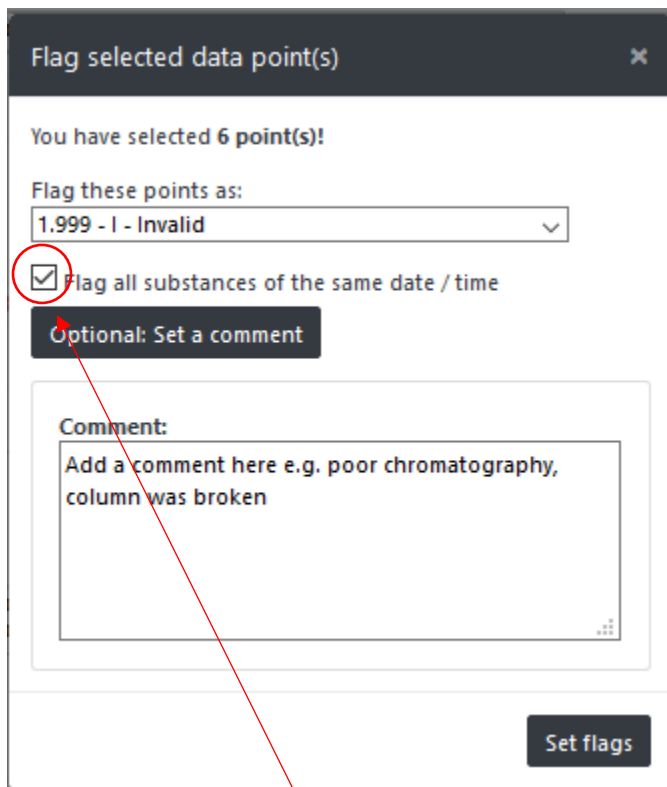
Flag date range

Flag selected points

This action only works when plots are rendered and at least 1 point is selected in the graphs. If no points are selected, it will display an error message.



After choosing a flag to set we can optionally set a comment that will be saved on all the selected points. To set a comment simply click on the “Optional: Set a comment” button and a text area will show where you can enter the comment.

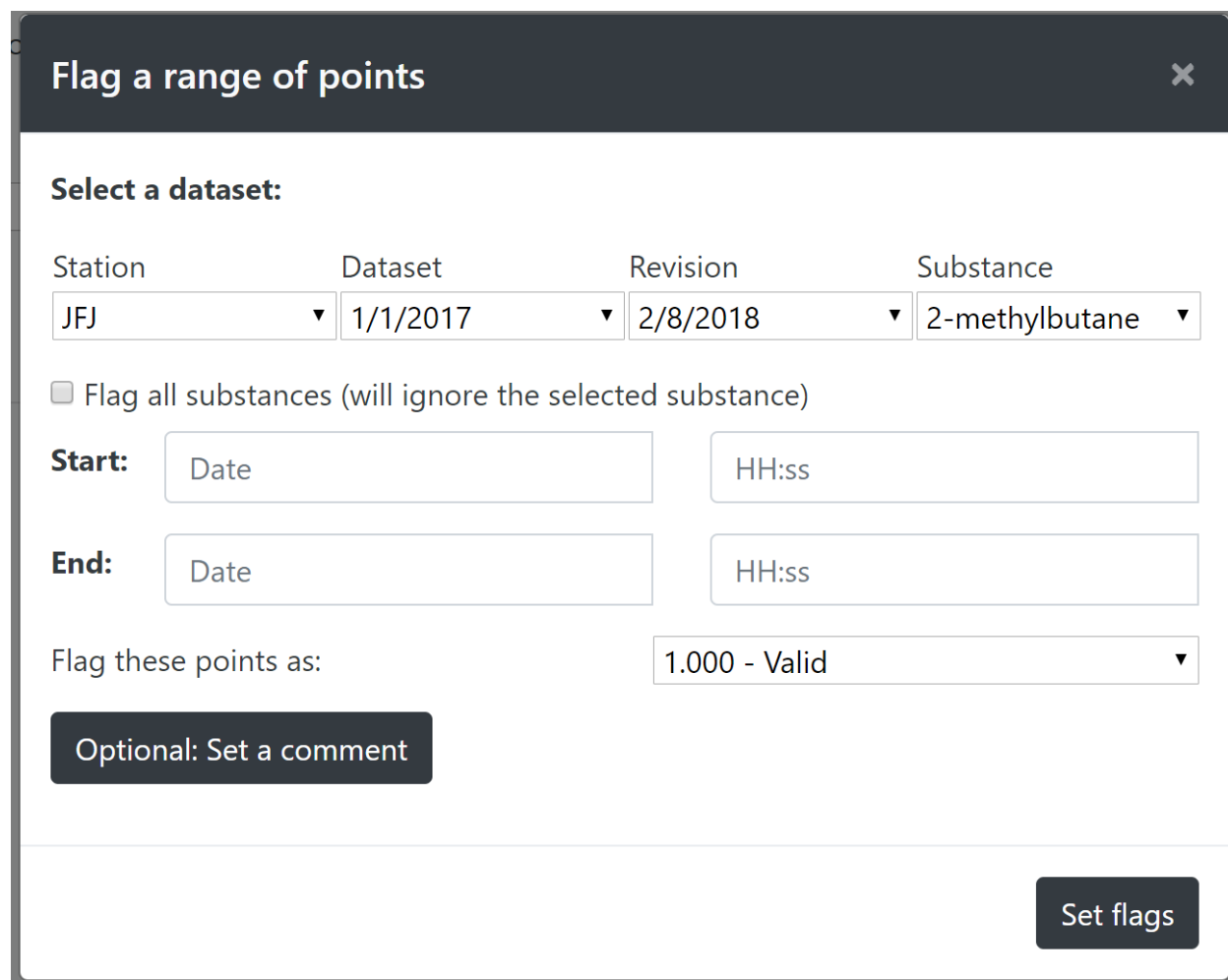


By setting a tick in this checkbox, the flag is applied to all substances

Note that the flags are between 1.000 and 2.000, this is to allow the code and the users to identify, which data points were flagged by source (flag 0.000 – 0.999) or through the QATool.

Flag data range

This action can be run / used without any data being rendered or points selected. A click on the “Flag data range” button will open this window:

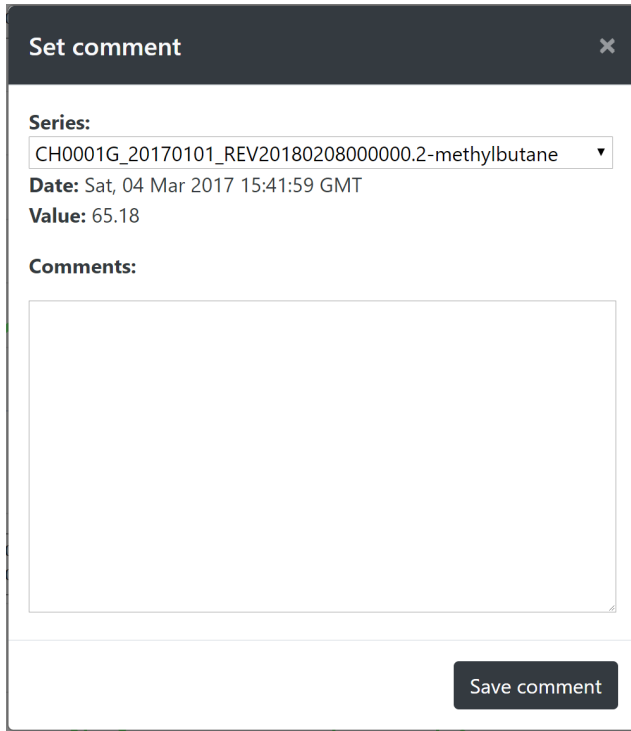


At the top, we can select which dataset / substance (or time series) we want to flag the data on. Alternatively, we can check the checkbox “Flag all substances”, which will ignore the selected substance and flag every data point inside that dataset matching the criteria’s. We can define a start and end date time to define the range that should get flagged.

Further, the flag can be selected in the dropdown and like the “flag selected points” dialog an optional comment can be set (which will overwrite any already existing comments).

Set comment on selected point

A slightly more “hidden” feature is, if one right clicks on a selected point on the rendered plot the following window opens:



Set comment [X]

Series:
CH0001G_20170101_REV20180208000000.2-methylbutane ▼

Date: Sat, 04 Mar 2017 15:41:59 GMT
Value: 65.18

Comments:

Save comment

As the title suggest, this is used to set comments on the point we just right clicked on. The dropdown is used when we right click on a point inside the scatter plot. Since in the scatter plot it is unclear which point / time series is meant.

The button “Save comment” will save the entered comment for the chosen point (doesn’t affect the other points, if more than one point is selectable in the dropdown).

Quick start

Reference station: HPB Change

Station	Dataset	Revision
CVO ▼	1.1.2019 ▼	31.3.2020 ▼

ethane vs propane

ethane vs n-butane

propane vs n-butane

2-methylpropane vs n-butane

2-methylpropane vs 2-methylbutane

2-methylbutane vs n-pentane

ethene vs propene

ethene vs ethyne

benzene vs ethyne

benzene vs toluene

In "Quick start" you will find preconfigured diagrams. Select the desired station and the revision date. @VOC@ automatically finds the correct graphs from the selected data set. Additionally you can select a reference station. It is recommended to compare the data not only against other stations but also against your own data set from last year. This way you can see immediately if the slope in the XY-plot has changed.

If you click on the desired substance pair, you will be forwarded to the QA Tool page. Here you can now edit the data set as described in the chapter QA Tool. When you are done with the substances, you don't have to search for new pairs, just click on Quick Start and select the next substance pair.

Import Data

The QATool used for ACTRIS, provided by Empa, supports the import of custom csv files. This import can be used to import custom datasets for visualization and flagging in the @VOC@, without having to go through the EBAS submission process.

These CSV files can contain any type of data (does not necessarily have to be VOC's listed as valid by EBAS).

With the 'Compare live data' there is a possibility to upload the VOC data from the current year. As soon as there are new data from your station they can be uploaded as well. @VOC@ will put the files together in one file and the new data are shown in a different color.

This chapter will explain how such a CSV file has to look like, for it to be imported into @VOC.

CSV Format

The CSV Importer of QATool will use “;” as delimiter between columns, so make sure your file uses this delimiter!

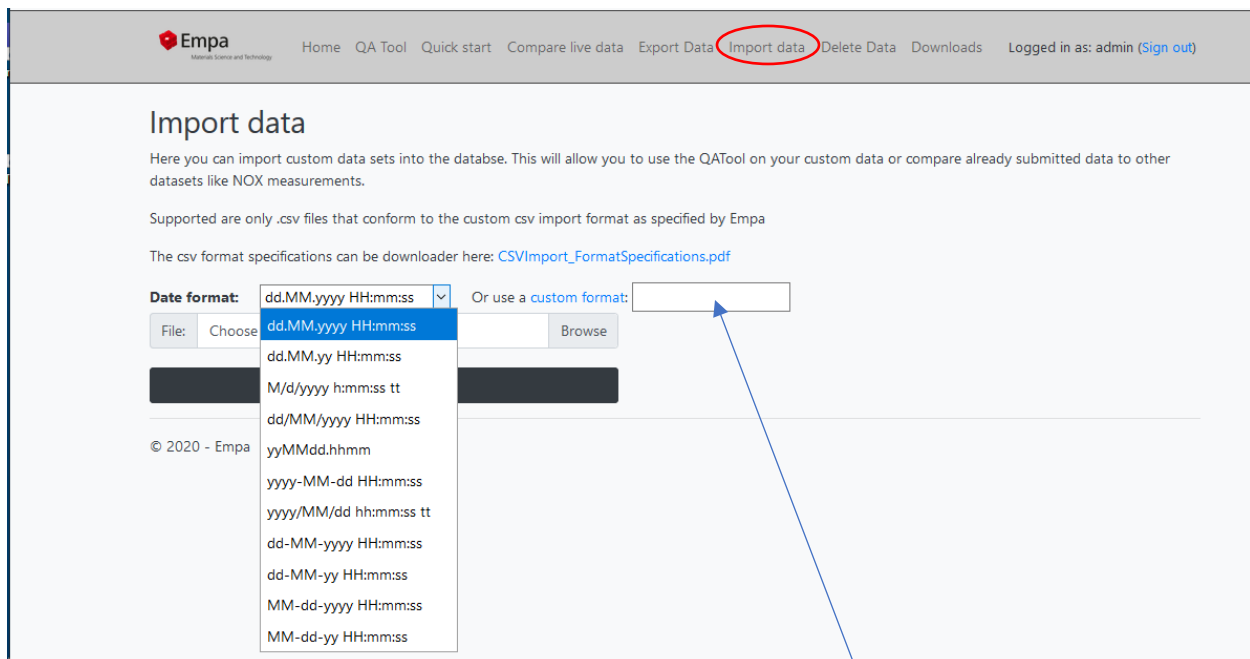
There are 2 main parts in the csv file: A mandatory part and an optional part

Mandatory part

The mandatory part includes these two columns:

- Starttime
- Endtime

These two columns are used to describe during which timeframe the data has been taken. If the measurement is a punctual (with just one date time, and no time range), just leave End time empty and enter your date in Start time. However, both columns should always be at the beginning of each row.



The screenshot shows the 'Import data' page of the Empa QATool. The navigation bar includes links for Home, QA Tool, Quick start, Compare live data, Export Data, Import data (circled in red), Delete Data, and Downloads. The user is logged in as admin. The main content area is titled 'Import data' and explains that users can import custom data sets. It provides a link to download CSV import format specifications. Below this, there is a 'Date format' dropdown menu with a list of predefined formats, and an 'Or use a custom format:' text input field. A blue arrow points from the 'custom format' field to the text below.

In the "Import data" section you will find some predefined data formats. Select the one you have used in your .csv file. If you do not find your format, you can type it into the **custom format**.

Also, mandatory of course is the header of the dataset, containing stuff like substance names, etc. For example (more further down in this document):

```
Start;End;CO-Value;CO-Flag;CO-Precision;CO-Accuracy;NO-Value;
```

Optional part

The optional part includes all the data. There are 4 types of data:

- **Value**
- **Precision**
- **Accuracy**

- **Flag**

These are represented as 4 columns each, per measured substance / object:

- [Substance]-Value (mandatory)
- [Substance]-Precision (optional)
- [Substance]-Accuracy (optional)
- [Substance]-Flag (optional)

So each substance has 1-4 columns. The measured value is obviously mandatory, while the others are optional and will be stored as NULL in the internal database when not provided.

Examples

1. CO₂ and NO_x data without end-time

Let's make an example. Lets assume we want to import a dataset containing CO₂ measurements (with a flag column) and NO_x measurements with no other optional columns: Where the data is a punctual measurement (End always empty)

```
Start;End;CO2-Value;CO2-Flag;NOX-Value;
2018-09-07 00:00:00;;234;1.000;42;
2018-09-08 00:00:00;;232;1.000;45;
2018-09-09 00:00:00;;242;1.000;37;
```

2. data with only mandatory fields, containing an end time

Lets assume one wants to import just some simple data, with start and end time defined for these substances:

```
Start;End;CO2-Value;NOX-Value;Wind-Speed-alue
2018-09-07 00:00:00; 2018-09-07 01:00:00;234;42;34
2018-09-08 00:00:00; 2018-09-07 02:00:00;232;45;36
2018-09-09 00:00:00; 2018-09-07 03:00:00;242;37;25
```

File name

The file name has to include station name, start date and revision date e.g.:

CH0005R_20170101_20180831.csv. This file can be imported with either the module

1. 'Compare live data' (see below, for checking data during the year continuously)
2. 'Import data' (used for checking for example VOCs with CO or other trace compounds at the end of the year)

Compare live data

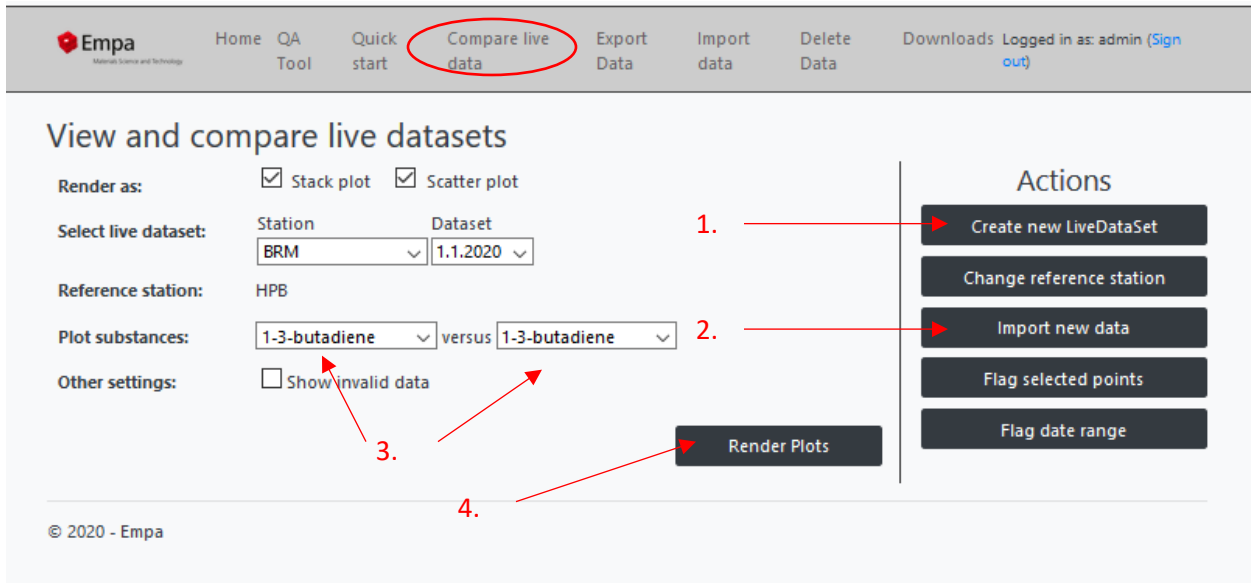
This module is used for quality check of near real-time data. One can compare a new dataset with an older one of the running year. One of the advantages is that you don't have to create a NASA AMES file. You can import a simple .csv file formatted as described above. **Important: Dates have to be in the format "JJJ-MM-TT HH:mm:ss", for example "2021-02-04 14:30:00"**

Example of a correct formatted .csv-file

	A	B	C	D	E	F	G	H
1	Start	End	ethane-Value	ethane-Flag	ethene-Value	ethene-Flag	propane-Value	propane-Flag
2	2021-02-04 02:14:00	2021-02-04 02:34:00	1960	0	162	0	730	0
3	2021-02-04 03:14:00	2021-02-04 03:34:00	1958	0	182	0	1110	0
4	2021-02-04 04:15:00	2021-02-04 04:35:00	1960	0	183	0	904	0
5	2021-02-04 05:15:00	2021-02-04 05:35:00	1936	0	179	0	835	0
6	2021-02-04 06:15:00	2021-02-04 06:35:00	1941	0	192	0	760	0
7	2021-02-04 07:15:00	2021-02-04 07:35:00	1932	0	189	0	674	0
8	2021-02-04 08:15:00	2021-02-04 08:35:00	1910	0	183	0	674	0
9	2021-02-04 09:15:00	2021-02-04 09:35:00	1871	0	178	0	651	0
10	2021-02-04 10:15:00	2021-02-04 10:35:00	1873	0	340	0	709	0
11	2021-02-04 11:16:00	2021-02-04 11:36:00	1859	0	279	0	695	0

Import of the first dataset

The import can be done within four steps. As soon as you have created your file, login in @VOC@ and click on 'Compare live data'



The screenshot shows the Empa web interface. The top navigation bar includes 'Home', 'QA Tool', 'Quick start', 'Compare live data' (circled in red), 'Export Data', 'Import data', 'Delete Data', 'Downloads', and 'Logged in as: admin (Sign out)'. The main content area is titled 'View and compare live datasets'. It features several input fields: 'Render as:' with checkboxes for 'Stack plot' and 'Scatter plot'; 'Select live dataset:' with 'Station' (BRM) and 'Dataset' (1.1.2020) dropdowns; 'Reference station:' (HPB); 'Plot substances:' (1-3-butadiene) versus (1-3-butadiene); and 'Other settings:' with a 'Show invalid data' checkbox. On the right, an 'Actions' panel contains buttons for 'Create new LiveDataSet', 'Change reference station', 'Import new data', 'Flag selected points', and 'Flag date range'. A 'Render Plots' button is located at the bottom of the form. Four numbered red arrows indicate the steps: 1. points to 'Create new LiveDataSet', 2. points to 'Import new data', 3. points to 'Show invalid data', and 4. points to 'Render Plots'.

1. First of all you have to create a new live dataset for your station (e.g. BRM2021_prelim) and a dataset start date. The station name has to be identical with the station name in your .csv-file!

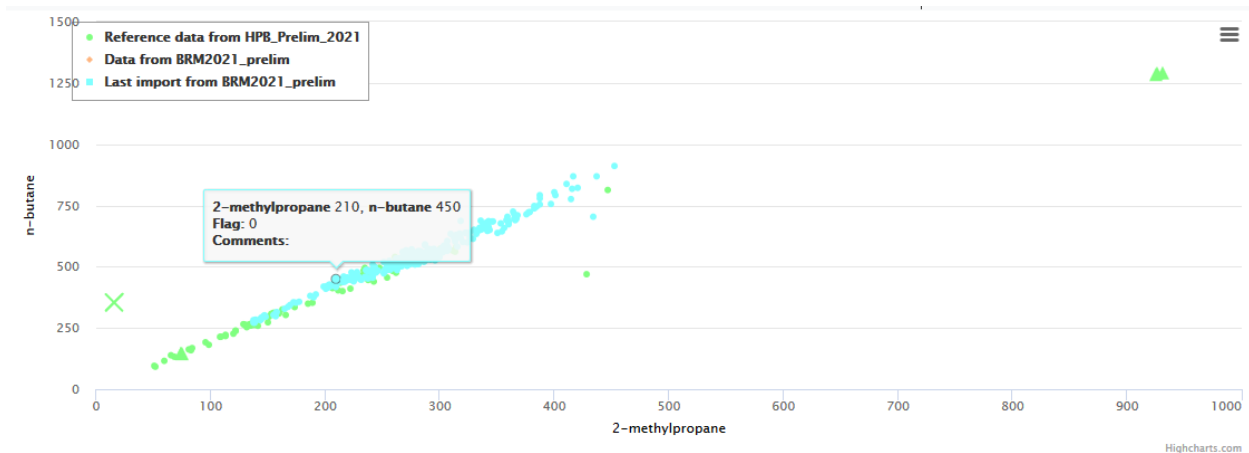
The dialog box titled "Create a new live dataset" has a close button (X) in the top right corner. It contains two input fields: "Station" with the value "BRM2021_prelim" and "Dataset start date" with the value "2021-01-01". A "Create" button is located at the bottom right.

There will be a message, which confirms the creation of the new dataset.

2. In a second step, one has to select the file which has to be uploaded by using the 'Import' link.

The dialog box titled "Import new data" has a close button (X) in the top right corner. It contains the text "Select file to import:" followed by a button labeled "Durchsuchen..." (circled in red) and the text "Keine Datei ausgewählt.". An "Import" button is located at the bottom right.

3. Choose two substances from the pull-down menu you are interested in.
4. Press 'Render scatter plot'. @VOC@ shows the scatter plot of your first dataset.



Because this is your first dataset all points are shown in light blue.

Adding a new dataset

As soon as there are new data available, one can upload them by using the 'Import' link. It doesn't matter if your new data set includes the old data or if you have a really new set, @VOC@ will only import the new data points.

Empa Materials Science and Technology

Home QA Quick start Compare live data Export Data Import data Delete Data Downloads Logged in as: admin (Sign out)

View and compare live datasets

Render as: Stack plot Scatter plot

Select live dataset: Station: BRM Dataset: 1.1.2020

Reference station: HPB

Plot substances: 1-3-butadiene versus 1-3-butadiene

Other settings: Show invalid data

Actions:

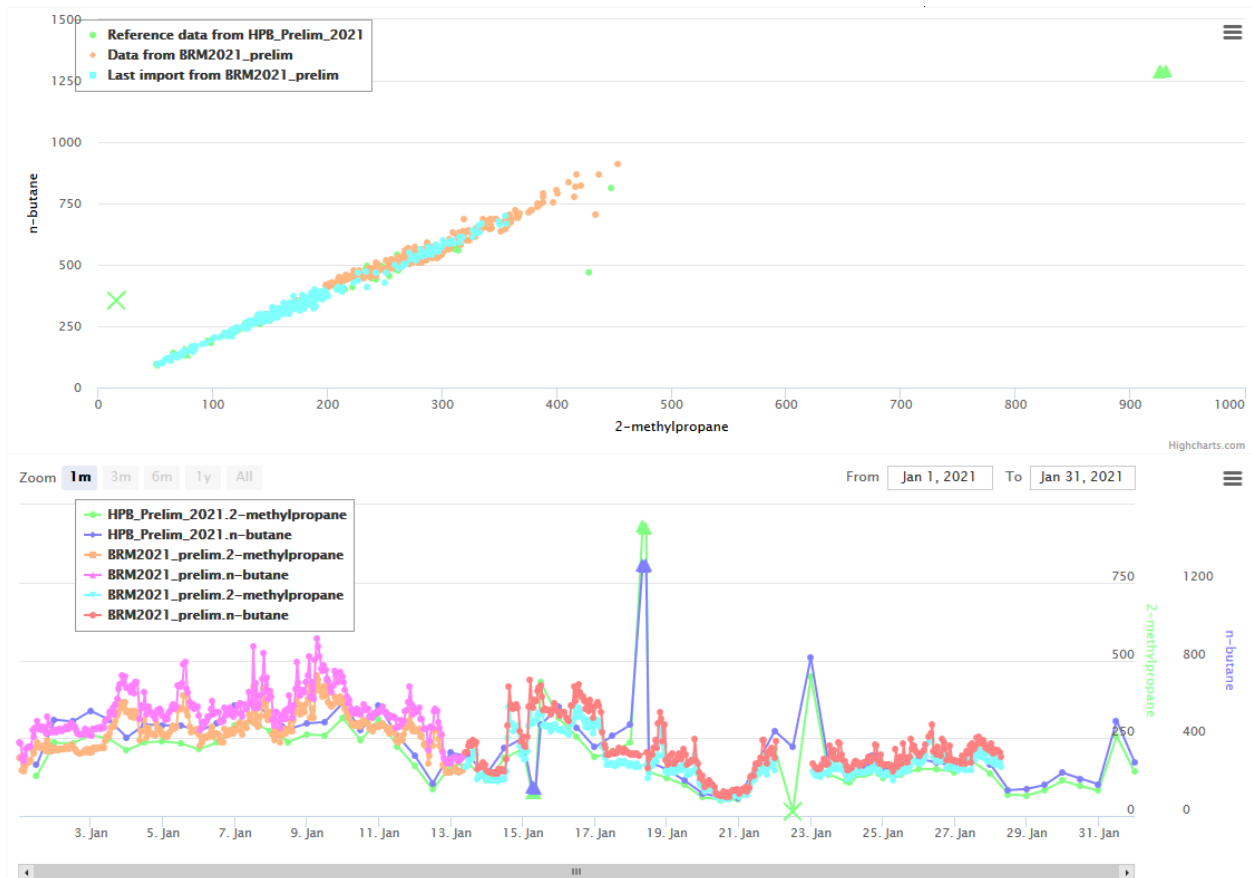
- Create new LiveDataSet
- Change reference station
- Import new data
- Flag selected points
- Flag date range

Render Plots

© 2020 - Empa

Choose two substances you are interested in and press 'Render scatter plot'.

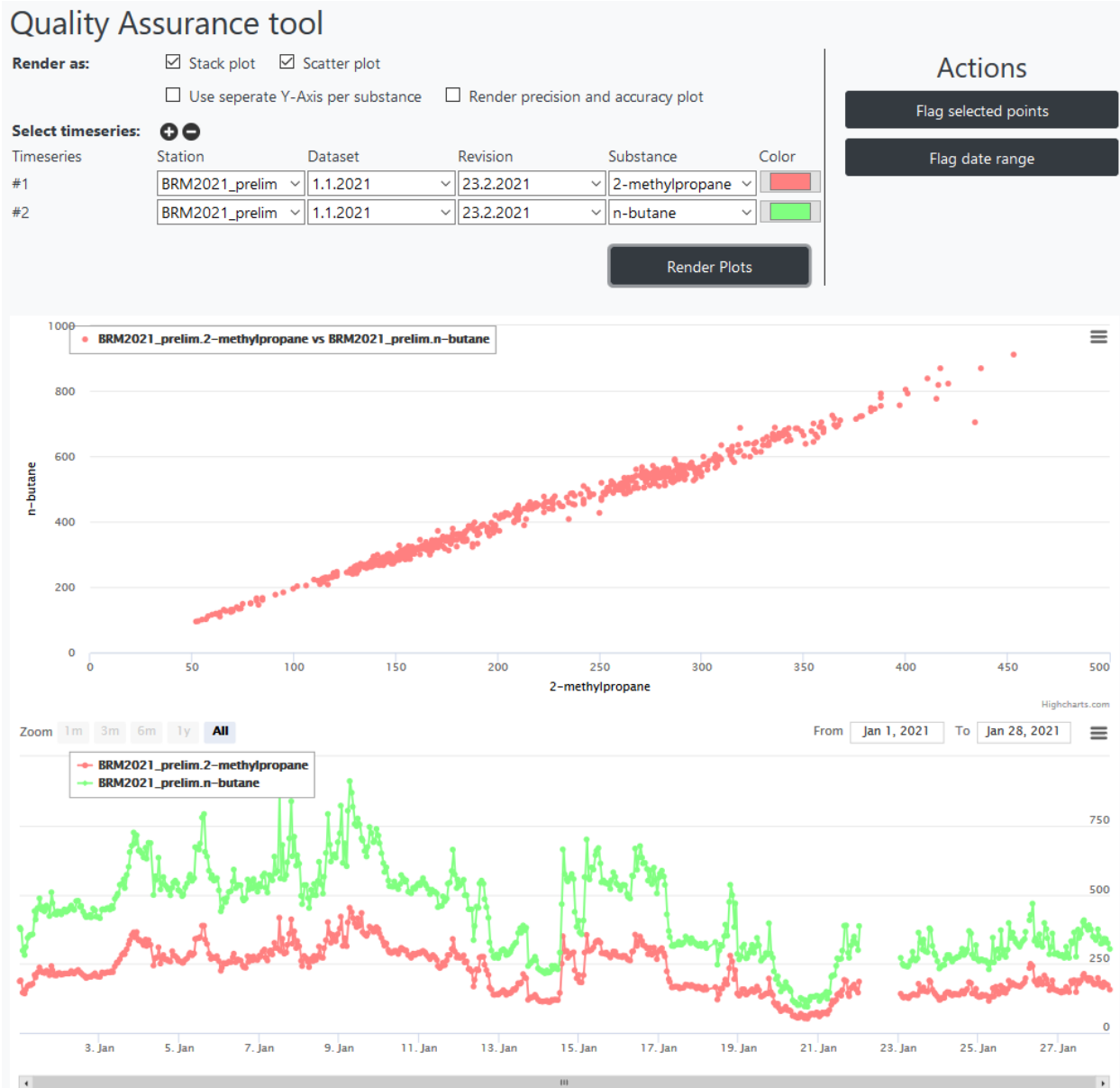
@VOC@ now shows the result of the second data import.



The results of the first import are shown in orange the freshly imported dataset is in light blue. In the background @VOC@ has merged the two files together in one file to prevent unnecessary files from being stored on the server.

You have also the possibility to flag some points as described in chapter XXX

The new merged file can also be seen and used in QA Tool of @VOC@ with all its advantages, such as adding a third substance or a third station.



Or just use the 'Quick start' where you will find this data set as well.

Export Data

In this section you can you can export your data as .csv file or, if you have imported your data in NASA AMES format into the QA tool, export it again. **Important: As long as values are marked with the flag 2.000 (Check), no NASA AIMS files can be exported. Flags made during the use of @VOC@ (i.e. flags > 1) are automatically converted to the correct flags.**