



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

Title	Report describing Time series analysis VRE
Work package n°	5
Deliverable n°	5.5
Lead beneficiary	CNRS
Author(s)	Pawel Wolff, Damien Boulanger, Valérie Thouret
Deliverable Type	Report
Dissemination Level	Public
Estimated delivery date	01/04/2023
First submission date	19/04/2023
Revised submission date	16/07/2024
Version	2
Reviewed by	Stefan Reimann, Ute Karstens
Accepted by	
Comments	





## Table of Content

<b>1. Introduction .....</b>	<b>3</b>
1.1. Purpose of the document.....	3
1.2. Scientific purpose .....	3
1.3. General overview of the work done in Task 5.3 .....	3
<b>2. Description of the web user interface .....</b>	<b>7</b>
2.1. Workflow .....	7
2.2. Description of the web user interface functionalities .....	10
2.2.1. Information .....	10
2.2.2. Search datasets .....	12
2.2.3. Select datasets .....	14
2.2.4. Filter data.....	19
2.2.5. Data analysis .....	21
2.2.6. Download .....	38
<b>3. Description of the notebook.....</b>	<b>39</b>
3.1. General overview.....	39
3.2. Comparison with the web user interface.....	39
3.3. Notebook workflow .....	40
<b>Appendix A: Glossary.....</b>	<b>42</b>
<b>Appendix B: Table of Figures.....</b>	<b>43</b>





## 1. Introduction

### 1.1. Purpose of the document

This document describes the **time series analysis VRE** developed within Task 5.3. Its purpose is to report on the work that has resulted in the development, deployment and documentation of a Virtual Research Environment (VRE) for analysis of time series of observational data published by the European Research Infrastructures ACTRIS, IAGOS and ICOS. This VRE is composed of:

- A **web user interface**: a web graphical interface for interactive time series analysis and data visualisation,
- A **notebook**: a jupyter notebook for expert users.

### 1.2. Scientific purpose

The objective of this development is to offer a large diversity of users (students, academics, private or public organisations, NGOs, communities of projects, etc...) a dedicated VRE for facilitating their exploration of multiple in situ datasets providing a long time series of different variables. The leading rationale is to make accessible the most commonly used basic metrics (e.g. means, percentiles) and statistical analysis (e.g. trends), as well as visualisation tools to look at e.g. 2D or 3D scatter plots. Such Virtual Access is meant to save the time of the users in order to concentrate on the scientific questions. This is also meant to be the first step of exploring cross-cutting issues in combining the different variables relevant for climate and air quality recorded by ACTRIS, IAGOS and ICOS.

### 1.3. General overview of the work done in Task 5.3

The purpose of the time series analysis VRE is to provide users with:

- a unified access to long-term observational data time series from the ACTRIS, IAGOS and ICOS Research Infrastructures,
- data filtering tools,
- tools for time series statistical analysis,
- tools for interactive data visualisation,
- download of the data produced by the service and citation.





The above goals are accomplished by each of the two kind of tools:

- the **web user interface** — a web application accessible in a user's web browser, dedicated to all users;
- the **notebook** — a Python software environment and a set of ready-to-go examples in the form of a Python jupyter notebook demonstrating possible use cases; it is dedicated to users having basic programming skills.

Both the web user interface and the notebook are driven by **data** consisting of:

- level 2 (i.e. final quality controlled observational data) data products based on observations performed at measurement stations (for data provided by ACTRIS and ICOS RIs),
- level 3 (i.e. derived products based on Level 2 data and/or external datasets) data products derived from observations made by IAGOS' instruments installed on commercial aircraft; the available data:
  - daily means of vertical profiles over airports frequently visited by the aircraft,
  - daily means of observations over specific geographical regions.

All data are provided under the Creative Commons Attribution 4.0 International licence (CC BY 4.0).

The web user interface and the notebook both rely on a bundle of Python packages (software libraries written in Python) developed within Task 5.3:

- **data access package:** allows for a unified access to the three RIs (ACTRIS, ICOS and IAGOS) long-term time series data and metadata;
- **data processing package:** provides high-level routines for processing time series such as:
  - datasets merging,
  - data filtering
  - data analysis (e.g. calculating means by period, moving average, estimating trend, etc.)



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

- **data visualisation package:** provides routines specialised in visualisation of data produced by the **data processing package**.

A general schema of the system components is presented in Figure 1.

The source code of

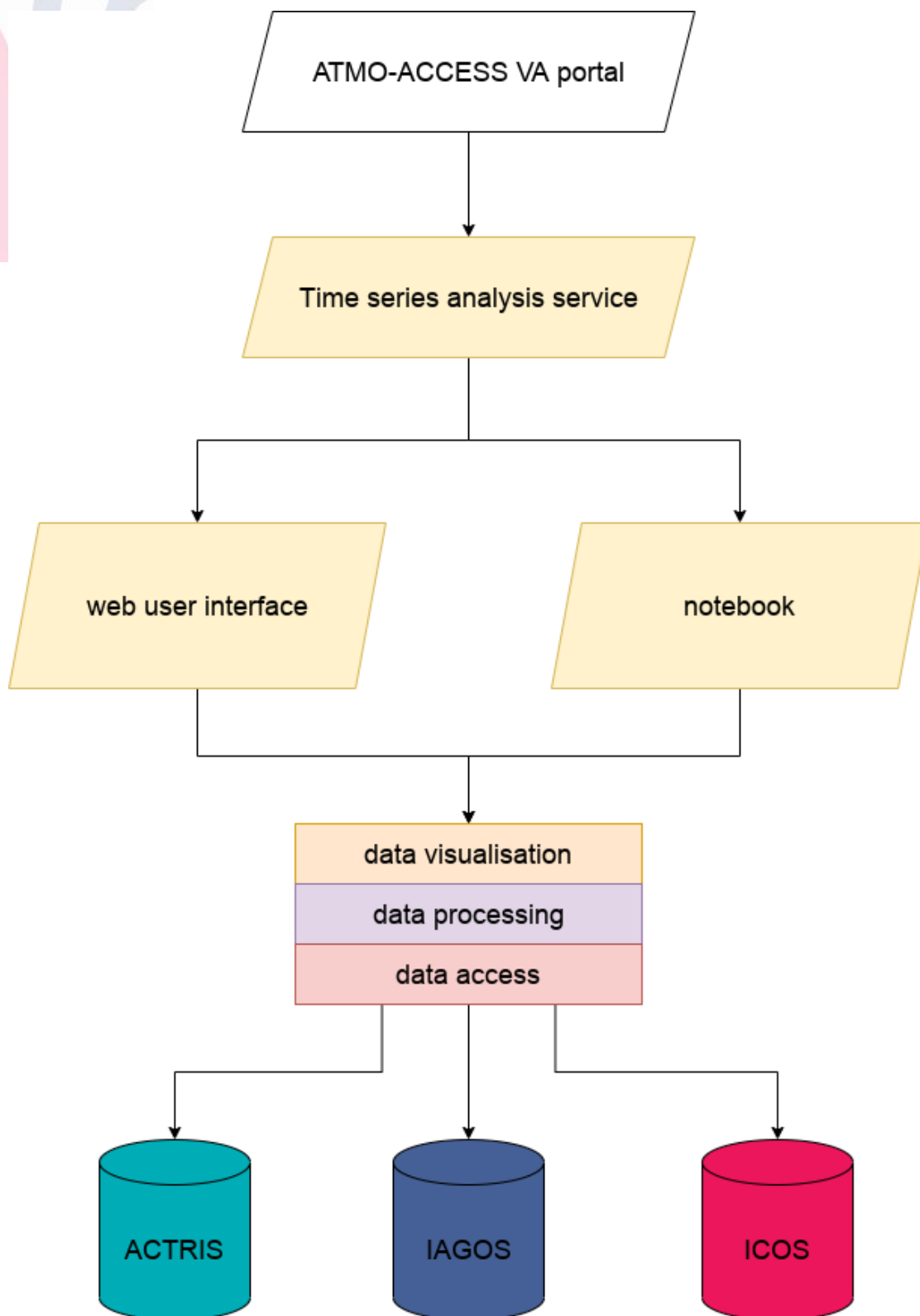
- the web user interface
- the notebook
- the underlying Python packages

is accessible at <https://github.com/iagos-dc/atmo-access-time-series>

A deployed version of the web user interface is available at <https://services.iagos-data.fr/atmo-access/timeseries>

The notebook can be downloaded by the users through the git repository and executed in their own environment. In the future the notebook will be accessible and executable through a Jupyter-Hub.





*Figure 1. General scheme of the system components (software and data)*



## 2. Description of the web user interface

### 2.1. Workflow

The user workflow of the web user interface consists of several steps:

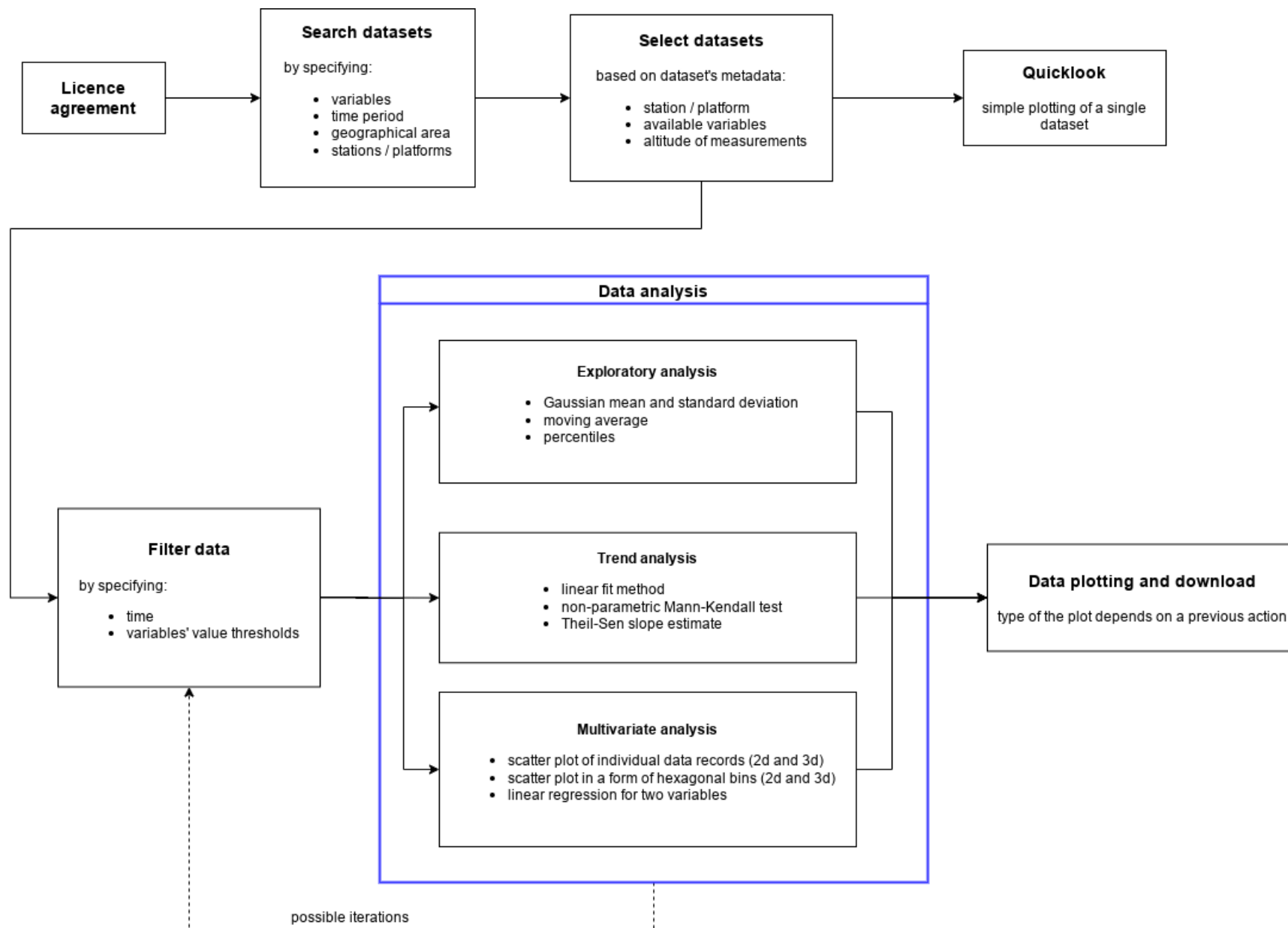
- General information about the application, available data and licence agreement
- Search datasets
- Select datasets (with a possibility to quick-look at a chosen dataset)
- Dataset analysis
- Data plot and download

The step “Dataset analysis” is consisted of two phases:

- Data filtering
- Computational action on filtered data

The user is guided by a Graphical User Interface (GUI) throughout the process. Whenever relevant, the user can navigate freely through all the steps of the workflow, in order to facilitate an iterative approach.

The user workflow diagram is presented in Figure 2.







**ATMO ACCESS**  
Access to Atmospheric Research Facilities

*Figure 2. The workflow from a user's perspective*



This work has received funding from the European Union's Horizon 2020 research and innovation programme through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004

[atmo-access.eu](https://atmo-access.eu)



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

## 2.2. Description of the web user interface functionalities

The functionalities are organised in the form of tabs:

- Information
- Search datasets
- Select datasets
- Filter data
- Data analysis

which closely reflects the workflow described in the previous section.

### 2.2.1. Information

The welcome page of the application, which is available under the “Information” tab (see Figure 3), contains general information on:

- the application,
- the RIs providing the data,
- available data,
- licence agreement.

On first use of the application, a user is asked to accept the licence agreement.





**ATMO ACCESS**  
Access to Atmospheric Research Facilities

## Time-series analysis



Information	Search datasets	Select datasets	Filter data	Data analysis
-------------	-----------------	-----------------	-------------	---------------

This tool allows you to search, analyse and visualise data from four Atmosphere European Research Infrastructures.

It has been implemented in the framework of the European Project ATMO-ACCESS to demonstrate interoperability within the Research Infrastructures

Only Essential Climate Variables are available. You can find more information about the provided datasets in the table below.




	ACTRIS is the pan-European research infrastructure producing high-quality data and information on short-lived atmospheric constituents and on the processes leading to the variability of these constituents in natural and controlled atmospheres. ACTRIS data from observational National Facilities means the ACTRIS variables resulting from measurements that fully comply with the standard operating procedures (SOP), measurement recommendations, and quality guidelines established within ACTRIS. Other data from the EBAS repository ( <a href="https://ebas.nilu.no">https://ebas.nilu.no</a> ) is also available through this service.	ACTRIS data are licensed under the Creative Commons Attribution 4.0 International licence (CC BY 4.0).
	The IAGOS datasets available are Level 3 data products derived from Level 2 products: Final quality controlled observational data. Monthly means timeseries have been calculated for all airports visited by the IAGOS fleet. Means are available for four pressure levels: surface (below 500 m), planetary boundary layer (PBL), free troposphere (FT), upper troposphere (UT). Concentrations of Ozone, Carbon Monoxide, H <sub>2</sub> O gas and relative humidity are provided as well as meteorological fields: air pressure, air temperature and wind. More information on the IAGOS Data Portal: <a href="https://iagos-data.fr">https://iagos-data.fr</a>	IAGOS data are licensed under the Creative Commons Attribution 4.0 International licence (CC BY 4.0).
	ICOS data are from the atmospheric network of ICOS Research Infrastructure for 36 stations and 90 vertical levels. The collection used contains the final quality controlled hourly averaged data for the mole fractions of CO <sub>2</sub> , CH <sub>4</sub> , N <sub>2</sub> O, CO and meteorological observations measured at the relevant vertical levels of the measurements stations. All stations follow the ICOS Atmospheric Station specification V2.0 (doi:10.18160/GK28-218) and are certified as ICOS atmospheric stations Class I or II. Data processing has been performed as described in Hazan et al., 2016	ICOS data are licensed under the Creative Commons Attribution 4.0 International licence (CC BY 4.0).

Figure 3. The welcome page of the application after the user's login with the active "Information" tab.



This work has received funding from the European Union's Horizon 2020 research and innovation programme through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004

[atmo-access.eu](https://atmo-access.eu)



### 2.2.2. Search datasets

At this step the user can search for datasets provided by ACTRIS, IAGOS and ICOS RIs. The search is based on the following filters:

- list of variables
- stations, airports
- geographical extent
- temporal extent

The choice of

- **ground stations** for ACTRIS and ICOS,
- **airports** or **geographical regions** for IAGOS

can be made via a dynamic map and later refined using a dropdown list.

The IAGOS' **geographical region** are pre-defined rectangular regions (on the longitude / latitude grid), such as

- Europe (from 15W 45N to 15E 55N),
- the North Atlantic (from 50W 50N to 20W 60N)
- Siberia (from 40E 50N to 120E 65N),
- etc.

and they are represented on the map in the form of a rectangle with a marker point in its centre.

Once finished with setting up the filters, the user can launch the search of datasets by clicking on the button **Search datasets** (see Figure 4).



## Time-series analysis

Information

Search datasets

Select datasets

Filter data

Data analysis

Select variable(s):  
☐ Select all / none  
☐ ACP - Aerosol Chemical Properties  
☐ APP - Aerosol Physical Properties  
☐ AT - Temperature (near surface)  
☒ CO<sub>2</sub> - Carbon Dioxide  
☒ CO - Carbon Monoxide  
☐ CIP - Cloud Properties  
☒ CH<sub>4</sub> - Methane  
☒ NO<sub>2</sub> - NO<sub>2</sub>  
☒ O<sub>3</sub> - Ozone  
☐ RH - Water Vapour (surface)  
☐ AP - Pressure (surface)  
☐ WSD - Surface Wind Speed and direction  
☐ N<sub>2</sub>O - Nitrous Oxide

SEARCH DATASETS

Date range: Start Date → 08/25/2017

Bounding box: 2.48 53.84 10.37 48.68

Stations map

Selected stations (you can refine your selection)

× CZI (Cabauw\_Zijdeweg, ACTRIS)

× REV (Revin, ACTRIS)

× AMS (Amsterdam, IAGOS)

× BRU (Brussels, IAGOS)

× CDG (Paris, IAGOS)

× DUS (Dusseldorf, IAGOS)

× FRA (Frankfurt, IAGOS)

× CBW (Cabauw, ICOS)

× HEI (Heidelberg, ICOS)

× JUE (Jülich, ICOS)

× KIT (Karlsruhe, ICOS)

× LUT (Lutjewad, ICOS)

× STE (Steinkimmen, ICOS)

Figure 4. The “Search datasets” tab.



### 2.2.3. Select datasets

At this stage the user chooses one or more datasets for a subsequent data analysis.

In order to facilitate the selection process, a plot showing data timeline coverage is displayed. The plot comes in two forms:

- **compact** — for each station or airport a single or multiple time periods for which data is available are shown in a form of horizontal bars (see Figure 5);
- **detailed** — for each variable or group of variables there is a plot similar to the compact plot (see Figure 6), but concerning data availability for that variable or the group of variables.

The plot allows (multi-)selection of horizontal bars representing the data coverage periods. Note that a single bar can correspond to multiple datasets (e.g. there might be several datasets, each covering some different time period, or, in the case of the compact version of the plot, these datasets might concern different variables). For example, the tooltip visible on Figure 4 indicates that a given horizontal bar represents two datasets with the variable CO<sub>2</sub> measured at the ICOS station Schauinsland covering (in total) the period Jan 1972–Feb 2022.

Selection of horizontal bar results in listing of corresponding datasets in a form of a table, displayed below the timeline coverage plots (see Figures 5 and 6).

The table contains metadata of the datasets:

- title,
- list of variables,
- name of the RI which provides the dataset
- station or airport name and its code
- time period covered by the data

Each individual dataset displayed in the table can be easily explored in a form of a quicklook plot (see Figure 7).





**ATMO ACCESS**  
Access to Atmospheric Research Facilities

The table itself allows for a multi-selection of datasets. Once the datasets are chosen, the user can pass to the next stage by clicking on the button **Select datasets**.



This work has received funding from the European Union's Horizon 2020 research and innovation programme through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004

[atmo-access.eu](https://atmo-access.eu)



# Time-series analysis

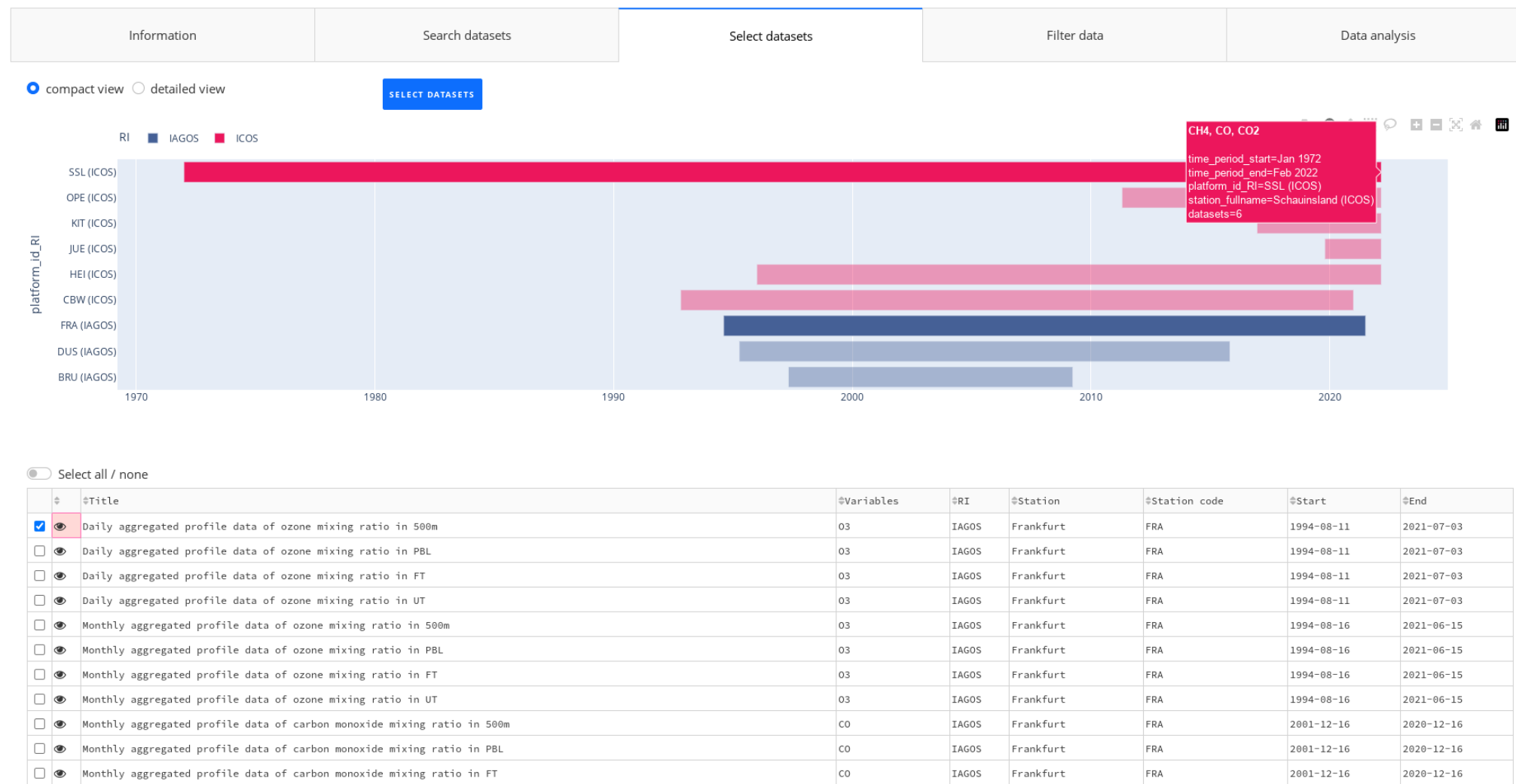


Figure 5. Select datasets: a timeline coverage plot in a compact form and a table with individual datasets. The selection of horizontal bars on the plot is synchronised with the table.







ATMO ACCESS  
Access to Atmospheric Research Facilities

## Time-series analysis

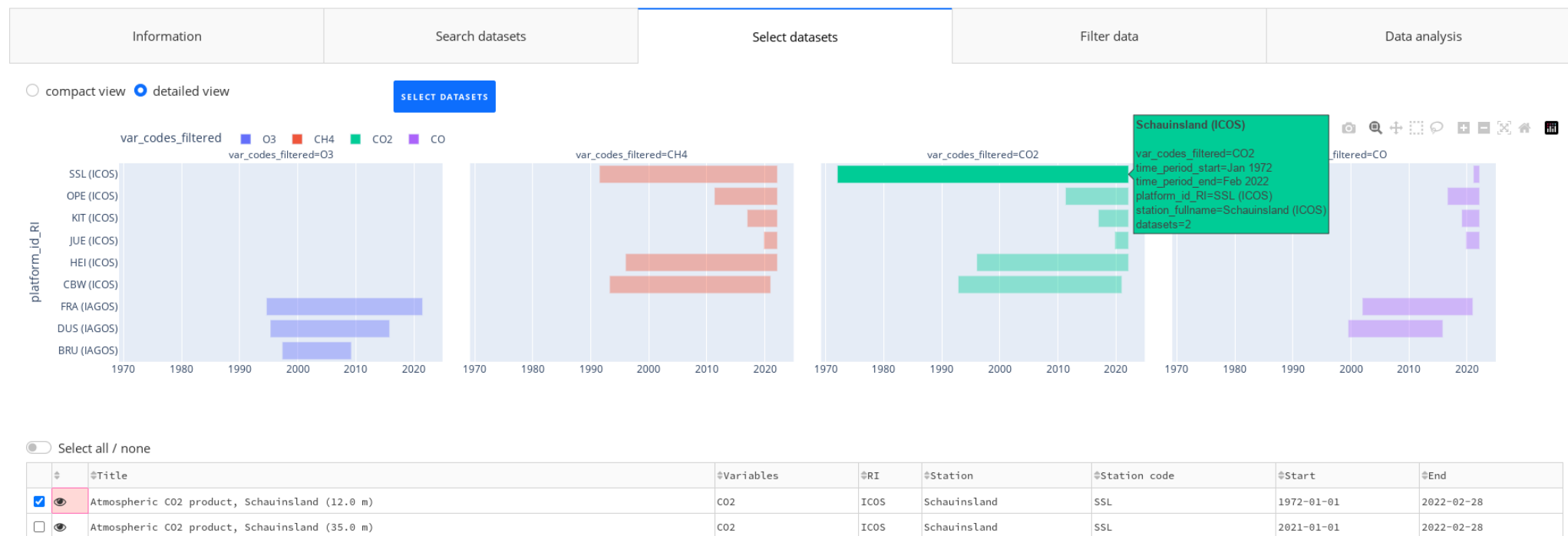


Figure 6. Select datasets: a timeline coverage plot in a detailed form and a table with individual datasets. The selection of horizontal bars on the plot is synchronised with the table.



This work has received funding from the European Union's Horizon 2020 research and innovation programme through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004

[atmo-access.eu](https://atmo-access.eu)



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

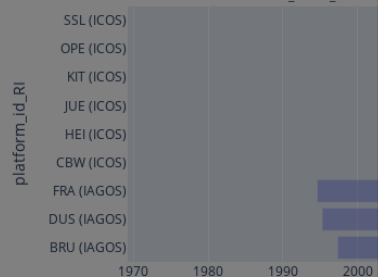
## Time-series analysis

Information

☐ compact view ☒ detailed view

var\_codes\_filtered

☒ O3 ☐ CO2

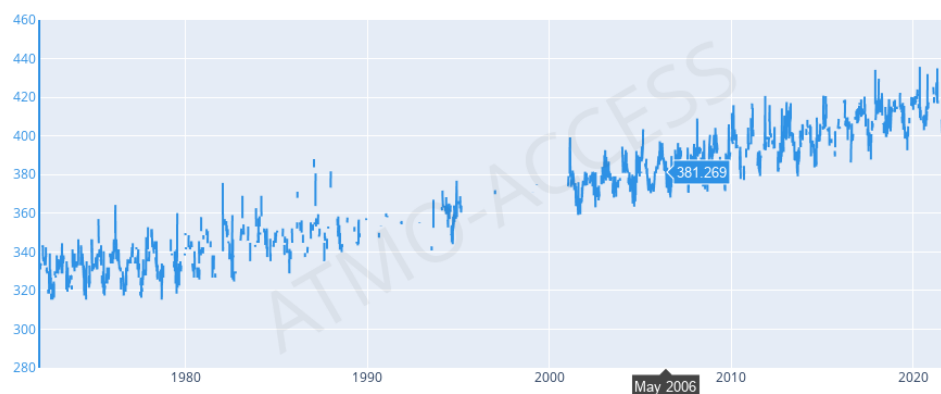


☐ Select all / none

#	#Title	#Variables	#RI	#Station	#Station code	#Start	#End
<input checked="" type="checkbox"/>	Atmospheric CO2 product, Schauinsland (12.0 m)	CO2	ICOS	Schauinsland	SSL	1972-01-01	2022-02-28
<input type="checkbox"/>	Atmospheric CO2 product, Schauinsland (35.0 m)	CO2	ICOS	Schauinsland	SSL	2021-01-01	2022-02-28

Atmospheric CO2 product, Schauinsland (12.0 m)

Atmospheric CO2 product, Schauinsland (12.0 m)



Data analysis

var\_codes\_filtered=CO

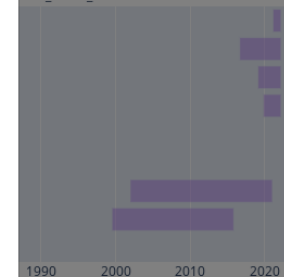


Figure 7. The quicklook plot of a single dataset



This work has received funding from the European Union's Horizon 2020 research and innovation programme through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004

[atmo-access.eu](https://atmo-access.eu)



#### 2.2.4. Filter data

Here the user is able to filter the time series of observations contained in datasets chosen in the “Select datasets” step. A filter can be composed of:

- a time filter: it filters on time period,
- a filter on variable's values: for each variable a filter defined by min and max value thresholds can be set up.

The latter kind of filters be applied either:

- as a “simple filter”: it means that a variable filter (defined by min and max thresholds) applies only to the variable in question;
- or as a “cross filter,” which works as follows: an ensemble of variables is considered jointly (as a time dependent multivariate observation) and only those observations which satisfy all the defined filters are selected. Since each variable can have different time resolution, the notation of joint observation is further parametrized by the user's choice of “observation coincidence time”, which can vary from 1 hour to 30 days.

The filters are accompanied with preview plots in a form of histograms showing distribution of variables' values as well as variables' data availability in time (see Figure 8).

The GUI provides the user with two ways to set up the filters:

- manual input of numerical values for min/max threshold filters or dates for the time filter;
- drag-and-drop type selection on the variable's histogram plot or on the plot representing data availability in time.

After setting up the filters, the user can pass to the next stage by clicking on the button **Apply filter to data**.

# Time-series analysis

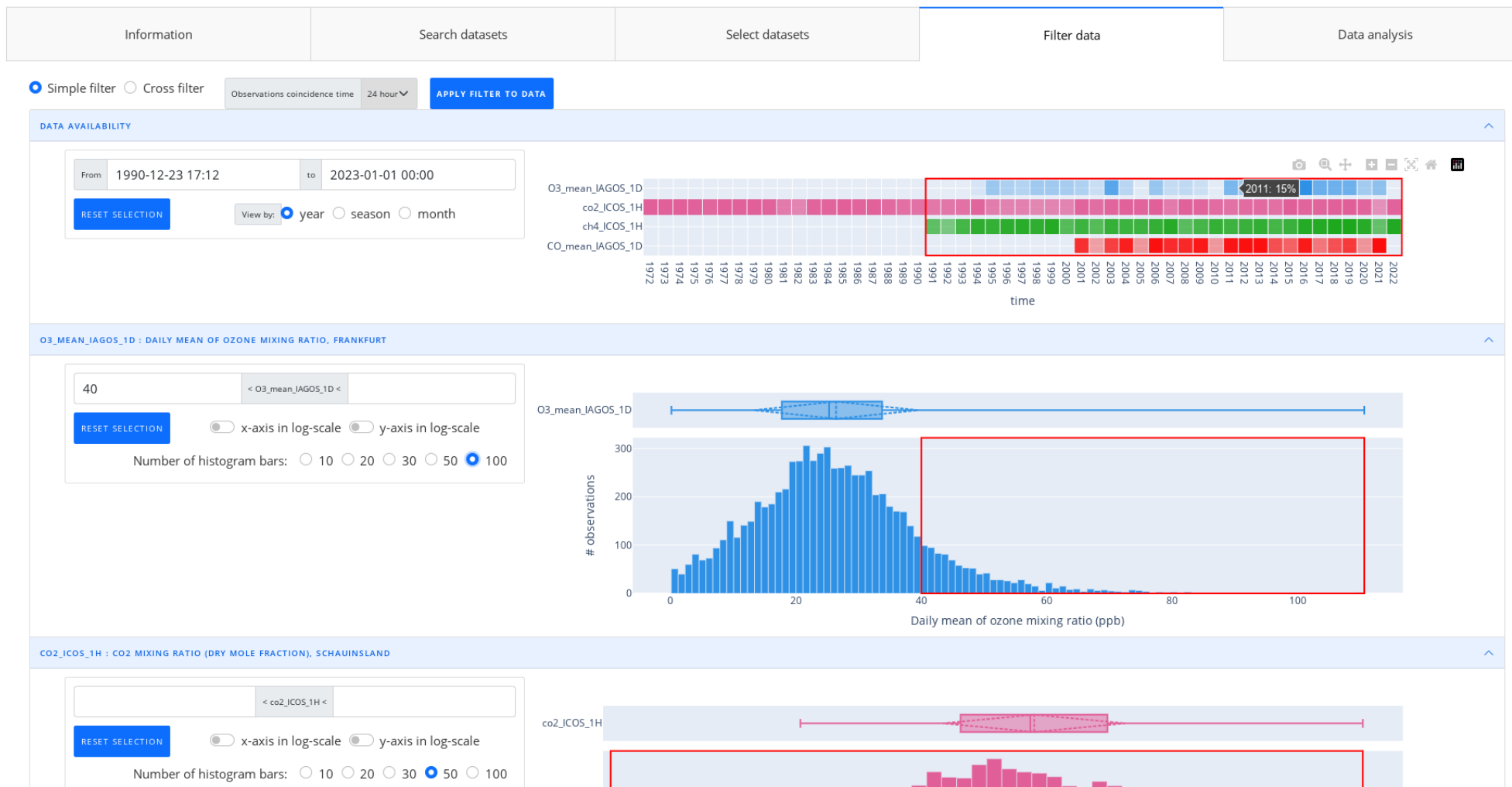


Figure 8. Filter data



### 2.2.5. Data analysis

Here the user performs an actual analysis of the data. A result of the analysis is displayed in a form of an **annotated plot** which is fully **interactive**: the user can freely zoom-in or pan onto an area of interest.

The figure and the data used to plot the figure can be downloaded. Both the data and the figure are provided under the Creative Commons Attribution 4.0 International licence (CC BY 4.0).

The analysis concerns one or several **variables**. The analysis itself is performed by using one of the available **methods**, each having its specific **parameters** to be set by the user. The methods are divided into three groups:

- **exploratory analysis**,
- **trend analysis**,
- **multivariate analysis**.

#### 2.2.5.1. *Exploratory analysis*

Analysis methods aiming at exploration of time series of individual variables are grouped into **exploratory analysis** methods. The methods are:

- **Gaussian means and standard deviations** (Figure 9)
- **Percentiles** (Figure 10)
- **Moving average** (Figure 11)

The user can choose one or several variables for the analysis. Despite the fact that the time series of each variable is processed independently from the others, the resulting plots are superposed in order to facilitate the user with a comparative analysis.

Parameters which are common for all three exploratory analysis methods are:

- **Aggregation period**: it defines the time periods over which the desired statistics (means, standard deviations, percentiles, etc.) of the time series should be calculated. In case of moving average, it is the size of the moving window.



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

- **Minimal sample size for period:** a minimal number of observations of a variable per aggregation period (or per a moving average window) required in order to calculate a statistic for a period.

Additionally, for the **percentile** method the user can choose which percentile(s) should be calculated and plotted.



# Time-series analysis

Information	Search datasets	Select datasets	Filter data	Data analysis
-------------	-----------------	-----------------	-------------	---------------

Exploratory analysis
Trend analysis
Multivariate analysis

Variables

☐ Select all / none  
☐ CO\_mean\_IAGOS\_1D : Daily mean of carbon monoxide mixing ratio  
☐ O3\_mean\_IAGOS\_1D : Daily mean of ozone mixing ratio  
☒ ch4\_ICOS\_1H : CH4 mixing ratio (dry mole fraction)  
☒ co2\_ICOS\_1H : CO2 mixing ratio (dry mole fraction)

Analysis method

☒ Gaussian mean and std  
☐ Percentiles  
☐ Moving average

Parameters

Aggregation period:  
☐ day  
☐ week  
☐ month  
☒ season  
☐ year

Minimal sample size for period: 5

☒ Show standard deviation with ☒ fill ☐ error bars

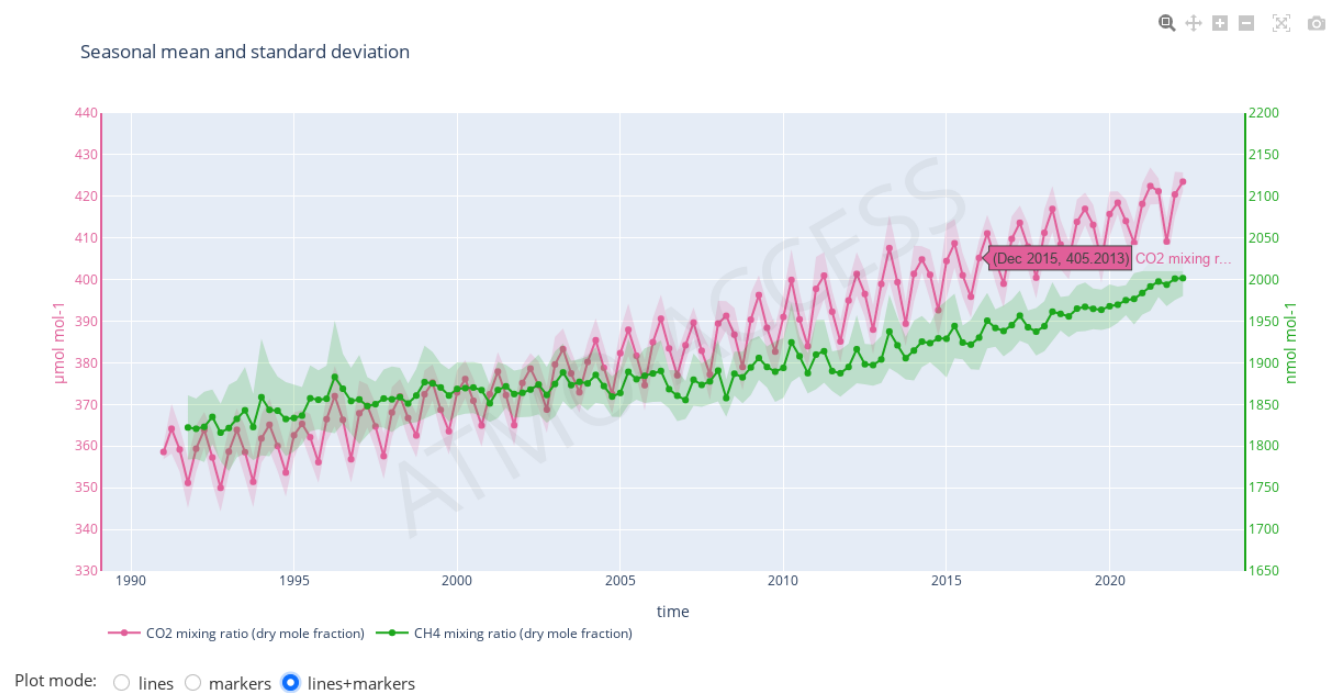


Figure 9. Exploratory analysis: Gaussian mean and standard deviation



ATMO ACCESS  
Access to Atmospheric Research Facilities

## Time-series analysis

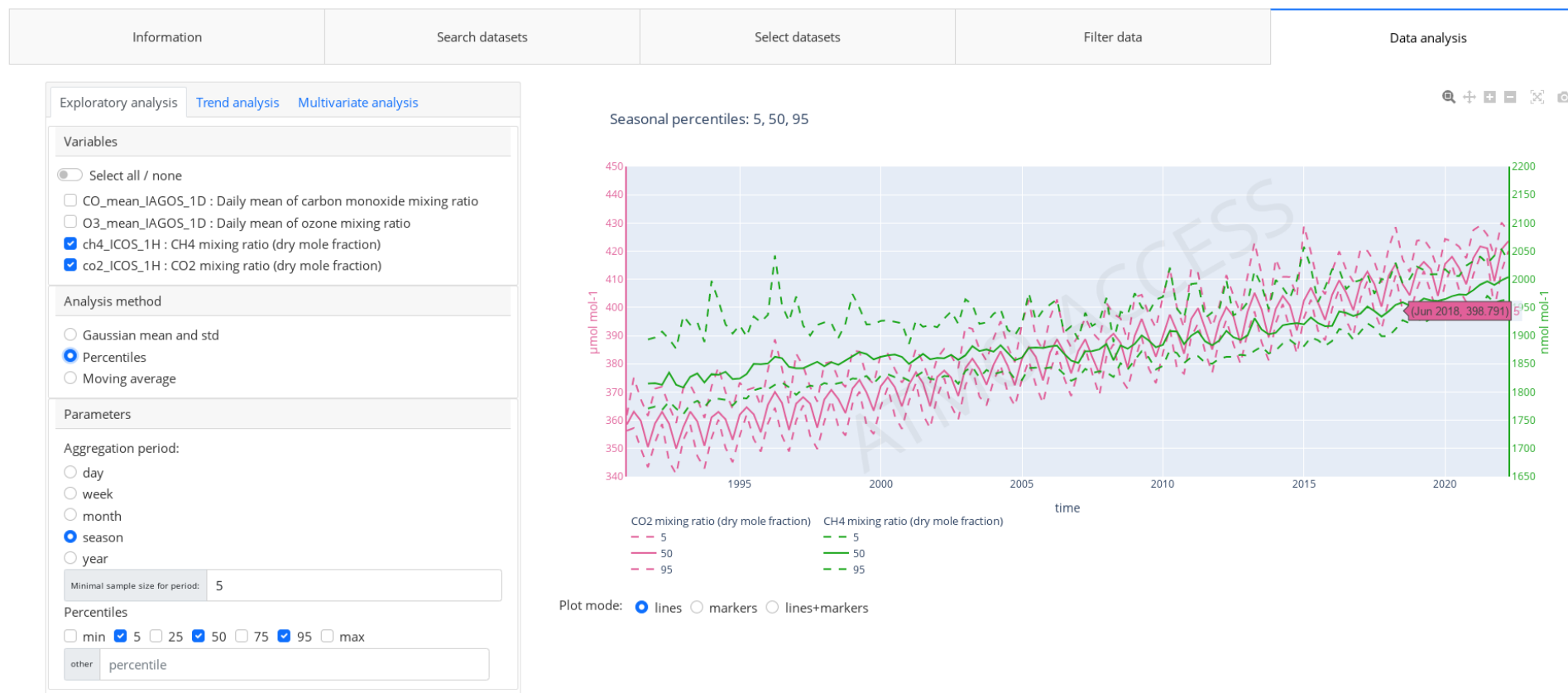


Figure 10. Exploratory analysis: percentiles





## Time-series analysis



Figure 11. Exploratory analysis: moving average



#### 2.2.5.2. *Trend analysis*

Analysis methods aiming at identifying and estimating trends in time series of individual variables are grouped as **trend analysis** methods (see Figure 12). The available trend estimation methods are:

- **Linear fit method**
- **Non-parametric Mann-Kendall test**
- **Theil-Sen slope estimate**

Similarly, to the Exploratory analysis, the user can choose one or several variables for the trend analysis.

Time series can be prepared for the trend estimation by applying operations from the following chain of operations:

- **Time filter** (allows to further refine the time domain)
- **Aggregation** of time series using daily / weekly / monthly / seasonal / yearly mean or median
- **Deseasonalising**
- **Moving average** with window size equal to day / week / month / season / year (available only if the deseasonalising is applied).

Note that the above chain of operations is applied to an original time series in the above order, e.g. if a weekly aggregation and the deseasonalising is chosen, the seasonal component will be removed from weekly aggregated time series and the resulting time series will be used for the subsequent trend estimation.

The results of the trend analysis are visualised using several figures:

- **Original time series** on which the user can drag-and-drop in order to set up a time filter.
- **Trend**, which shows
  - **time series** which is used as an input for the trend estimation (e.g. the time series resulted in applying weekly mean and the removal of seasonal component, if those operations are chosen)



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

- **trend line** as estimated for the input time series if linear fit or Theil-Sen slope estimation is used.
- In the case of
  - Linear fit and Theil-Sen slope estimate:
    - **Trend rate and 95 % confidence interval of the trend rate estimation**
    - **Autocorrelation function** of the detrended time series (i.e. the time series used as an input for the trend estimation minus the estimated trend).
  - Non-parametric Mann-Kendall test:
    - **A result of the test.**





**ATMO ACCESS**  
Access to Atmospheric Research Facilities

## Time-series analysis

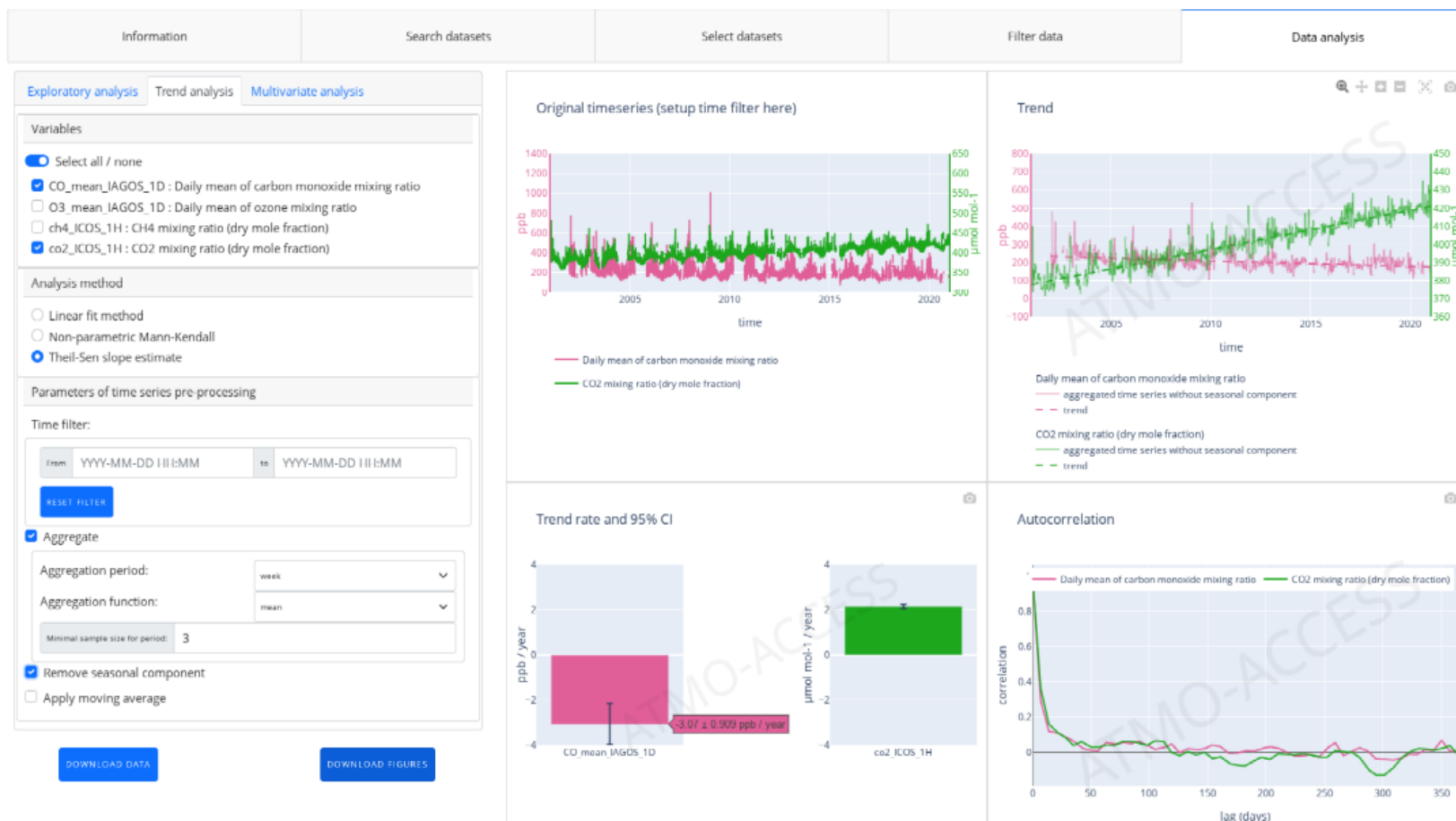


Figure 12. Trend analysis

[atmo-access.eu](https://atmo-access.eu)



#### 2.2.5.3. *Multivariate analysis*

Analysis methods aiming at exploring dependencies between two or three variables are grouped as **multivariate analysis** methods. The input data for these methods are prepared as follows:

- The user chooses two or three variables among those selected and filtered in the previous steps. The variables are labelled as X, Y (and C, in case of the choice of three variables), which determines how the variables are presented on a plot (see e.g. Figure 13).
- For each of the two or three variables to be analysed, a time series of observations of the variable issued from the “Filter data” step (see Section 2.2.4) are identified.
- The two or three time series identified above are merged together in a form of a multivariate time series; since each variable time series could possibly have a different time resolution, a time series upsampling method is used if necessary.
- Finally, the time axis is ignored; this results with an ensemble of multivariate observations, which constitutes data for a subsequent multivariate analysis.

The available multivariate analysis methods are:

- **Scatter plot** (see Figures 13, 14, 15 and 16)
- **Linear regression** (see Figure 17)

The **Scatter plot** method allows to **visually** explore a dependence between two or three variables.

In case of **two variables**, the dependence can be visualised using:

- 2D density plot in the form of **hexagonal bins** — each hexagonal bin on the XY-plane (where X and Y refer to the variables chosen by the user at the start of the analysis) is coloured according to a number of observations of X and Y variables falling into the bin (Figure 13)



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

- 2D plot of **individual observations** (Figure 14; this type of plot may be not very suitable in the case of a large number of observations).



# Time-series analysis

Information	Search datasets	Select datasets	Filter data	Data analysis
-------------	-----------------	-----------------	-------------	---------------

Exploratory analysis Trend analysis Multivariate analysis

Variables (1.125e+05 observations)

X	CO_mean_IAGOS_1D_layer=500m : Daily mean of carbon monoxide mixing ratio	▼
Y	O3_mean_IAGOS_1D_layer=500m : Daily mean of ozone mixing ratio	▼
C	—	▼

Analysis method

☒ Scatter plot  
☐ Linear regression

Parameters

Plot type: ☒ hexagonal bins ☐ individual observations

Hex-bin plot resolution

10 20 30 40 50 60

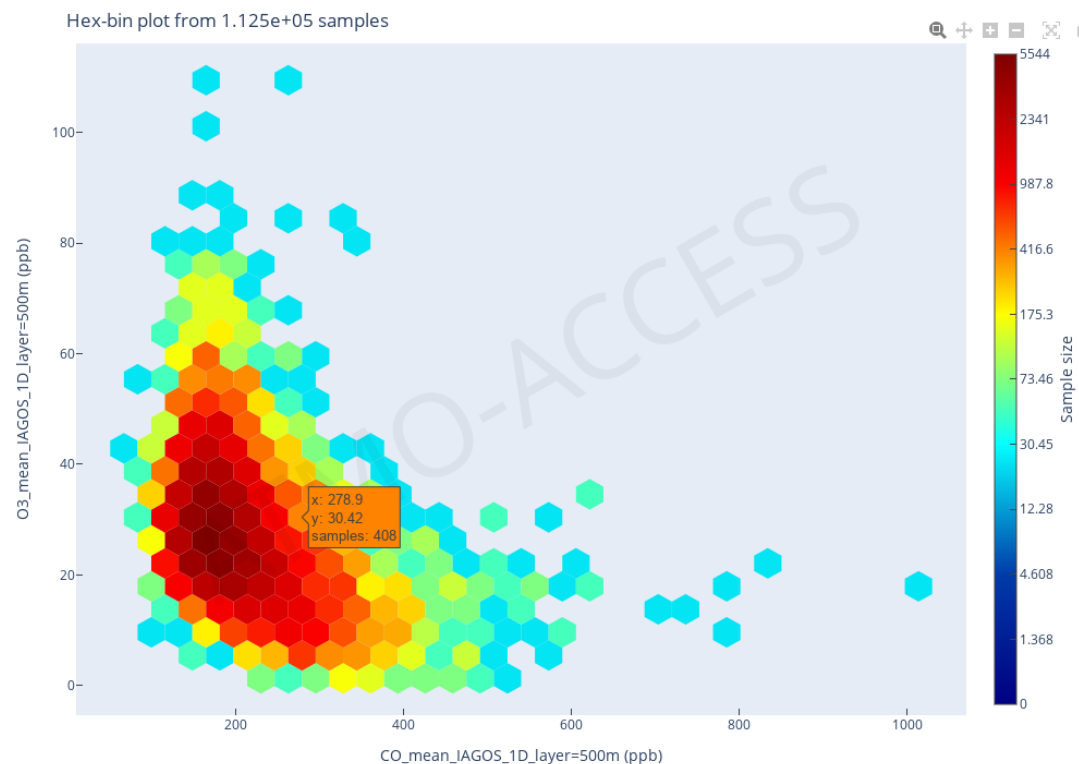


Figure 13. Multivariate analysis: 2D hex-bin density plot.





ATMO ACCESS  
Access to Atmospheric Research Facilities

## Time-series analysis

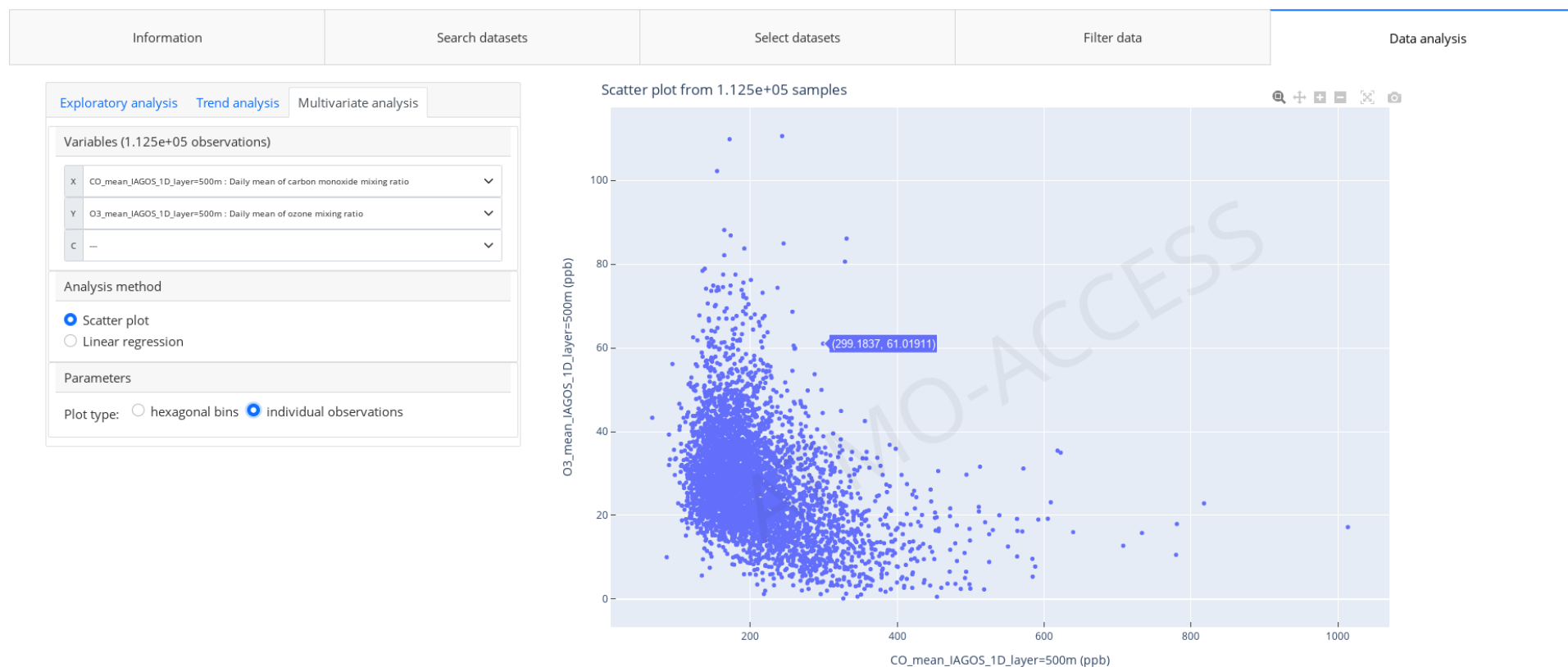


Figure 14. Multivariate analysis: 2D plot of individual observations



This work has received funding from the European Union's Horizon 2020 research and innovation programme through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004

[atmo-access.eu](https://atmo-access.eu)





**ATMO ACCESS**  
Access to Atmospheric Research Facilities

In the case of **three variables**, the available plots are similar:

- 3D plot in the form of hexagonal bins — each hexagon bin on the XY-plane is coloured according to an **aggregated value** (the mean or other statistics) of the **component C** of observations of X, Y and C falling into the bin. Moreover, the hexagon itself is scaled according to a number of observations of X, Y and C in the bin (i.e. smaller the hexagon, smaller the sample size); see Figure 15 for an example.
- 3D plot of **individual observations**, where observation of X, Y and C is represented as a point on the XY-plane with a colour corresponding to a value C of the observation (see Figure 16).

In the case of hex-bin type plots a resolution of the hexagonal lattice can be adjusted (from 10 to 60 hexagons along the X-axis).

Additionally, in the case of the 3D hex-bin plot, the aggregation function of C-variable can be chosen among:

- mean,
- mean +/- the standard deviation,
- 5th or 95th percentile,
- min or max.





## Time-series analysis

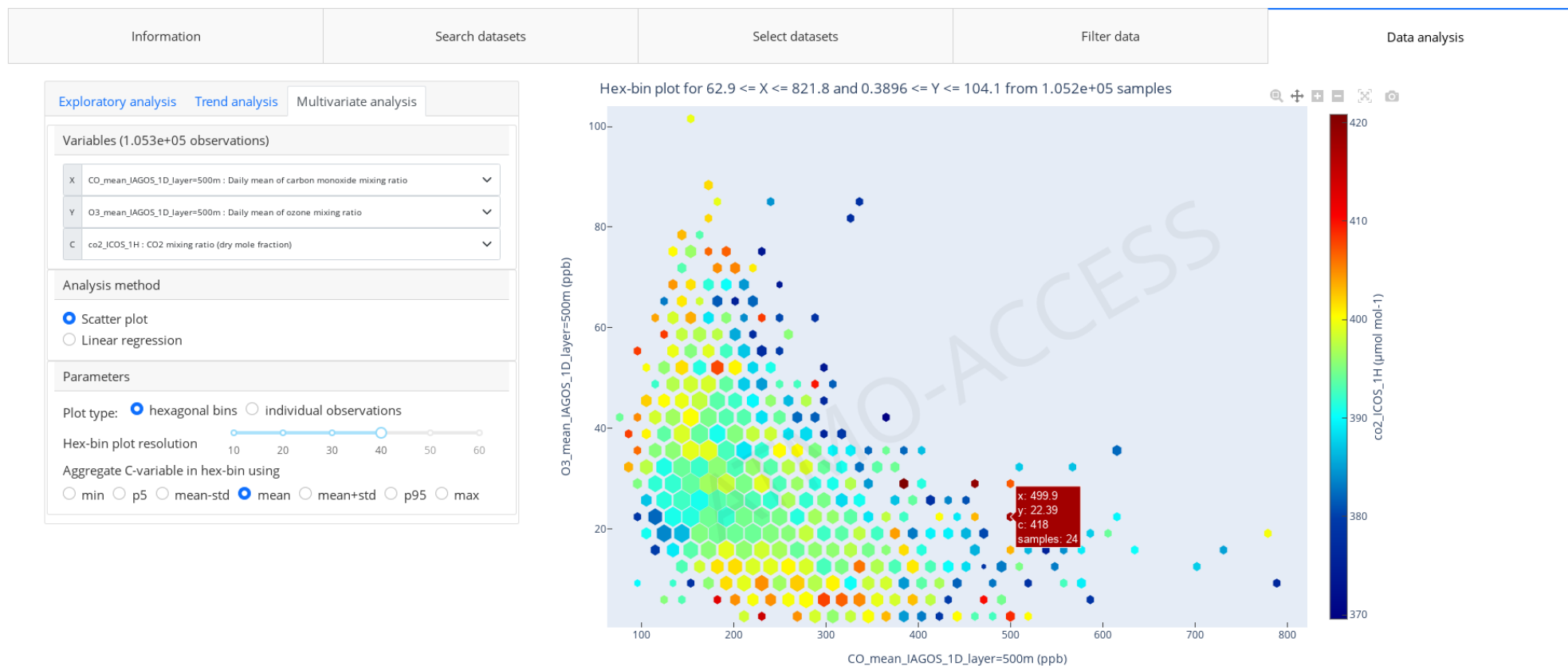


Figure 15. Multivariate analysis: 3D plot hex-bin plot



## Time-series analysis

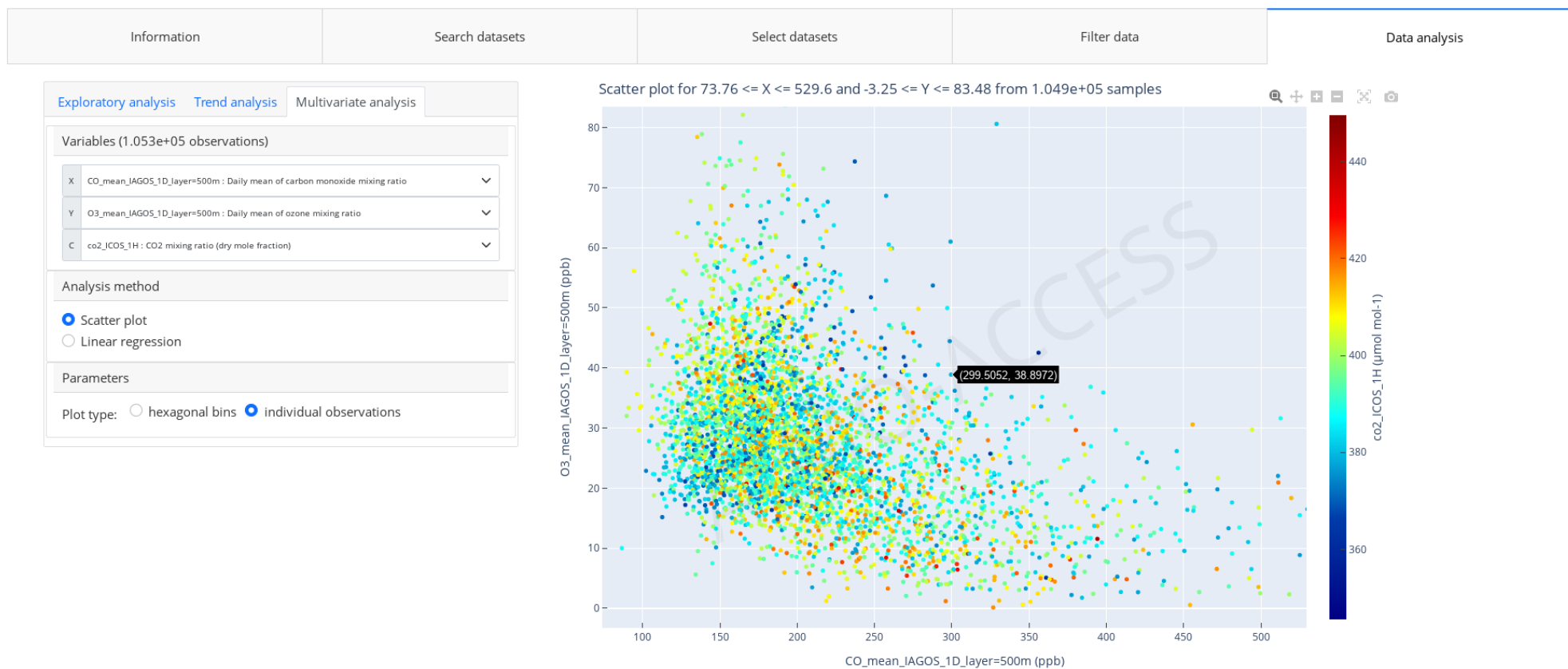


Figure 16. Multivariate analysis: 3D plot of individual observations.



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

The **linear regression** method quantifies a linear dependence of the variable Y on the variable X applying the ordinary least squares method. The result of such analysis is visualised in the form of either 2D hex-bin plot or 2D plot of individual observations (as described above) annotated with a linear equation of the form

$$Y = a X + b$$

which represents an estimated linear fit of Y to X. Moreover, the coefficient of determination  $R^2$  is provided. In the plot, the linear fit of Y to X is shown in the form of a thick red line; see Figure 17 for an example.





ATMO ACCESS  
Access to Atmospheric Research Facilities

## Time-series analysis

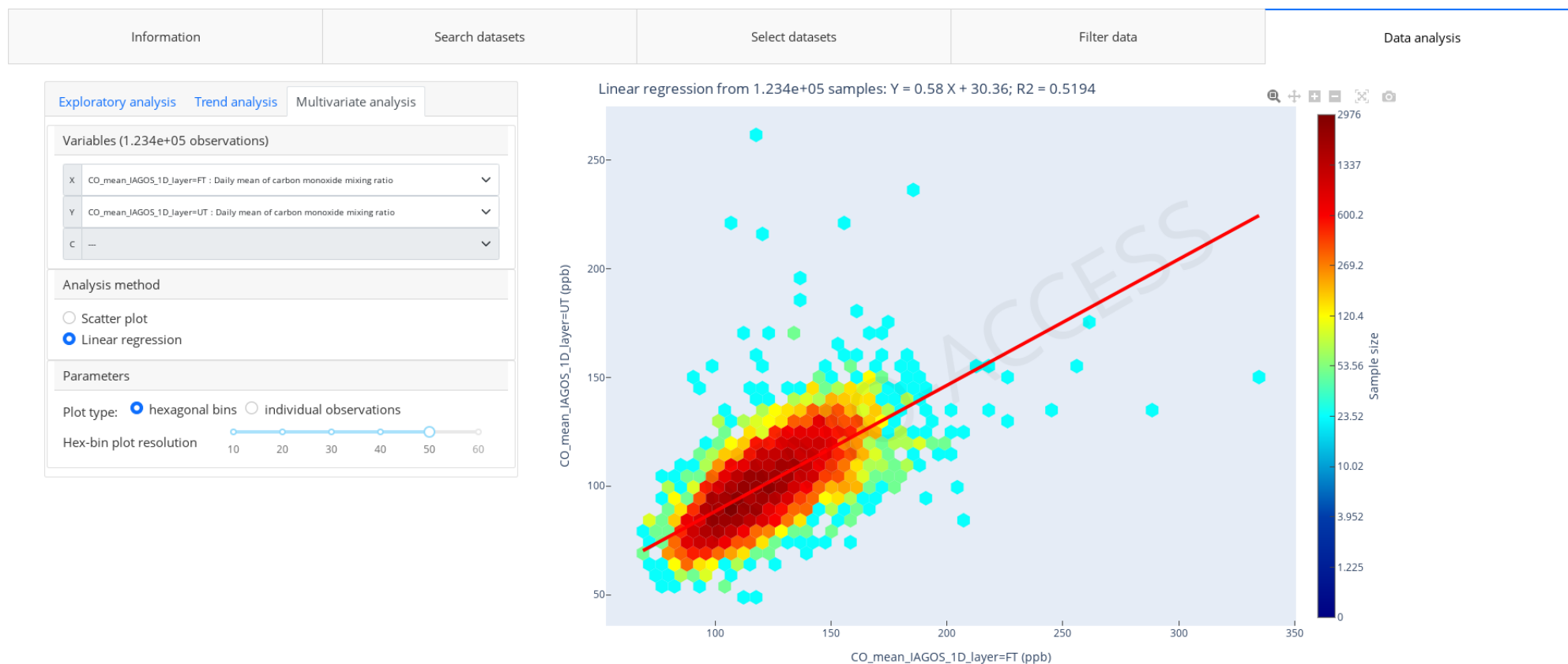


Figure 17. Multivariate analysis: linear regression



**ATMO ACCESS**  
Access to Atmospheric Research Facilities

## 2.2.6. Download

### 2.2.6.1. *Save figures*

Every figure visible in the **web user interface** can be downloaded in the PNG format by clicking on the toolbar icon  of the figure.

Moreover, the **Data analysis** tab has a **Download figure(s)** button (see e.g. Figure 12), which allows the user to download an interactive version of figure(s) in HTML format.

### 2.2.6.2. *Download results*

Data used to generate plot(s) of **Data analysis** can be downloaded in the NetCDF format by clicking on the button **Download data** (see e.g. Figure 12). Metadata containing information on:

- provenance of source datasets (i.e. the datasets chosen for the analysis),
- applied data filters,
- an analysis method and its parameters

is included.



### 3. Description of the notebook

#### 3.1. General overview

The **notebook** is an ensemble of programming recipes for accessing, analysing and visualising time series data provided by ACTRIS, IAGOS and ICOS RIs. The recipes are implemented in **Python** and they use the same stack of Python packages as the web user interface (see Figure 18). The notebook can be arbitrarily modified by the user in an ad hoc manner, or it can serve as a code example for the user's own Python scripts aiming at processing RIs data.

Along with the notebook itself, a Python software environment for running the notebook is provided. Technically speaking, it is a specification of Python packages which allows the user to easily install the environment using one of the popular tools for managing Python software environments such as conda or venv (Python virtual environment).

Typically, the user consult, run, edit and overall experiment with the notebook using **jupyter**, a web-browser based environment for notebook execution. Jupyter is included in the Python software environment mentioned above.

#### 3.2. Comparison with the web user interface

Although the notebook and the web user interface provide the access to the same data and rely on the same data processing, analysis and visualisation routines, the interest in using the notebook instead of only the web user interface comes from:

- flexibility in customising the workflow according to user's needs,
- possibility of the workflow automation,
- possibility of analysing external datasets using the data processing and data visualisation routines provided with the notebook,
- possibility of extending the existing toolbox of data analysis and data visualisation routines (e.g. the user can add her/his own alternative routines for trend analysis, result visualisation, etc.).

The flexibility and reusability of the notebook follows from the underlying design principles which assume consistent usage of well-established data containers as interfaces between different data access, data processing and data visualisation routines. The data containers are:

- **DataFrame** container for tabular-like data (e.g. tables with datasets' metadata), implemented by pandas, a popular Python package for data analysis.





**ATMO ACCESS**  
Access to Atmospheric Research Facilities

- **Dataset** container for NetCDF-like data, implemented by xarray (a Python package for processing NetCDF-like data).

All the advantages of the notebook come with the cost of a steeper learning curve in comparison to using the web user interface and assume that the user has basic programming skills.

### 3.3. Notebook workflow

The following diagram presents the workflow which is implemented in the notebook. The workflow contains various data access, data analysis and data visualisation tasks and is meant to be an example for similar user's workflows.





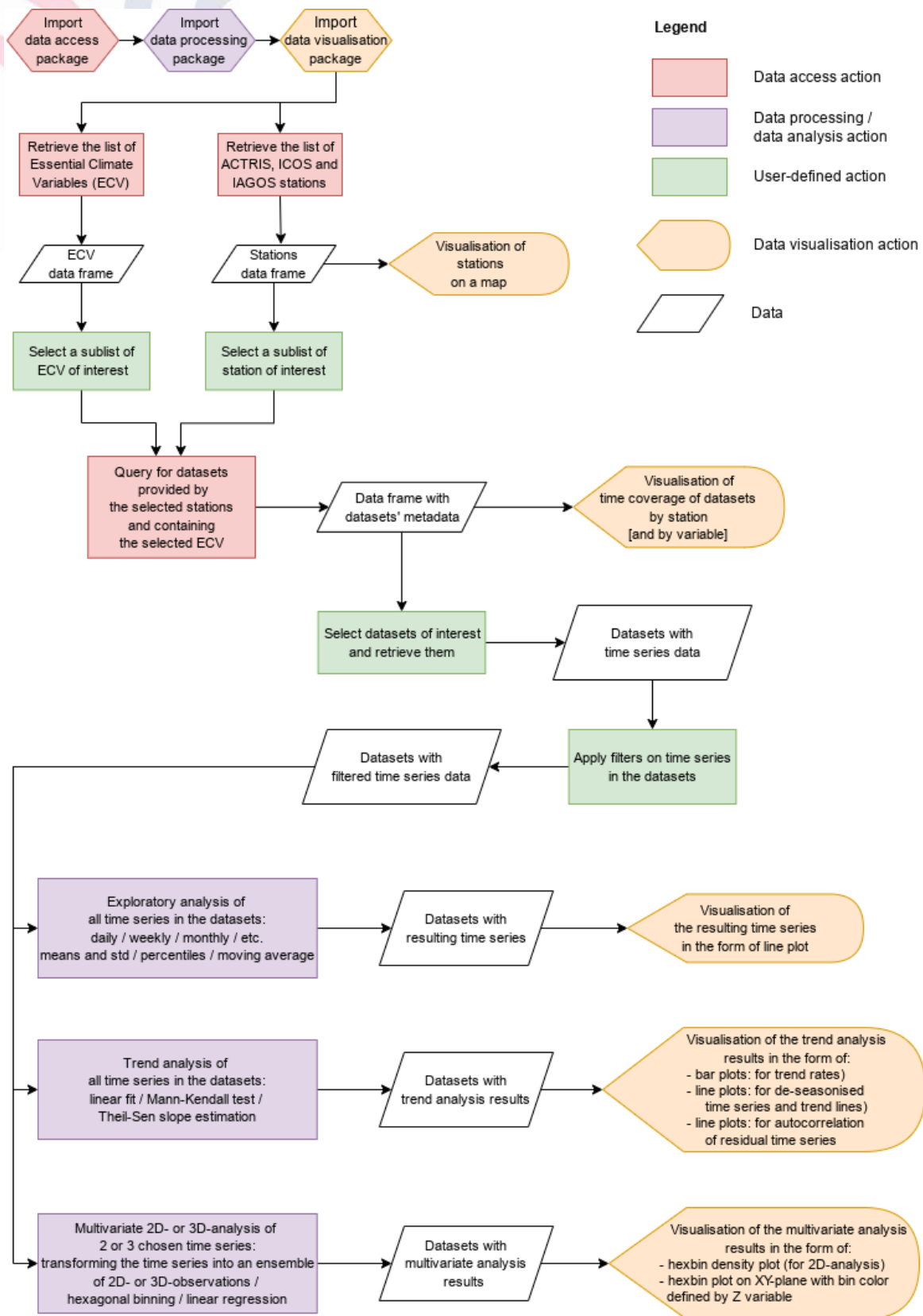


Figure 18. Notebook workflow





## Appendix A: Glossary

ICOS	Integrated Carbon Observation System
ACTRIS	The Aerosol, Clouds and Trace Gases Research Infrastructure
IAGOS	In-service Aircraft for a Global Observing System
PID	Persistent Identifier
VRE	Virtual Research Environment
RI	Research Infrastructure
ENVRI	Environmental Research Infrastructures
FAIR	Findability, Accessibility, Interoperability, Reusability
API	Application Programming Interface
GUI	Graphical User Interface
NetCDF	Network Common Data Form — a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.



## Appendix B: Table of Figures

Figure 1. General scheme of the system components (software and data).....	6
Figure 2. The workflow from a user's perspective .....	9
Figure 3. The welcome page of the application after the user's login with the active "Information" tab. ....	11
Figure 4. The "Search datasets" tab.....	13
Figure 5. Select datasets: a timeline coverage plot in a compact form and a table with individual datasets. The selection of horizontal bars on the plot is synchronised with the table. .....	16
Figure 6. Select datasets: a timeline coverage plot in a detailed form and a table with individual datasets. The selection of horizontal bars on the plot is synchronised with the table. .....	17
Figure 7. The quicklook plot of a single dataset.....	18
Figure 8. Filter data.....	20
Figure 9. Exploratory analysis: Gaussian mean and standard deviation .....	23
Figure 10. Exploratory analysis: percentiles .....	24
Figure 11. Exploratory analysis: moving average.....	25
Figure 12. Trend analysis .....	28
Figure 13. Multivariate analysis: 2D hex-bin density plot.....	31
Figure 14. Multivariate analysis: 2D plot of individual observations.....	32
Figure 15. Multivariate analysis: 3D plot hex-bin plot .....	34
Figure 16. Multivariate analysis: 3D plot of individual observations.....	35
Figure 17. Multivariate analysis: linear regression .....	37
Figure 18. Notebook workflow .....	41