Barcelona Supercomputing Center
Centro Nacional de Supercomputación

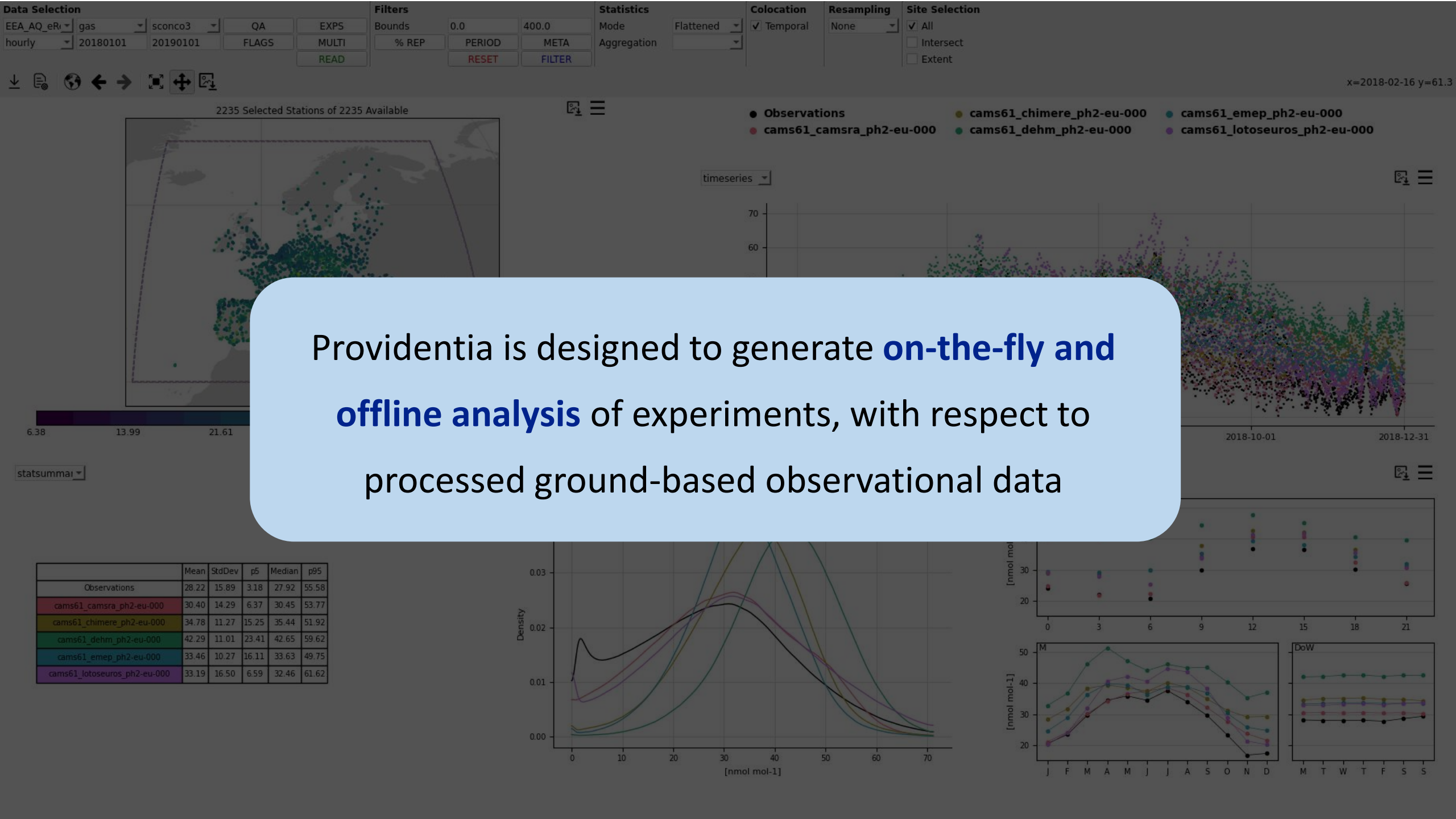
Providentia v2.2
Training Session

16/10/2023

Alba Vilanova | Dene Bowdalo

Introduction

Barcelona Supercomputing Center
Centro Nacional de Supercomputación



	Mean	StdDev	p5	Median	p95
Observations	28.22	15.89	3.18	27.92	55.58
cams61_camsra_ph2-eu-000	30.40	14.29	6.37	30.45	53.77
cams61_chimere_ph2-eu-000	34.78	11.27	15.25	35.44	51.92
cams61_dehm_ph2-eu-000	42.29	11.01	23.41	42.65	59.62
cams61_emep_ph2-eu-000	33.46	10.27	16.11	33.63	49.75
cams61_lotoseuros_ph2-eu-000	33.19	16.50	6.59	32.46	61.62

Data Selection

EEA_AQ_eR | gas | sconco3 | QA | EXPS

hourly | 20180101 | 20190101 | FLAGS | MULTI | READ

Filters

Bounds | 0.0 | 400.0

% REP | PERIOD | META

RESET | FILTER

Statistics

Mode: Flattened

Aggregation:

Colocation

☑ Temporal

Resampling

None

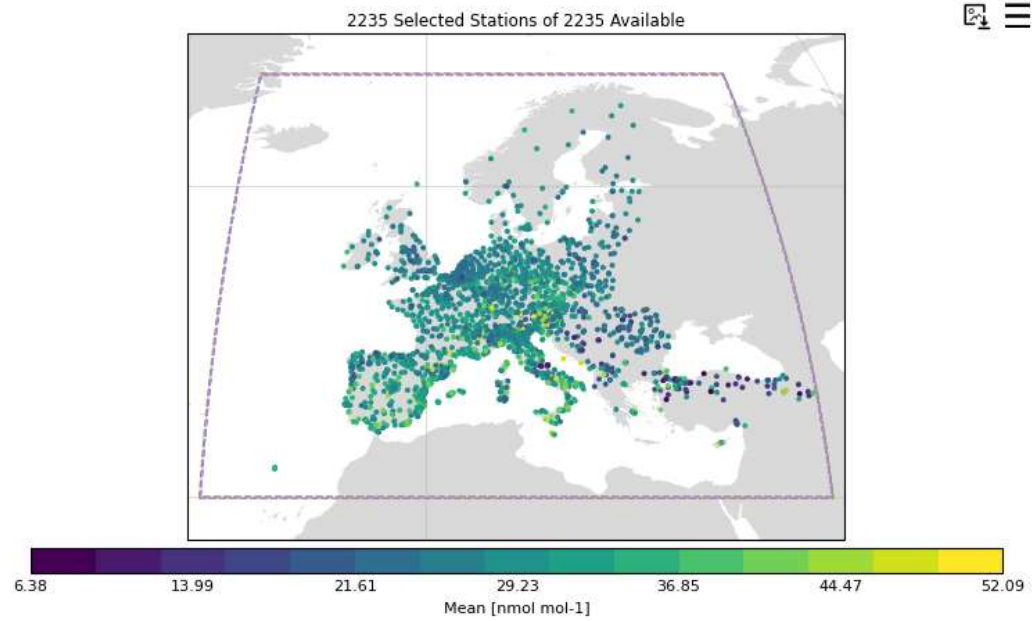
Site Selection

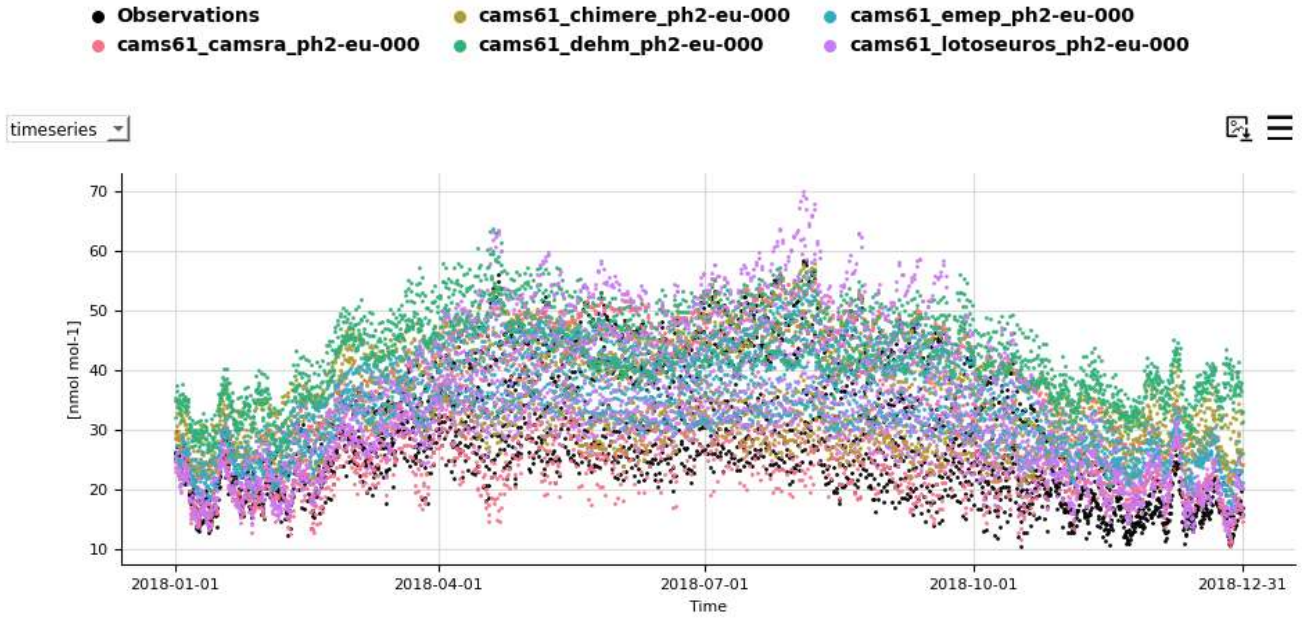
☑ All

☐ Intersect

☐ Extent

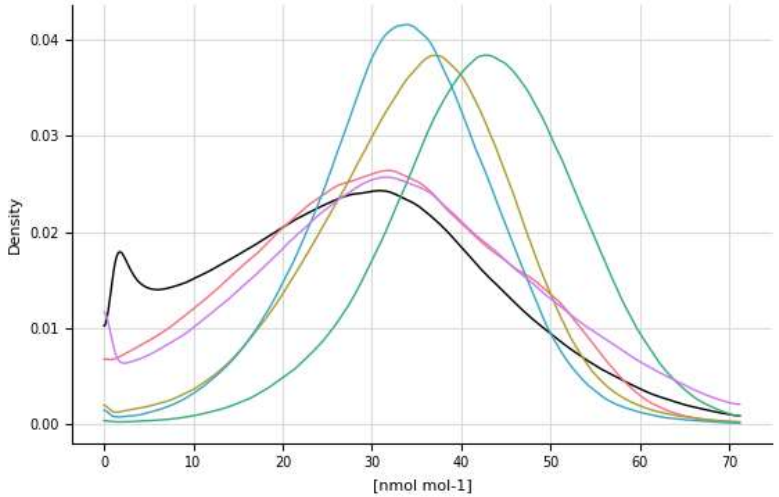
x=2018-02-16 y=61.3

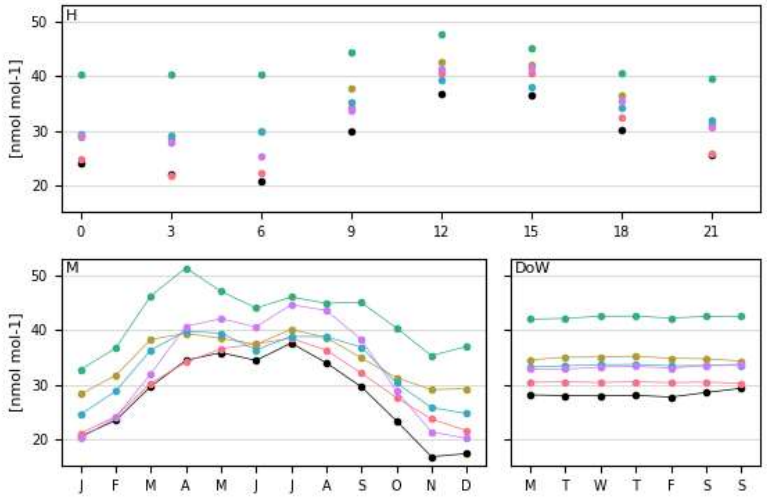




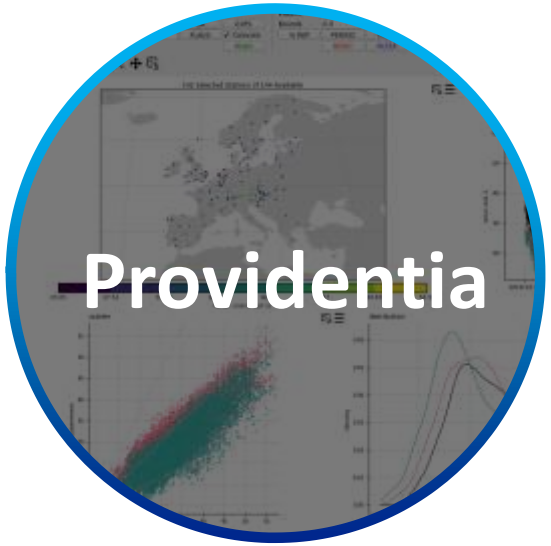
statsummar

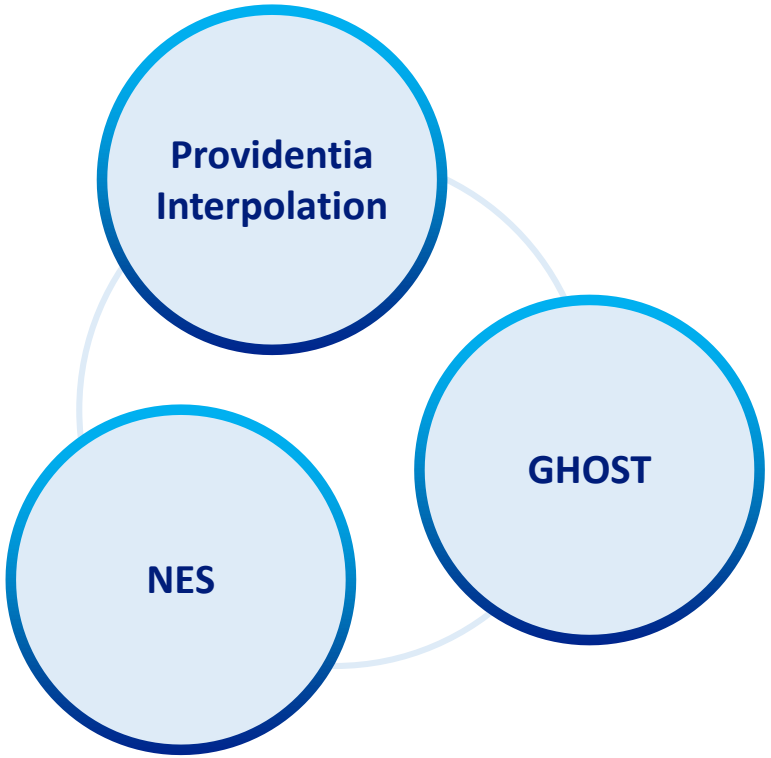
	Mean	StdDev	p5	Median	p95
Observations	28.22	15.89	3.18	27.92	55.58
cams61_camsra_ph2-eu-000	30.40	14.29	6.37	30.45	53.77
cams61_chimere_ph2-eu-000	34.78	11.27	15.25	35.44	51.92
cams61_dehm_ph2-eu-000	42.29	11.01	23.41	42.65	59.62
cams61_emep_ph2-eu-000	33.46	10.27	16.11	33.63	49.75
cams61_lotoseuros_ph2-eu-000	33.19	16.50	6.59	32.46	61.62





Related tools





Providentia Interpolation

Providentia Interpolation

To allow an experiment to be evaluated in Providentia, the model grid needs to be first interpolated to observations

Details: https://earth.bsc.es/gitlab/ac/providentia-interpolation

GHOST

Project dedicated to the harmonisation of global surface observations (most notably air quality pollutants) and metadata

Details: https://earth.bsc.es/gitlab/ac/GHOST

GHOST

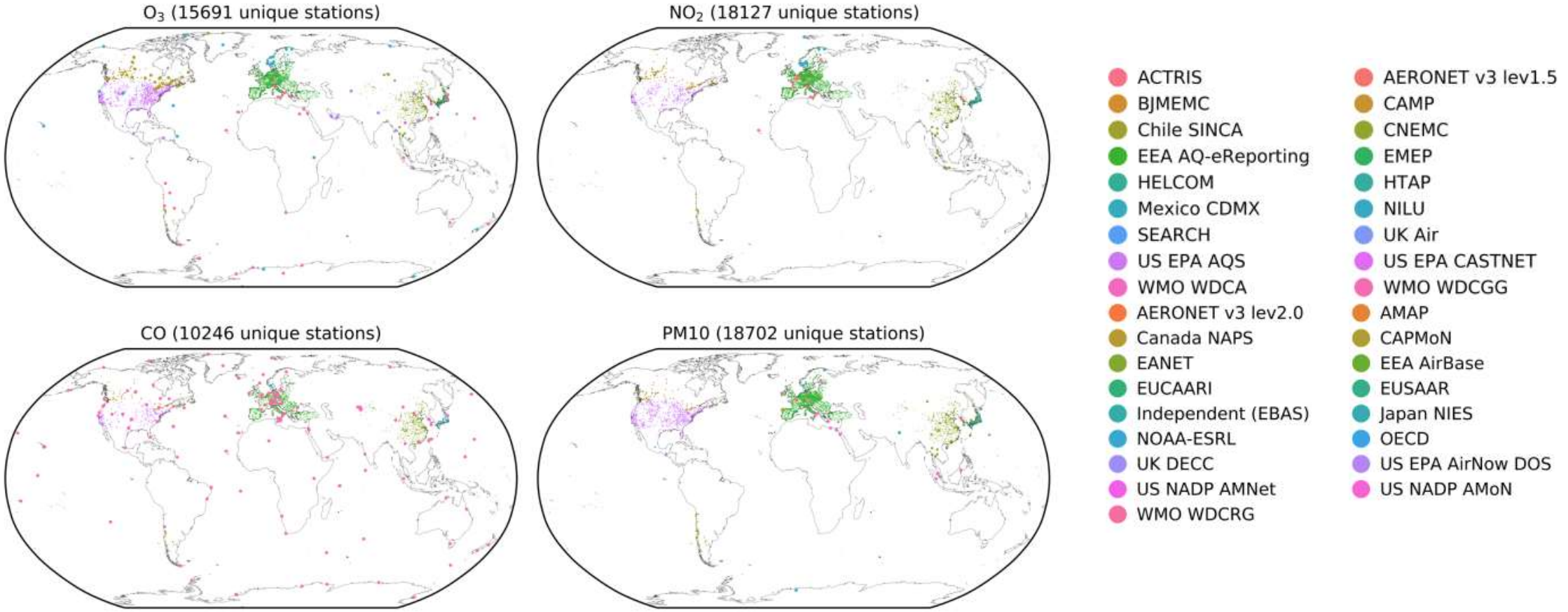
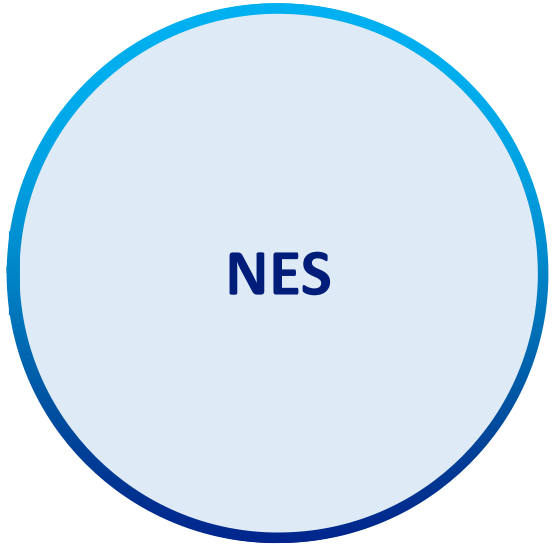


Figure 1. Unique stations per pollutant and network that have been harmonized using GHOST.

NES



Used to create observational networks from scratch that can be read by Providentia (XVPCA, CSIC, Port Barcelona, etc.)

Details: https://earth.bsc.es/gitlab/es/NES

Set up


BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Steps

1. Connect to HPC machines
2. Clone Providentia Interpolation and Providentia
3. Interpolate your experiments
4. Launch the dashboard or offline reports

Step 1: Connect to HPC machines

In your local machine**, open and edit the SSH configuration file** with:

```
$ vi .ssh/config
```

```
Host mn1
        HostName mn1.bsc.es
        User bsc32XXX
        IdentityFile ~/.ssh/id_rsa
        ForwardX11Trusted yes
        ForwardX11 yes
        Compression yes
        Ciphers aes128-gcm@openssh.com
        ForwardX11Timeout 7d
```

```
Host nord3v2
        HostName nord4.bsc.es
        User bsc32781
        IdentityFile ~/.ssh/id_rsa
        ForwardX11Trusted yes
        ForwardX11 yes
        Compression yes
```

Note that the options for Nord3v2 do not include the line with Ciphers. Note also that compression is turned on.

Step 2: Clone Providentia Interpolation

1. Enter the project's GitLab page:

   https://earth.bsc.es/gitlab/ac/providentia-interpolation

2. In Clone copy link from Clone with HTTPS
3. In your terminal, clone using:

```
$ git clone --recurse-submodules https://earth.bsc.es/gitlab/ac/providentia-interpolation.git
```

4. Transfer the folder to the path you usually work. (e.g. your gpfs scratch)

```
$ scp -r providentia-interpolation bsc32XXX@dt01.bsc.es:/gpfs/scratch/bsc32/bsc32XXX/
```

Step 2: Clone Providentia

1. Enter the project's GitLab page:

   https://earth.bsc.es/gitlab/ac/Providentia

2. In Clone copy link from Clone with HTTPS
3. In your terminal, clone the repo:

```
$ git clone https://earth.bsc.es/gitlab/ac/Providentia.git
```

4. Change to the production branch:

```
$ git checkout production
```

5. Transfer the folder to the path you usually work. (e.g. your gpfs scratch)

```
$ scp -r Providentia bsc32XXX@dt01.bsc.es:/gpfs/scratch/bsc32/bsc32XXX/
```

Step 3: Interpolate your experiments

1. Add the experiment path to **defined_experiments.py**

```
defined_experiments_dictionary = {# ------------------------------------------------------------ #

                                  # MONARCH experiments
                                  'b007': {'esarchive': '/esarchive/exp/nmmb-bsc-ctm/b007/'},
                                  'a1wd': {'esarchive': '/esarchive/exp/monarch/a1wd/'},
                                  'a1vv': {'esarchive': '/esarchive/exp/monarch/a1vv/'},
                                  'a1vw': {'esarchive': '/esarchive/exp/monarch/a1vw/'},
                                  'a1x8': {'esarchive': '/esarchive/exp/monarch/a1x8/'},
                                  'a1xa': {'esarchive': '/esarchive/exp/monarch/a1xa/'}, ...
                                  }
```

Step 3: Interpolate your experiments

2. Edit the configuration inside **configuration.py**

```
qos = 'default’
GHOST_version = 'default’
n_neighbours_to_find = 'default’
start_date = '201801' # YYYYMM START FROM THIS POINT
end_date = '201802' # YYYYMM GO UP TO THIS POINT
experiments_to_process = ['cams61_emep_ph2’]
species_to_process = ['sconco3’]
grid_types_to_process = ['default’]
ensemble_options = ['default’]
networks_to_interpolate_against = ['EBAS’]
temporal_resolutions_to_output = ['hourly']
```

Step 3: Interpolate your experiments

3. Submit the interpolation job to the queue

```
$ sbatch experiment_interpolation_submit.sh
```

4. Check if the job has successfully finished in **management_logs**
5. Check the outputs in:

/gpfs/projects/bsc32/AC_cache/recon/exp_interp

Due to inconsistencies in a dependency module, the interpolation will only run in parallel in Power9 and MN4, but in serial in Nord3v2.

Step 4: Launch Providentia

Launch the dashboard using:

```
$ ./bin/providentia
```

It is also possible to launch the dashboard with a configuration file:

```
$ ./bin/providentia --config=configurations/training.conf
```

The modules will automatically load and the allocation in the machine (either MN4 or Nord3) will be requested. When we are granted the allocation, the dashboard of Providentia will initialize.

Launch the offline reports by just adding **--offline** as an argument:

```
$ ./bin/providentia --config=configurations/training.conf --offline
```

Step 4: Launch Providentia

Using the debug mode you can avoid losing the job allocation when you close Providentia or the tool crashes unexpectedly. To do this you can:

1. Activate the debug mode

```
$ ./bin/providentia --debug
```

2. Run tool as usual

```
$ ./bin/providentia
```

Dashboard

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Data Selection

EEA_AQ_eR | gas | sconco3 | QA | EXPS

hourly | 20180101 | 20190101 | FLAGS | MULTI | READ

Filters

Bounds | 0.0 | 400.0

% REP | PERIOD | META

RESET | FILTER

Statistics

Mode: Flattened

Aggregation:

Colocation

☑ Temporal

Resampling

None

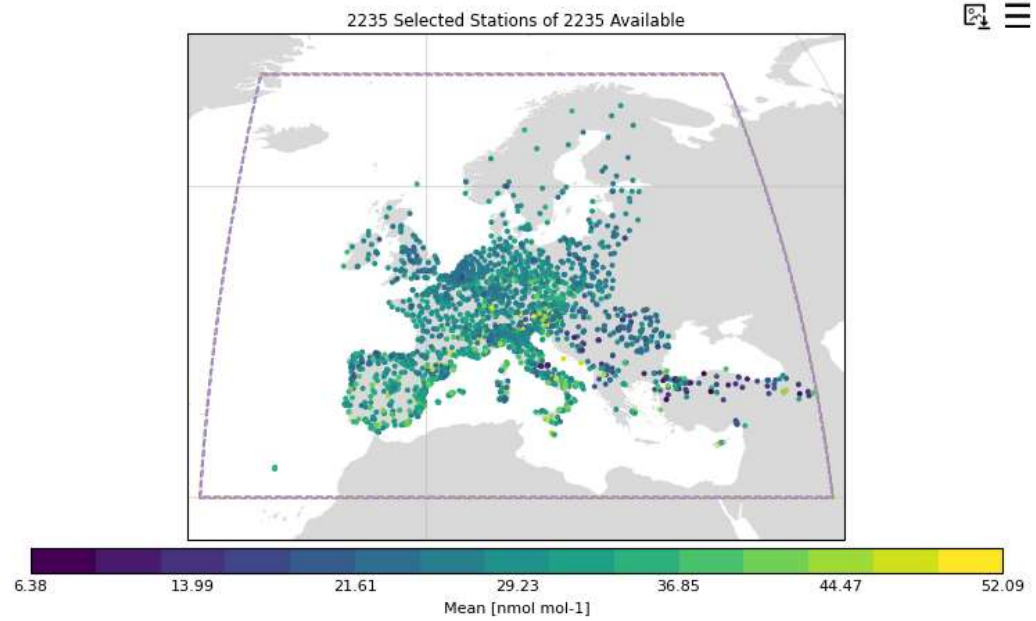
Site Selection

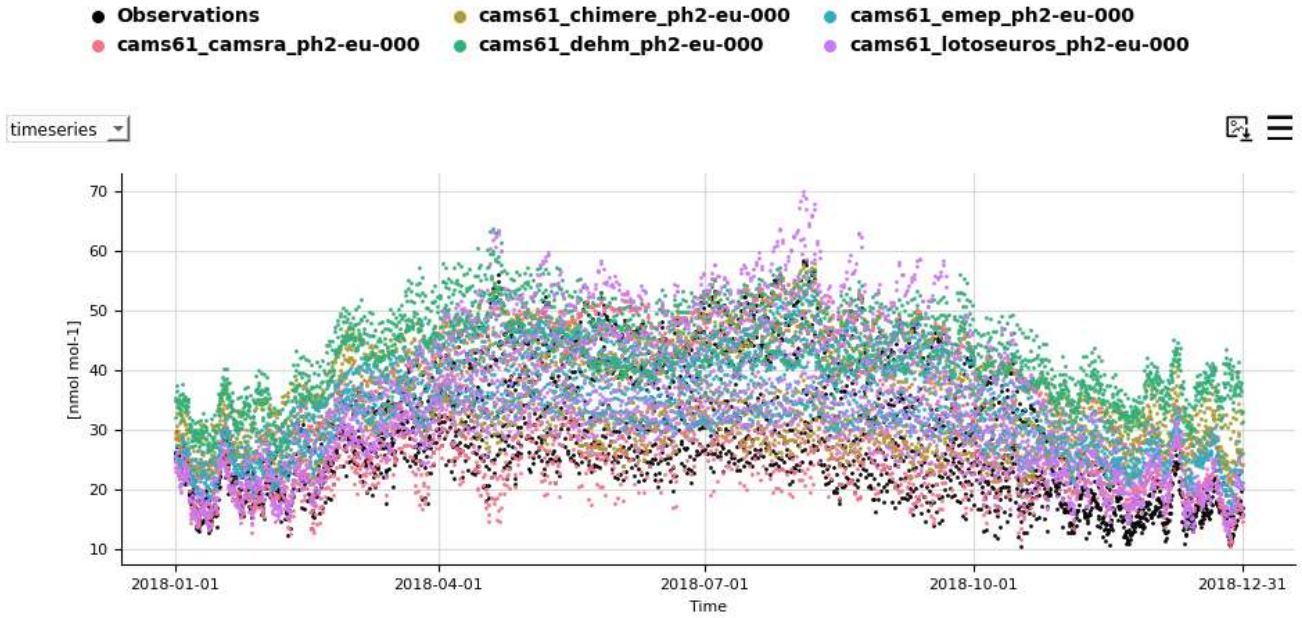
☑ All

☐ Intersect

☐ Extent

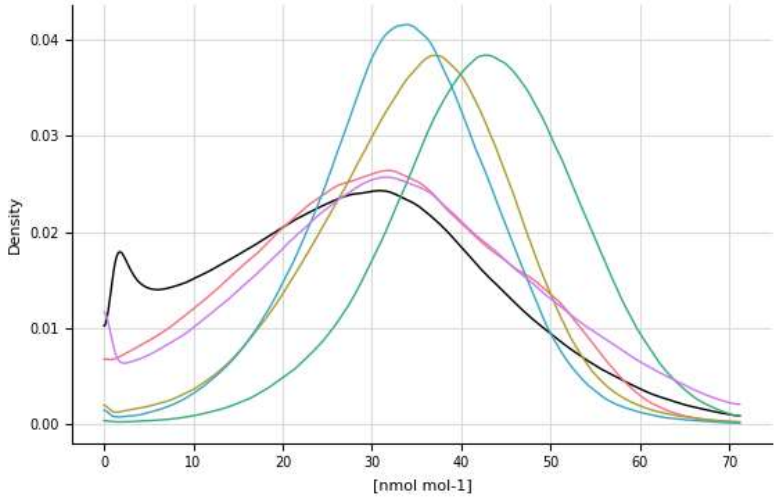
x=2018-02-16 y=61.3

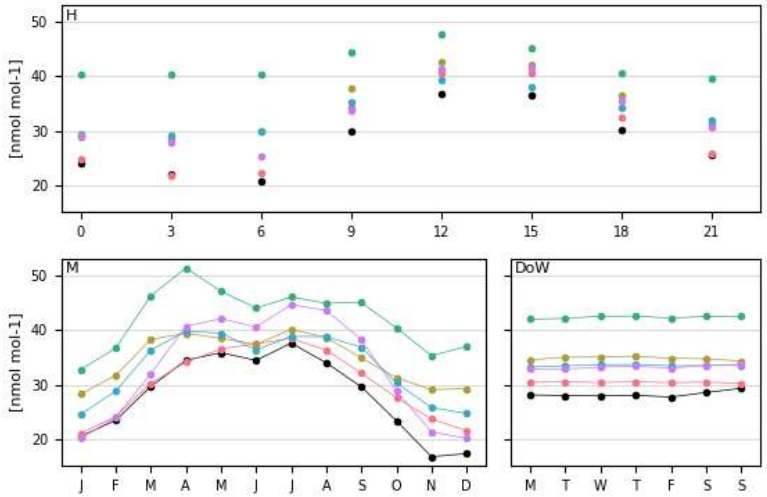




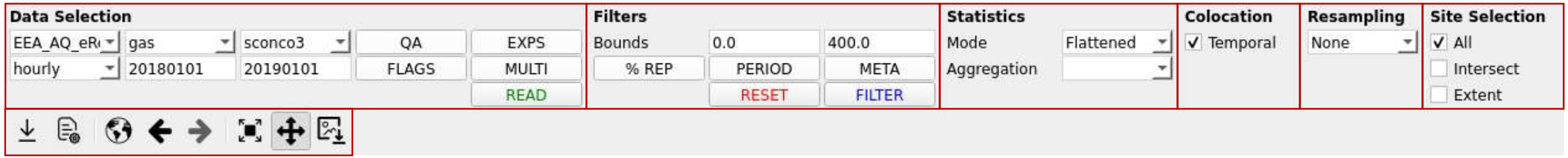
statsummar

	Mean	StdDev	p5	Median	p95
Observations	28.22	15.89	3.18	27.92	55.58
cams61_camsra_ph2-eu-000	30.40	14.29	6.37	30.45	53.77
cams61_chimere_ph2-eu-000	34.78	11.27	15.25	35.44	51.92
cams61_dehm_ph2-eu-000	42.29	11.01	23.41	42.65	59.62
cams61_emep_ph2-eu-000	33.46	10.27	16.11	33.63	49.75
cams61_lotoseuros_ph2-eu-000	33.19	16.50	6.59	32.46	61.62

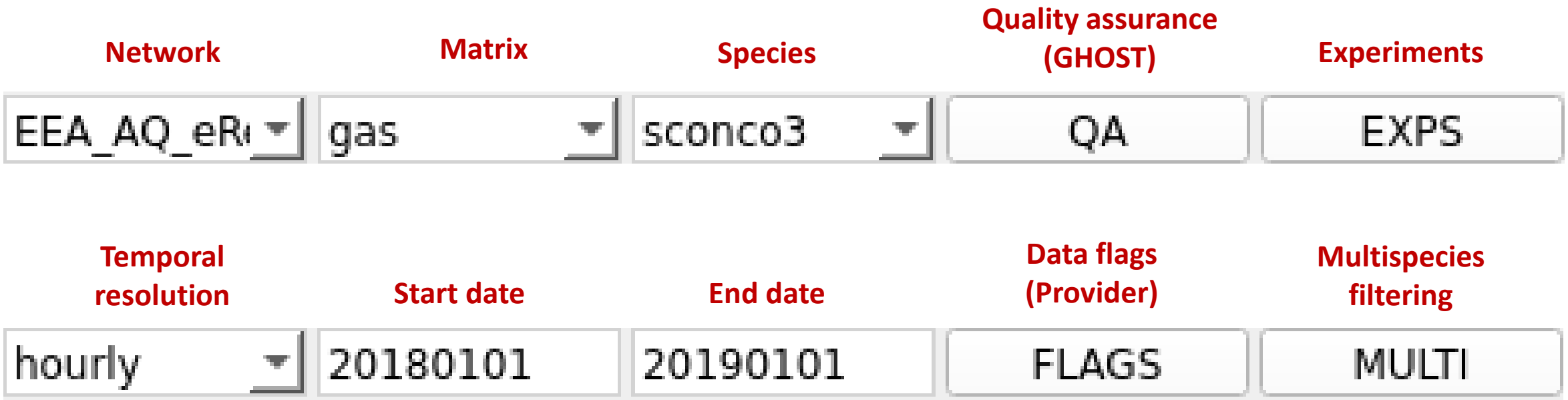




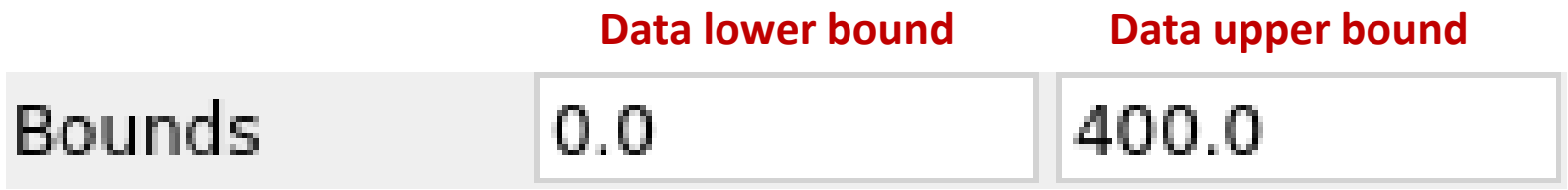
Main menu


Data Selection
EEA_AQ_eR
gas
sconco3
QA
EXPS
hourly
20180101
20190101
FLAGS
MULTI
READ
Filters
Bounds
0.0
400.0
% REP
PERIOD
META
RESET
FILTER
Statistics
Mode
Flattened
Aggregation
Colocation
Temporal
Resampling
None
Site Selection
All
Intersect
Extent

Data selection menu



Filters menu

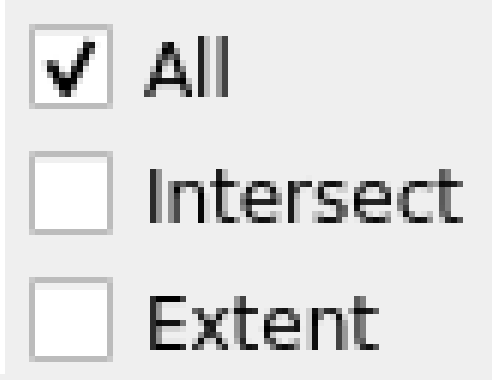


Stations selection menu

Select all stations

Select intersecting stations within all model domains

Select stations on current map view



Colocation menu

Colocate experiments vs. observations, removing temporal gaps

Data Selection

EBAS | gas | sconco3 | QA | EXPS
hourly | 20190101 | 20200701 | FLAGS | MULTI | READ

Filters

Bounds | 0.0 | 400.0
% REP | PERIOD | META | RESET | FILTER

Statistics

Mode: Flattened
Aggregation:

Colocation

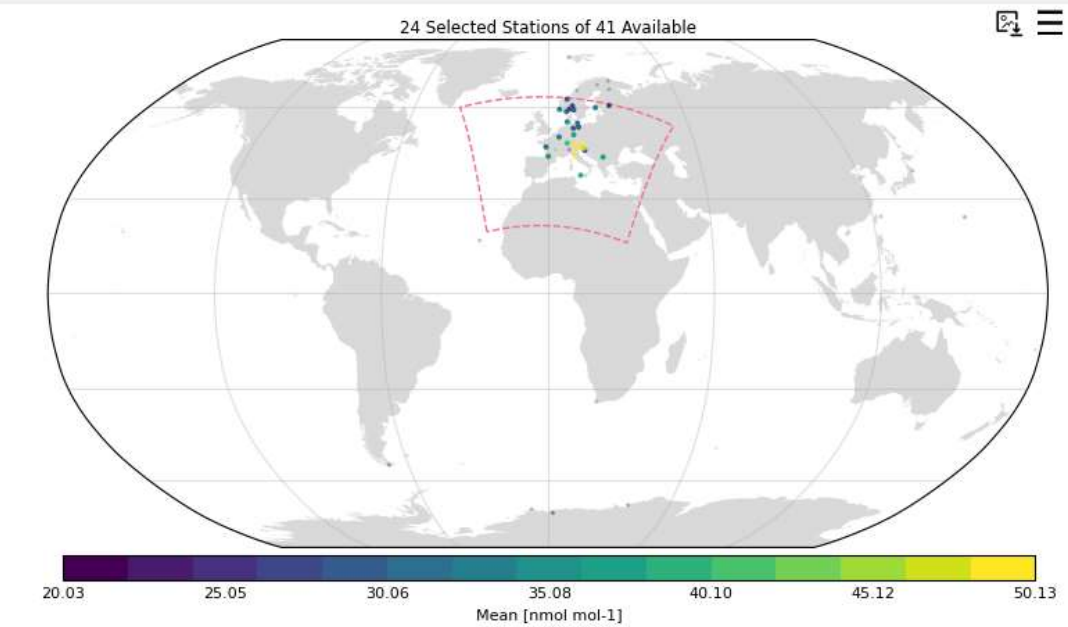
Temporal

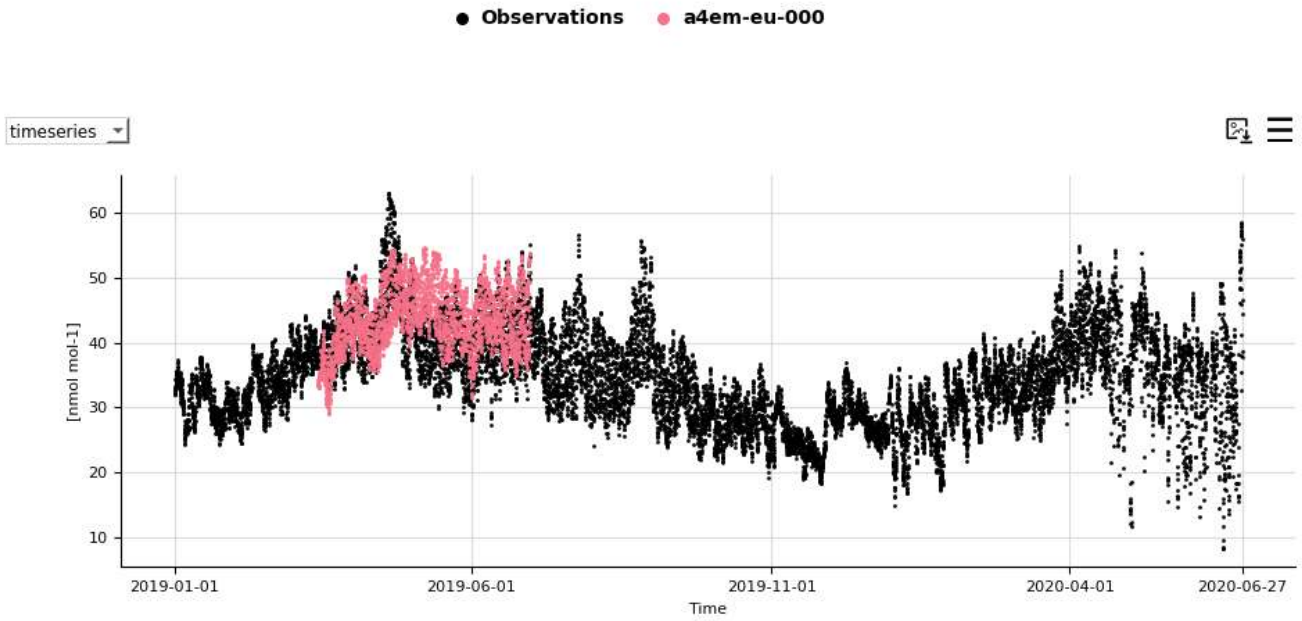
Resampling

None

Site Selection

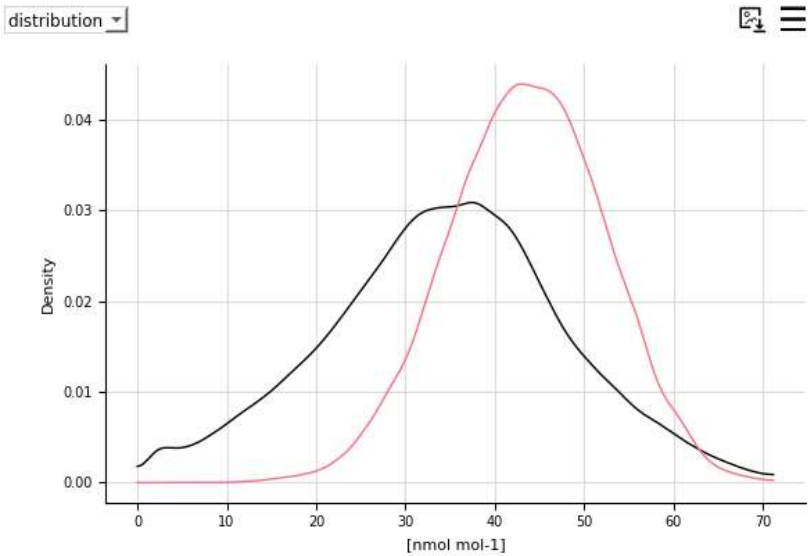
All
Intersect
Extent

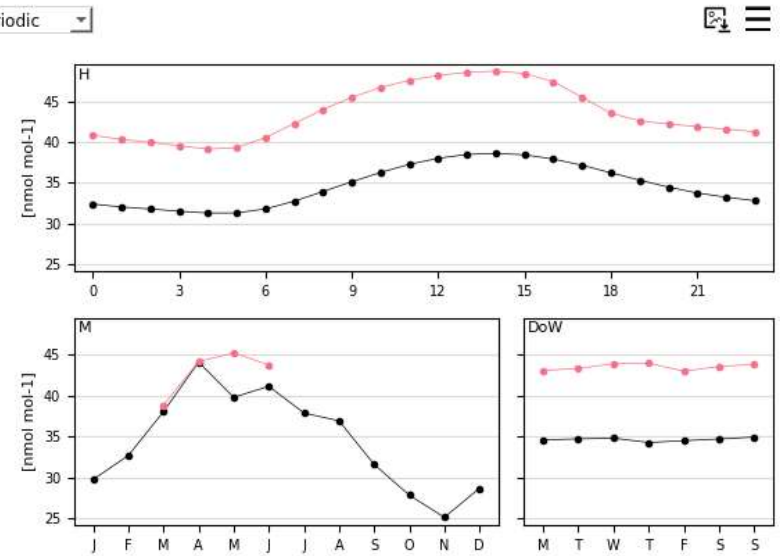




statsummar

	Mean	StdDev	p5	Median	p95
Observations	34.63	13.54	11.40	34.88	57.20
a4em-eu-000	43.54	8.78	28.88	43.64	57.66





Data Selection

EBAS | gas | sconco3 | QA | EXPS
hourly | 20190101 | 20200701 | FLAGS | MULTI
READ

Filters

Bounds | 0.0 | 400.0
% REP | PERIOD | META
RESET | FILTER

Statistics

Mode: Flattened
Aggregation:

Colocation

☑ Temporal

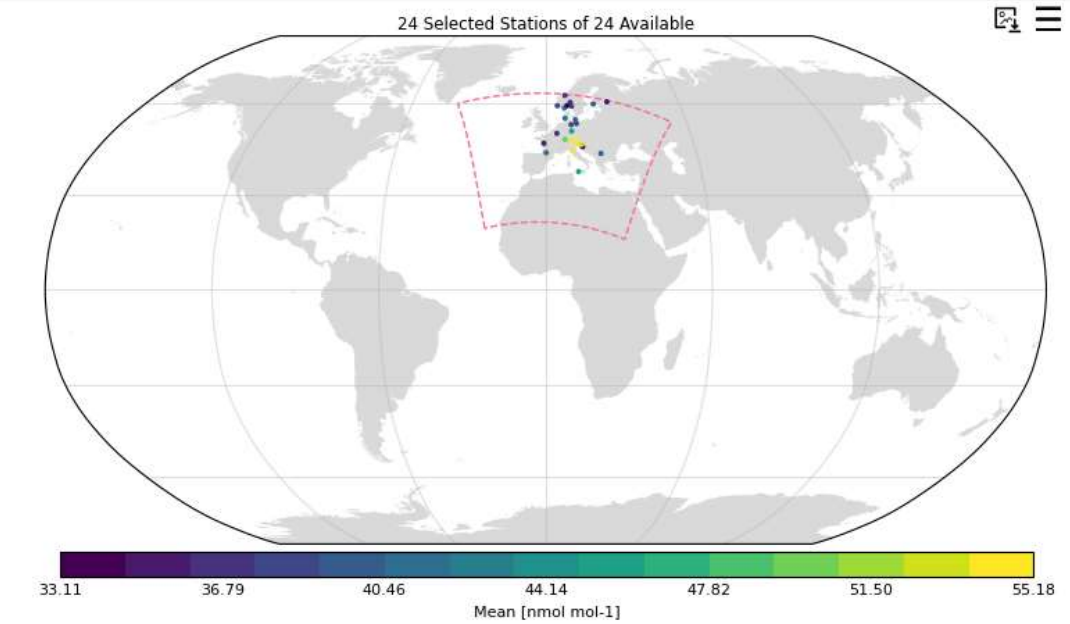
Resampling

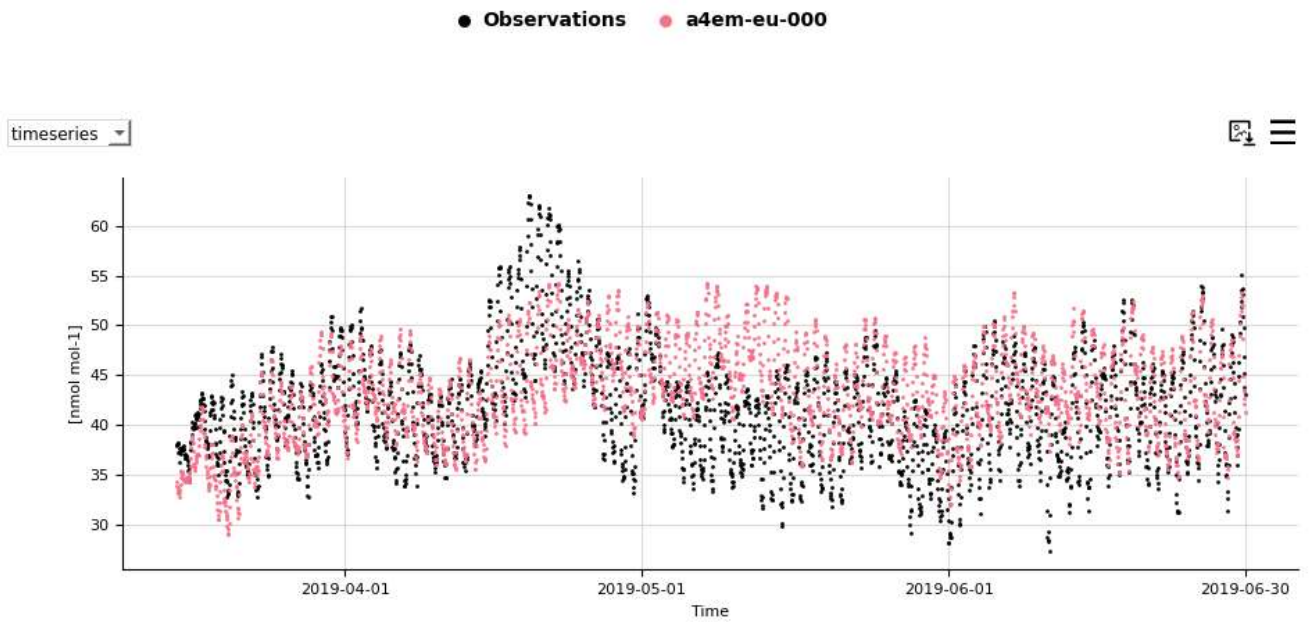
None

Site Selection

☐ All
☑ Intersect
☐ Extent

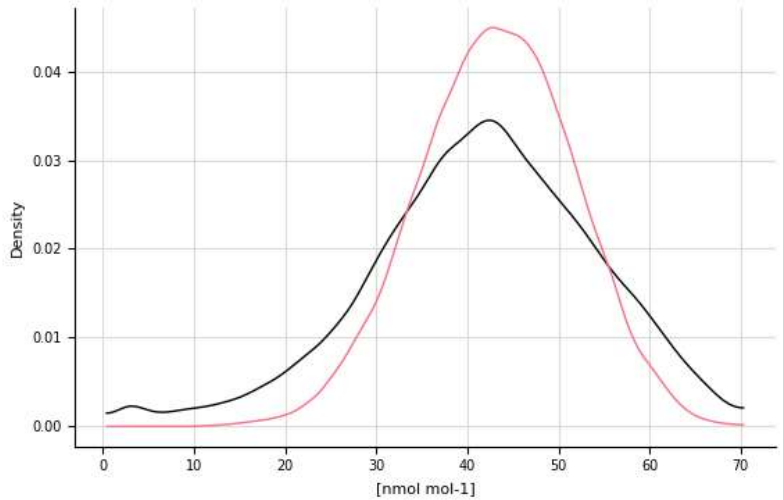
x=0.098 y=0.49

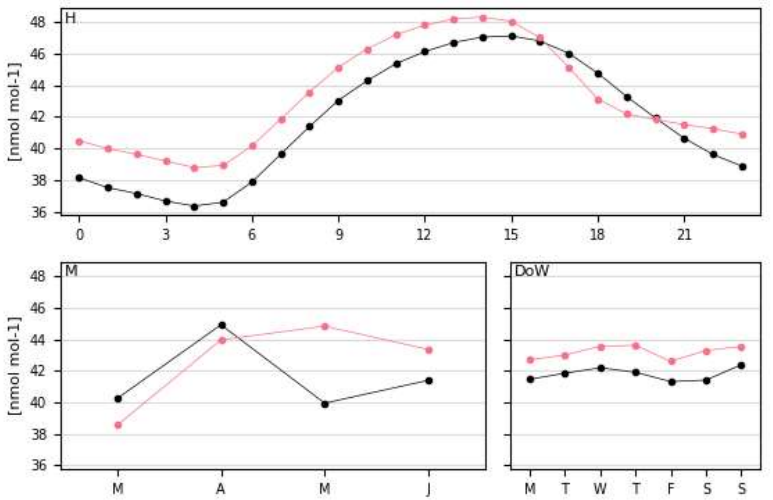




statsummar

	Mean	StdDev	p5	Median	p95
Observations	41.78	12.92	19.59	42.12	62.00
a4em-eu-000	43.18	8.61	28.76	43.31	56.97





Resampling menu

Temporal resolution to resample your data to (always lower than the selected resolution)

Data Selection

EBAS | gas | sconco3 | QA | EXPS
hourly | 20190101 | 20200701 | FLAGS | MULTI
READ

Filters

Bounds | 0.0 | 400.0
% REP | PERIOD | META
RESET | FILTER

Statistics

Mode: Flattened
Aggregation:

Colocation

☑ Temporal

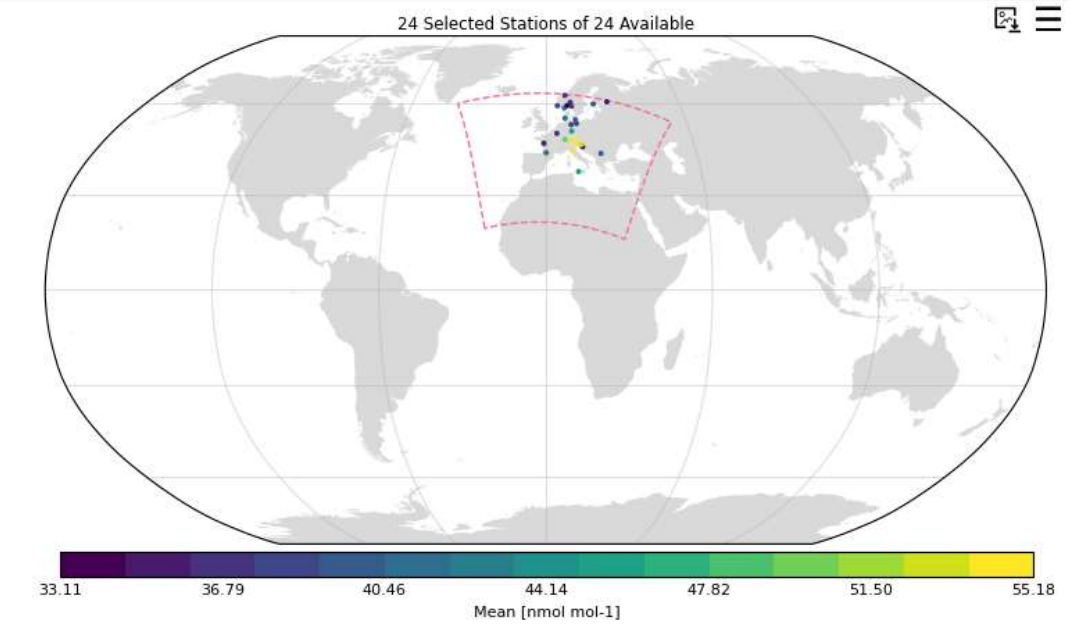
Resampling

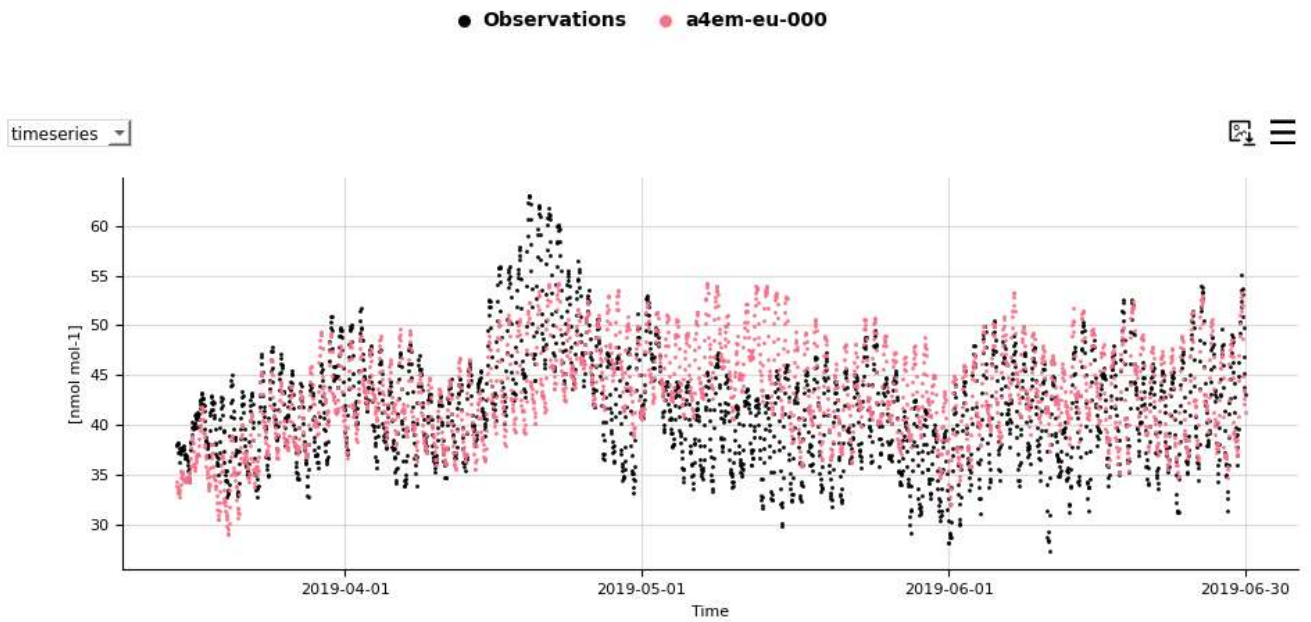
None

Site Selection

☐ All
☑ Intersect
☐ Extent

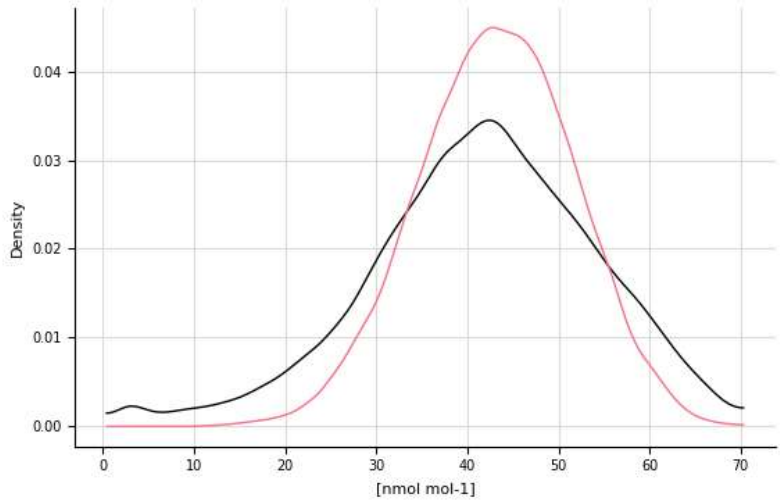
x=0.098 y=0.49

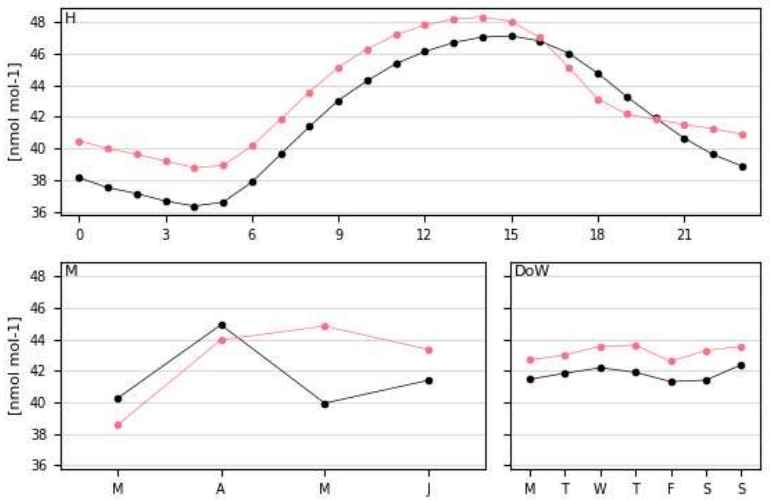




statsummar

	Mean	StdDev	p5	Median	p95
Observations	41.78	12.92	19.59	42.12	62.00
a4em-eu-000	43.18	8.61	28.76	43.31	56.97





Data Selection

EBAS | gas | sconco3 | QA | EXPS
hourly | 20190101 | 20200701 | FLAGS | MULTI
READ

Filters

Bounds | 0.0 | 400.0
% REP | PERIOD | META
RESET | FILTER

Statistics

Mode: Flattened
Aggregation:

Colocation

☑ Temporal

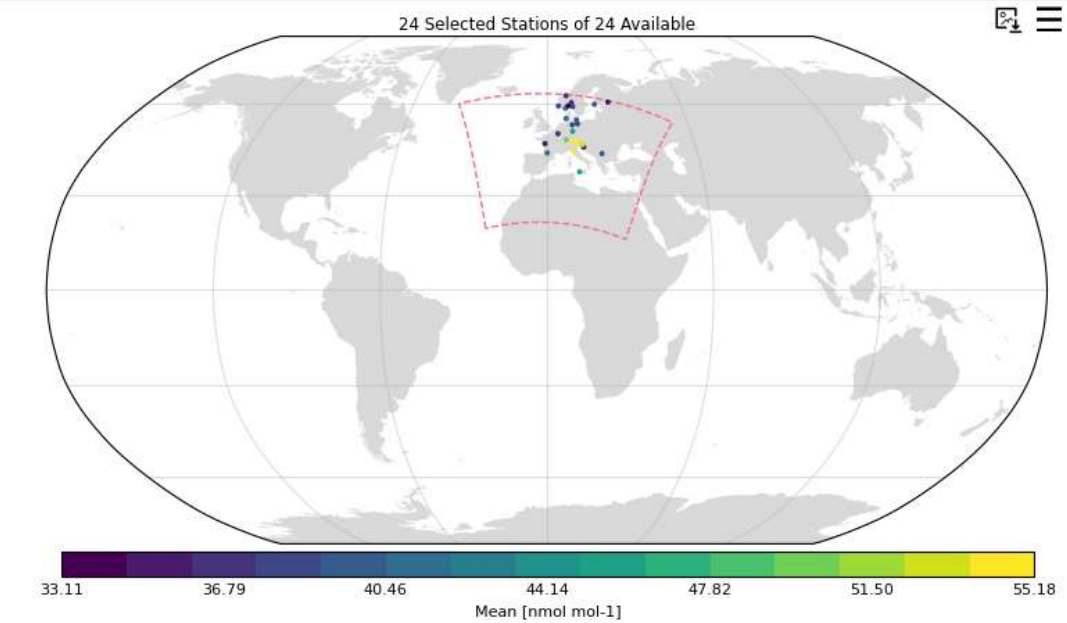
Resampling

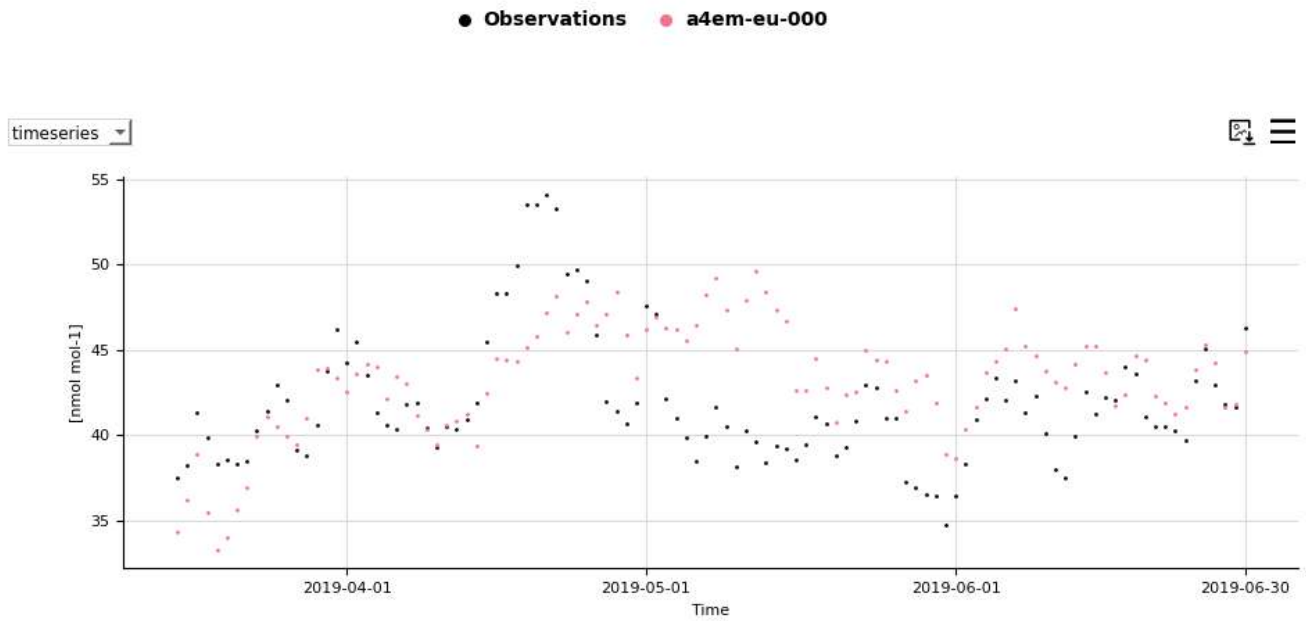
daily

Site Selection

☑ All
☐ Intersect
☐ Extent

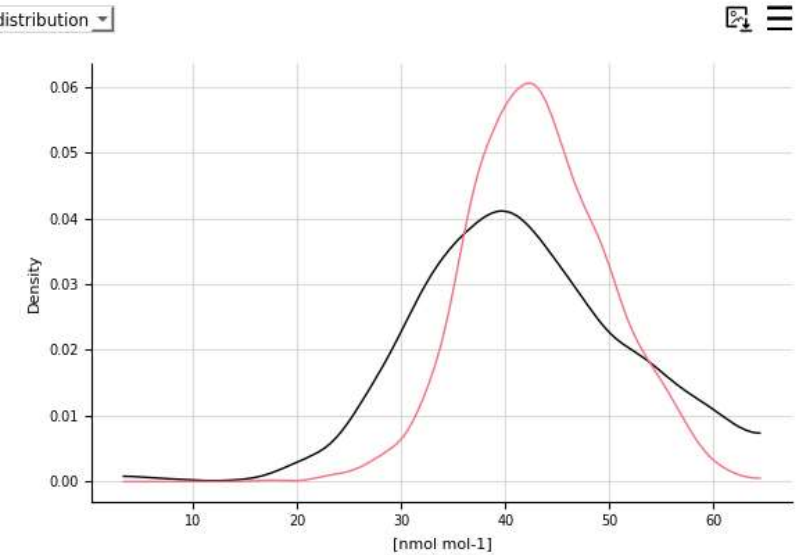
x=0.234 y=0.61

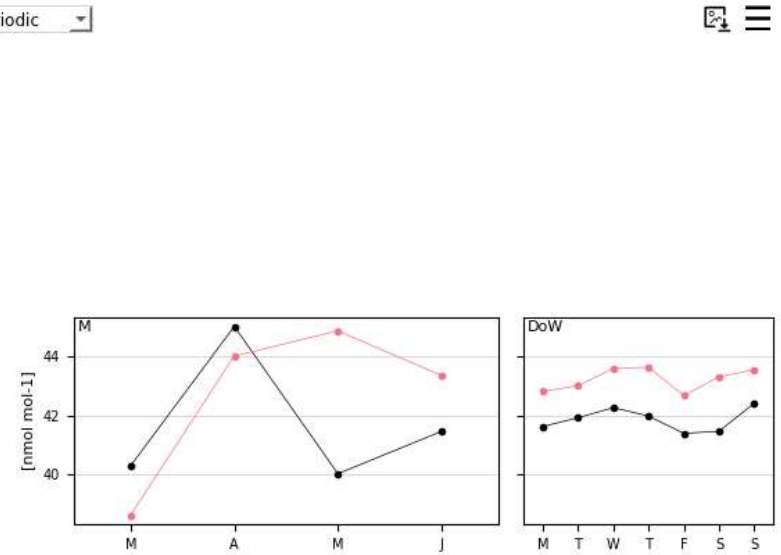




statsummar

	Mean	StdDev	p5	Median	p95
Observations	41.86	10.33	26.49	41.05	60.11
a4em-eu-000	43.23	6.69	32.93	42.91	55.06





Icons menu

Upload configuration files

Back

Zoom to rectangle

Save canvas

Download data and configuration files

Back to initial screen

Forward

Pan

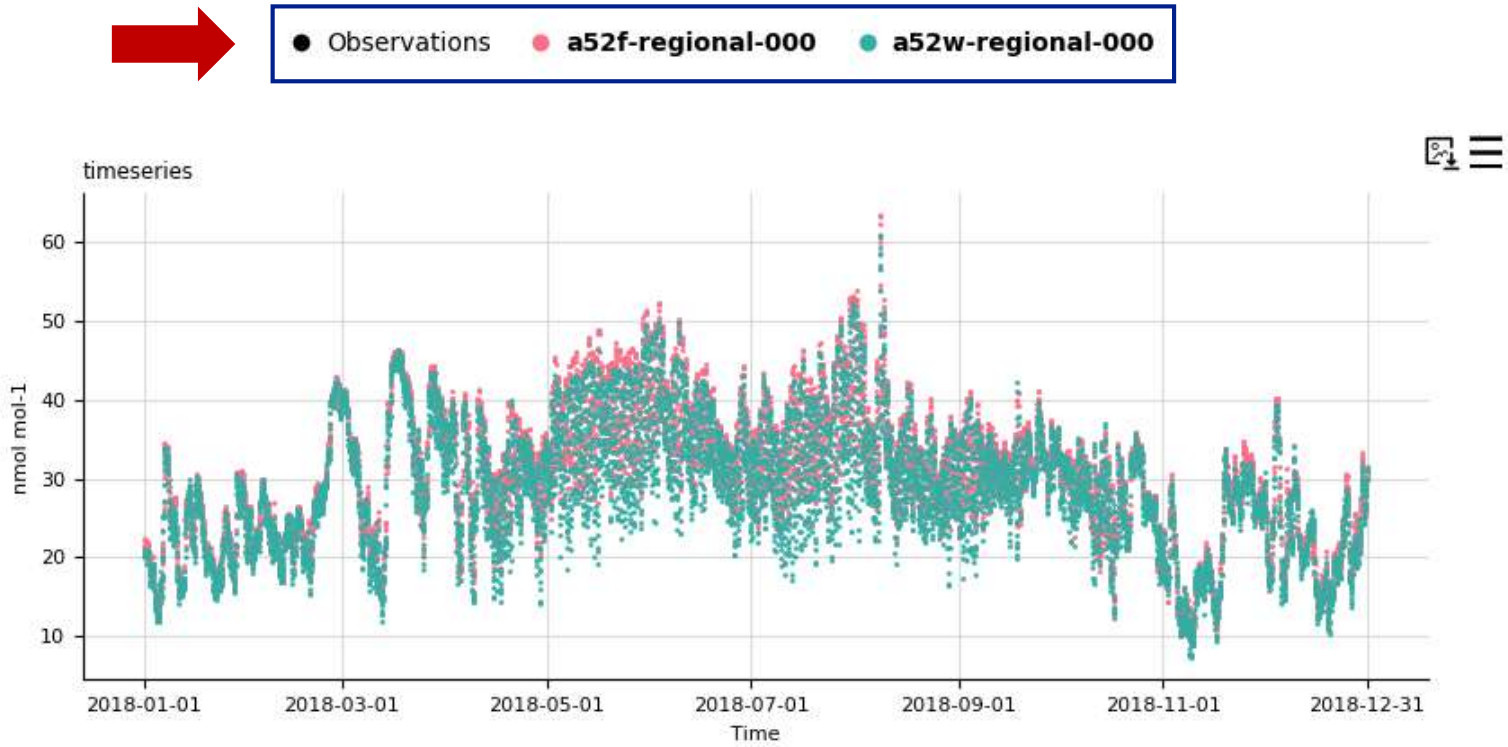
Interactive features

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

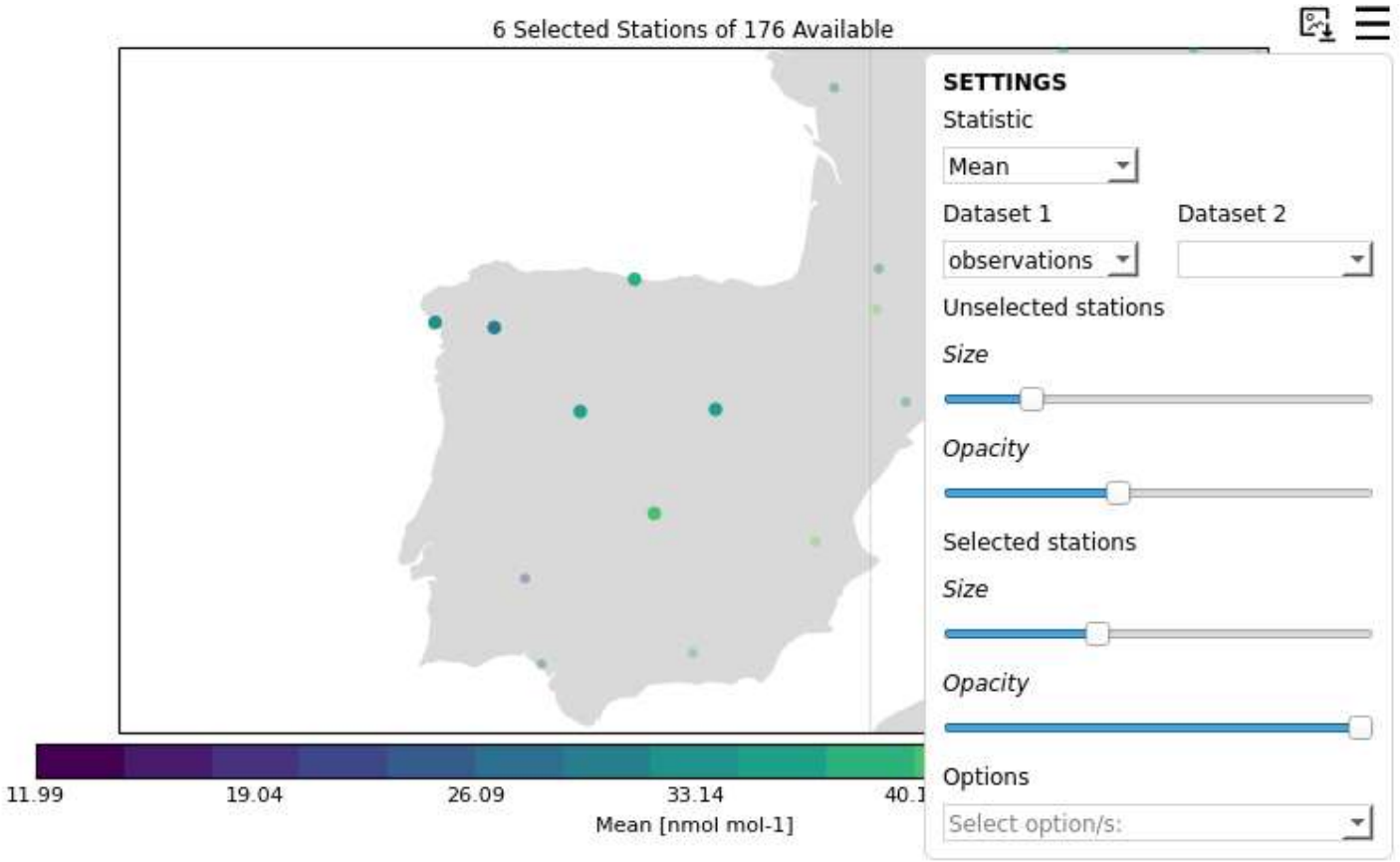
Legend picking

Clicking on the legend labels will remove or add data to each of the plots

Bold = Visible
Roman = Invisible



Changing the plot style



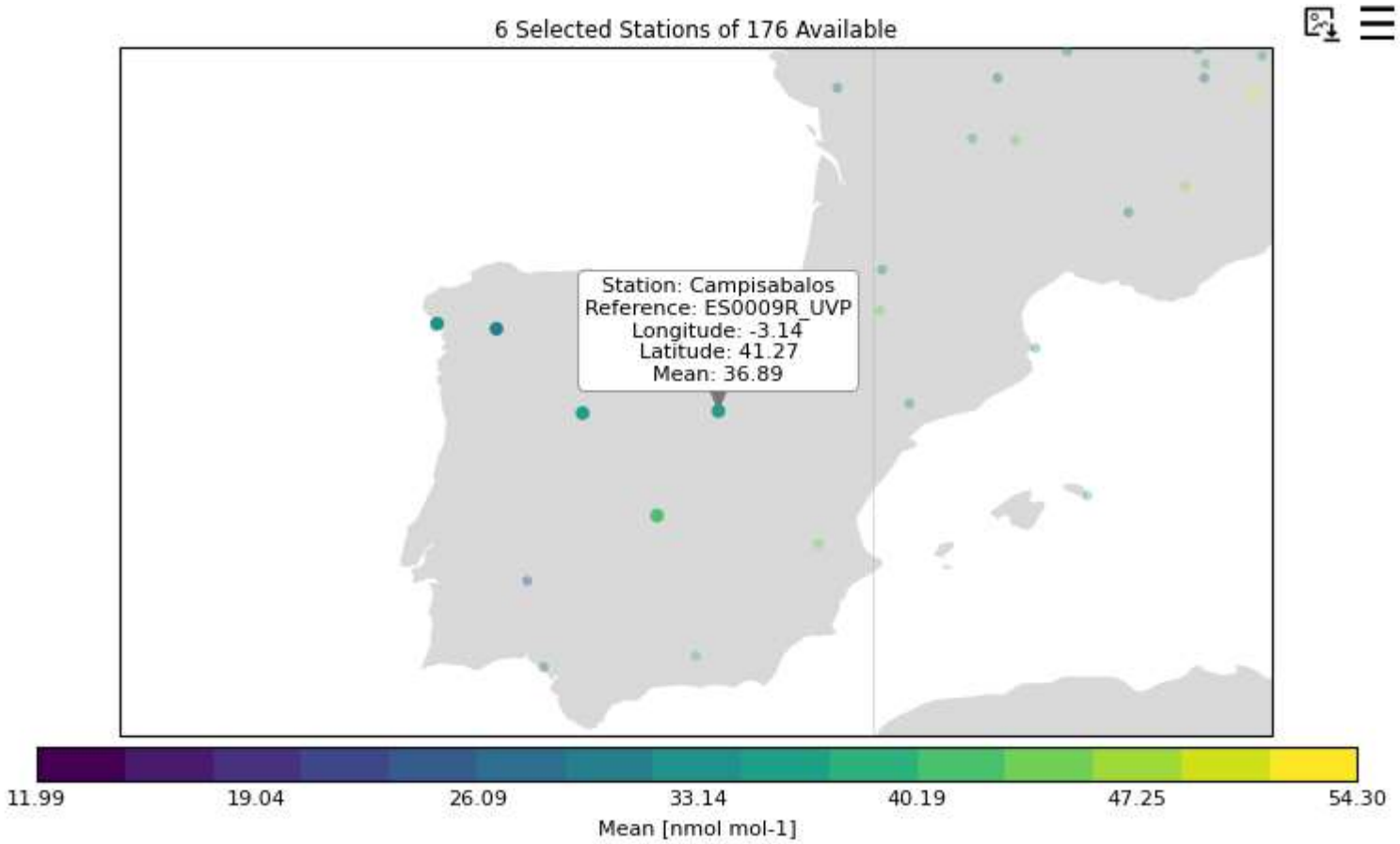
The style of the plots can be edited by clicking on the **burger menus** and changing the settings

Changing the statistics

The statistics in the statsummary can be updated from the burger menu

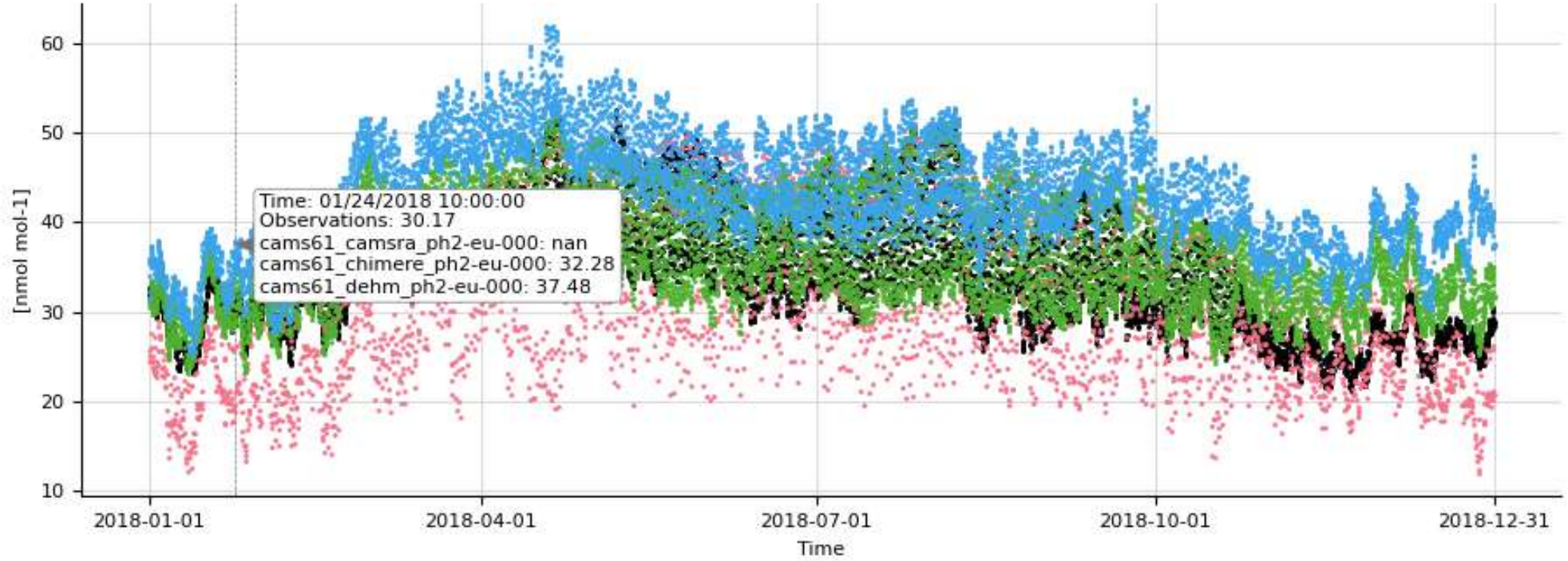
Information on hover

We can see the stations details and data by hovering on the map



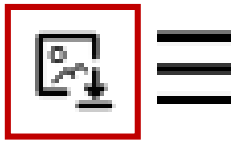
Information on hover

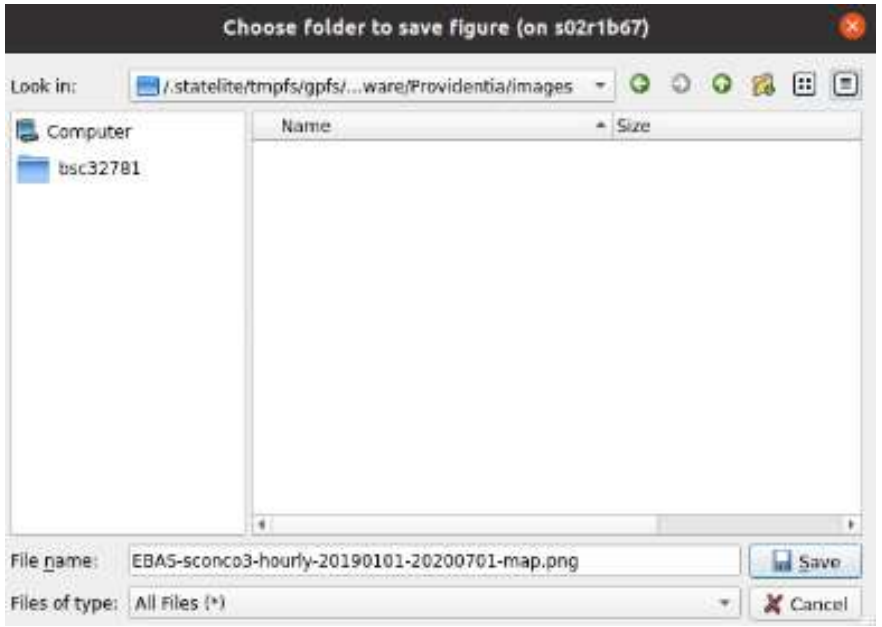
We can also check the values of each dataset by hovering on the other plot types

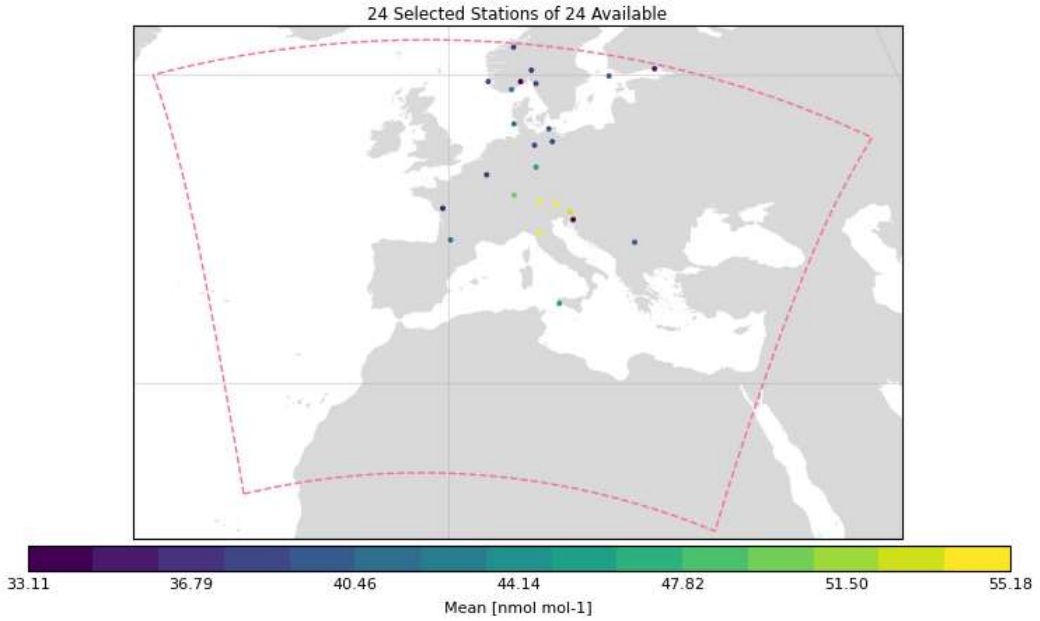


Save plot figures

Saving plot figures is now possible by clicking on the image icons next to the burger menus

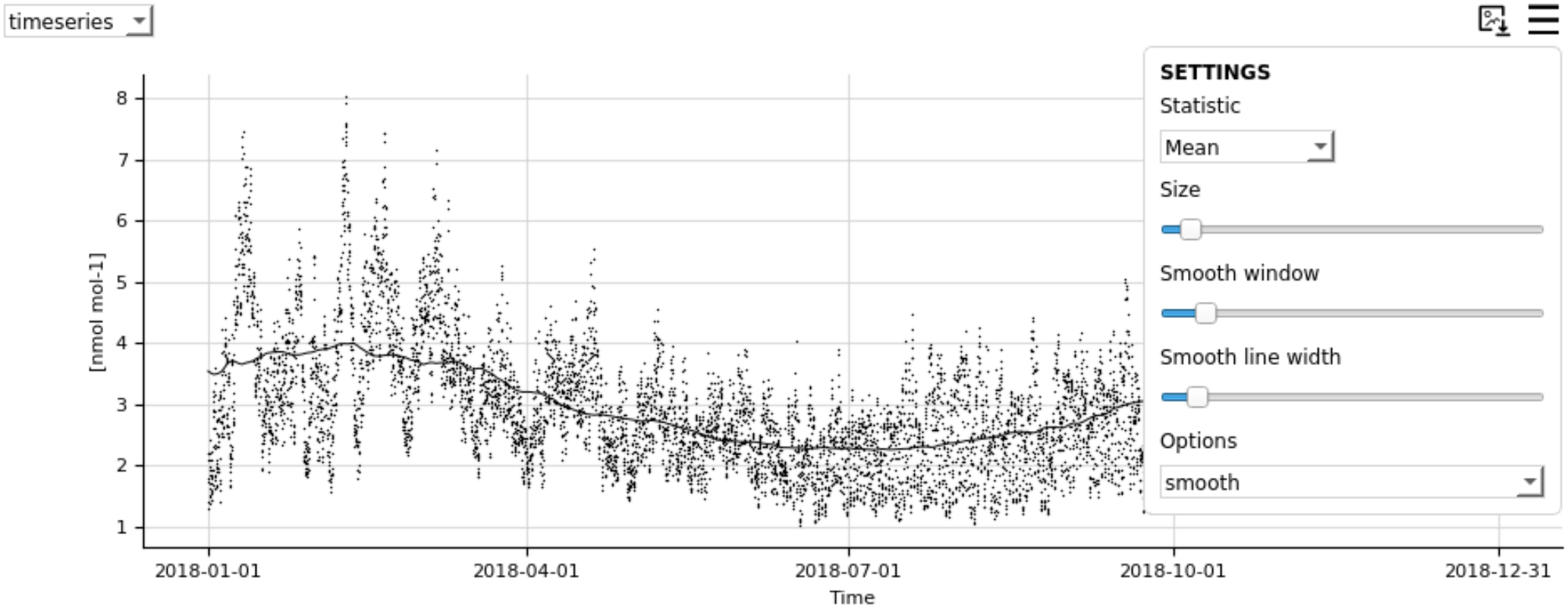






Smoothing

It is possible to add a smooth line in the timeseries plot by editing the smooth window, the data points can be then hidden by reducing their size to 0



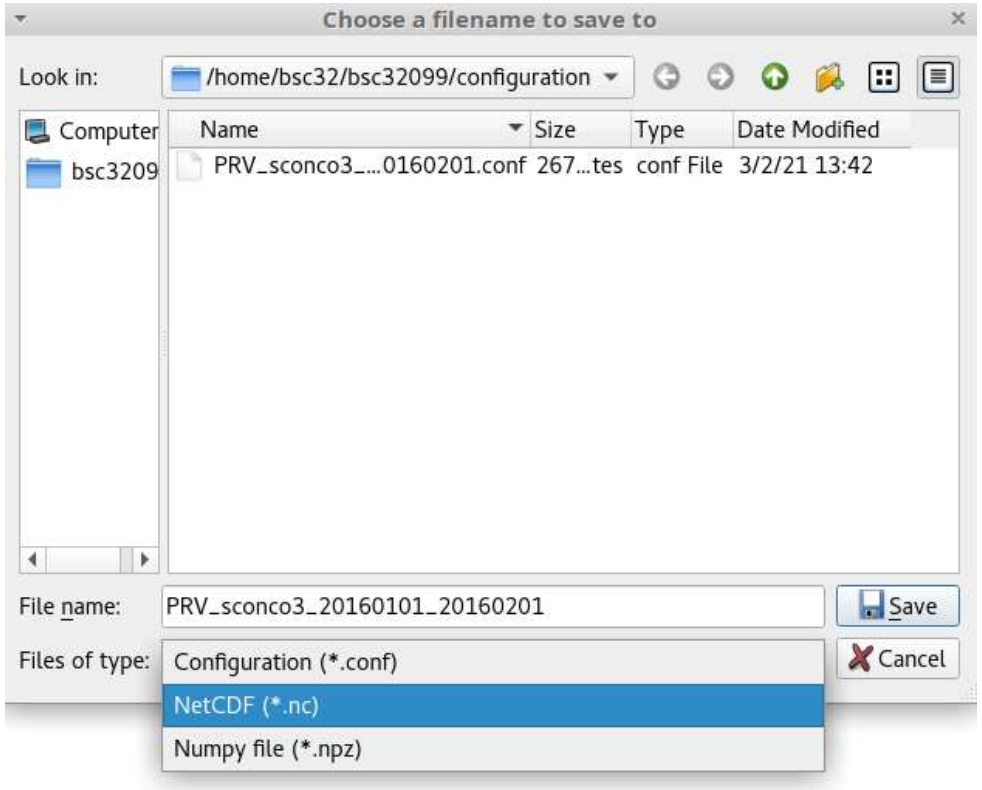
Export and load

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Export data and configuration files

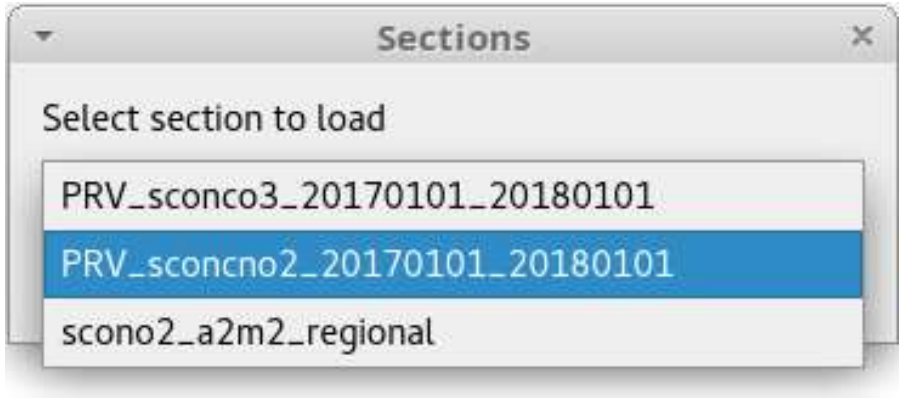
If we wish to export the data that we used during our evaluation session, we can do it using the save button on the general menu. The formats are **Numpy and NetCDF**.

We can also export **configuration files**, useful to launch the dashboard and create offline reports. It is possible to change the name as well as select its destination path.



Load data using configuration files

You can select the configuration file which you want to load using the load button on the general menu. After selecting a file and clicking Open, an extra dialog will appear in which you can select which section of your configuration you want to load.



Remember that you can also load the dashboard using a configuration file as an argument as in:

```
$ ./bin/providentia --config=configurations/training.conf
```

Statistics

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Statistics

- **Statistical modes:**
  - Flattened (new default)
  - Spatial|Temporal (previous Providentia versions)
  - Temporal|Spatial (CAMS / AEROVAL)

- **Calculation of statistics per periodic cycle, 2 modes:**
  - Independent
  - Cycle

Statistical modes

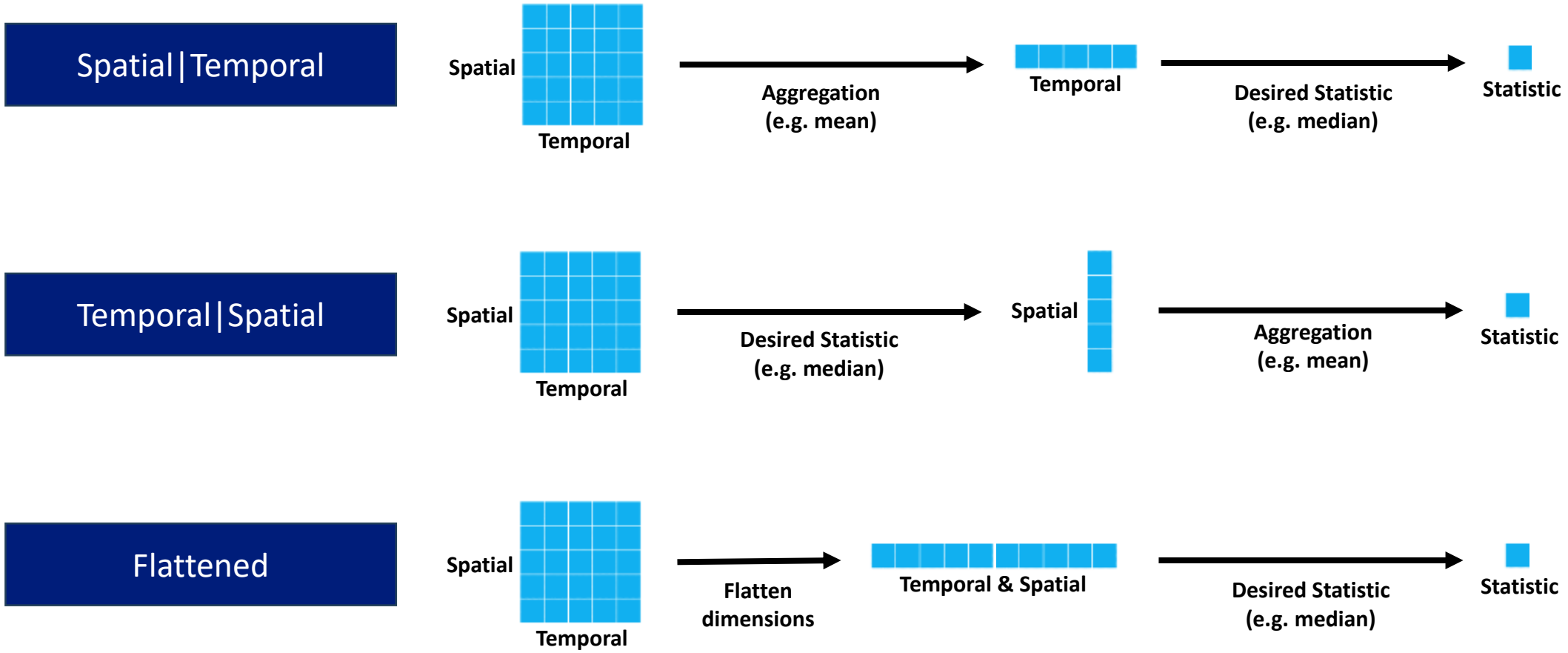
The name of each statistical mode relates to how the dimensions of the selected data are reduced to calculate the statistical metrics, e.g. mean, median etc, going from 2D to 0D.

When selecting data across multiple stations, it has 2 dimensions:

Spatial and **Temporal**.

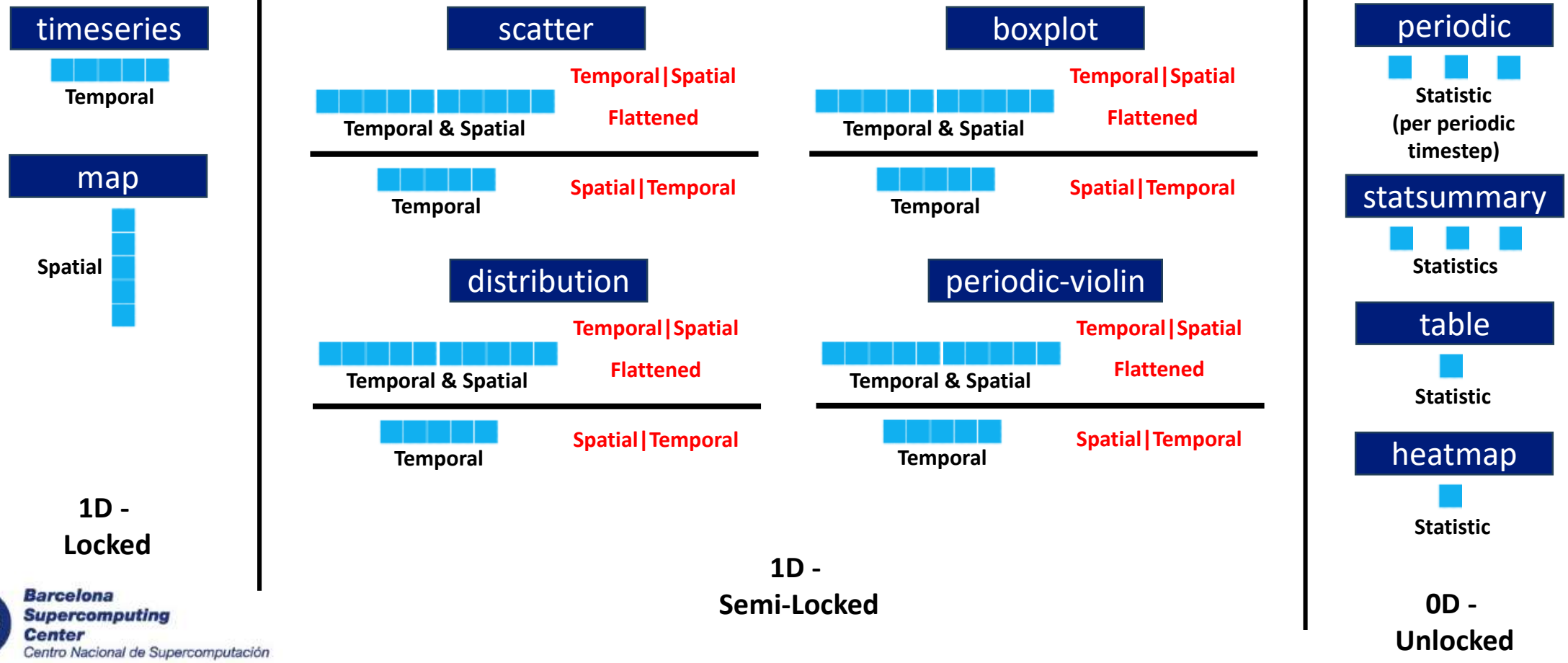
In the previous versions of Providentia, the mode was always **Spatial|Temporal**, with aggregation performed across stations (e.g. taking mean across stations per timestep), going to 1D, before calculating the desired statistic across the aggregated timesteps (e.g. median), going to 0D.

Statistical modes



Dimensional reduction per plot type

For some plot types the full dimensional reduction is not possible, e.g map. Therefore, they are locked in certain reduction configurations. This is illustrated per plot type:



Statistical modes

In both the dashboard and offline versions of Providentia, the statistical modes and aggregation statistics can be set.

On the dashboard, in the **statistics** tab at the top, the **mode** and **aggregation** can be selected via the dropdown menus. In the .conf file these can be set like so:

statistic_mode = Temporal|Spatial

statistic_aggregation = Median

Note: For the **Flattened** mode, there is no aggregation statistic.

Periodic statistics

The periodic plot gives statistical information for grouped data in individual timesteps. Thus it can be seen how each individual timesteps compare for observations vs experiment/s.

However, when looking to evaluate the agreement across the whole of the periodic cycle, Providentia was previously lacking statistics to enable such a comparison.

Statistics can now be calculated which assess the available periodic cycles, i.e. diurnal, weekly, monthly. These statistics are available via the **statsummary**, **table** and **heatmap** plot types.

Periodic statistics

There exist 2 modes for calculating these periodic statistics:

- Independent (default)
- Cycle

Independent works by calculating the desired statistic per timestep (i.e. as seen in periodic plot), before aggregating across the timesteps.

Cycle works by aggregating the grouped data per timestep (e.g. mean), before then calculating the desired statistic across the timesteps.

Periodic statistics

Again, in both the dashboard and offline versions of Providentia, the periodic statistical modes and aggregation statistic can be set.

On the dashboard, in the **plot options** of the **statsummary** plot, the **periodic statistic mode** and **aggregation** can be selected via the dropdown menus. Additionally, periodic statistics can be added to the statsummary plot also via the dropdown menus.

In the .conf file these can be set like so:

periodic_statistic_mode = Independent

periodic_statistic_aggregation = Median

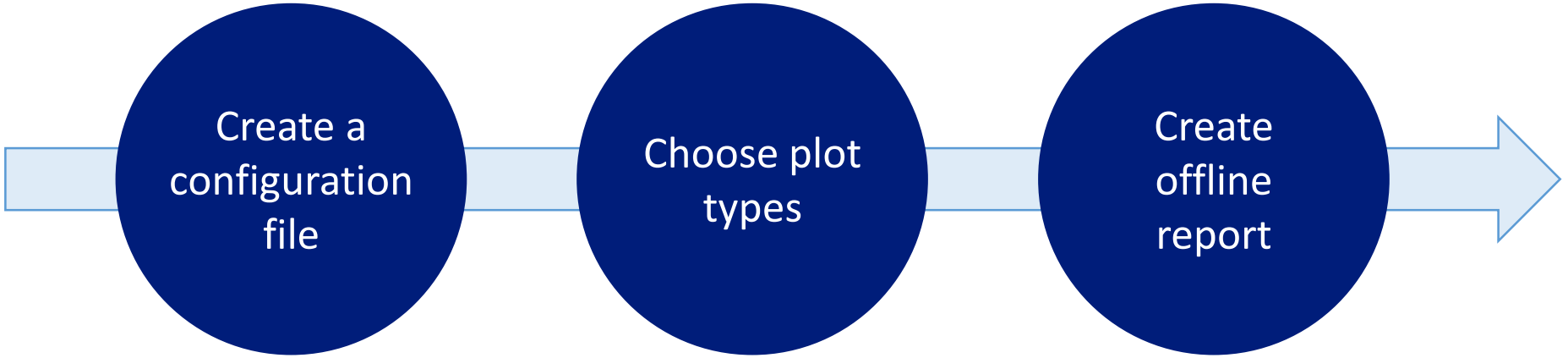

Q&A

Offline reports

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Characteristics of offline reports

- PDFs as complete reports for an **in-depth analysis** of the experiments
- More available plot types than in the dashboard
- Preparation:



Configuration files

SECTION
Defined with [ ]

SUBSECTION
Defined with [[ ]]

SUBSECTION
Defined with [[ ]]

```
[All]
network = AERONET_v3_lev1.5
species = vconcaerobin*
resolution = hourly_instantaneous
start_date = 20200601
end_date = 20200701
experiments = a55u-global-av_an (a55u)
temporal_colocation = True
spatial_colocation = True
filter_species = AERONET_v3_lev1.5:ae440-870aero (>0.6, :, nan)
report_type = sizedist
report_summary = True
report_stations = True
report_title = Providentia Offline Report
report_filename = dod_size

  [[All]]
  map_extent = -180, 180, -90, 90

  [[Mediterranean]]
  longitude = -20, 50
  latitude = 35, 50
  map_extent = -20, 50, 35, 50
```

Mandatory fields

Field	Description
network	Network you want to load observations from. Can be multiple (e.g. CAPMoN, EBAS).
species	Species to load. Can be multiple (e.g. sconco3, sconcno2). Adding a wild card (*) is going to expand to certain variables (vconc* → vconc1, vconc2, etc.).
resolution	Resolution of the observations you want to load (e.g. 3hourly).
start_date	Comparison start date in YYYYMMDD format (e.g. 20170101).
end_date	Comparison end date in YYYYMMDD format (e.g. 20180601).

Optional fields

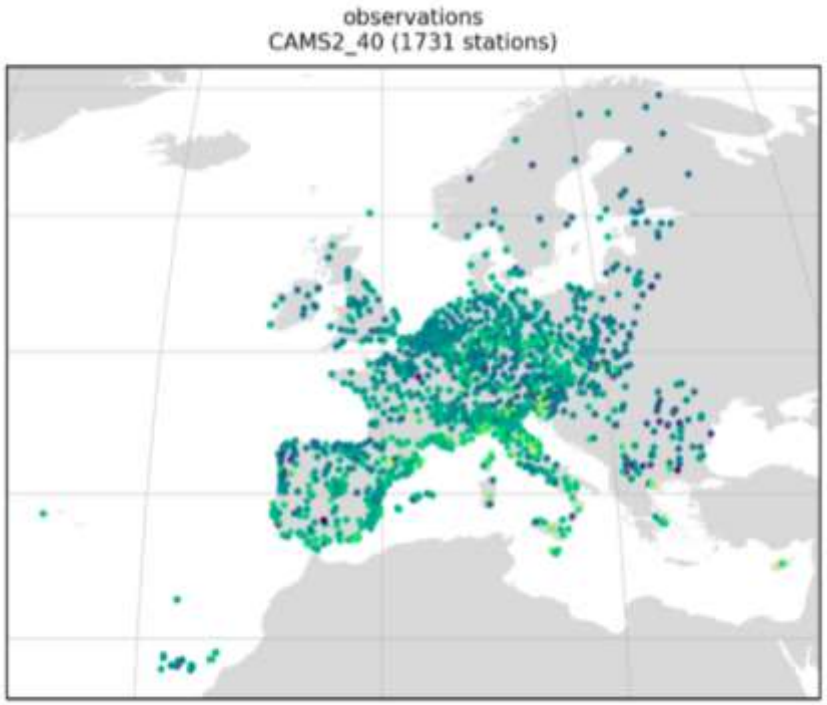
Field	Description
statistic_mode	Statistic mode: Flattened (default), Spatial|Temporal or Temporal|Spatial.
statistic_aggregation	Aggregation statistic, e.g. Median.
periodic_statistic_mode	Periodic statistic mode: Independent (default), Cycle.
periodic_statistic_aggregation	Periodic aggregation statistic, e.g. Mean (default).
temporal_colocation	Boolean variable to set if you want to temporally colocate the observation and experiment data.
spatial_colocation	Boolean variable to set if you want to spatially colocate the observation and experiment data across multiple species.

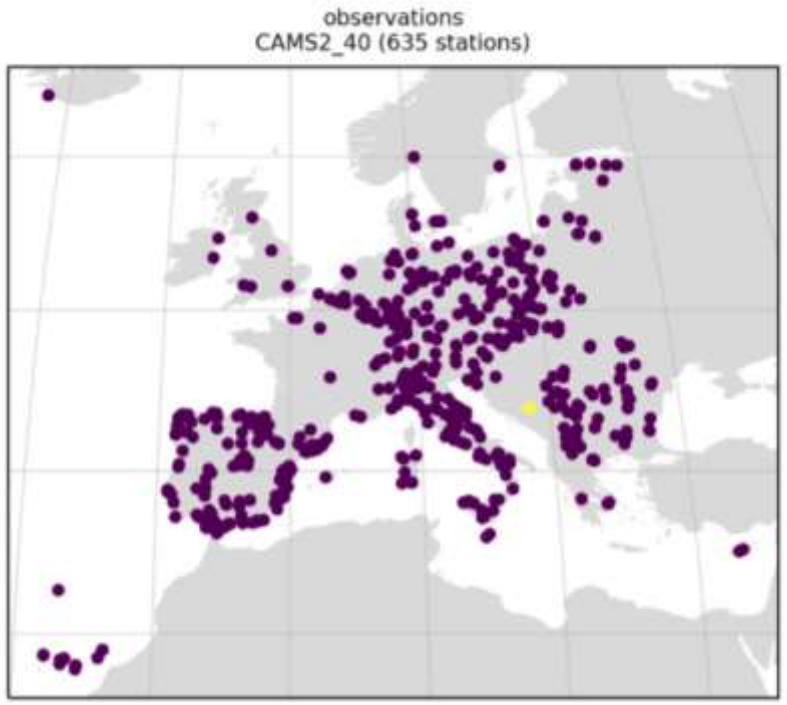
Spatial colocation

Surface O_3 (mean)
1731 stations

Surface CO (mean)
635 stations

Without spatial colocation

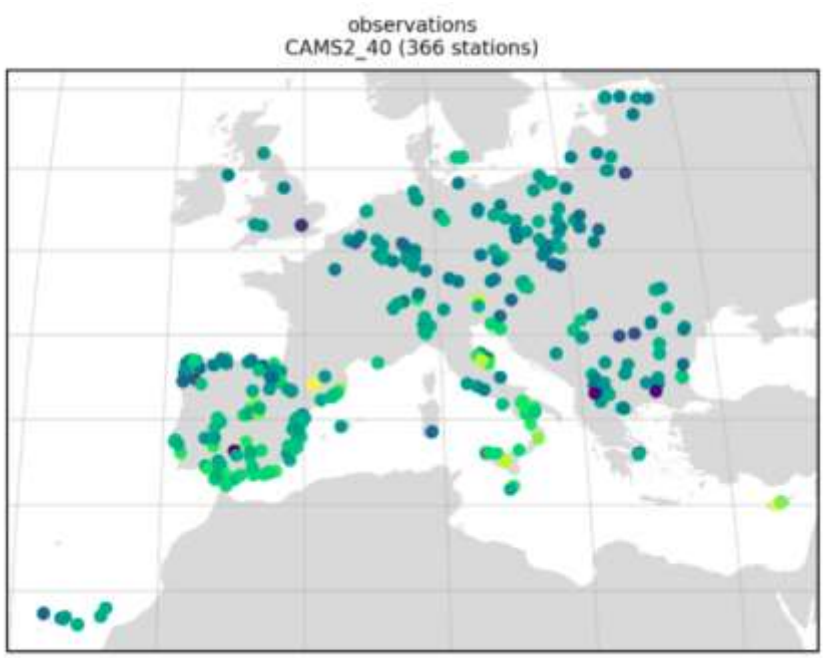




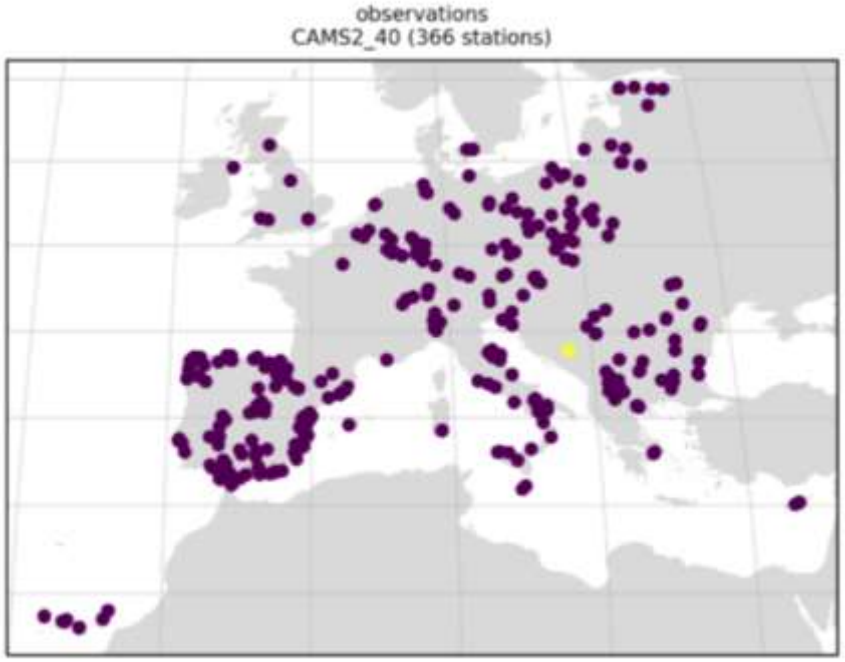
Spatial colocation

With spatial colocation

Surface O_3 (mean)
366 stations



Surface CO (mean)
366 stations



Optional fields

Field	Description
report_type	Type of report to generate that defines which plots the report will contain, from the options given in report_plots.json.
report_summary	Boolean variable to set if you wish to make specific plots for each station in subsection.
report_stations	Boolean variable to set if you wish to make summary plots across station subsection.
report_title	The header in the first page of the report (as in the PDF).
report_filename	The filename of the report or the path to create the report (as in the PDF).
plot_characteristics_filename	The paths to the files containing the plot characteristics in format dashboard:path_dashboard, offline:path_offline.

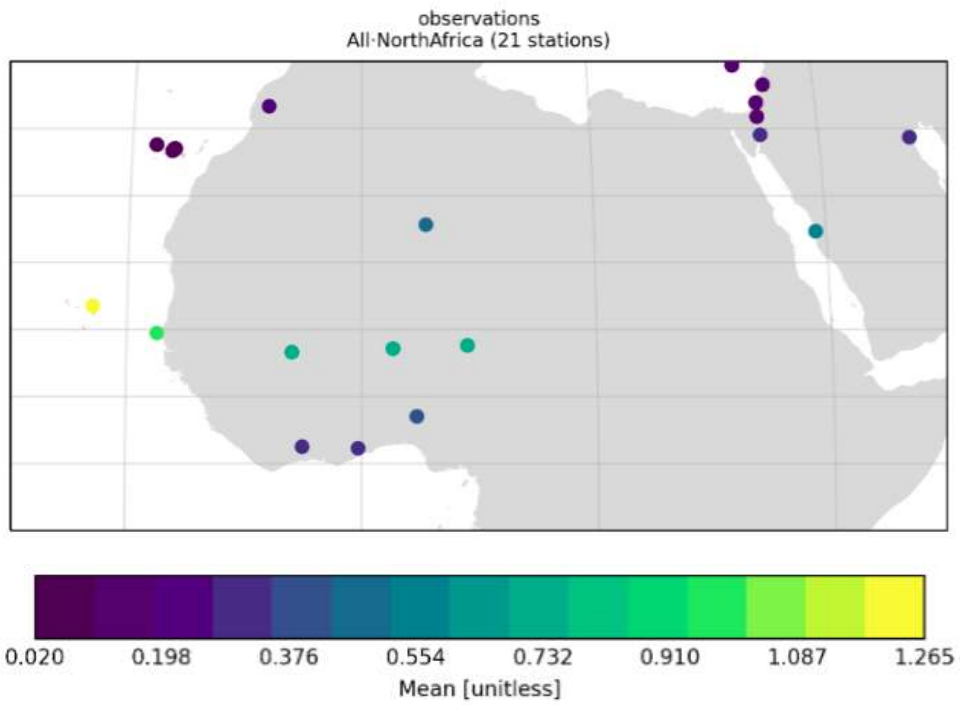
Optional fields

Field	Description
experiments	ID of interpolated experiment using providentia-interpolation. The experiment IDs can be mapped to different names by adding a **list of alternative names** after the experiment IDs e.g. exp1, exp2 (altexp1, altexp2).
map_extent	Set the map plot extents with the syntax: minimum longitude, maximum longitude, minimum latitude, maximum latitude.
resampling	Boolean variable to set if you wish to resample your data using the resampling resolution.
resampling_resolution	Temporal resolution to resample your data to.
calibration_factor	A number will be added or subtracted to the experiment data, or the data will multiplied or divided by a number, e.g. a54s-regional-000 (*0.62)
filter_species	Filter read species by other species data within a data range (can be multiple), e.g. nasa-aeronet/directsun_v3-lev15:ae440-870aero (>0.75, <=1.2, nan), nasa-aeronet/directsun_v3-lev15:ae440-870aero (>1.2, :, 0).

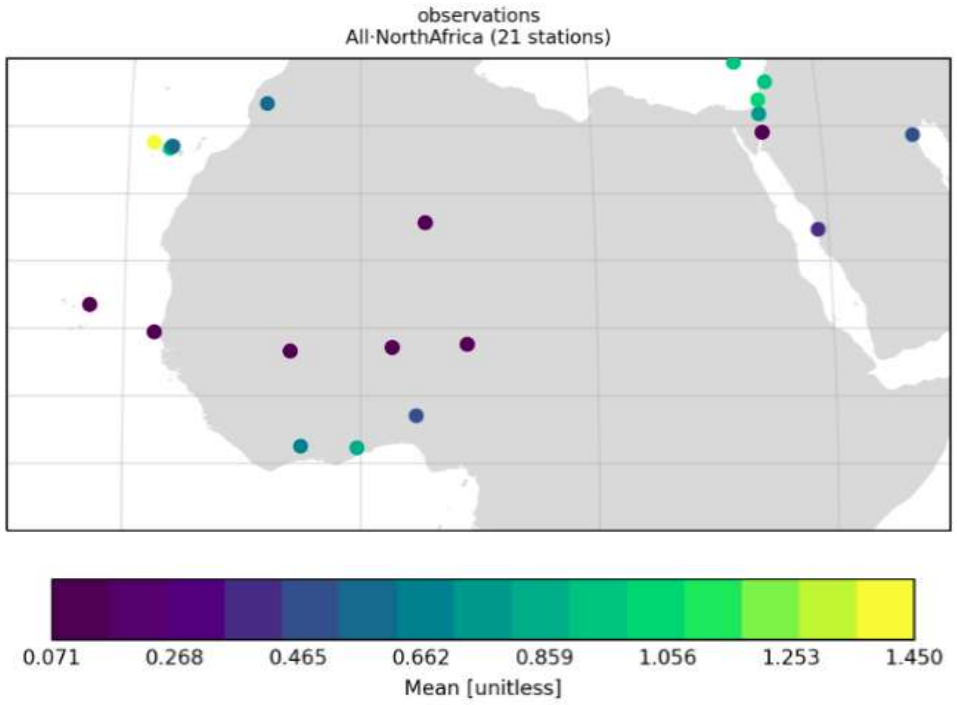
Multispecies filtering

In the reports created to study the dust in the atmosphere it is common practice to filter the optical depth by the Angstrom exponent to know which values are associated with dust

Optical depth (od550aero, mean)



Angstrom exponent (ae440-870aero, mean)

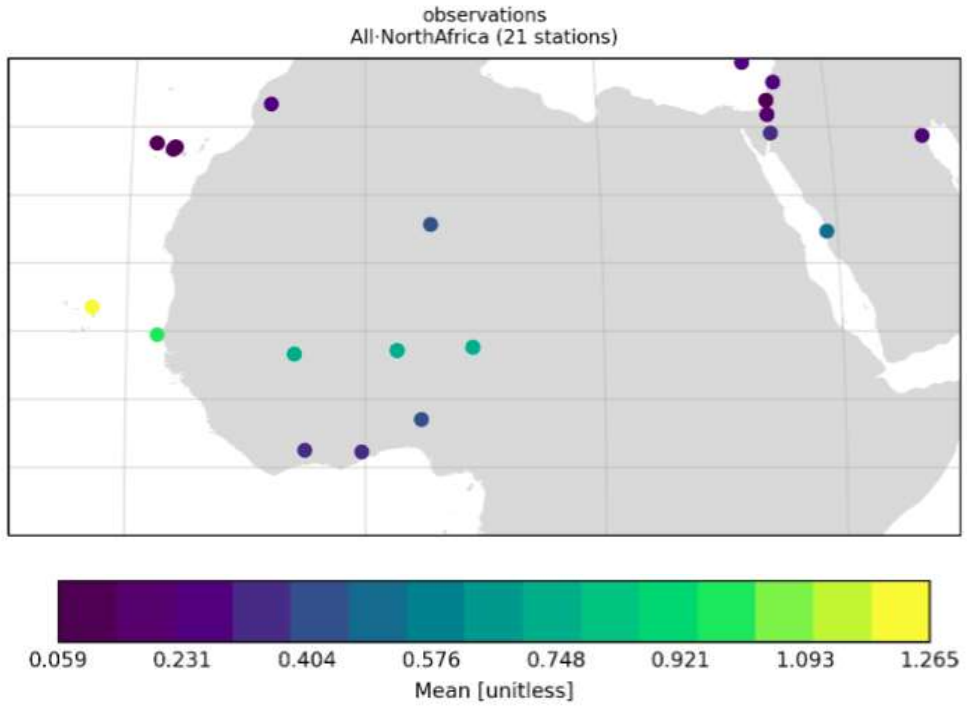


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Applying a multispecies filter…

Optical depth (od550aero, mean) filtered by ae440-870aero (>0.6, :, nan)

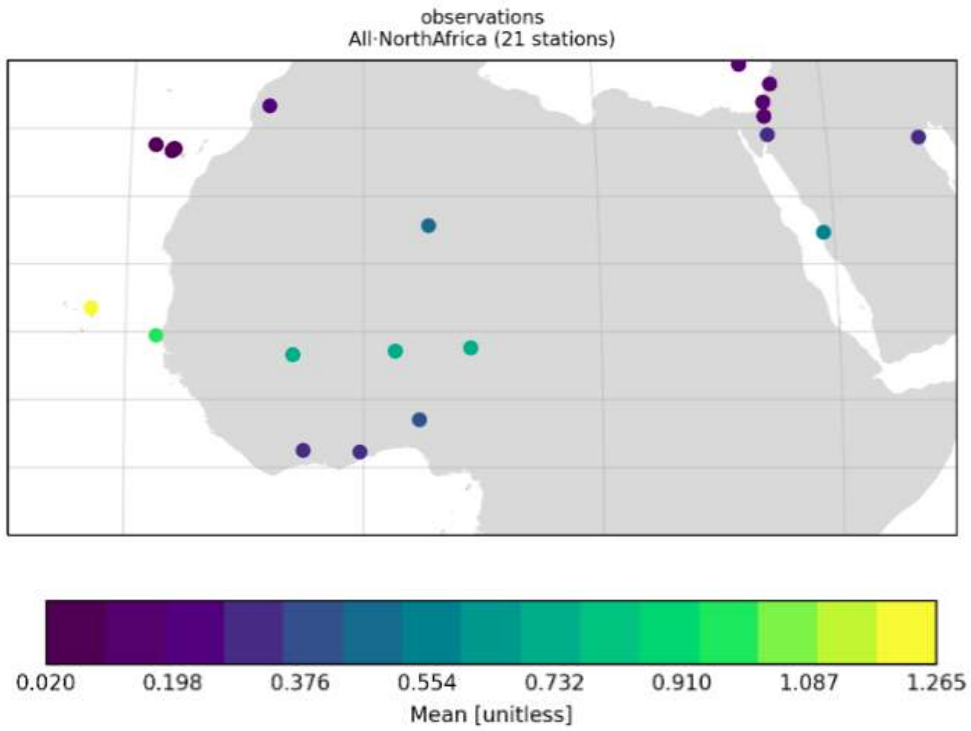
Every value of the Angstrom exponent above 0.6 has been removed



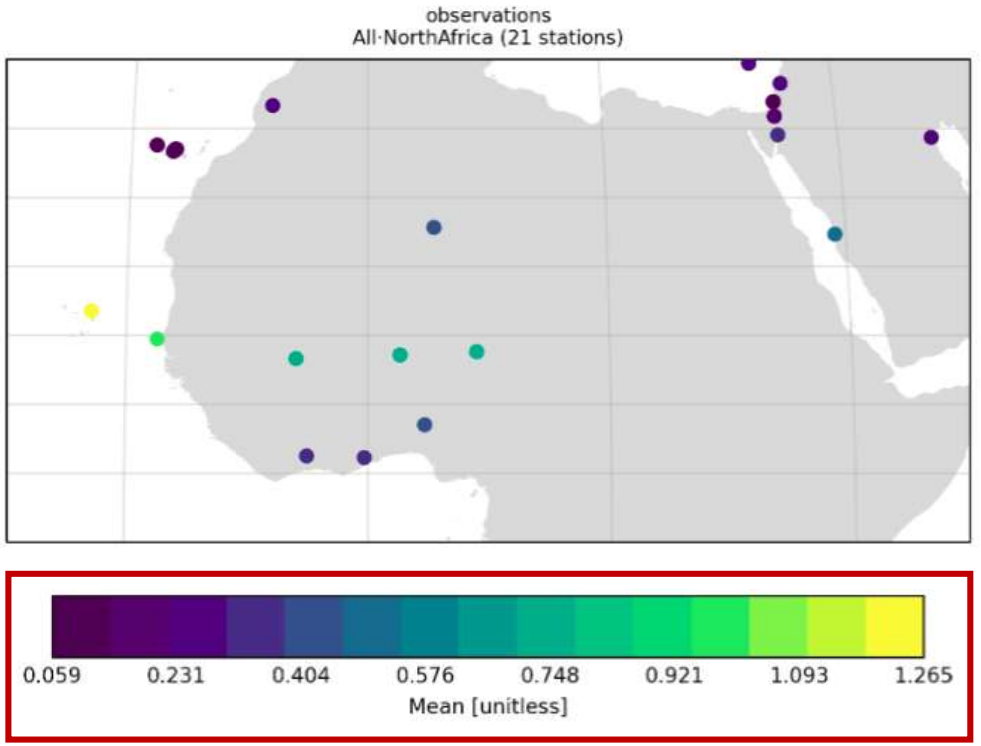
Have you noticed any change?

Let’s put the maps side by side

Before



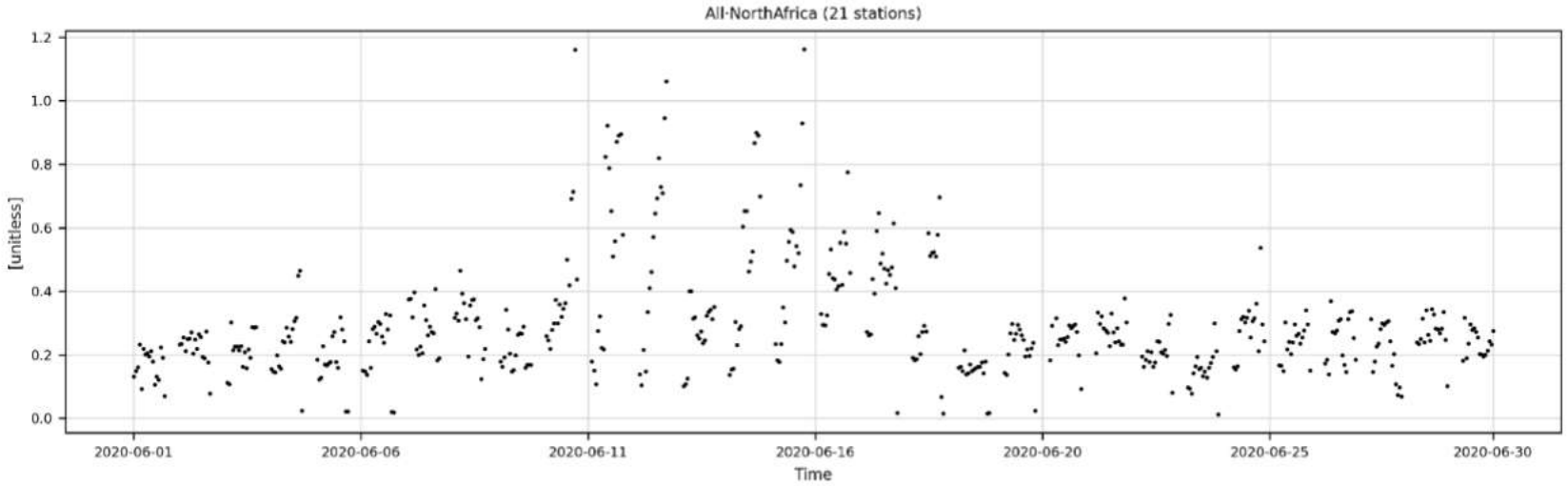
After



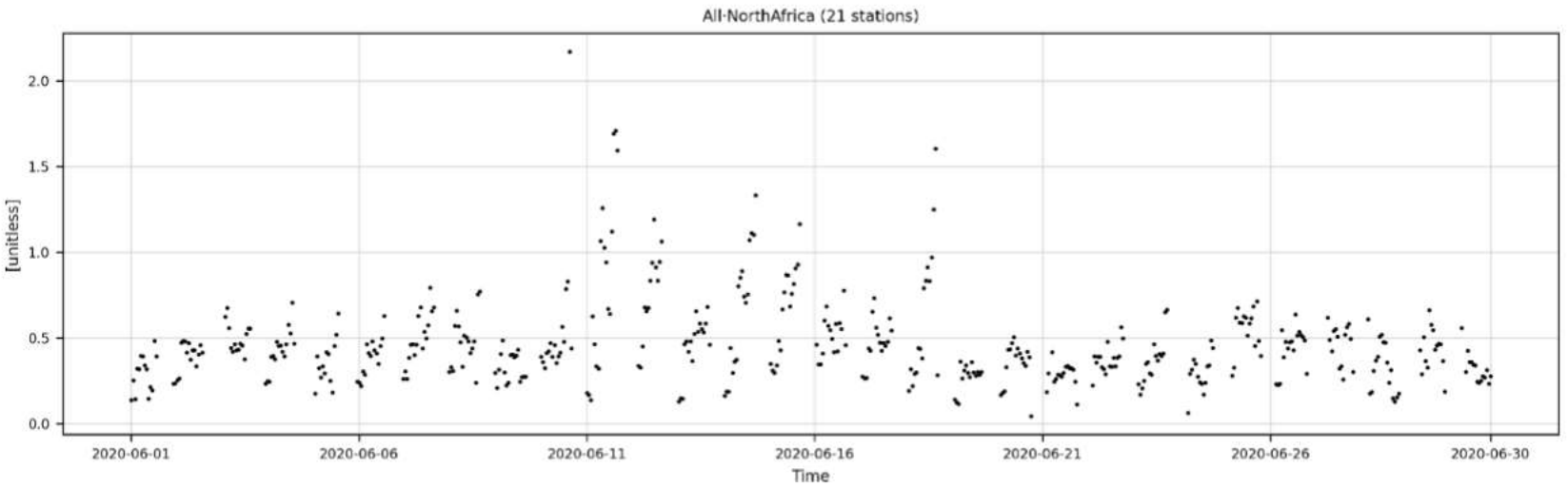
The bounds in the colorbar have changed!

Timeseries for all stations

AOD without multispecies filtering



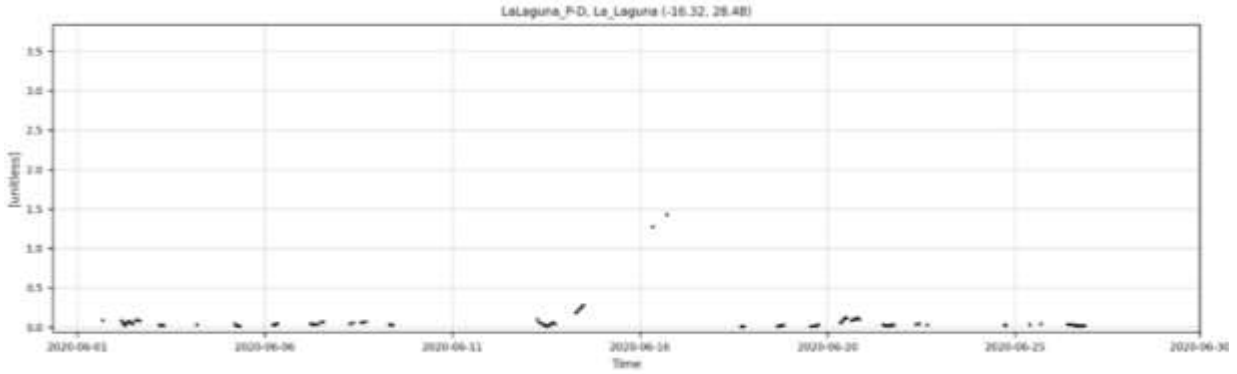
AOD with multispecies filtering



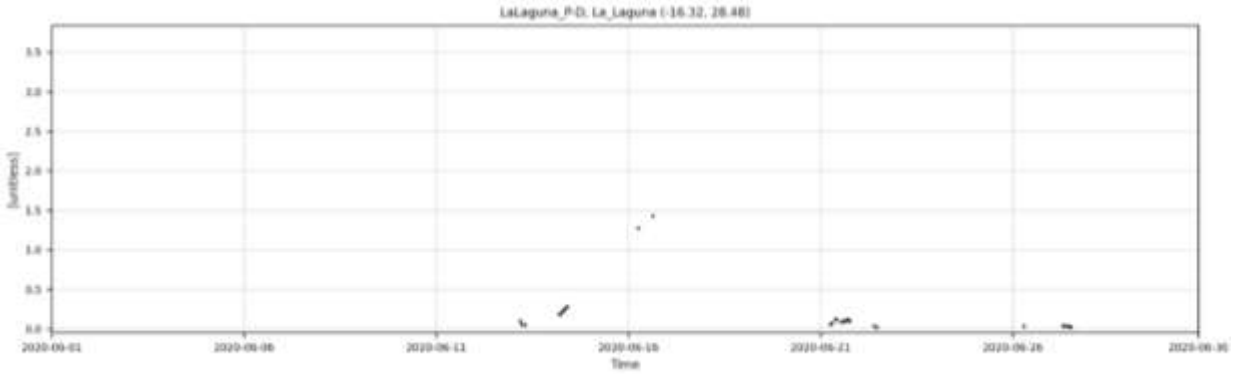
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Timeseries for La Laguna station

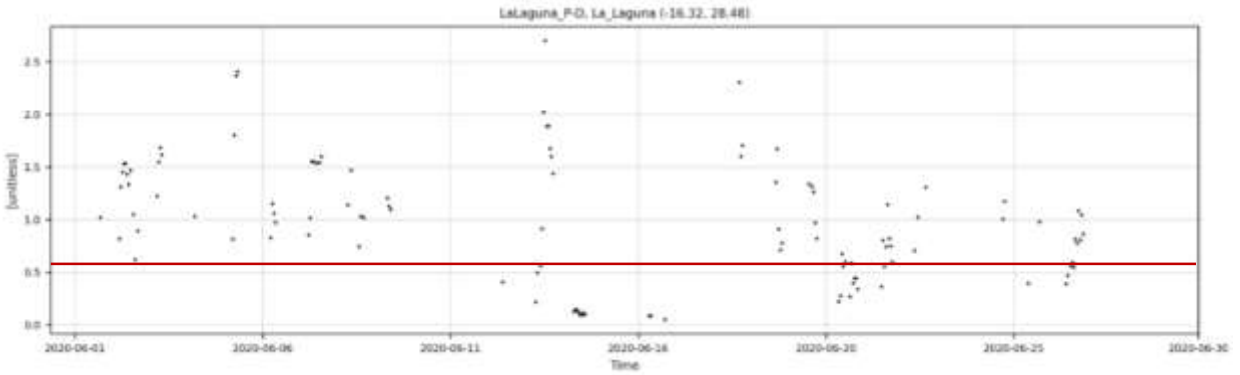
AOD without multispecies filtering



AOD with multispecies filtering



Angstrom exponent



Optional fields – Data filtering

Any data field that exists in GHOST observational files can be used to filter data in sections/subsections.

This extends to periodic variables, representativity variables, QA and flags.

```
[[Barcelona]]
 period = keep: Winter, Daytime || remove: Weekend
 QA = 0,1,2,3

[[Madrid]]
flags = 1
```

Optional fields – Metadata filtering

In a similar vein, any metadata field that is available can be used to filter data.

This also applies to a limited selection of non-GHOST metadata fields (e.g. longitude, latitude). But for GHOST, there is an exhaustive list to choose from.

[[Barcelona]]
latitude = 39.8, 41.8
longitude = 1.5, 2.5

[[Madrid]]
latitude = 39.57, 42.2
longitude = -4.57, -2.42

Plot types

	-[stat]	_bias	_obs	_individual	_annotate	_regression	_multispecies	_logx	_logy	_smooth
map	■	■	■		■					
timeseries		■	■	■	■				■	■
periodic	■	■	■	■	■				■	
periodic-violin			■	■	■				■	
distribution		■	■	■	■			■	■	
scatter				■	■	■		■	■	
heatmap	■	■			■		■			
table	■	■					■			
boxplot			■	■	■		■		■	
statsummary		■					■			
metadata										

■ Available

□ Unavailable

Report plots

Users need to define the **report_type** in the configuration file, which will be linked to the plot types that appear in **settings/report_plots.json**

Configuration file

```
[CAMS2_40]
network = ineris/eionet-cams2_40-ira
species = sconco3,sconcno2,sconcco,sconcso2,pm10,pm2p5
resolution = hourly
start_date = 20230101
end_date = 20230215
experiments = a59g-regional-000, a59j-regional-006 (Forecast, Analysis)
temporal_colocation = True
spatial_colocation = False
report_type = operational
report_summary = True
report_stations = False
report_title = CAMS2_40 Forecast and Analysis Report
report_filename = operational_report
statistic_mode = Temporal|Spatial
statistic_aggregation = Median
periodic_statistic_mode = Independent
periodic_statistic_aggregation = Mean
```

Inside report_plots.json

```
"operational": ["statsummary_multispecies",
"statsummary_multispecies_bias",  "map-Mean", "map-MB",
"map-RMSE", "map-r", "timeseries_annotate",
"timeseries_bias_annotate", "distribution", "scatter_annotate",
"periodic-Mean", "periodic-MB", "periodic-RMSE", "periodic-r"]
```

Plot option –[stat]

It must be used to create maps, periodic plots, heatmaps and tables to indicate the statistic to plot

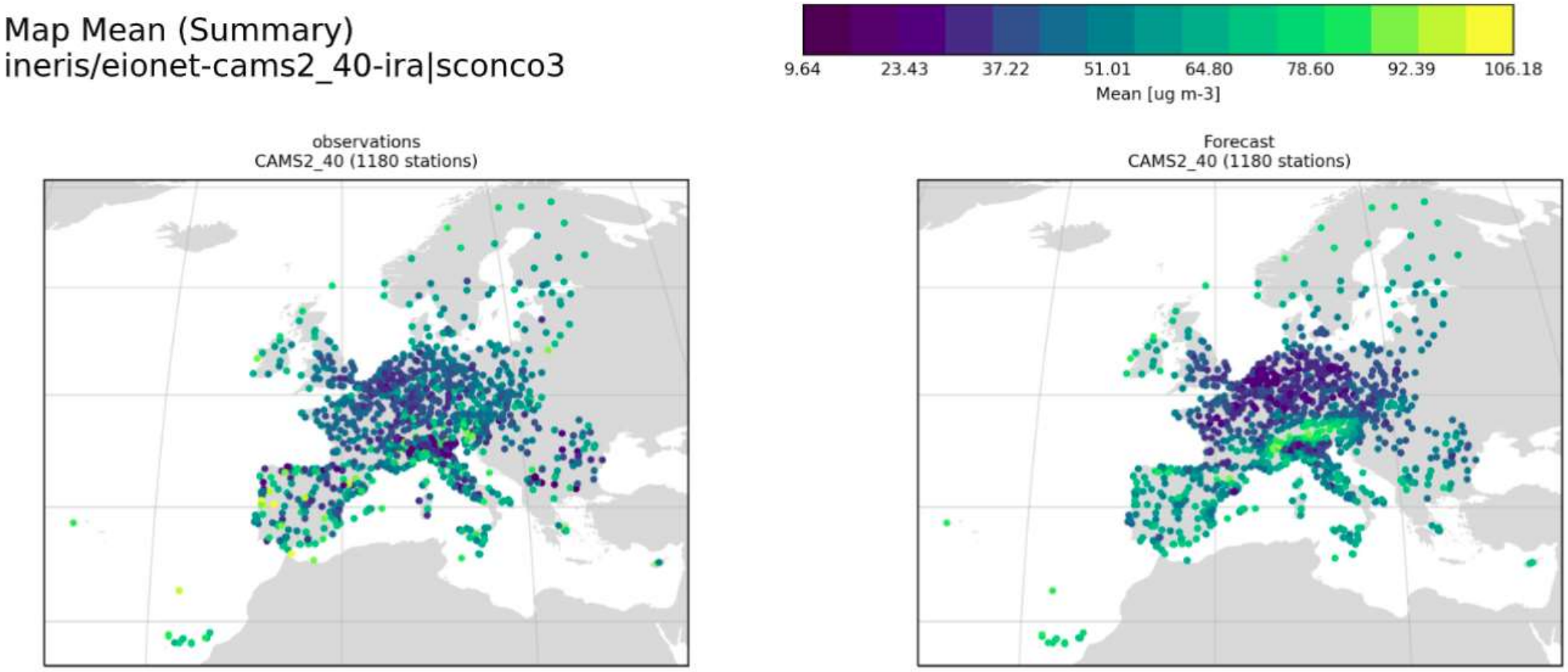
Basic statistics

Statistic	Meaning
Mean	Mean
StdDev	Standard deviation
Var	Variance
Min	Minimum
Max	Maximum
Data%	Data availability
Exceedances	Number of exceedances
p1, p5, p10, p25, p50, p75, p90, p95, p99	Percentiles

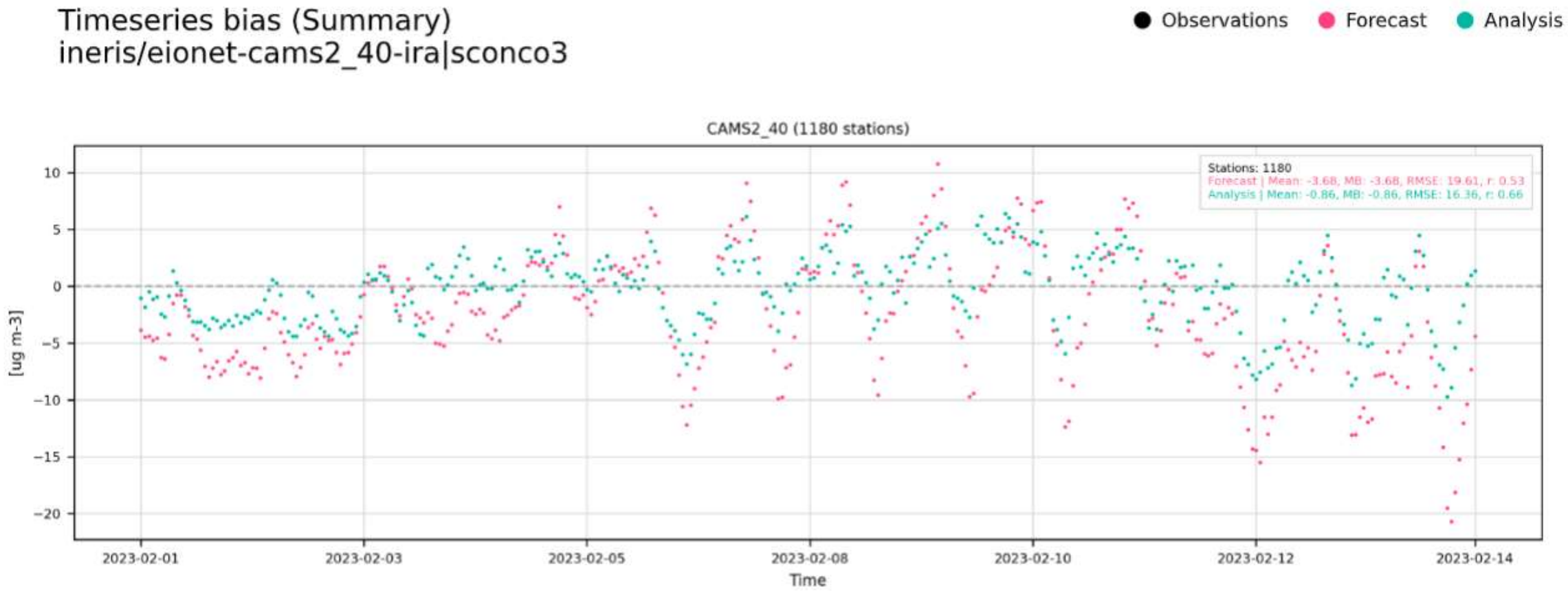
Experiment bias statistics

Statistic	Meaning
MB	Mean bias
NMB	Normalized mean bias
ME	Mean error
NME	Normalized mean error
MNB	Mean normalized bias
MNE	Mean normalized error
MFB	Mean fractional bias
MFE	Mean fractional error
RMSE	Root mean square error
NRMSE	Normalized root mean square error
COE	Coefficient of efficiency
FAC2	Fraction of experiment values within a factor of two of observed values
IOA	Index of agreement
R	Pearson correlation coefficient
R^2	Coefficient of determination
UPA	Unpaired peak accuracy

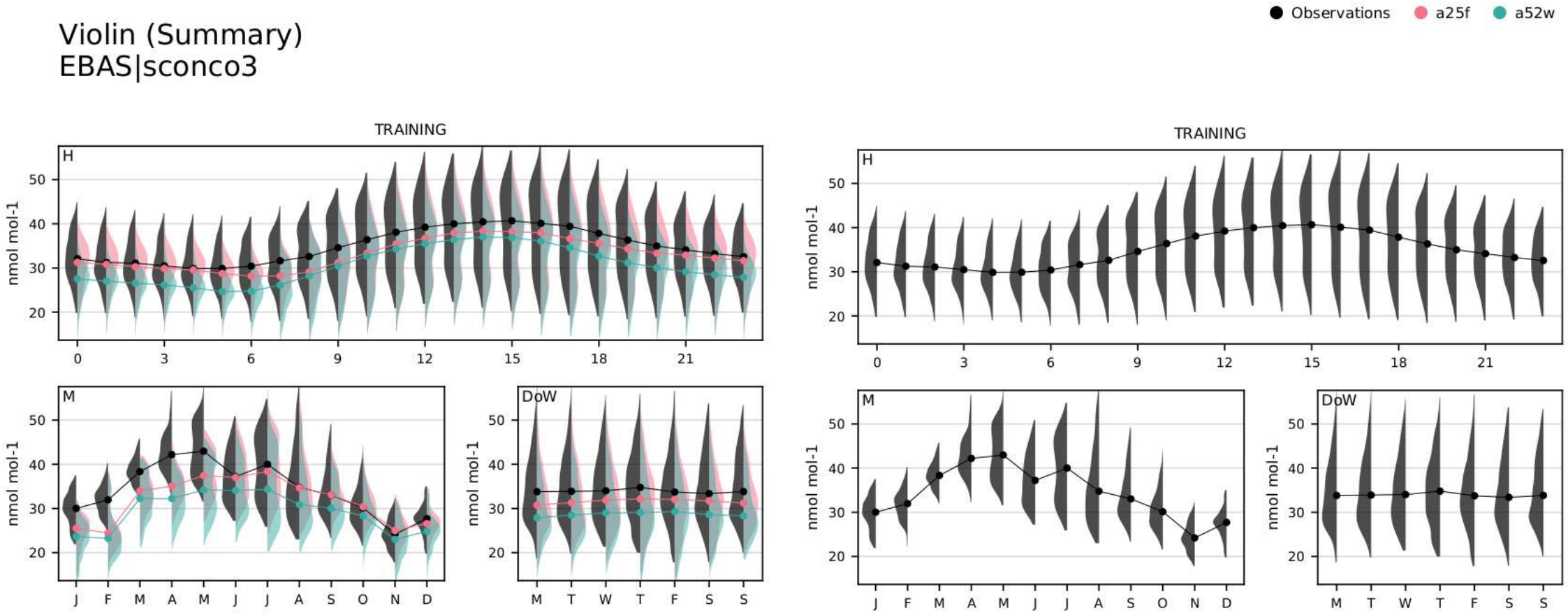
Plot option –[stat]



Plot option _bias

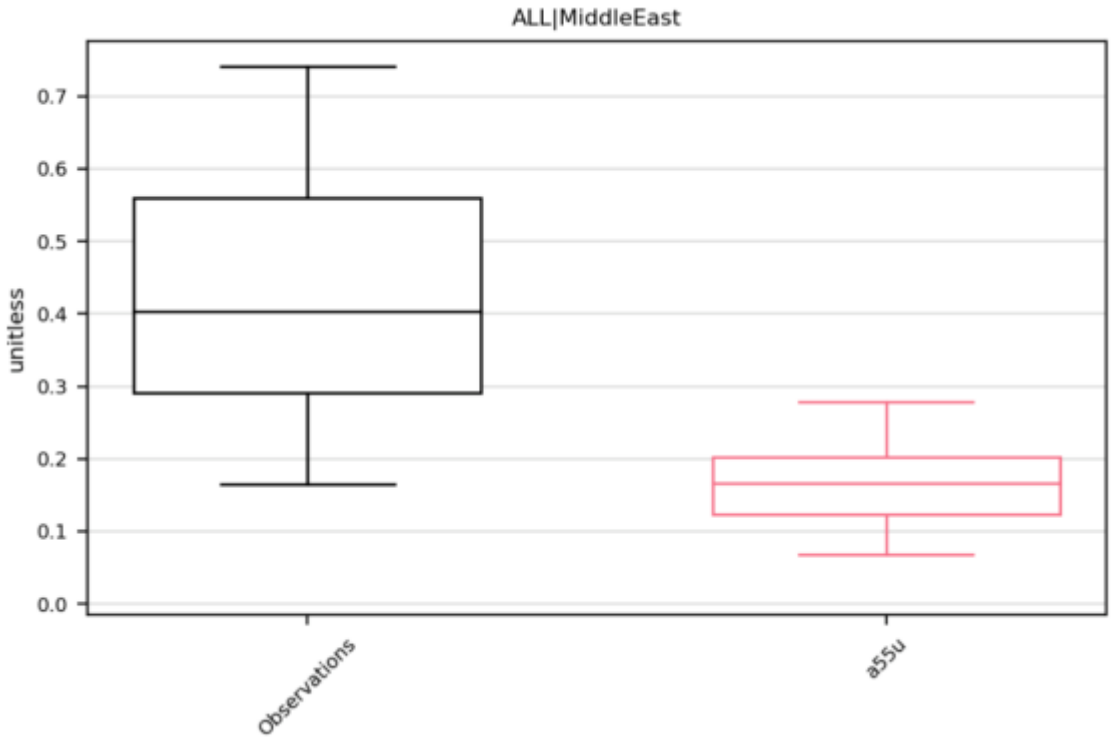


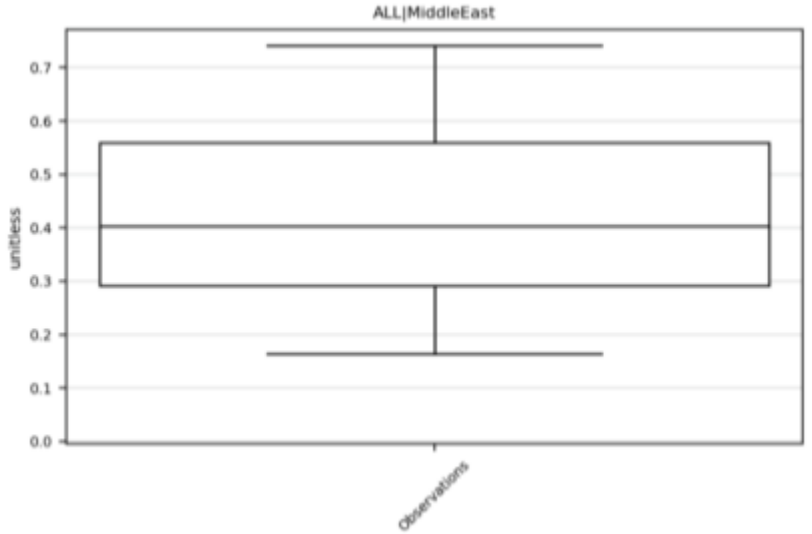
Plot option _obs

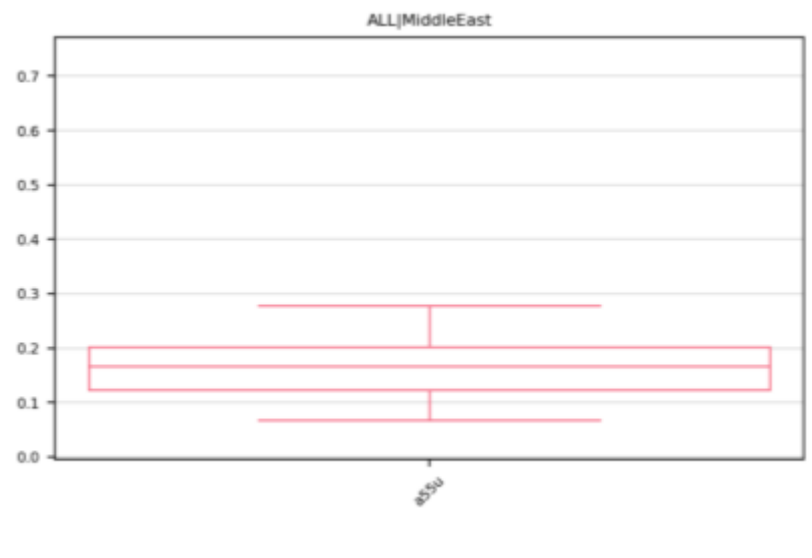


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

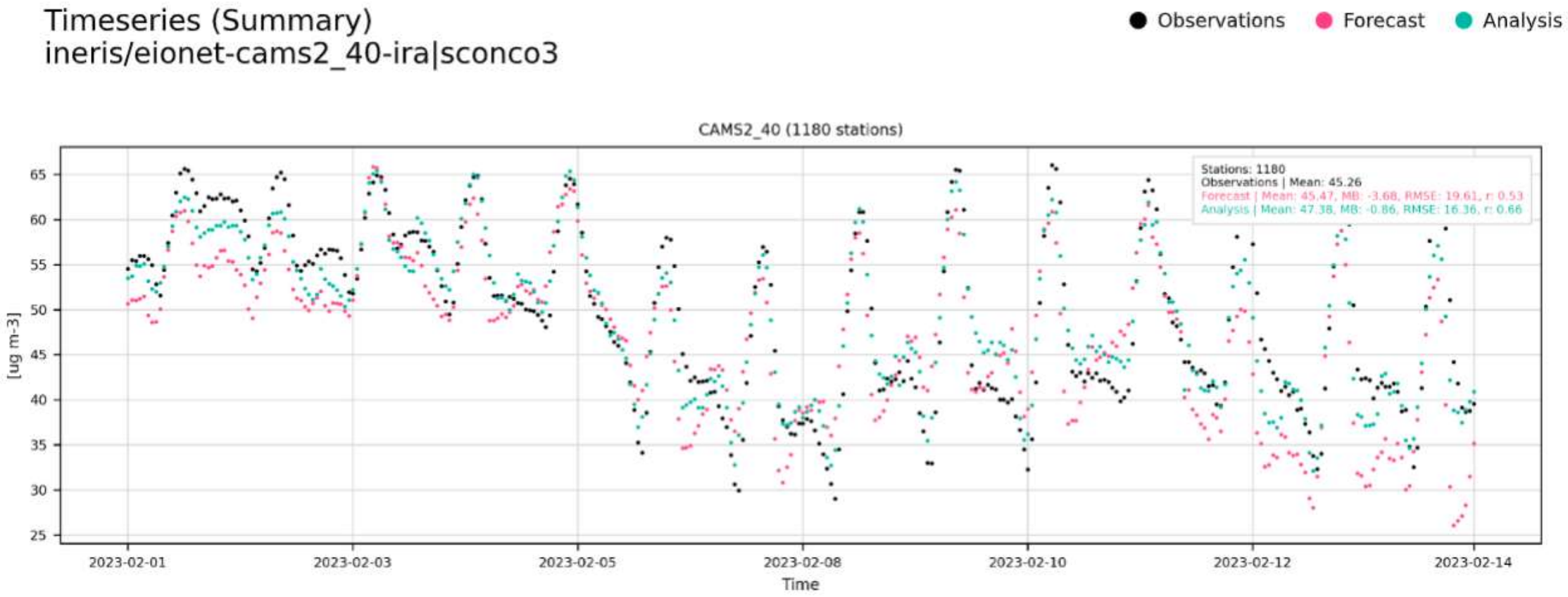
Plot option _individual



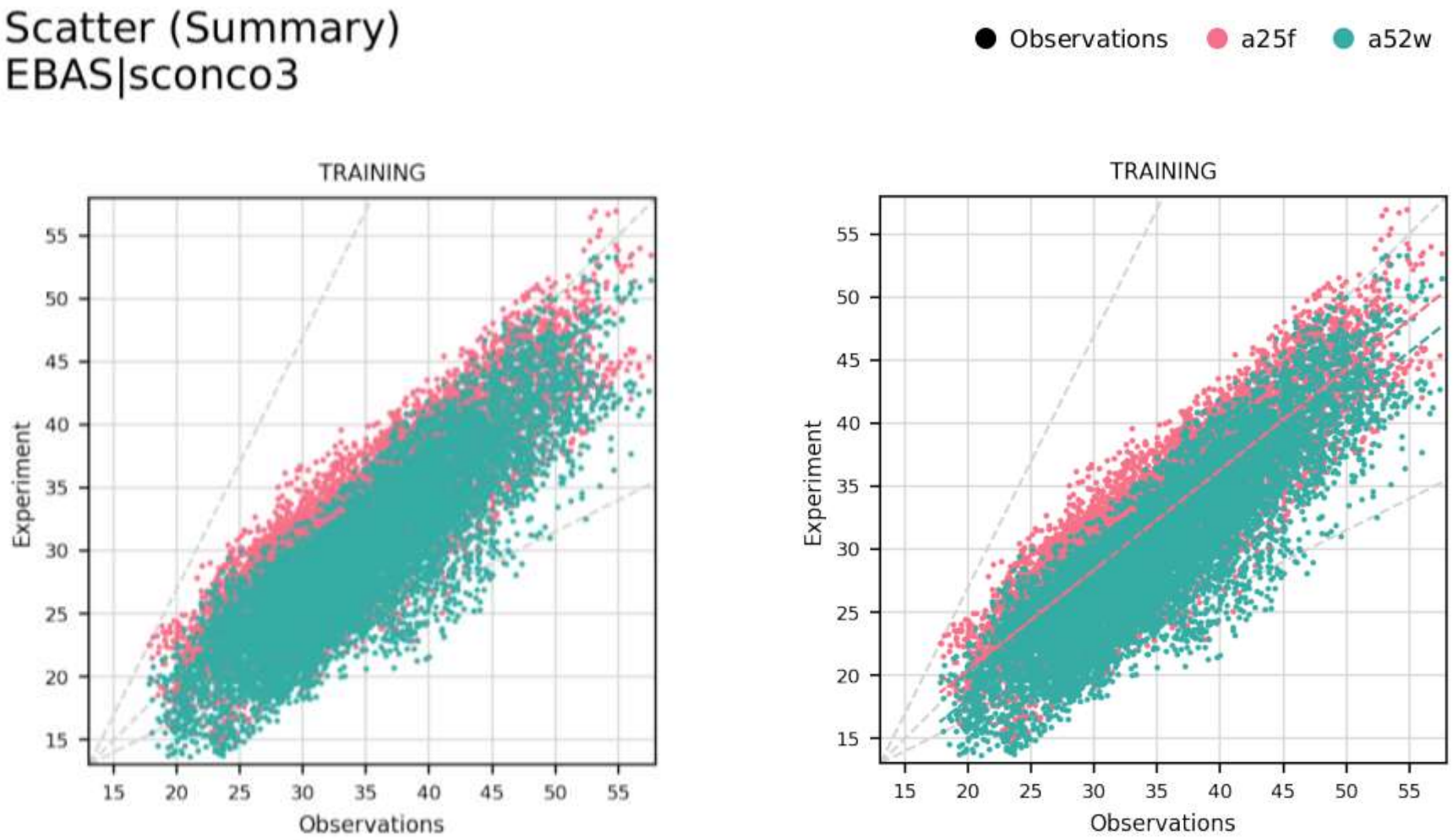




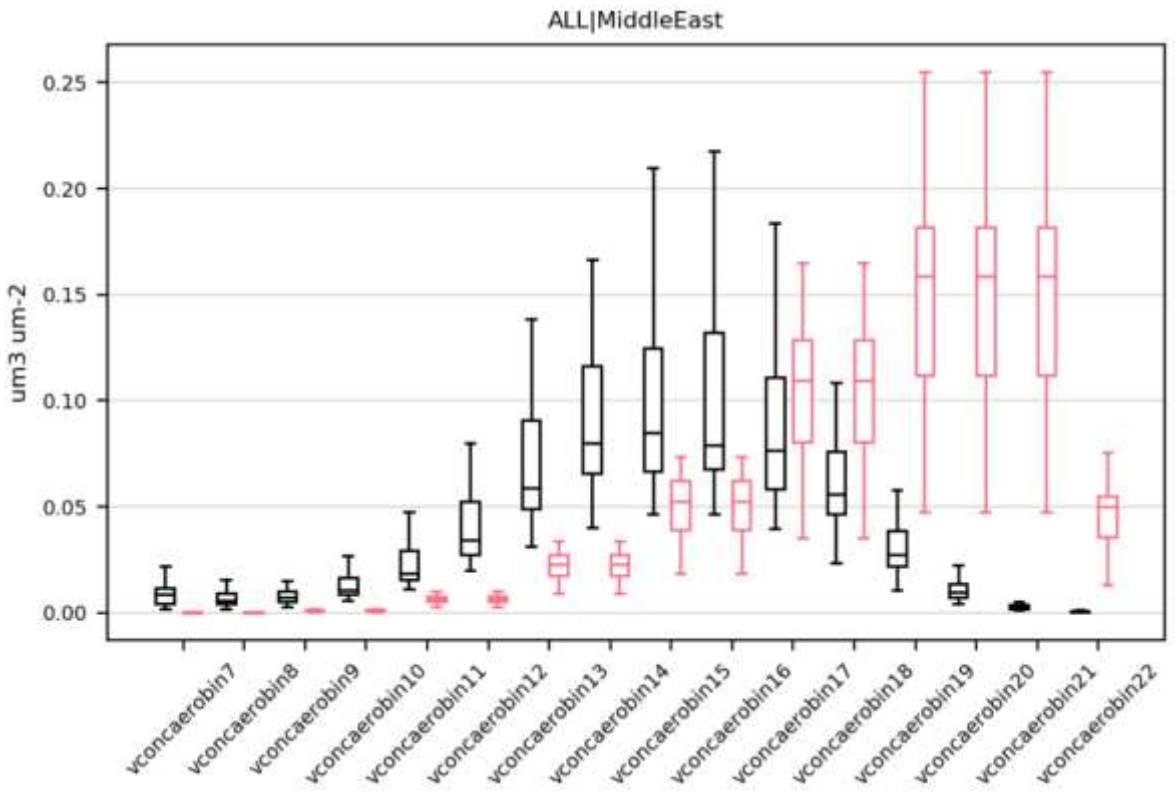
Plot option _annotate



Plot option _regression



Plot option _multispecies



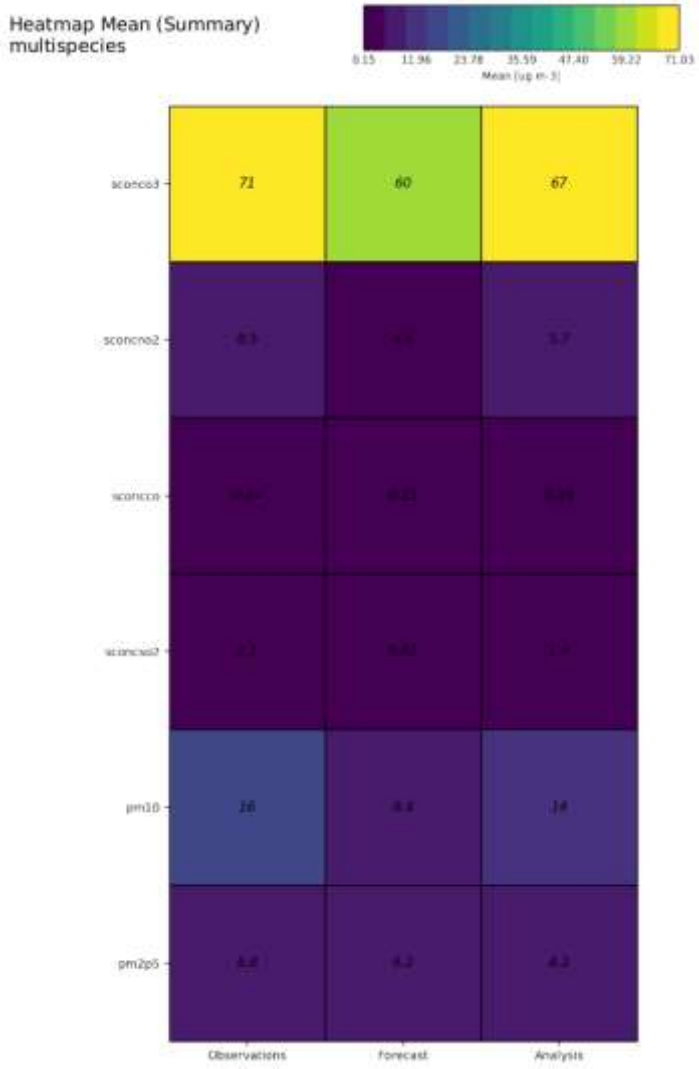
StatSummary (Summary)
multispecies

		p5	Mean	StdDev	p50	p95
sconco3	Observations	50.0	71.03	14.59	69.42	96.0
	Forecast	40.44	60.34	13.72	60.12	83.31
	Analysis	47.14	67.25	13.7	65.81	89.43
sconcno2	Observations	5.64	8.5	2.21	8.0	13.08
	Forecast	1.66	4.59	2.34	4.31	9.09
	Analysis	2.92	5.65	2.09	5.34	9.69
sconcco	Observations	0.2	0.24	0.02	0.24	0.28
	Forecast	0.13	0.15	0.01	0.14	0.17
	Analysis	0.16	0.18	0.01	0.18	0.2
sconcso2	Observations	2.0	2.1	0.16	2.0	2.49
	Forecast	0.45	0.81	0.24	0.82	1.2
	Analysis	0.95	1.45	0.29	1.52	1.82
pm10	Observations	12.01	15.91	2.43	16.19	19.46
	Forecast	5.99	8.44	1.6	8.17	10.96
	Analysis	10.43	13.91	2.27	14.06	17.3
pm2p5	Observations	5.7	8.83	1.84	9.02	11.55
	Forecast	4.18	6.17	1.33	6.03	8.52
	Analysis	5.12	8.15	1.72	8.29	10.84

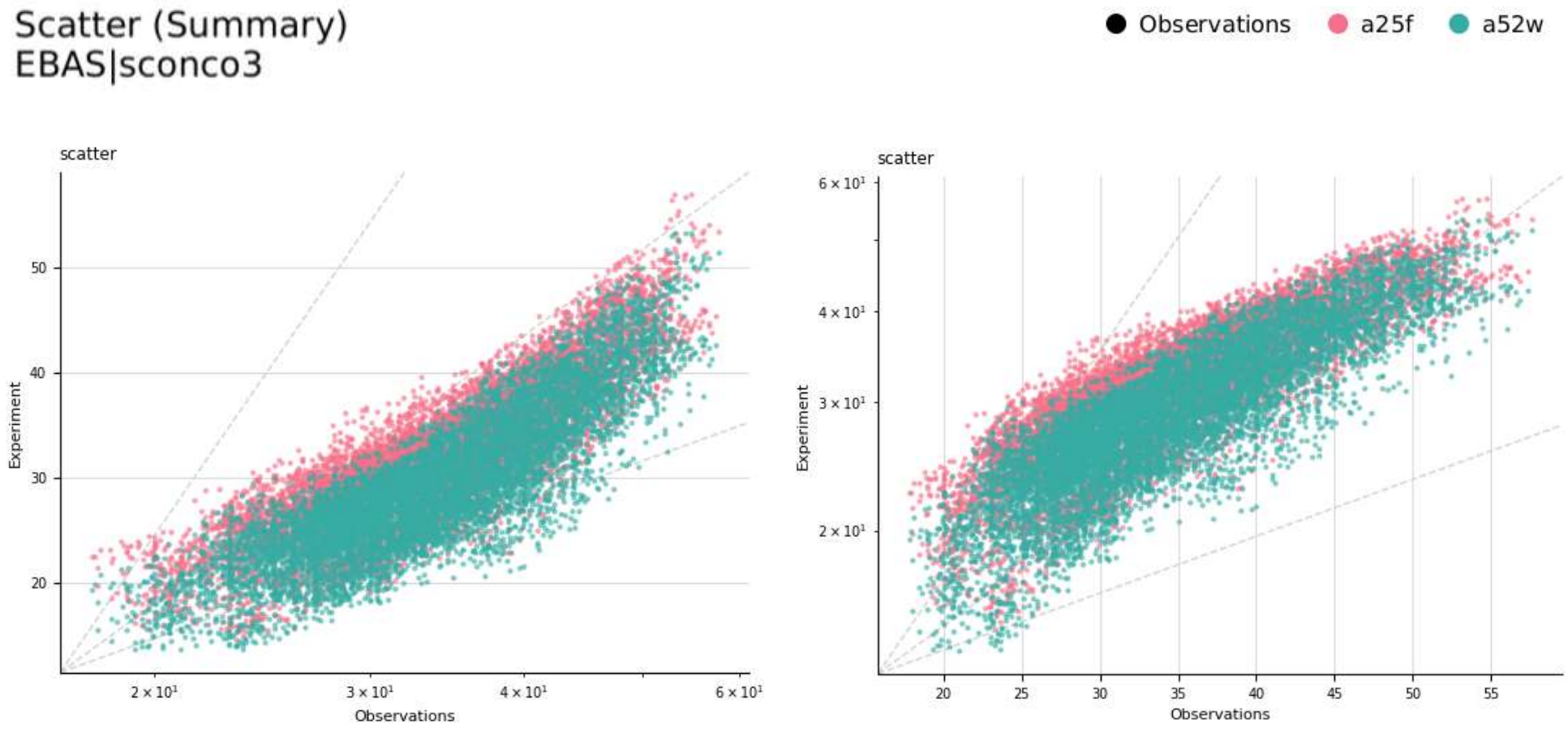
Plot option _multispecies

Table Mean (Summary)
multispecies

	Observations	Forecast	Analysis
sconco3	71.03	60.34	67.25
sconcno2	8.5	4.59	5.65
sconcco	0.24	0.15	0.18
sconcso2	2.1	0.81	1.45
pm10	15.91	8.44	13.91
pm2p5	8.83	6.17	8.15

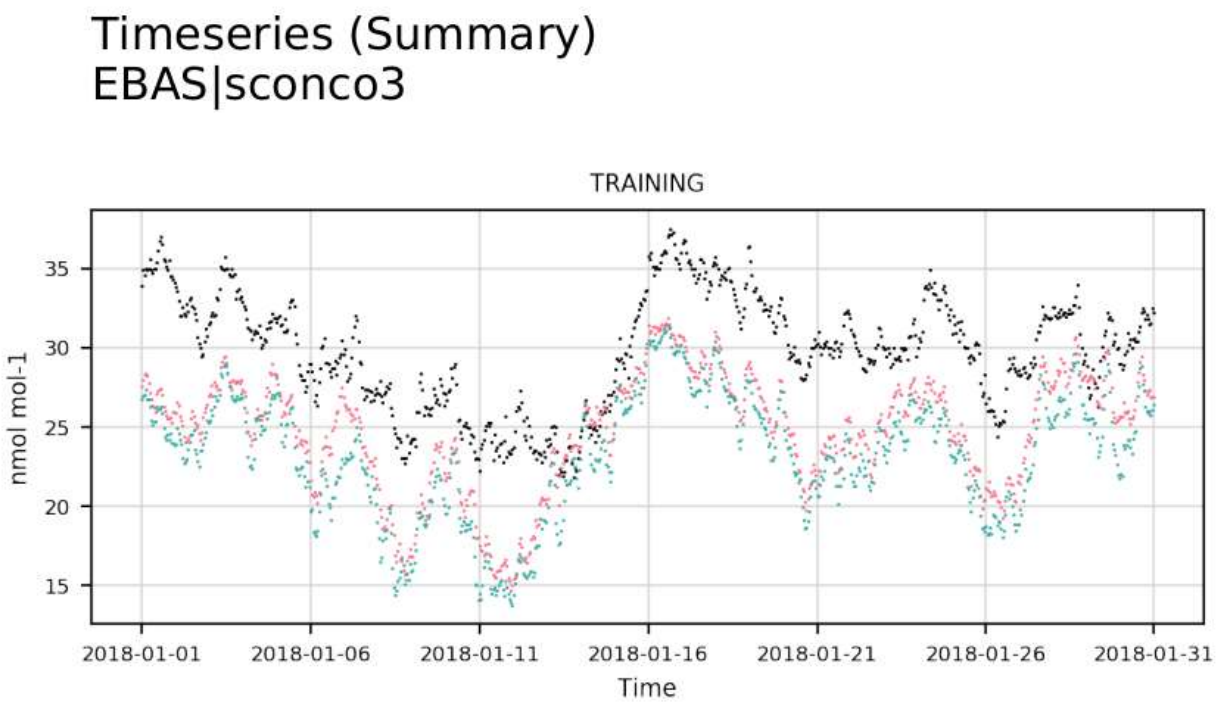


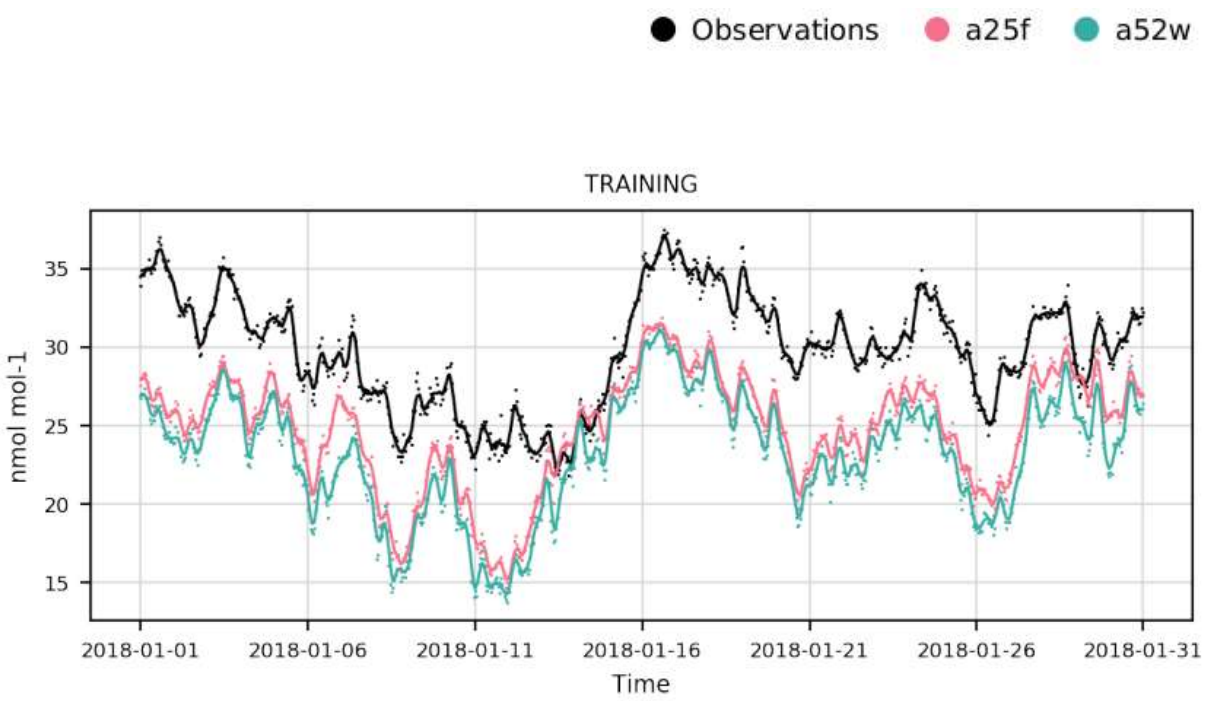
Plot options _logx and _logy



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Plot option _smooth





Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Plot customization

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Plot customization

If you want to edit the plot style and statistics, you will need to edit the file **settings/plot_characteristics_offline.json**. Most parameters are based in Matplotlib 3.1.1 and have been summarized in:

https://earth.bsc.es/gitlab/ac/Providentia/-/wikis/Plot-customization

Inside the same folder, you will find the file **plot_characteristics_dashboard.json**, which you can use to style the dashboard.

Customize the colorbar

There are two ways to change the colorbar bounds:

- If you want to set the same bounds for all statistics, you can edit the parameters *vmin_absolute*, *vmax_absolute*, *vmin_bias* and *vmax_bias* under *map* in the plot characteristics files.
- If you want to set the bounds for each statistic (recommended), you should edit the same parameters in **settings/basic_stats.json** and **settings/experiment_bias_stats.json**.

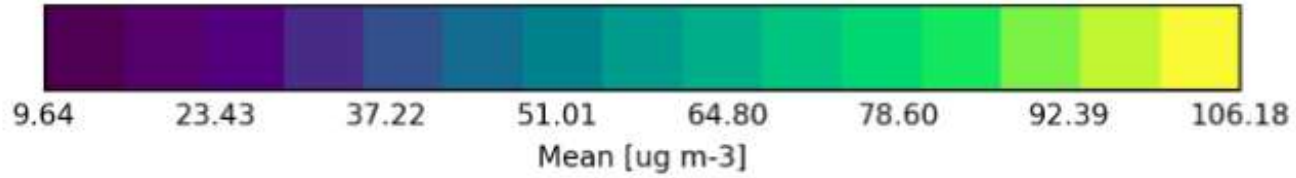
To change the colors, you need to edit the cmap:

- If you want to set the same colors for all statistics, edit the parameters *cmap_absolute* and *cmap_bias* under *map* in the plot characteristics files.
- If you want to set the colors for each statistic (recommended), you should edit them in **settings/basic_stats.json** and **settings/experiment_bias_stats.json**.

Customize the colorbar

You might also want to change the number of breaks if you have a discrete colorbar. For this you will need to change the number of ticks (*n_ticks*) and number of discrete colors (*n_discrete*) under *map* in the plot characteristics files.

In this example, we have set *n_ticks* to be 8 and *n_discrete* to be 15.

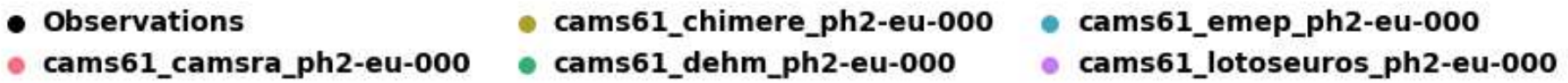


If you prefer to have a continuous colorbar and want to remove the breaks, you need to set *discrete* to be false.

Customize the legend

If you want to change the color of the legend, you can edit the *legend_color_palette* under *general* in the plot characteristics files.

The default palette is *husl*:

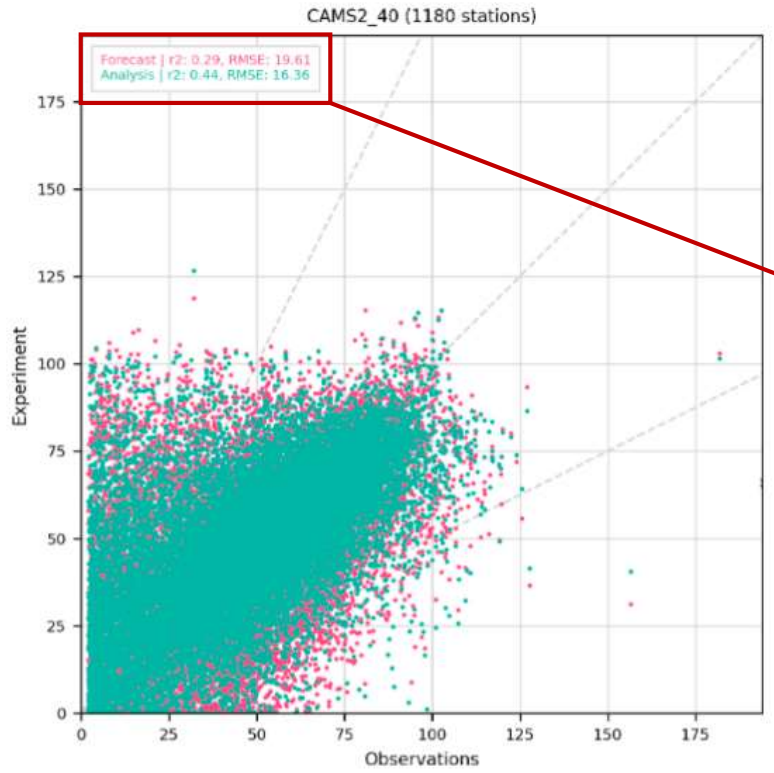


If you want to edit the names of the experiments, you can only do it launching Providentia from a configuration file, where you will define the alternative names as in:

experiments = cams61_chimere_ph2-eu-000, cams61_emep_ph2-eu-000 (CHIMERE, EMEP)

Edit the statistics

If you want to change the statistics that appear when you make annotations on the plots, you will need to go to each of the plot types in the plot characteristics files and change *annotate_stats*.



"annotate_stats" = ["r2", "RMSE"]


Q&A

Future plans

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

What to expect in Providentia 2.3.0

Grouping

Forecast with more than 24 hours

New plot types

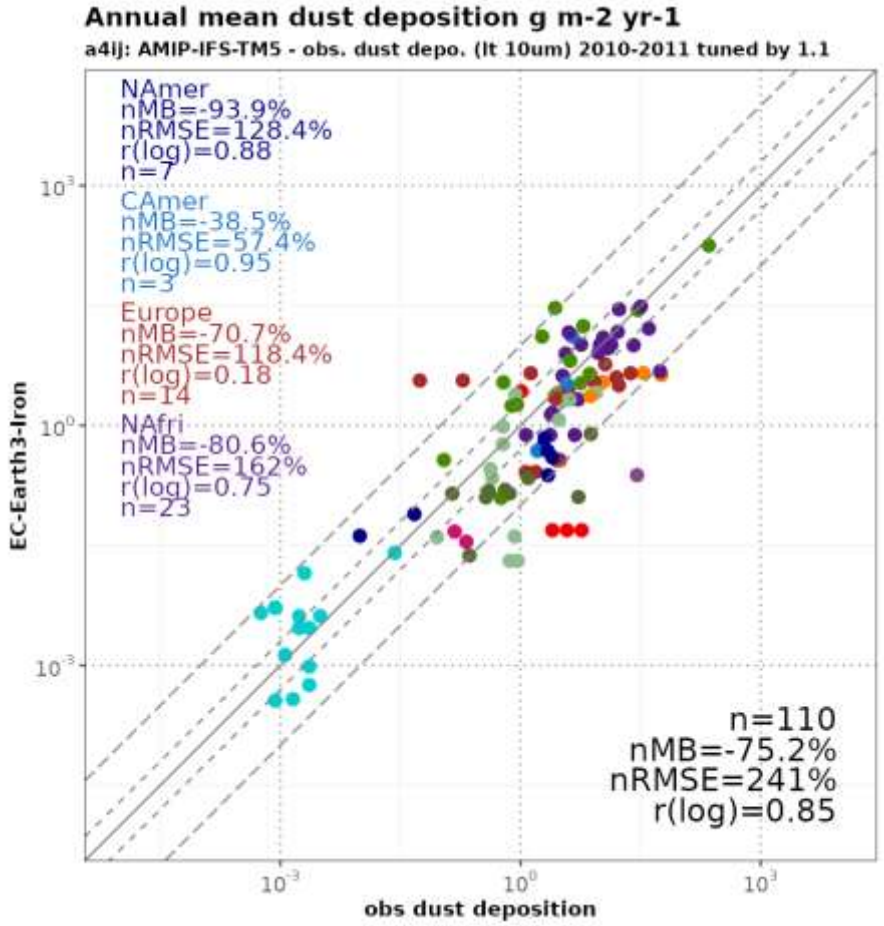
Grouping

Create and use custom subsets per plot type to highlight data per subset.

These subsets could be GHOST metadata fields or be defined in a JSON file.

They will be activated from the main menu or through the configuration file.

Example: We could define dust regions using their map extent and show them in a scatter plot as in the image on the right.



Grouping: Proposed options

- Assimilated vs. validated stations.
- Dust regions (by coordinates).
- Common GHOST metadata:
  - Countries
  - Continents
  - Land use
  - Distance to coast
- Station classifications.
- Area classifications.

Add your ideas here:
https://earth.bsc.es/gitlab/ac/Providentia/-/issues/280

Discussion

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Thank you for your attention!

More information at:
https://earth.bsc.es/gitlab/ac/Providentia

Join the #providentia Slack channel!

alba.vilanova@bsc.es | dene.bowdalo@bsc.es