Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Providentia v2.2
User Meeting

03/08/2023

Alba Vilanova | Dene Bowdalo

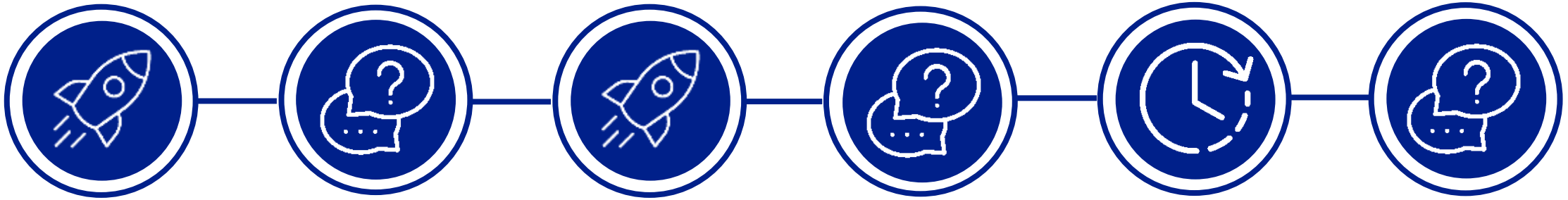
Timeline

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Statistical features
15 min.

Other features
10 min.

Future plans
10 min.



Q & A
5 min.

Q & A
5 min.

Discussion

Statistical features

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Statistical features

- **Optimisation to statistical calculations.**

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

Statistical modes

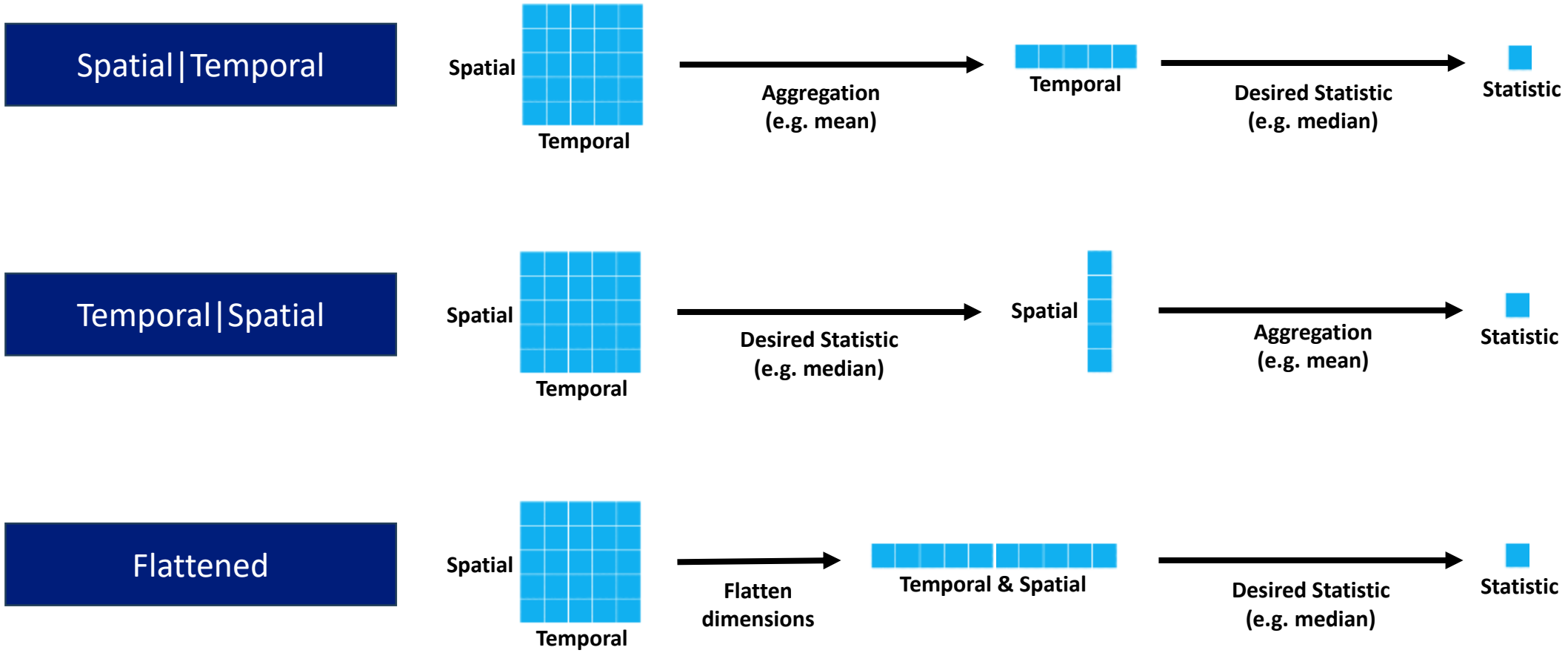
The name of each statistical mode relates to how the dimensions of the selected data are reduced to calculate the statistical metrics, e.g. mean, median etc, going from 2D to 0D.

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

Spatial and **Temporal**.

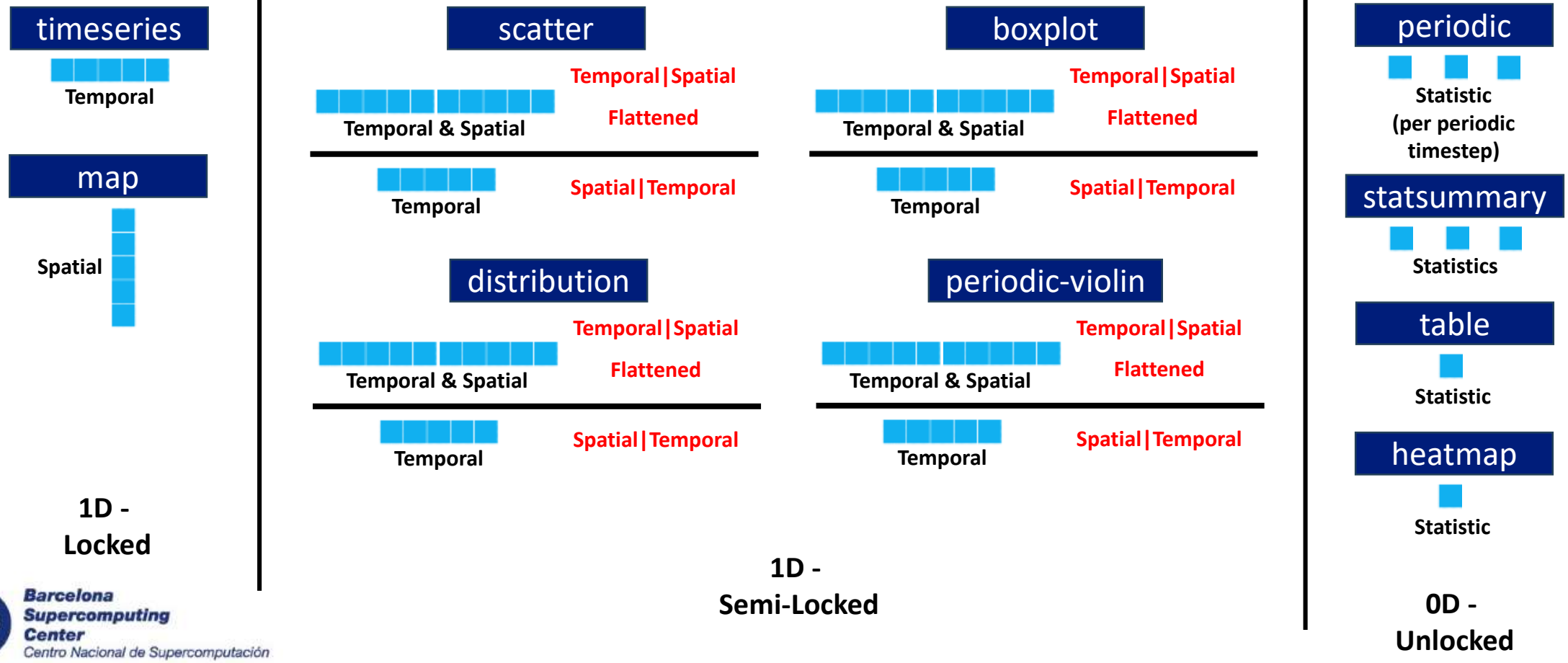
In the previous versions of Providentia, the mode was always **Spatial|Temporal**, with aggregation performed across stations (e.g. taking mean across stations per timestep), going to 1D, before calculating the desired statistic across the aggregated timesteps (e.g. median), going to 0D.

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.

Statistical features

- **Calculation of statistics per periodic cycle (2 modes):**
  - Diurnal
  - Weekly
  - Monthly

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

Other features

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Other features

- **Visual changes to the dashboard.**
- **Creation of debug mode.**

  1. Open session with: ./bin/providentia –debug
  2. Run tool as usual: ./bin/providentia

- **Export data with active temporal colocation.**
- **Compression of reports.**
- **Stacking heatmaps, tables and statistical summaries.**

  Adding “multispecies” to the plot type, i.e. “statsummary_multispecies”.

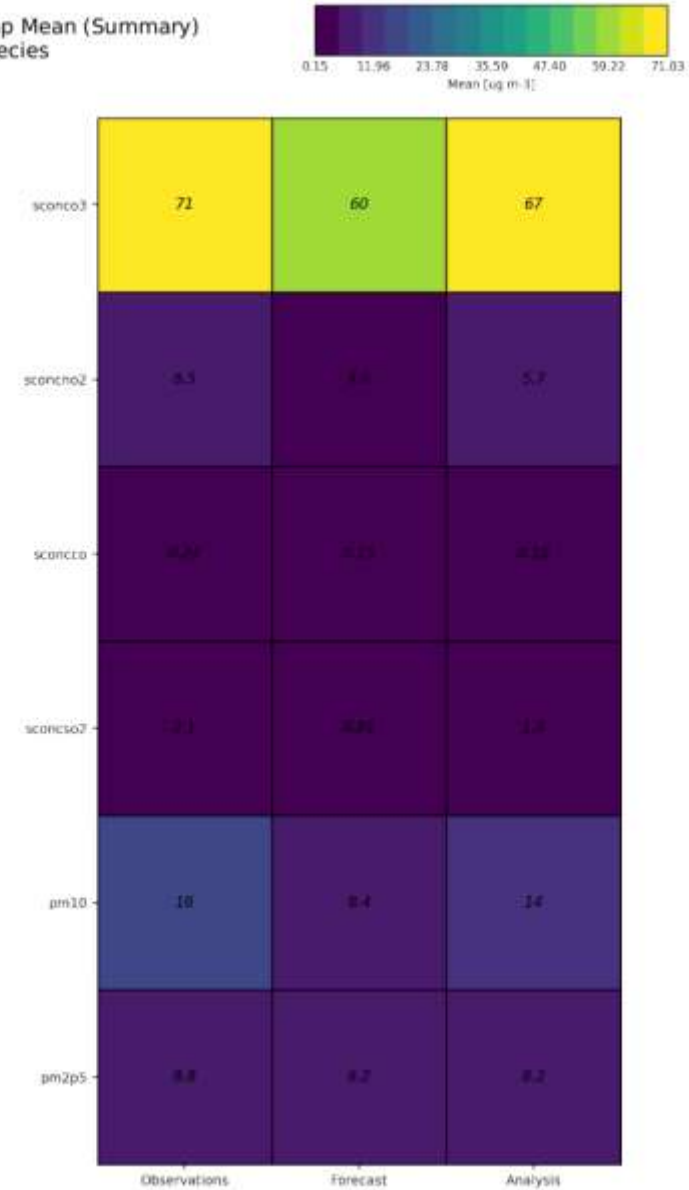
Table Mean (Summary) multispecies

	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

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

Heatmap Mean (Summary) multispecies



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Other features

- **Use external files for custom plot characteristics.**

  In your configuration file:

  plot_characteristics_filename = dashboard:path_dashboard, offline:path_offline

- **Multiple filters for the same variable.**

  In your configuration file, for example:

  filter_species = nasa-aeronet/directsun_v3-lev15:ae440-870aero (>0.75, <=1.2, nan),

  nasa-aeronet/directsun_v3-lev15:ae440-870aero (>1.2, :, 0)

  This will convert all values between 0.75 and 1.2 of ae440-870aero to NaN, and those above 1.2 to 0.

Other features

- Create settings folder with content inside the old conf.
- Ability to select/deselect statistics in statsummary.
- Fix bugs.


Q&A

Future plans

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

WHAT TO EXPECT IN PROVIDENTIA v2.3.0

Grouping

Forecast with more than 24 hours

New plot types

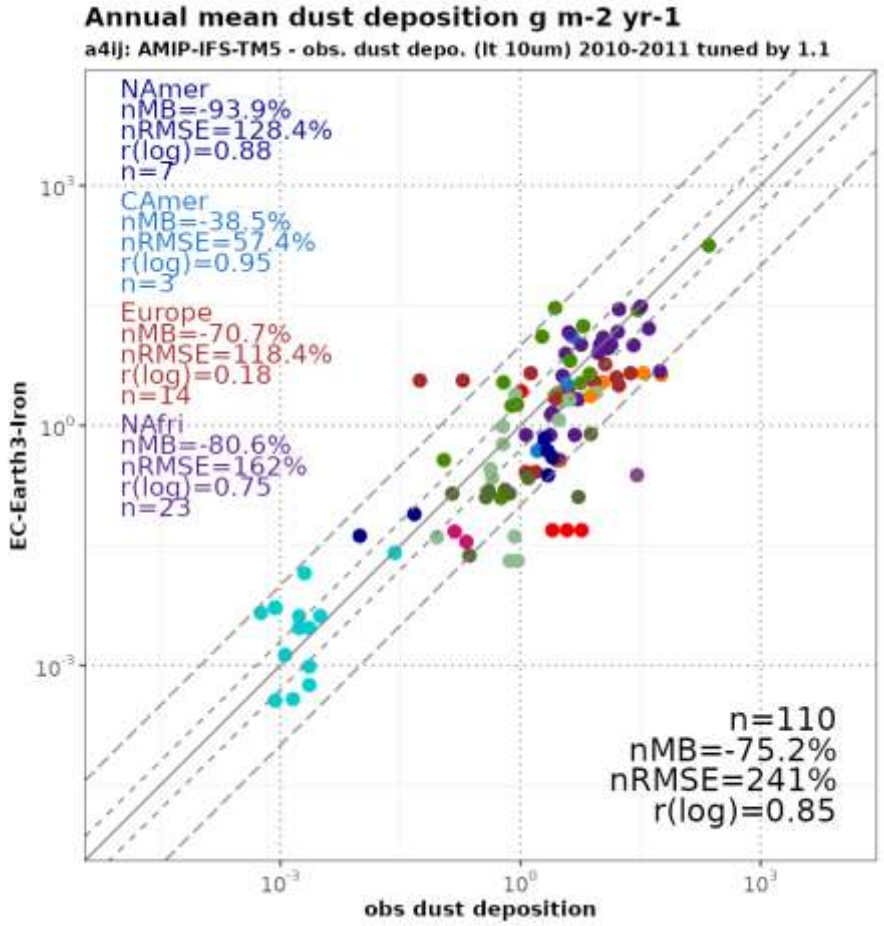
Grouping

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