Barcelona Supercomputing Center
Centro Nacional de Supercomputación

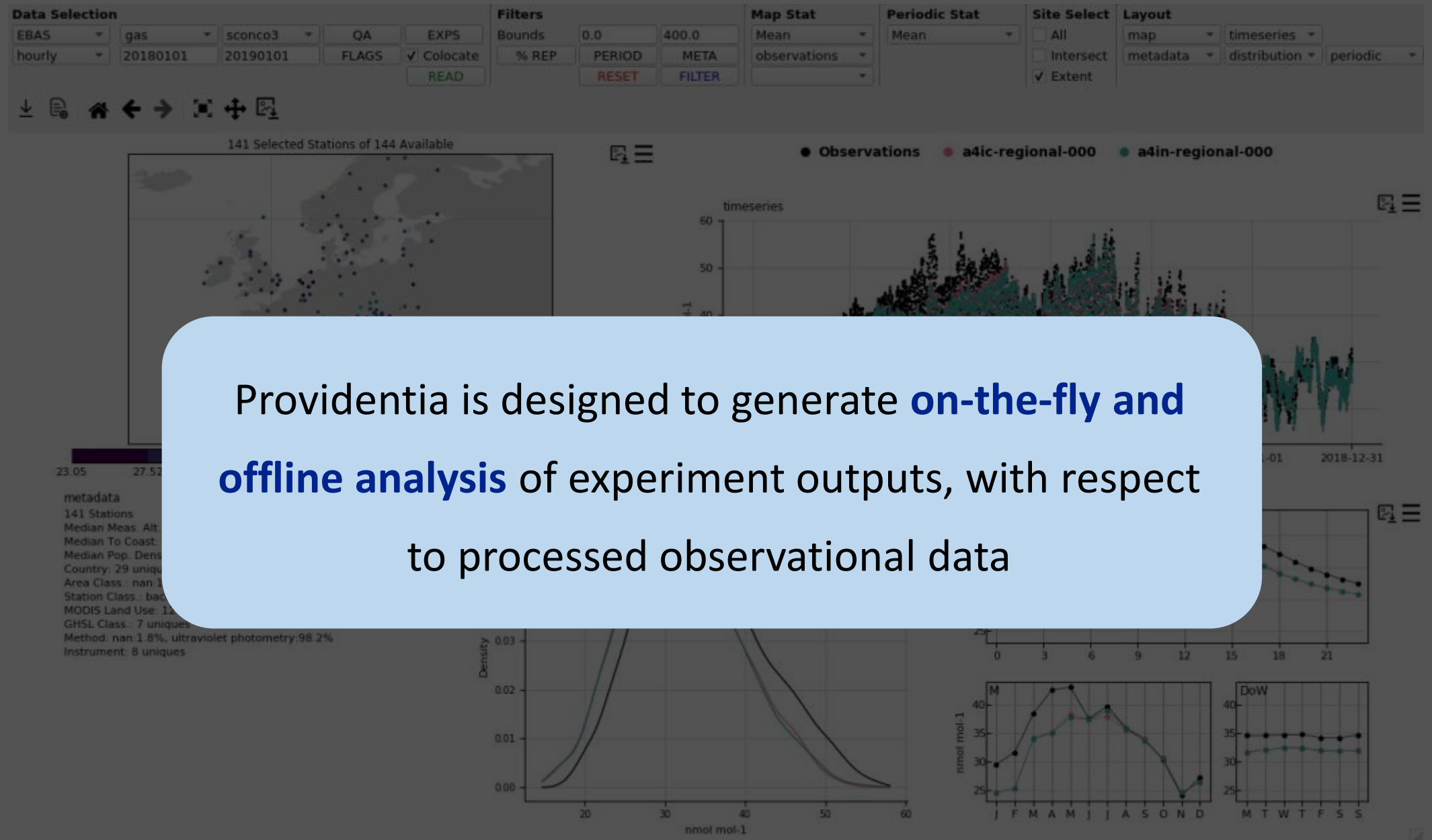
Providentia v2.0
Training Session 2

29/09/2022

Alba Vilanova | Dene Bowdalo

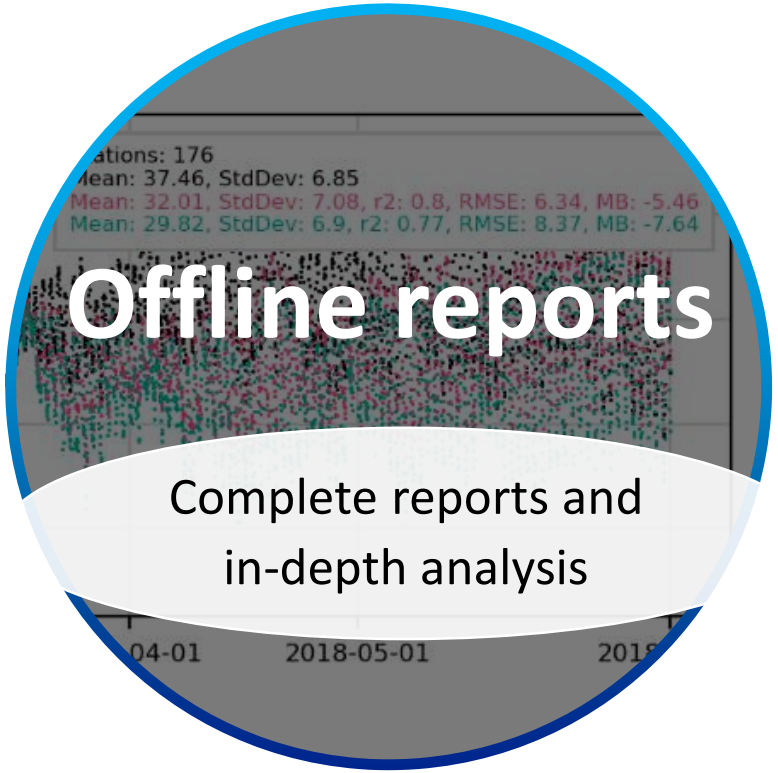
Introduction

Barcelona Supercomputing Center
Centro Nacional de Supercomputación



Providentia is designed to generate **on-the-fly and offline analysis** of experiment outputs, with respect to processed observational data

VISUALIZATION TOOLS



Report

Network = ['EEA_AQ_eReporting']
Temporal Resolution = hourly
Species = ['sconcno2']
Date Range = 20190101 - 20200101
Experiments = ['a4eh', 'a4ez', 'a4f1', 'a4m3', 'a4m4']
Subsections = ['SPAIN|Spain', 'SPAIN|Barcelona']

In PDF!

STEPS

1. Set up your connection to BSC machines
2. Clone Providentia Interpolation and Providentia
3. Interpolate your experiments

TRAINING SESSION 1

4. Create configuration files
5. Define report plots
6. Launch the offline reports
7. Edit the plots style

4

Create configuration files

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXAMPLE OF A CONFIGURATION FILE

SECTION
Defined with [ ]

SUBSECTION
Defined with [[ ]]

SUBSECTION
Defined with [[ ]]

```
[TRAINING]
network = EBAS
resolution = hourly
matrix = gas
species = sconco3
start_date = 20180101
end_date = 20190101
temporal_colocation = True
spatial_colocation = True
experiments = a52f-regional-000, a52w-regional-000 (a52f, a52w)
report_filename = Providentia_Report
report_title = Report
report_type = training
report_summary = True
report_stations = False

   [[Barcelona]]
   latitude = 39.8, 41.8
   longitude = 1.5, 2.5
   period = keep: Spring, Daytime ||

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

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
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 path to the file containing the plot characteristics.

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.
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.
filter_species	Filter read species by other species data within a data range (can be multiple), e.g. sconcno (5.0, 15.0), sconcco (20.0, 200.0)

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
```

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

5

Define report plots

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

PLOT SELECTION

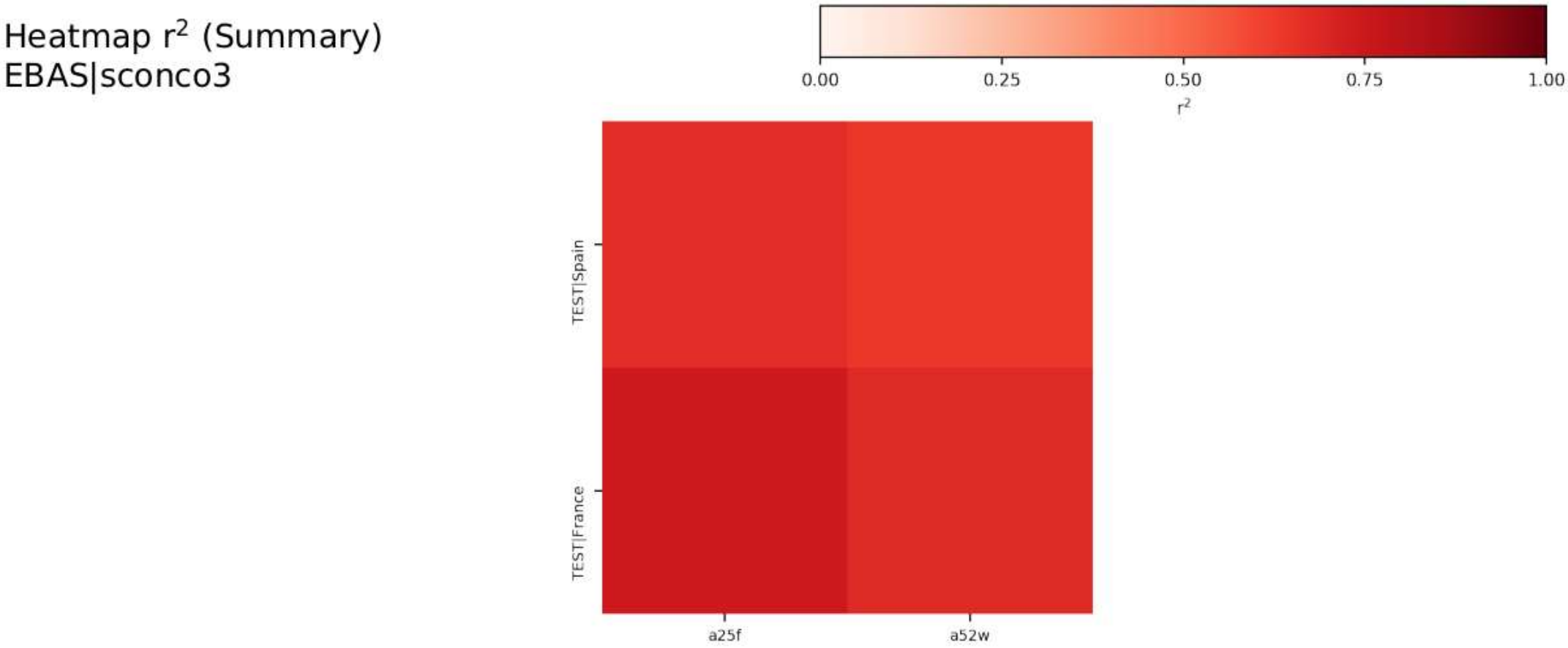
Users should list the plots they want to have in the report in **conf/report_plots.json** under a report type name, which then will appear in the configuration file

Available

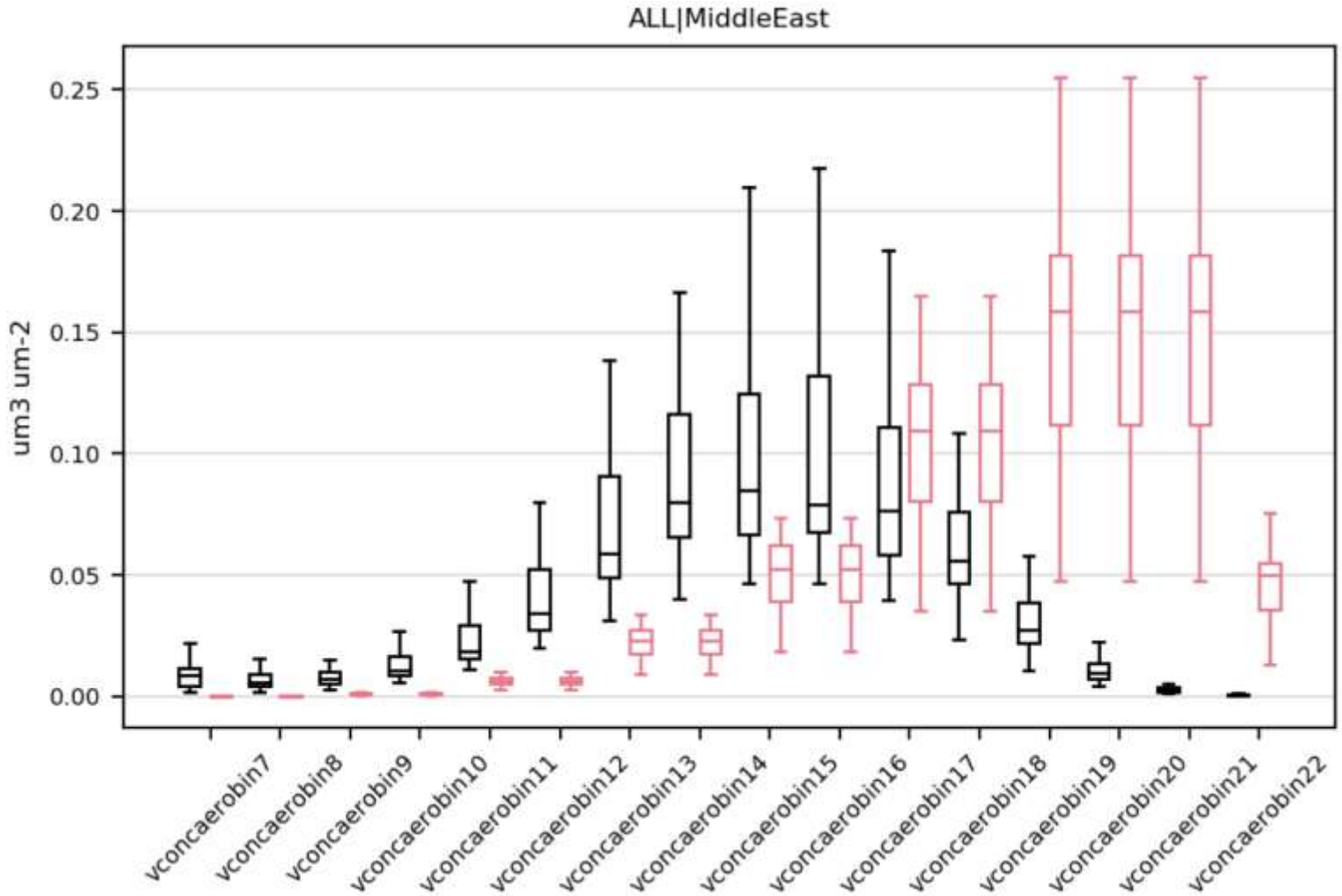
Unavailable

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

ADDITIONAL PLOTS: HEATMAPS



ADDITIONAL PLOTS: BOX PLOTS



PLOT OPTION -[stat]

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

Basic statistics

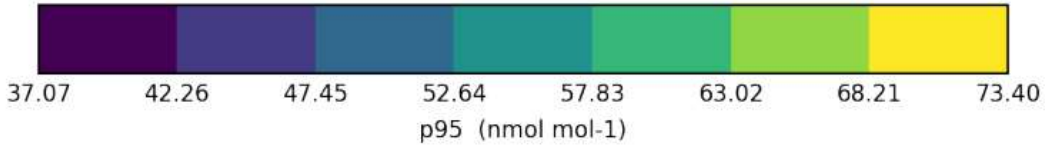
Statistic	Meaning
Mean	Mean
StdDev	Standard deviation
Var	Variance
Min	Minimum
Max	Maximum
Data%	Data availability
Exceedances	Number of exceedances
p1, p5, p10, p25, p50, p75, p90, p95, p99	Percentiles

Experiment bias statistics

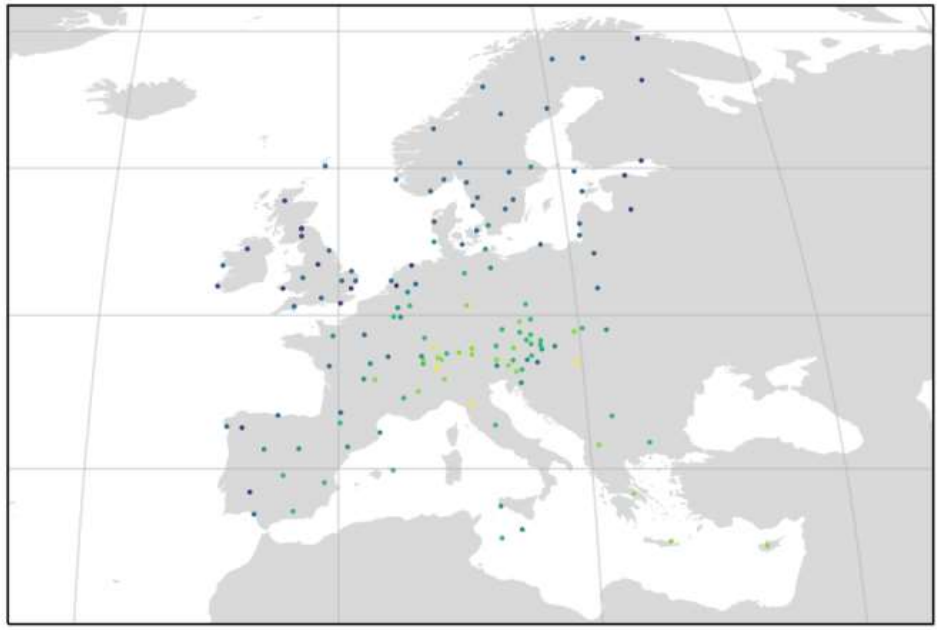
Statistic	Meaning
MB	Mean bias
NMB	Normalized mean bias
MAE	Mean absolute error
NMAE	Normalized mean absolute error
MNB	Mean normalized bias
MNAE	Mean normalized absolute error
MFB	Mean fractional bias
MAFB	Mean absolute fractional bias
RMSE	Root mean square error
NRMSE	Normalized root mean square error
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]

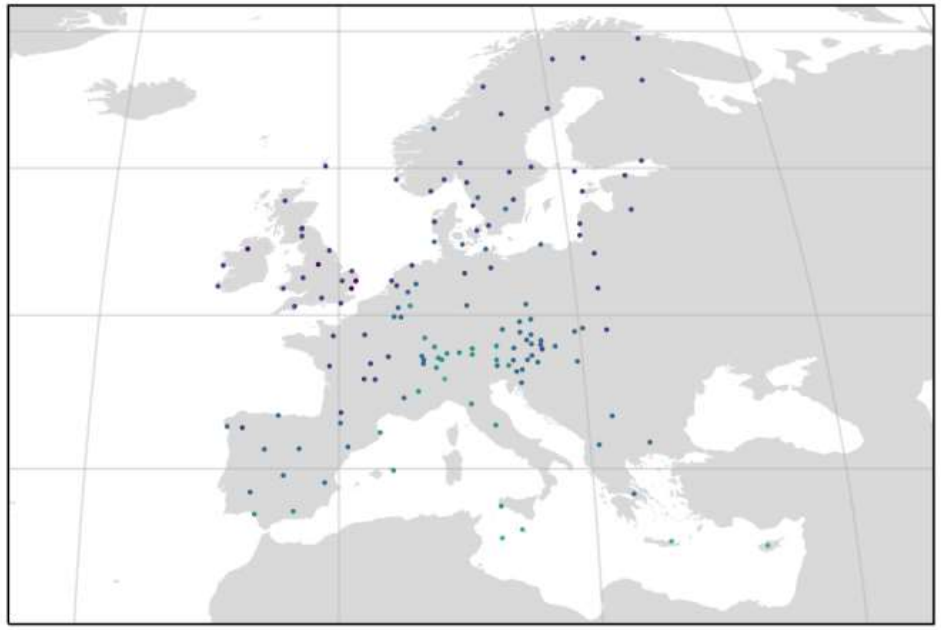
Map p95 (Summary)
EBAS|sconco3



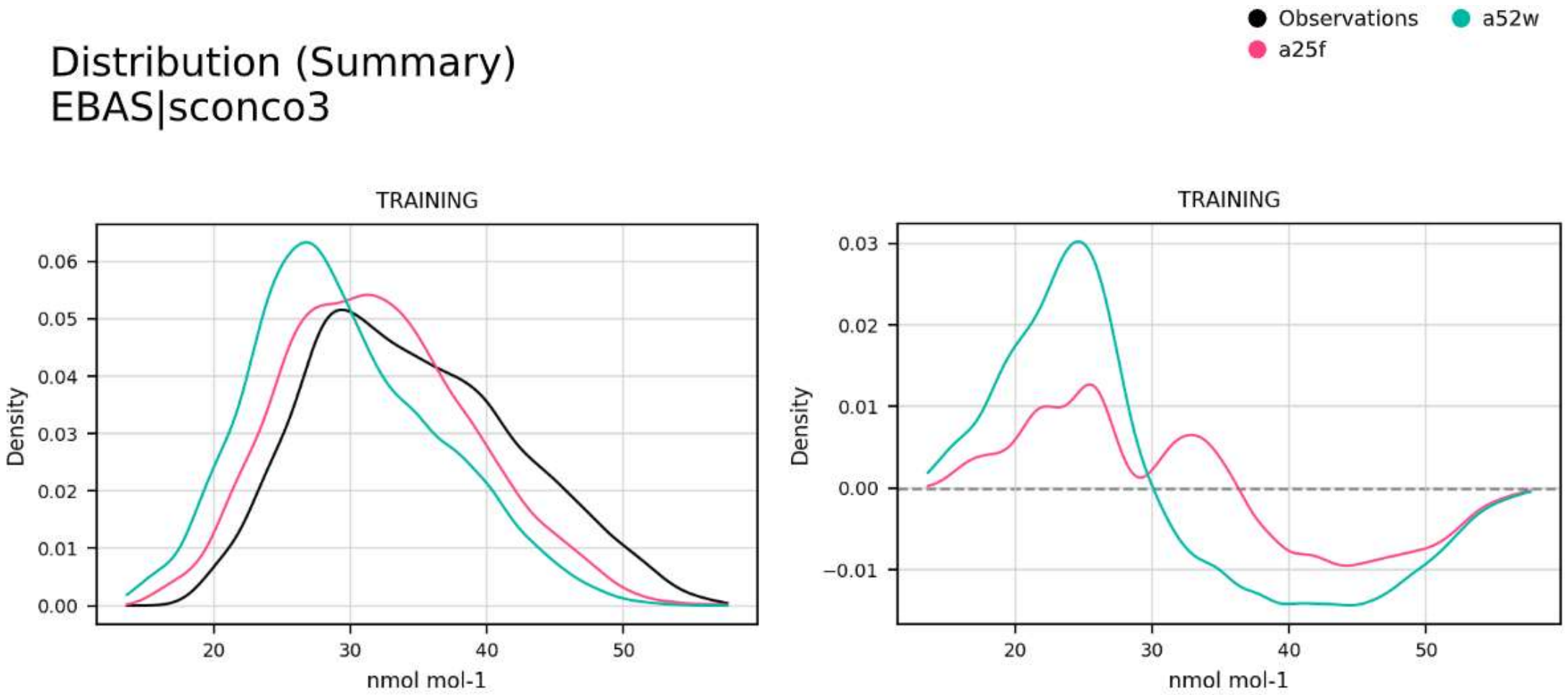
observations
TRAINING



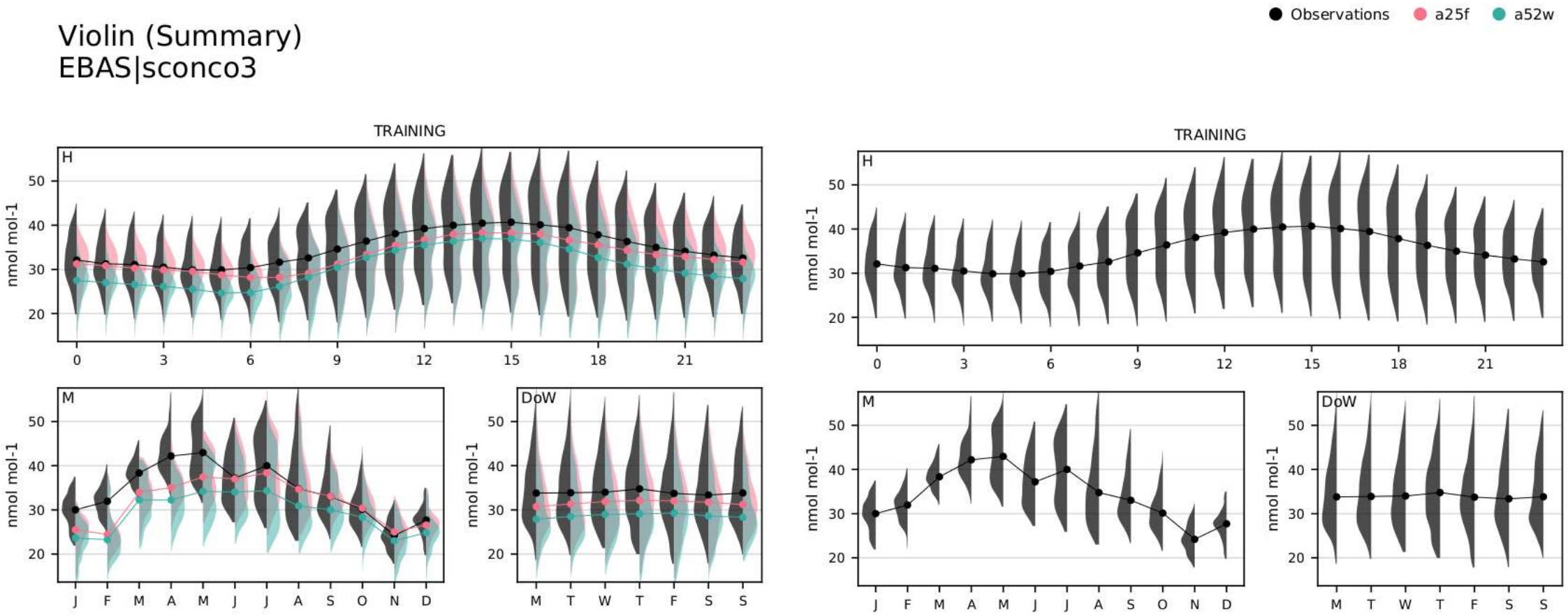
a52f-regional-000
TRAINING



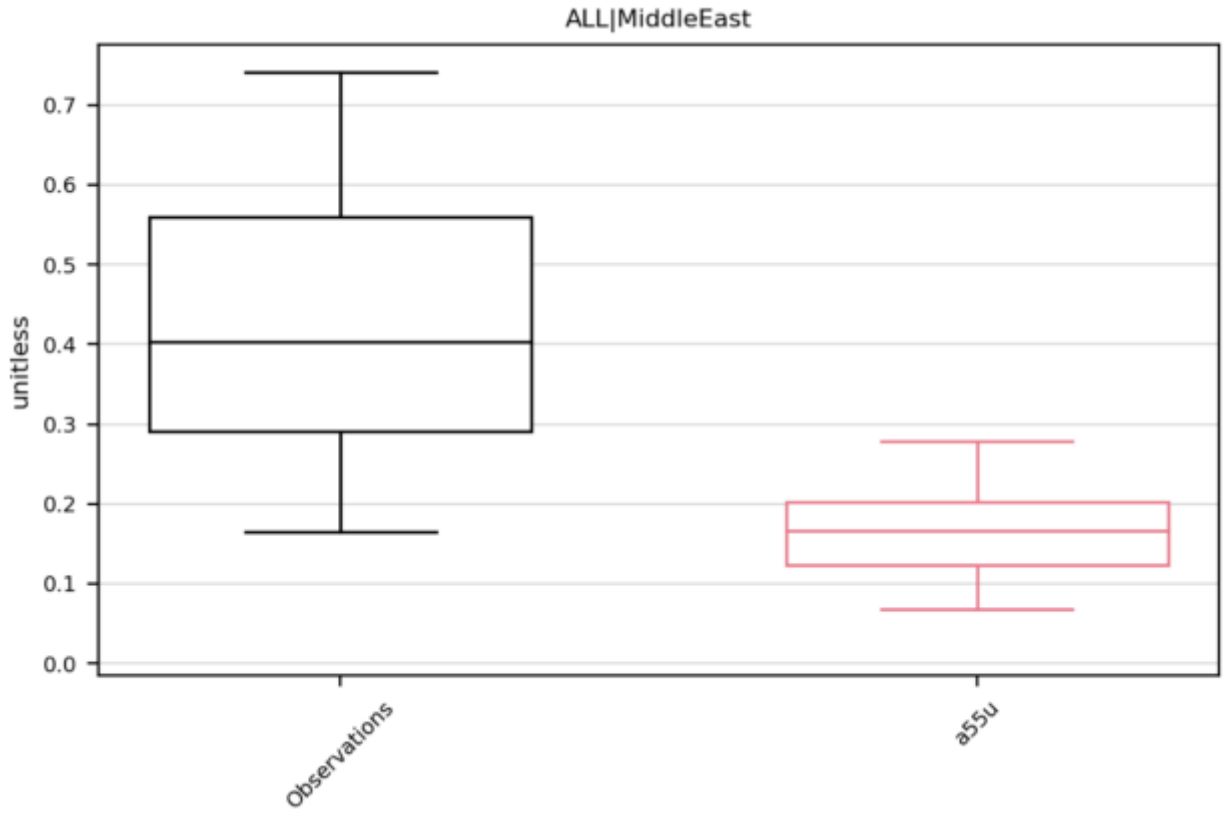
PLOT OPTION _bias

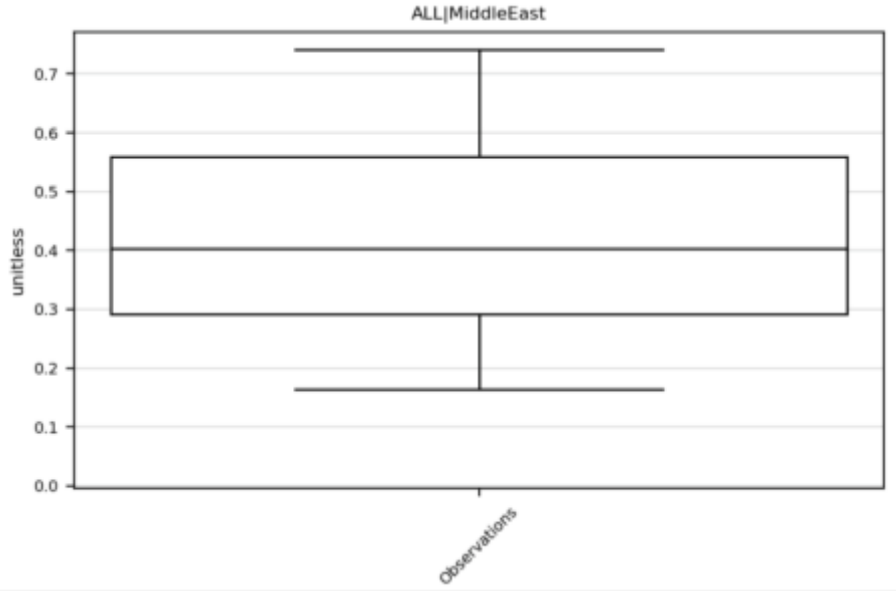


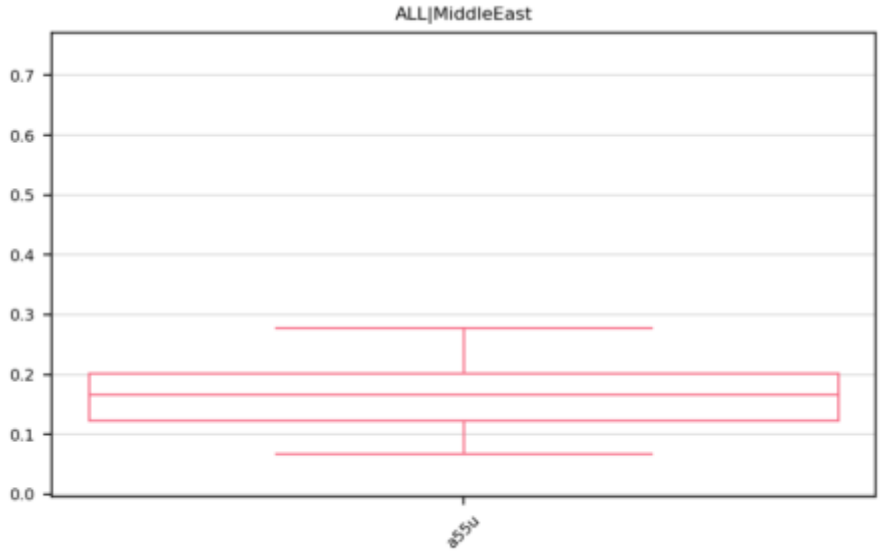
PLOT OPTION _obs



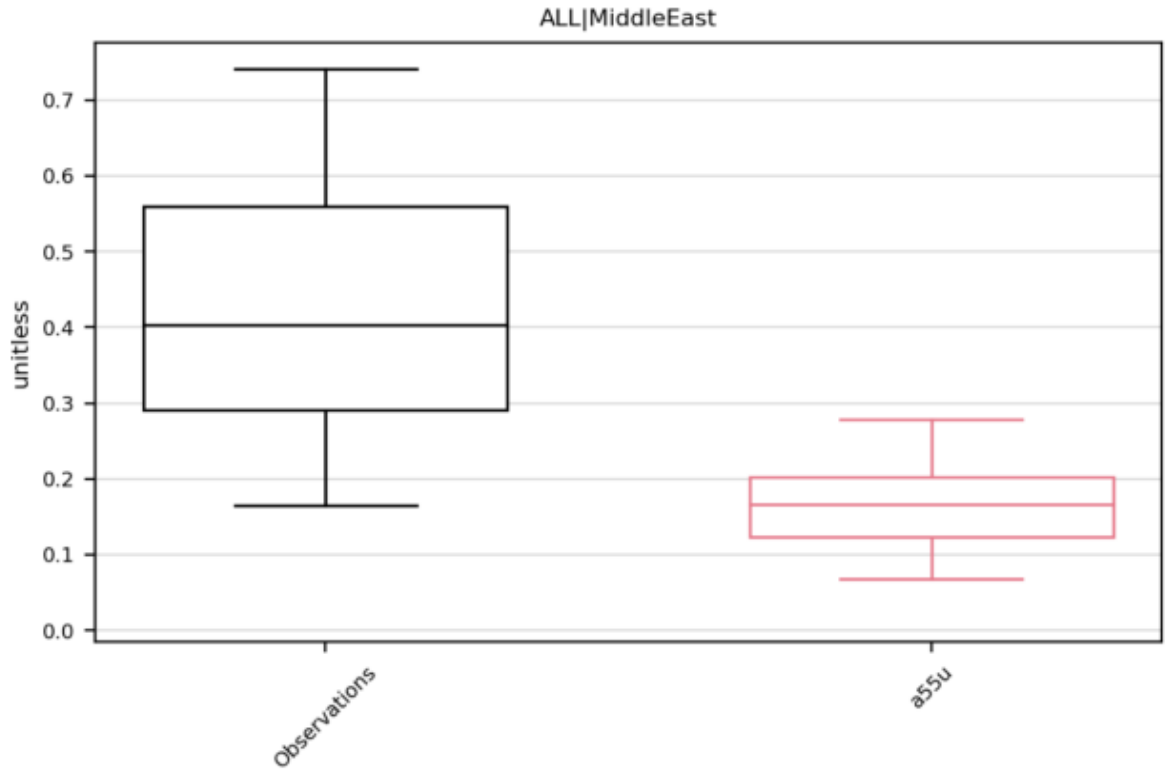
PLOT OPTION _individual

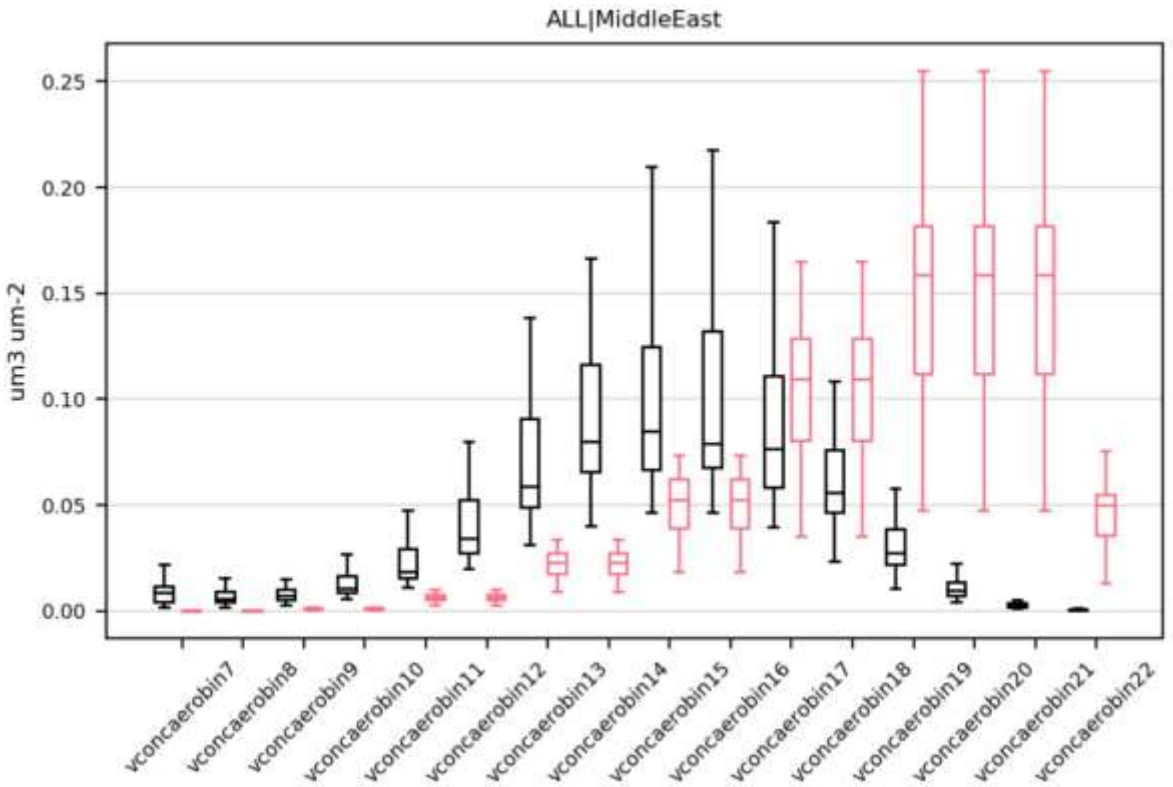




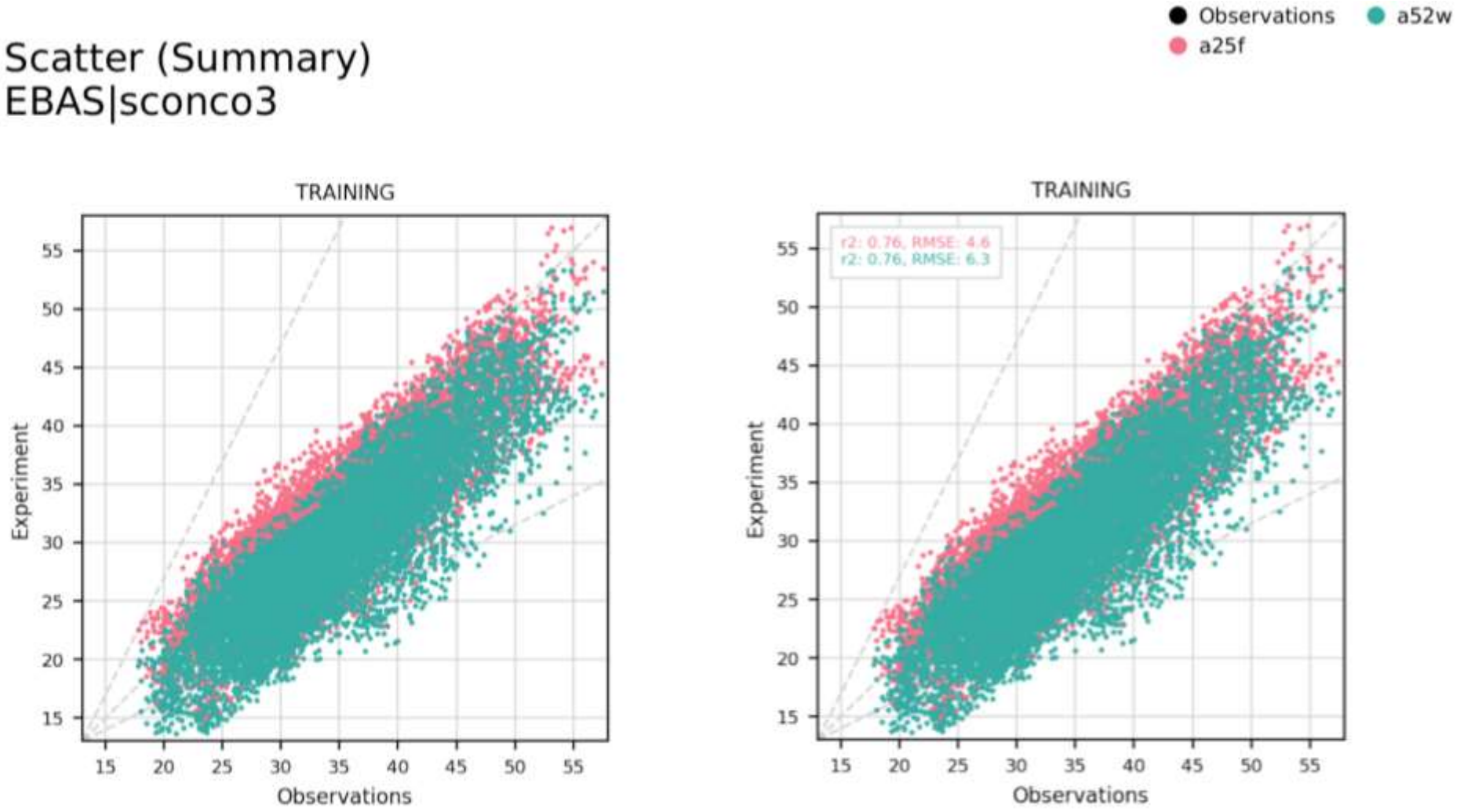


PLOT OPTION _multispecies

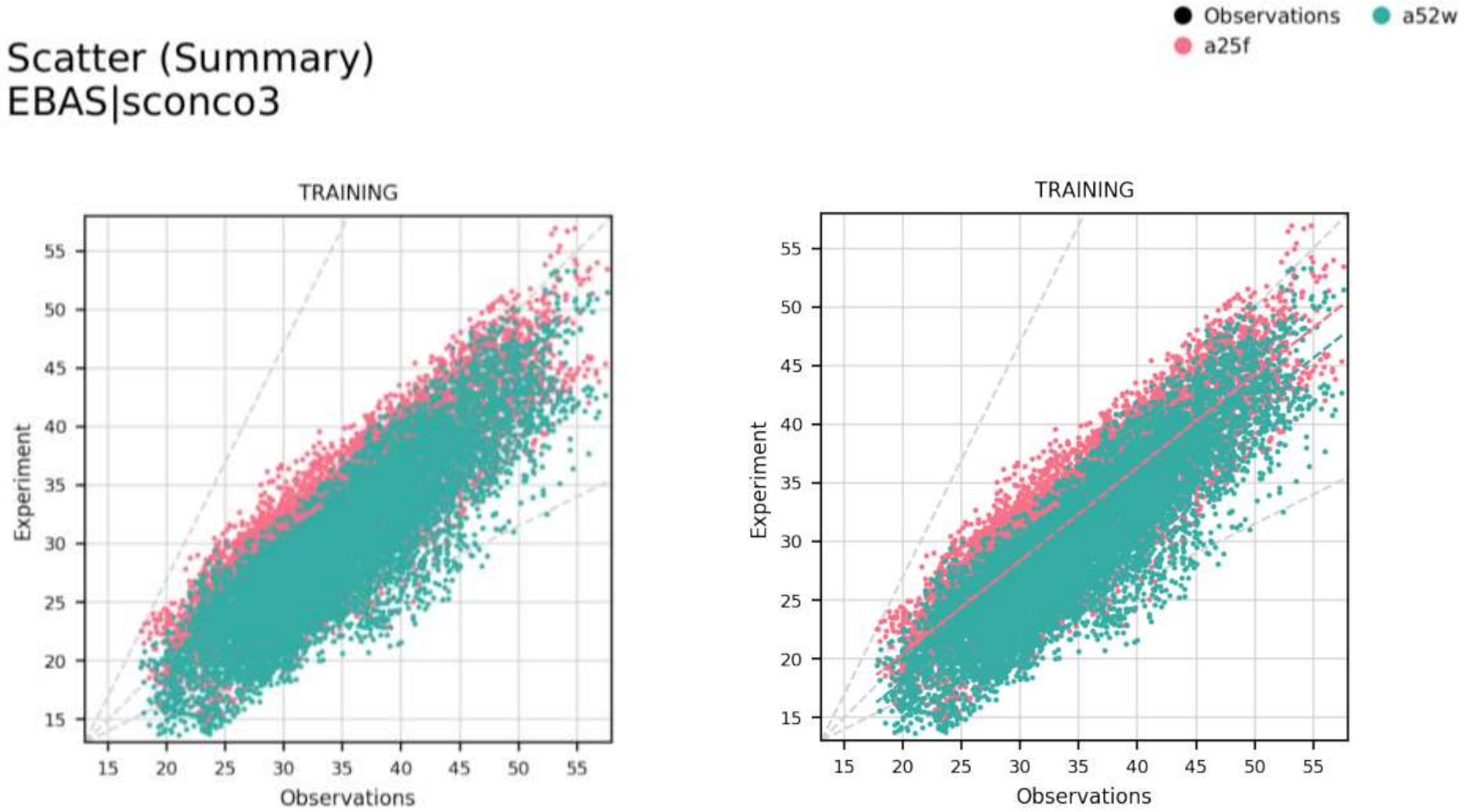




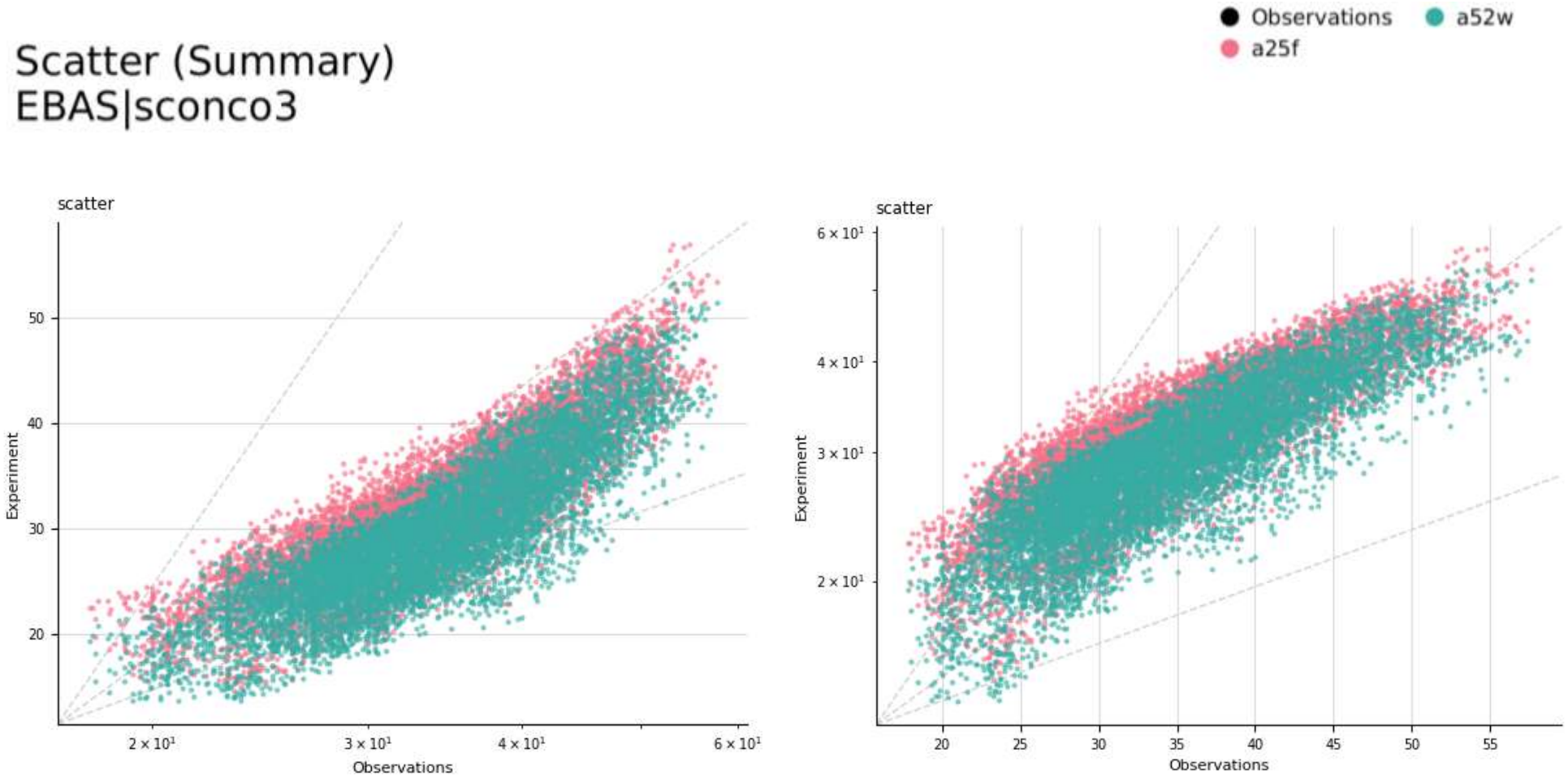
PLOT OPTION _annotate



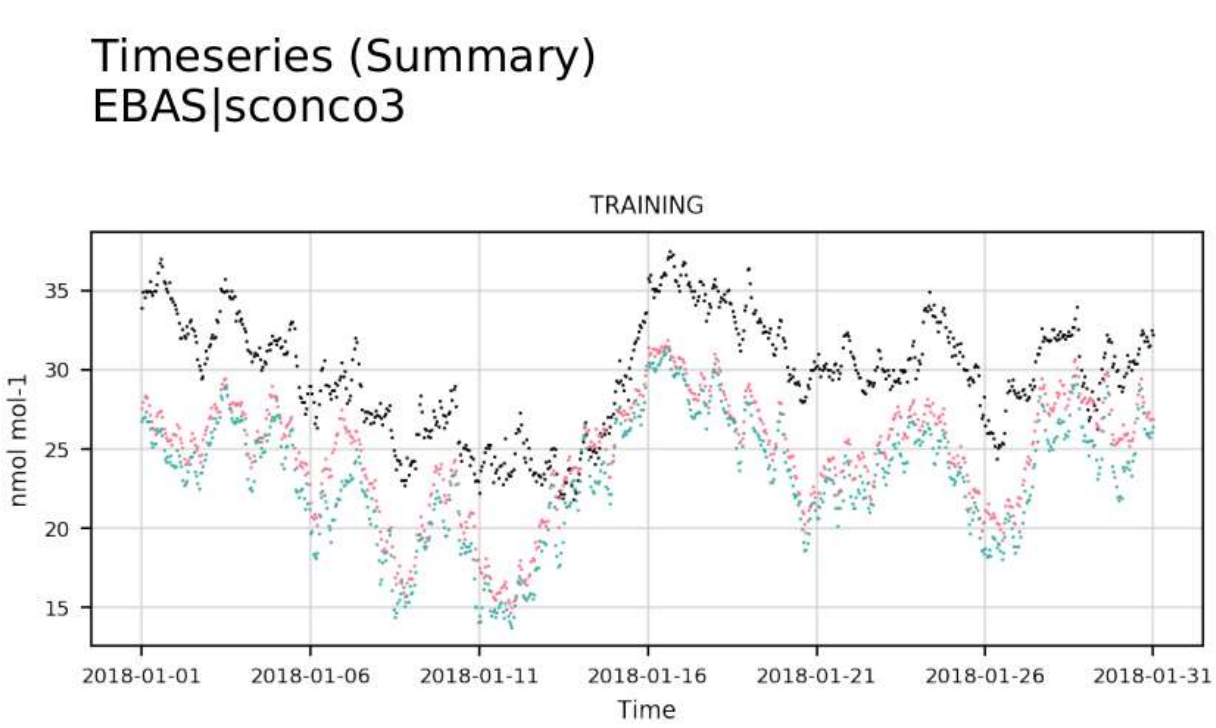
PLOT OPTION _regression

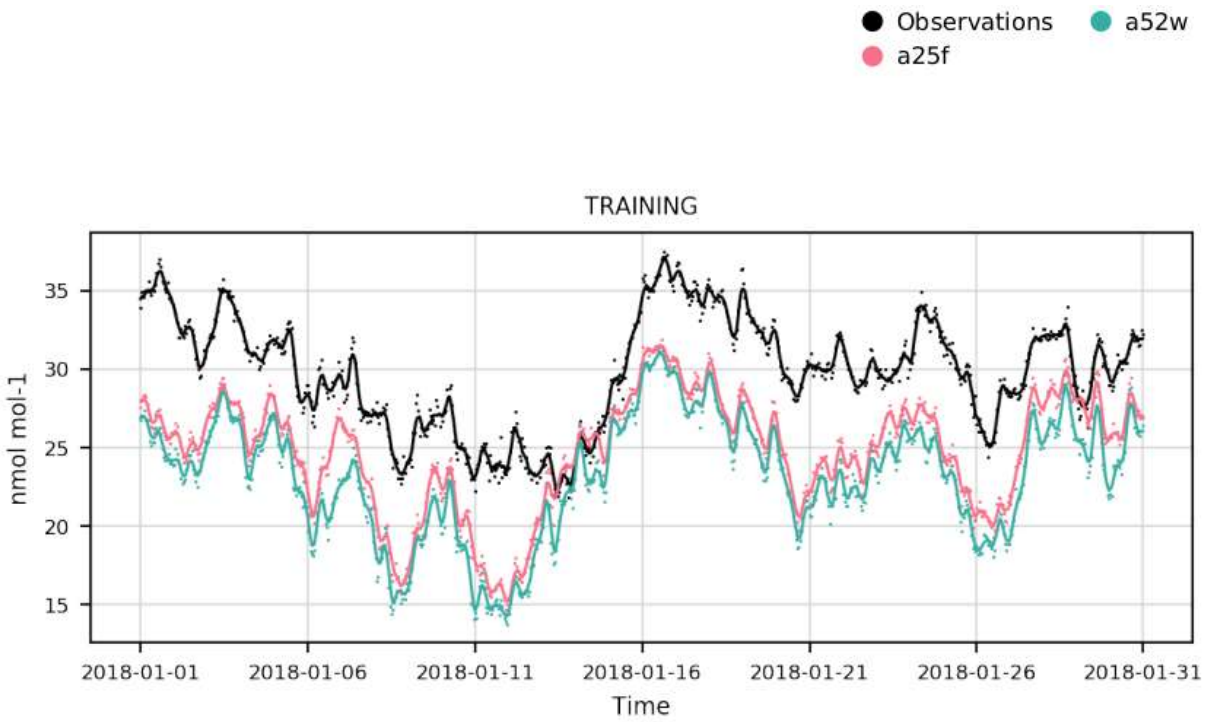


PLOT OPTION _logx / _logy



PLOT OPTION _trend





6

Launch the offline reports

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

LAUNCH THE OFFLINE REPORTS

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

```
$ ./bin/providentia --config=" /esarchive/scratch/avilanova/software/Providentia/configurations/training.conf“
  --offline
```

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.

7

Edit the plots style

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EDIT THE PLOTS STYLE

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

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

You can choose the statistics that you want to show for each plot also in the file **plot_characteristics_offline.json**.

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