Validation and Evaluation Tools

Amy Stidworthy, Cambridge Environmental Research Consultants (CERC)

Air Quality Workshop

Hôpital de l'Archet 2, Centre Hospitalier Universitaire, Nice, France

12 March 2013

Validation and Evaluation Tools

Two toolkits developed during PASODOBLE:

1. Myair Toolkit for Model Evaluation

   Developed by CERC

2. AUTH Toolkit for retrieval and pre-processing of satellite and in situ data

   Developed by AUTH

Validation and Evaluation Tools

Two toolkits developed during PASODOBLE:

1. **Myair Toolkit for Model Evaluation**

   Developed by CERC

2. AUTH Toolkit for retrieval and pre-processing of satellite and in situ data

   Developed by AUTH

Myair Toolkit for Model Evaluation

- What is it?
- What can it do?
- What do you get?
- Demonstration of its use in evaluating the PASODOBLE AIRSHEDS products

Myair Toolkit for Model Evaluation

What is it?

- The Myair Toolkit is a free, open-source toolkit for evaluating air quality models, focussed on requirements for standardised evaluation of local air quality forecast models
- The Toolkit was designed following an extensive review of the state-of-the-art in air quality model evaluation, and builds on existing tools and initiatives (e.g. FAIRMODE, openair)

Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- Help you investigate model performance at individual stations using openair graphs
- Run in batch mode too, for easy automation

Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What can it do?

- **Assess your model's forecast skill**
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- Help you investigate model performance at individual stations using openair graphs
- Run in batch mode too, for easy automation

Forecast index scales and alert thresholds are set by the user

Myair Pasodoble

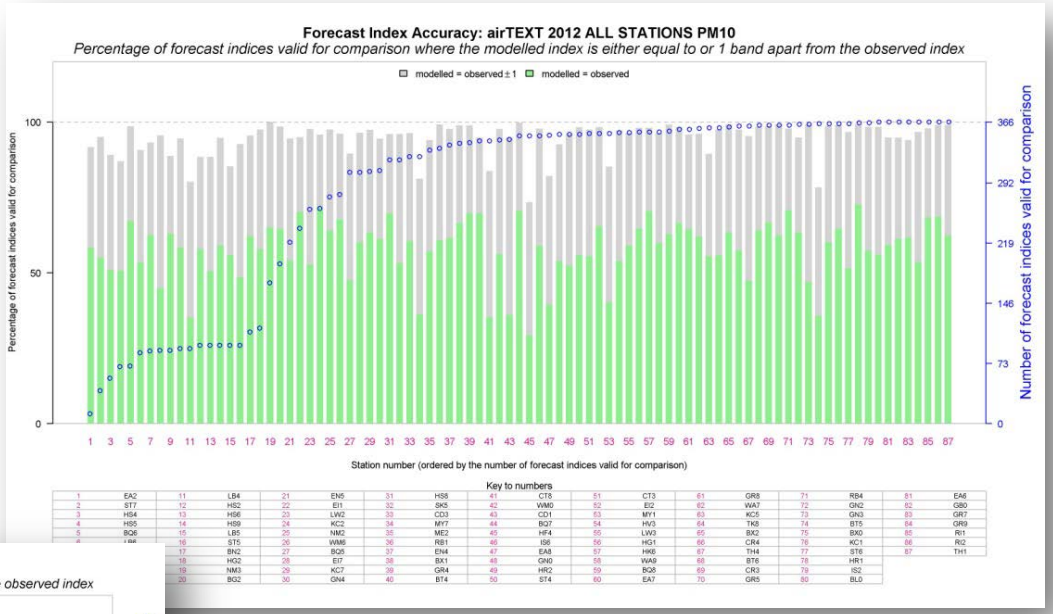
GMES

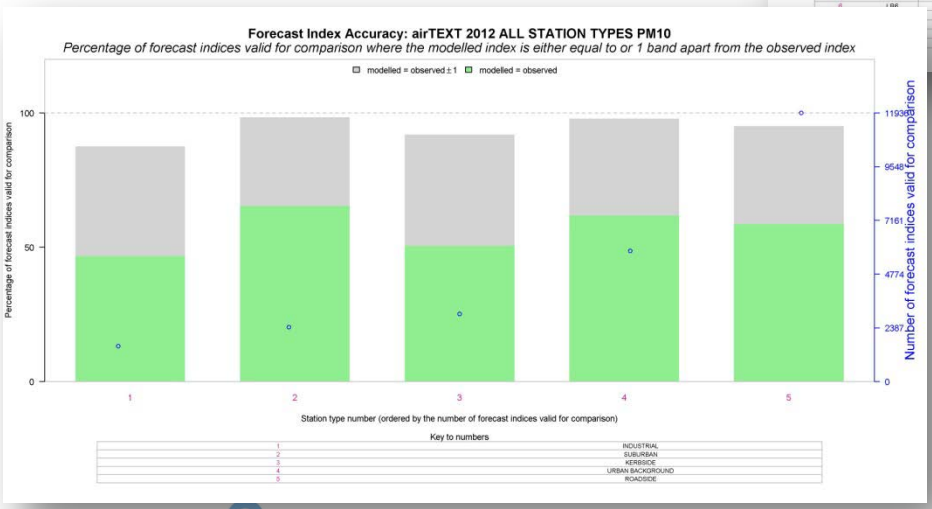
SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

Look at the percentage of forecast indices within one of observed (should be close to 100%) for each pollutant, grouped by station...





... or grouped by station type (e.g. roadside, urban background, rural etc)...

Myair Pasodoble

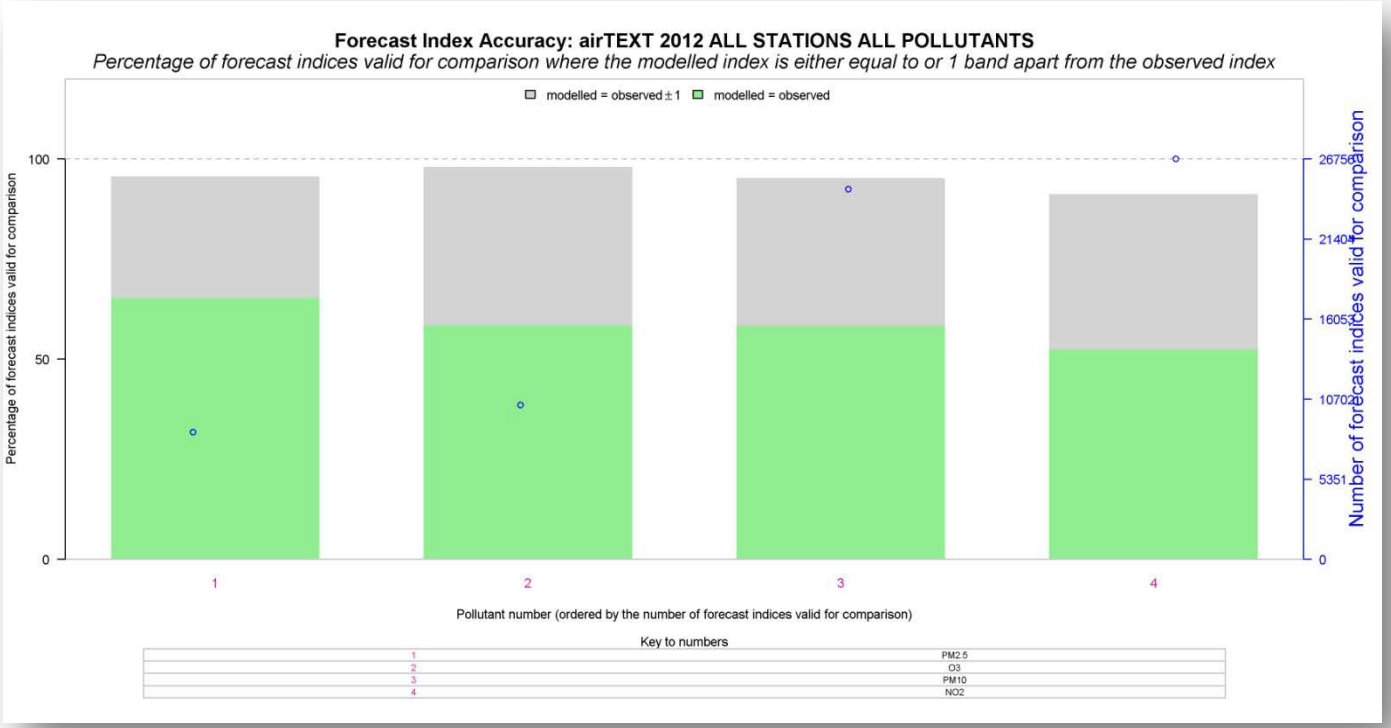
GMES Because we need to know

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

...or grouped by pollutant



Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

Look at model's skill at predicting alert threshold exceedences (i.e. pollution episodes) in different ways:

		Alert modelled?	
		Yes	**No**
Alert observed?	**Yes**	a	b
	No	c	d

a, b, c and d are counts of the number of days where alerts were or were not modelled and were or were not observed

$$Odds\ Ratio\ Skill\ Score\ (ORSS) = \frac{ad - bc}{ad + bc}$$

ORSS gives equal weighting to correct non-prediction and to correct prediction

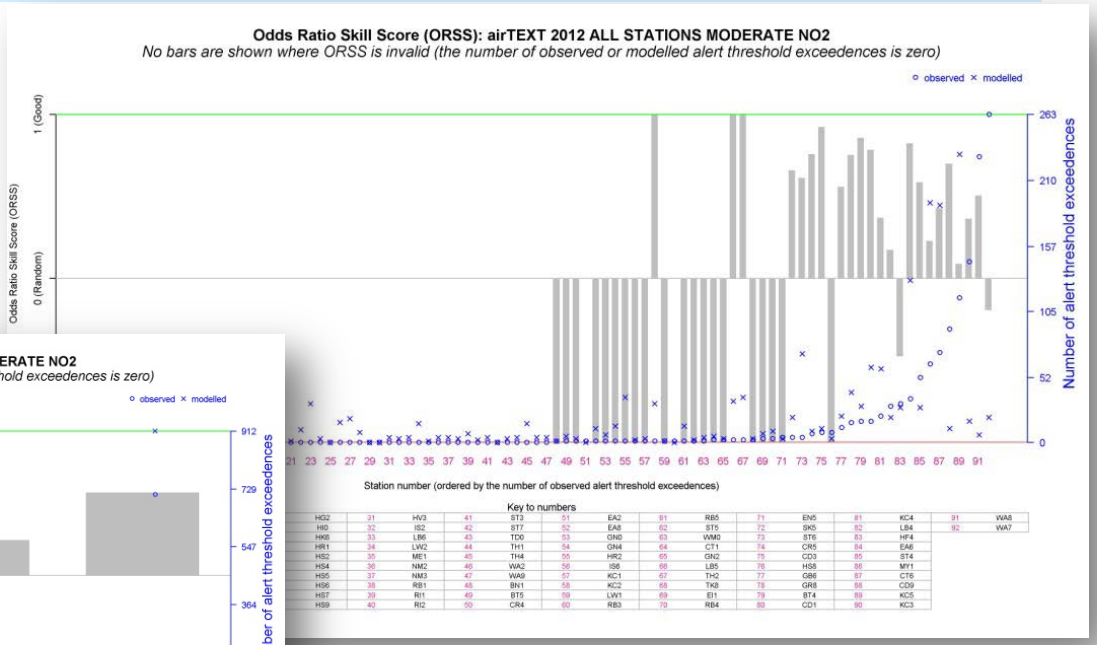
Perfect score:	b=c=0	ORSS=1
Good score:	ad>bc	ORSS>0
Bad score:	bc>ad	ORSS<0
Fail score:	a=d=0	ORSS=-1

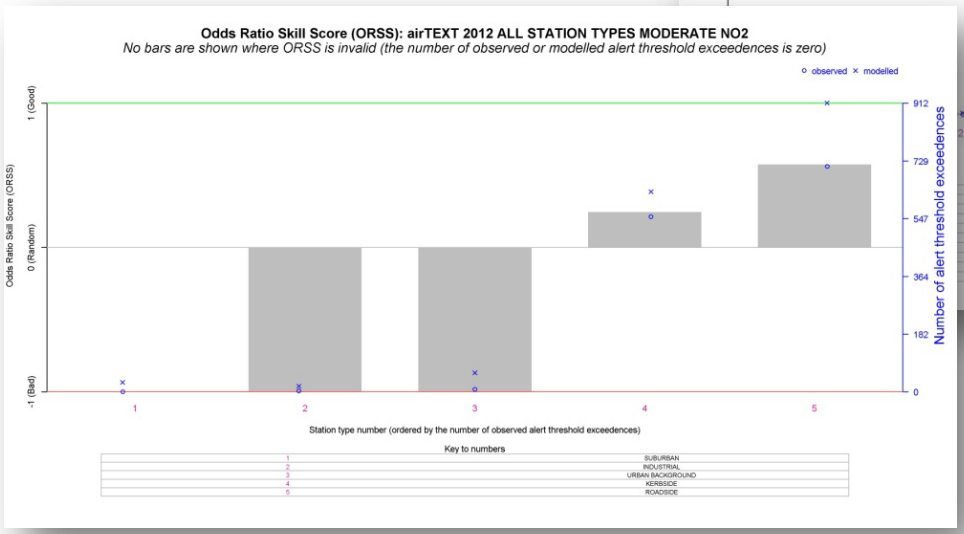
Myair Pasodoble

GMES SEVENTH FRAMEWORK PROGRAMME EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

ORSS grouped by station...





... or grouped by station type

ORSS is a good measure if a lot of episodes are measured, but note that it's easy to get a good score if there are few episodes compared to the number of forecasts because d will be high

Myair Pasodoble

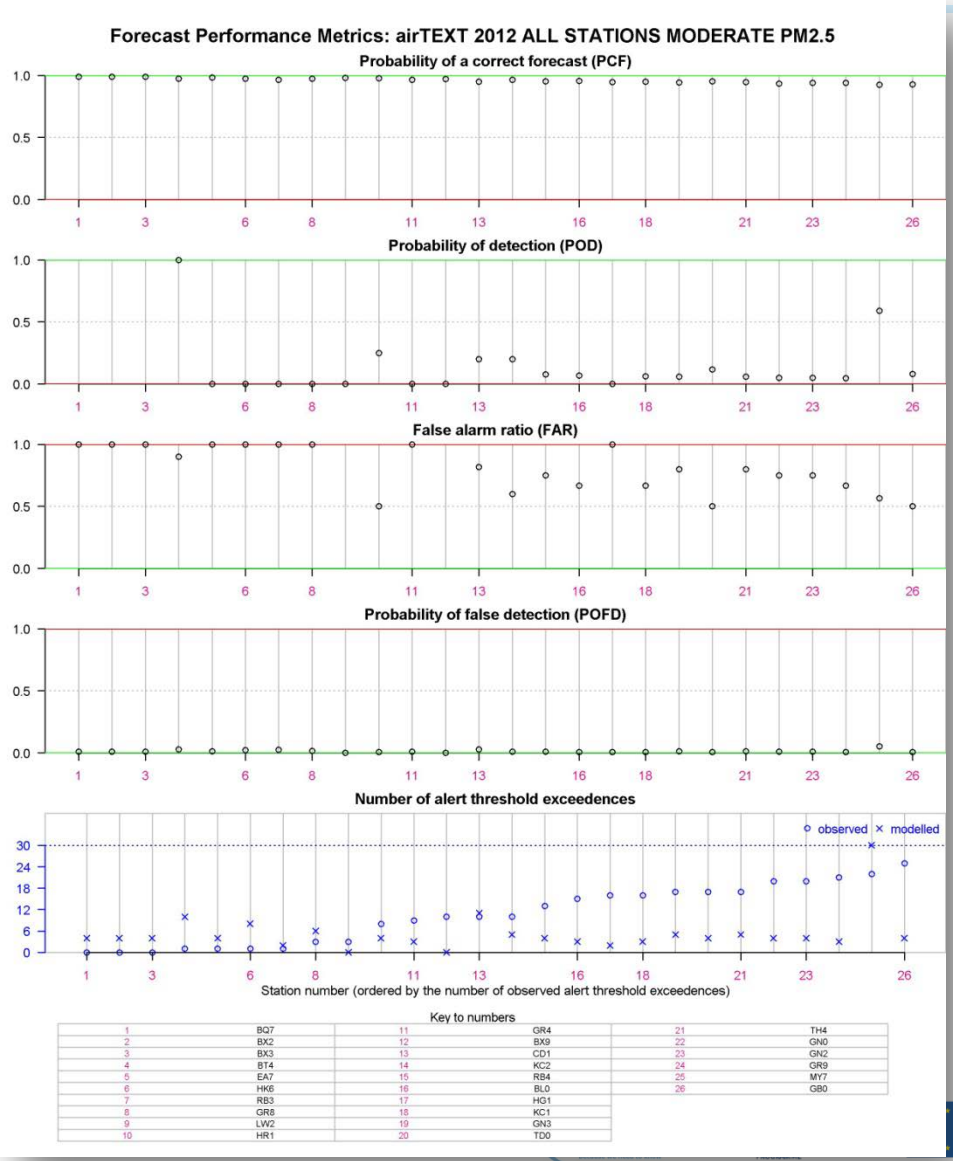
GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

Using the Toolkit you can also look at other measures of model skill, for example the ‘probability of detection’ and the ‘false alarm ratio’ for different alert thresholds...



Myair Pasodoble

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- **Assess your model's concentration predictions**
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- Help you investigate model performance at individual stations using openair graphs
- Run in batch mode too, for easy automation

Output averaging times and units for the concentration assessment are set by the user

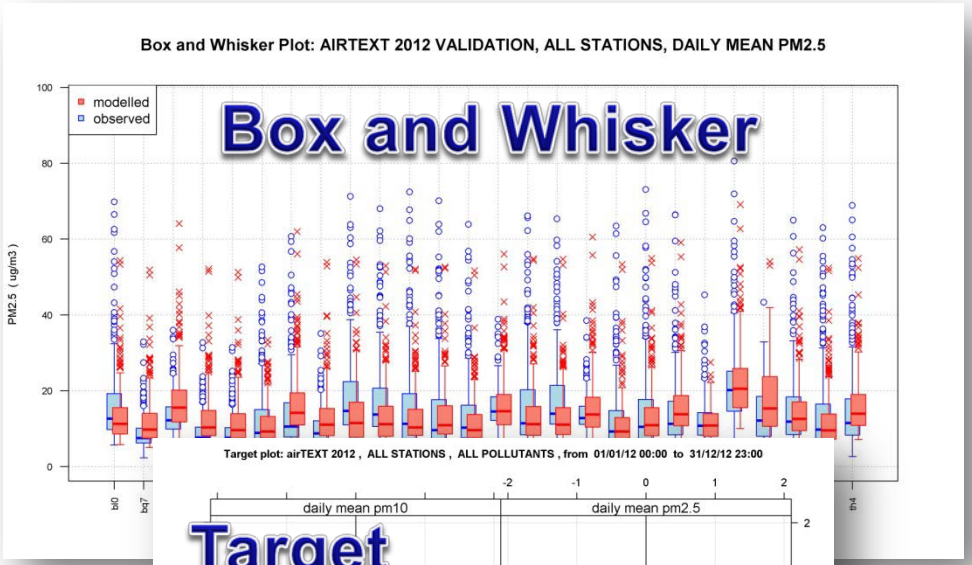
Myair Pasodoble

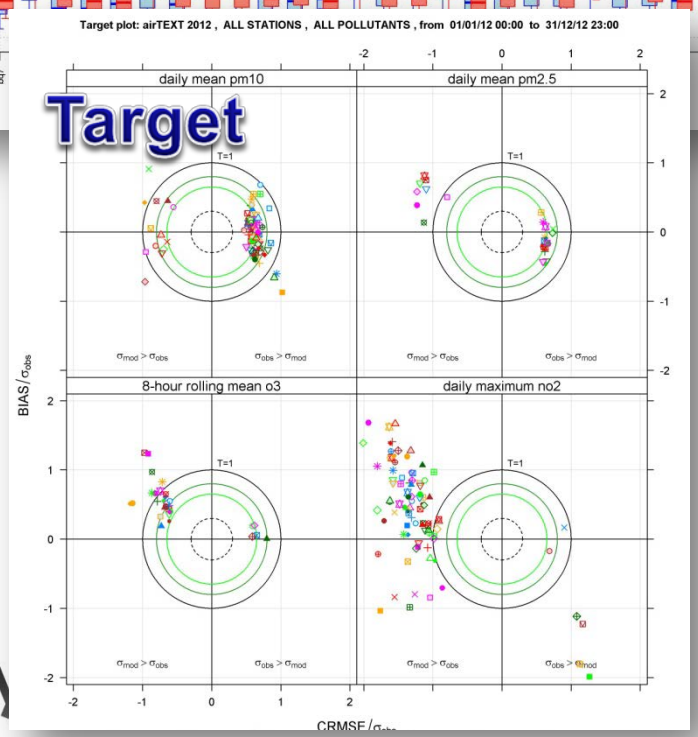
GMES

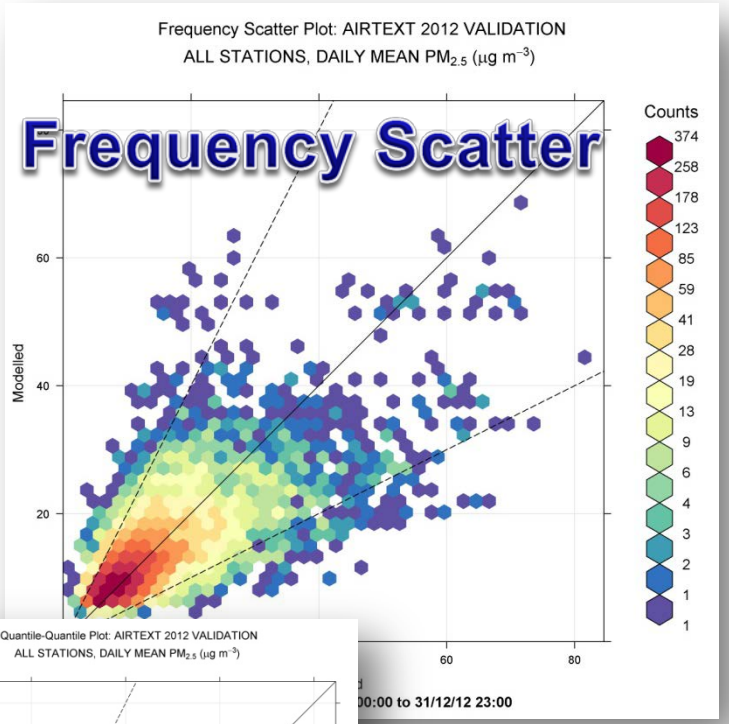
SEVENTH FRAMEWORK PROGRAMME

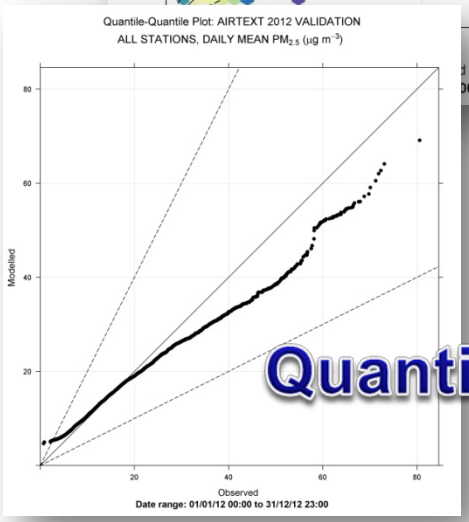
EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation









My

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

Statistical output includes standard results such as mean, bias, standard deviation and more (this table is an extract from the User Guide)

Name	Description	Equation
Num.valid.values	Number of values	
obs.mean mod.mean	Mean	$1/n \sum C$
SDO SDM	Standard Deviation	$\sqrt{1/n \sum (C - \overline{C})^2}$
MB	Mean Bias	$\overline{(C_p - C_o)}$
NMSE	Normalised Mean-Square-Error	$\overline{(C_p - C_o)^2} / \overline{C_o C_p}$
R	Pearson's Correlation Coefficient	$\mathrm{cov}(C_p, C_o)/\sigma_{C_p}\sigma_{C_o}$
Fac2	Factor of 2	Fraction of data where $0.5 \leq C_p/C_o \leq 2$ (when $C_o = 0$, $C_p/C_o \rightarrow \infty$ and the data pair is not counted)
Fb	Fractional Bias	$(\overline{C_p} - \overline{C_o})/0.5(\overline{C_o} + \overline{C_p})$
Fs	Fractional Standard Deviation	$(\sigma_{C_p} - \sigma_{C_o})/0.5(\sigma_{C_o} + \sigma_{C_p})$
obs.max mod.max	Maximum	$\max C$
obs.RHC mod.RHC	Robust Highest Concentration	$\chi(n) + (\chi - \chi(n))\ln\left(\frac{3n-1}{2}\right)$, where n is the number of values used to characterise the upper end of the concentration distribution, χ is the average of the $n-1$ largest values, and $\chi(n)$ is the n^{th} largest value; n is taken to be 26.

Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- Assess your model's concentration predictions
- **Easily import a wide range of gridded and point modelled data formats**
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- Help you investigate model performance at individual stations using openair graphs
- Run in batch mode too, for easy automation

Myair Pasodoble

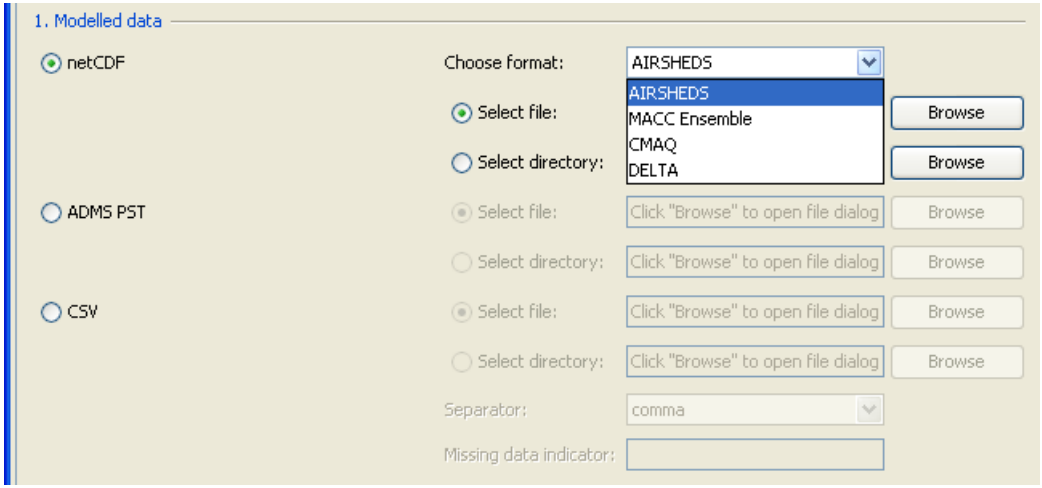
GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

- Supported modelled data formats:
  - Gridded netCDF
    - AIRSHEDS
    - MACC Ensemble
    - CMAQ
  - Point data
    - ADMS PST
    - Generic CSV
    - DELTA netCDF



- The Toolkit interpolates gridded data to station locations
- You can import a single file or a whole directory of files

Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- **Download and import in situ monitoring data for the UK (also CSV files)**
- Save graphical and statistical output to your computer
- Help you investigate model performance at individual stations using openair graphs
- Run in batch mode too, for easy automation

Myair Pasodoble

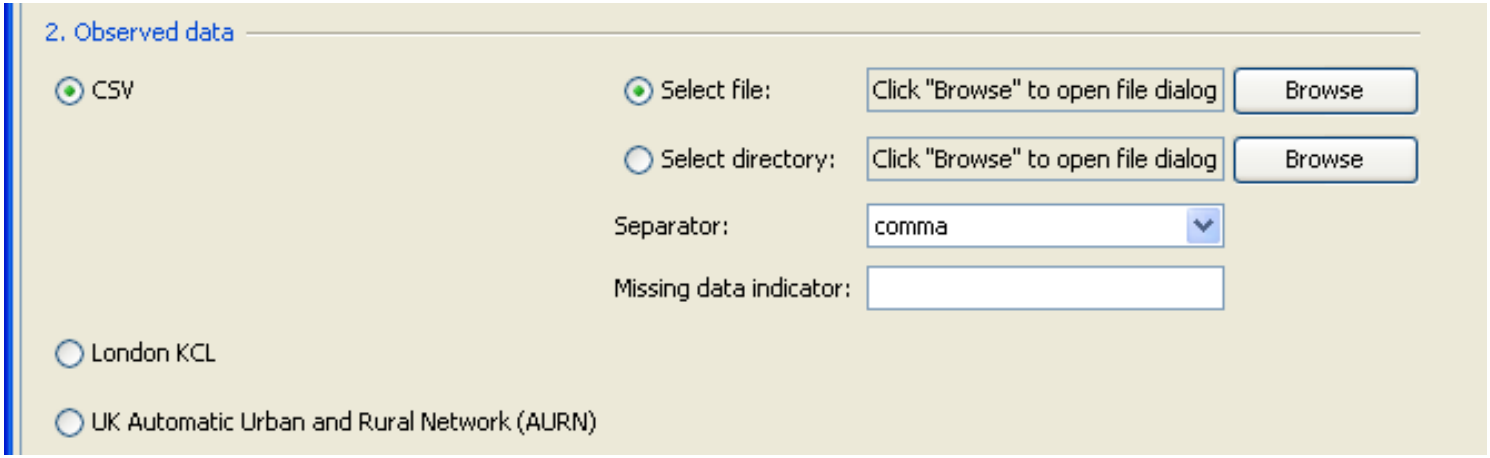
GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

- In situ observed data for 2 UK networks can be downloaded and imported automatically (London KCL and UK AURN)
- Observed data in a generic CSV format can be imported from a single file or directory of files



Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- **Save graphical and statistical output to your computer**
- Help you investigate model performance at individual stations using openair graphs
- Run in batch mode too, for easy automation

Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

- Saves graphs as image files (JPG, PNG) or PDFs for importing into documents
- Saves data (raw, processed and statistics) in CSV files, to provide an audit trail and for further analysis

One of the CSV files output by the Toolkit

airTEXT 2012_Validation_forecast_alert_stats.csv

	A	B	C	D	E	F	G	H	I	J	K	L
1	pollutant	type	alert.name	num.obs.alerts	a	b	c	d	FB	POD	PCF	FAR
2	no2	suburban	moderate	0	0	30	0	3265	NA	NA	0.990895296	
3	no2	suburban	very high	0	0	0	0	3295	NA	NA	1	NA
4	no2	suburban	high	0	0	0	0	3295	NA	NA	1	NA
5	no2	urban background	moderate	8	0	60	8	6077	7.5	0	0.988934093	
6	no2	urban background	high	0	0	0	0	6145	NA	NA	1	NA
7	no2	urban background	very high	0	0	0	0	6145	NA	NA	1	NA
8	no2	kerbside	moderate	553	170	462	383	1718	1.142857143	0.307414105	0.690815953	0.7310126
9	no2	kerbside	high	34	3	39	31	2660	1.235294118	0.088235294	0.97438712	0.9285714
10	no2	kerbside	very high	0	0	0	0	2733	NA	NA	1	NA
11	no2	roadside	moderate	712	135	777	577	12311	1.280898876	0.189606742	0.901884058	0.8519736
12	no2	roadside	high	2	0	4	2	13794	2	0	0.999565217	
13	no2	roadside	very high	0	0	0	0	13800	NA	NA	1	NA

airTEXT 2012_Validation_forecas

Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- **Help you investigate model performance at individual stations using openair graphs**
- Run in batch mode too, for easy automation

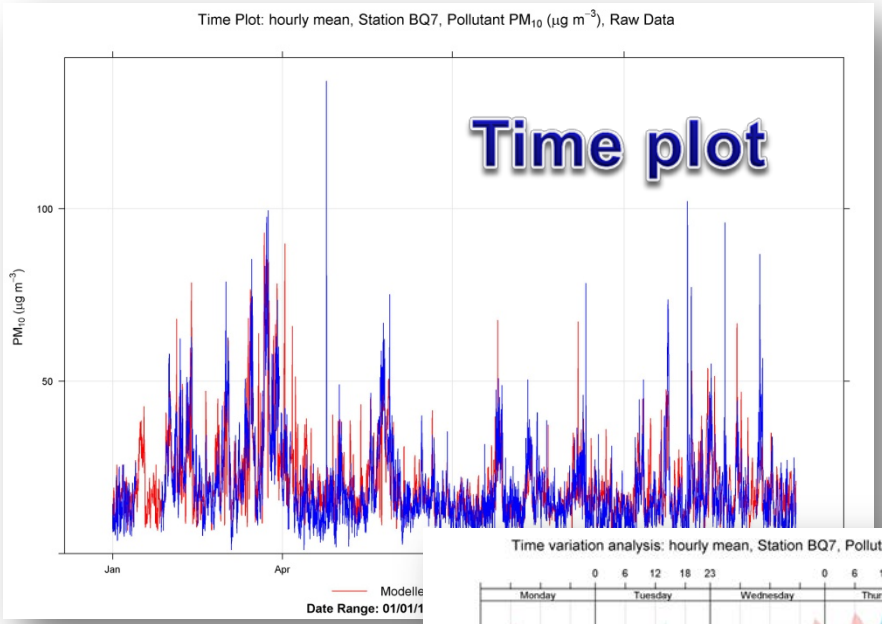
Myair Pasodoble

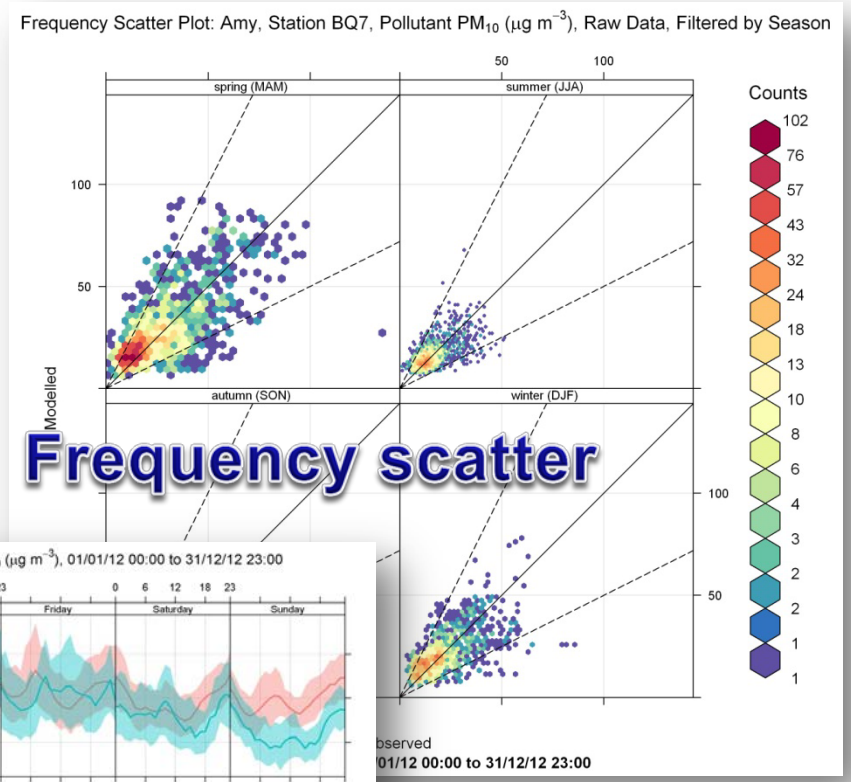
GMES

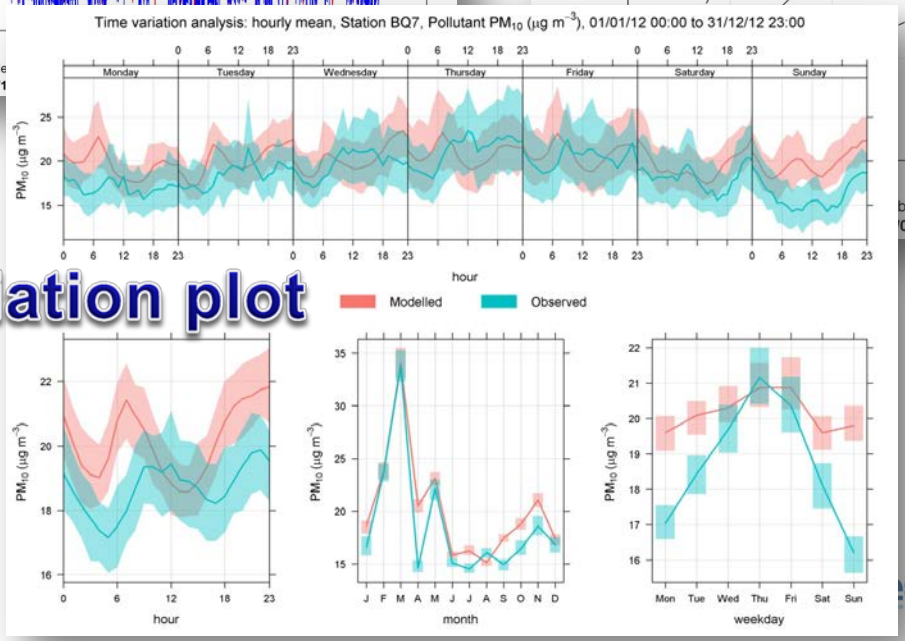
SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation







Time variation plot

Myair Pasodoble

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What can it do?

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- Help you investigate model performance at individual stations using openair graphs
- **Run in batch mode too, for easy automation**

Batch mode allows easy integration of model evaluation into automatic processes, and also easy re-generation of results with new data

Myair Pasodoble

GMES

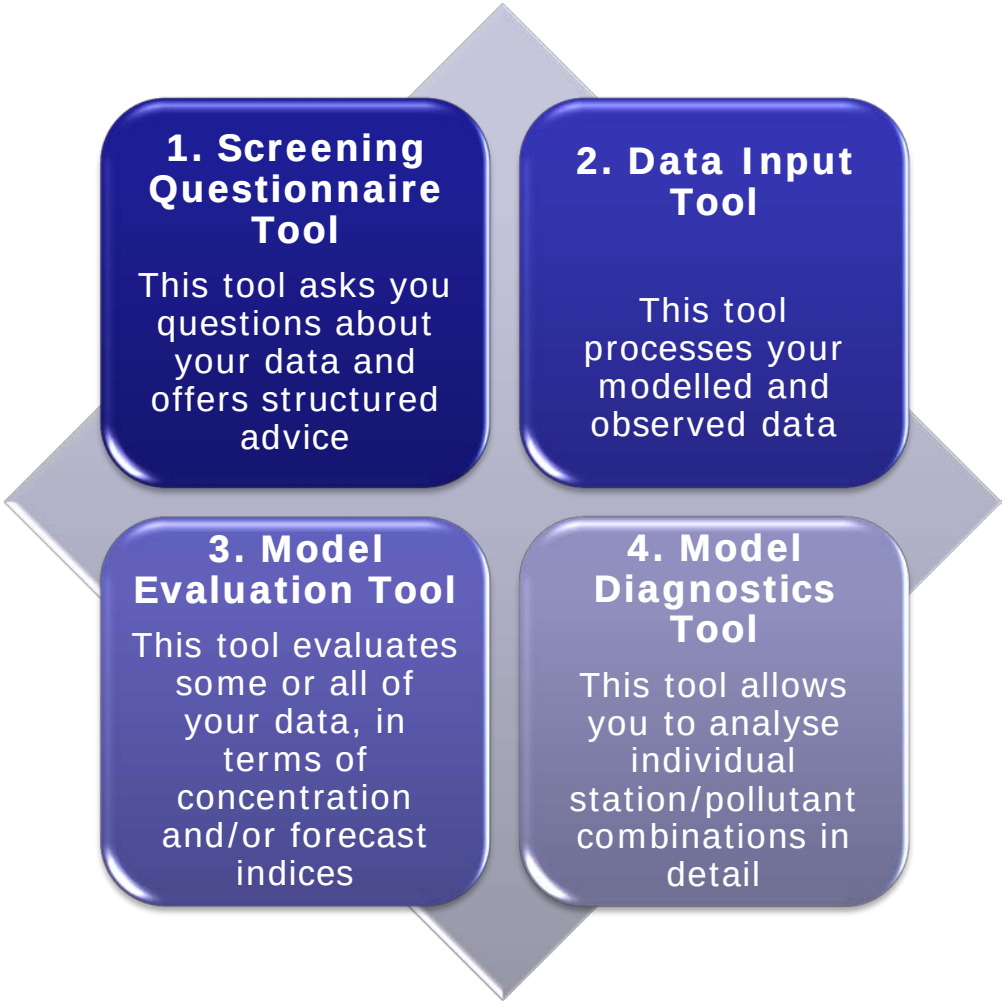
SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Myair Toolkit for Model Evaluation

What do you get?

- 4 tools
- Runs on most commonly-used platforms, including Windows, Linux, Mac
- Requires you to download and install some free software, which only takes a few minutes
- Comprehensive User Guide included



Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

AIRSHEDS - Domains

Intermediate airsheds at ~ 7x7km resolution

Bridge gap between MACC- ENS and local nests

covering most of the local/urban services within PASODOBLE

Myair Pasodoble

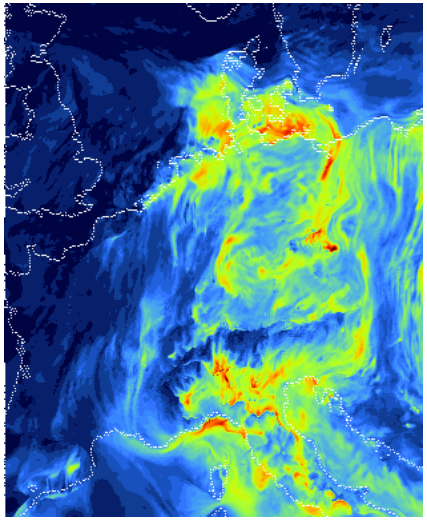
GMES

SEVENTH FRAMEWORK PROGRAMME

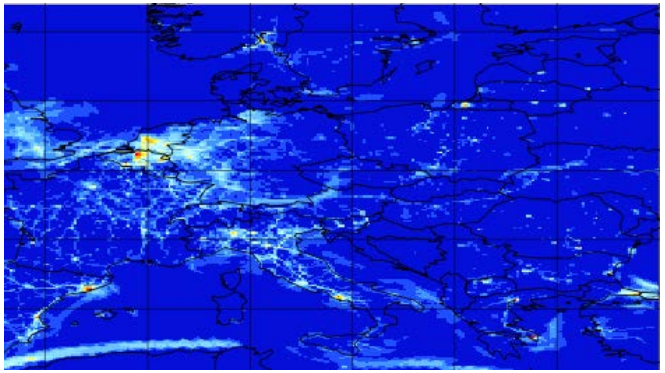
EUROPEAN COMMISSION

AIRSHEDS - examples

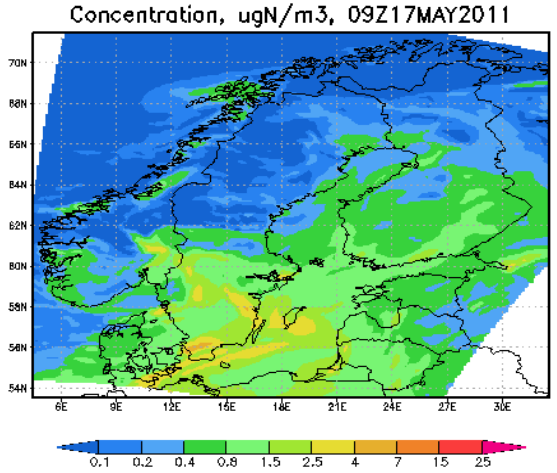
FRIUUK
EURAD



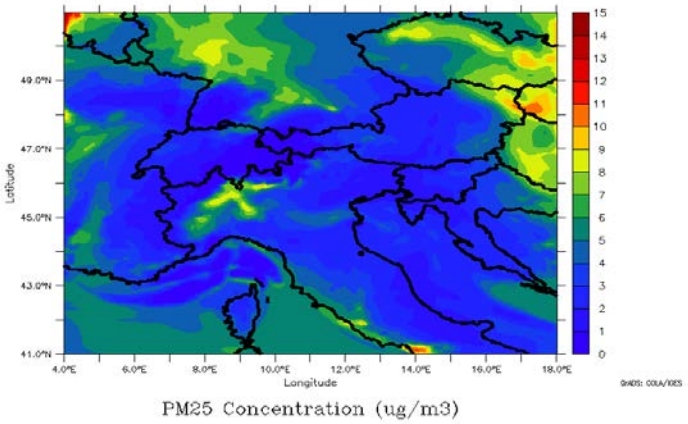
TNO/KNMI/RIVM
LOTOS-EUROS

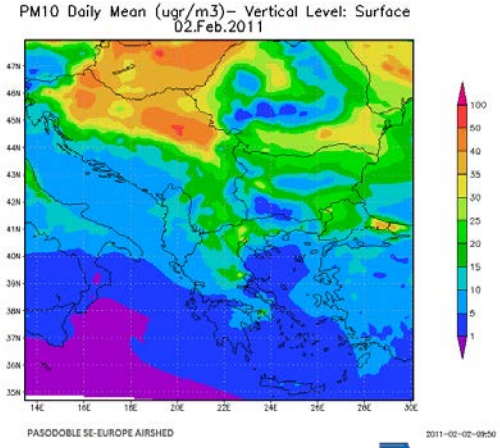


FMI
SILAM



ACRI
CHIMERE





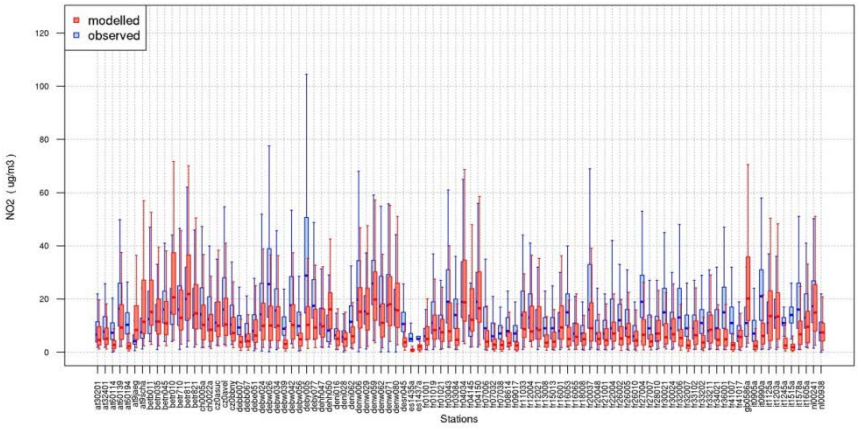
AUTH
WRF-CAMx

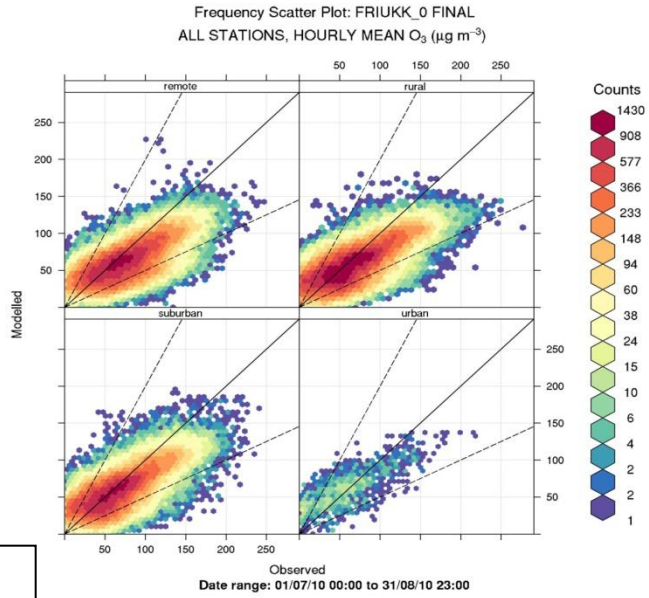
Myair Pasodoble

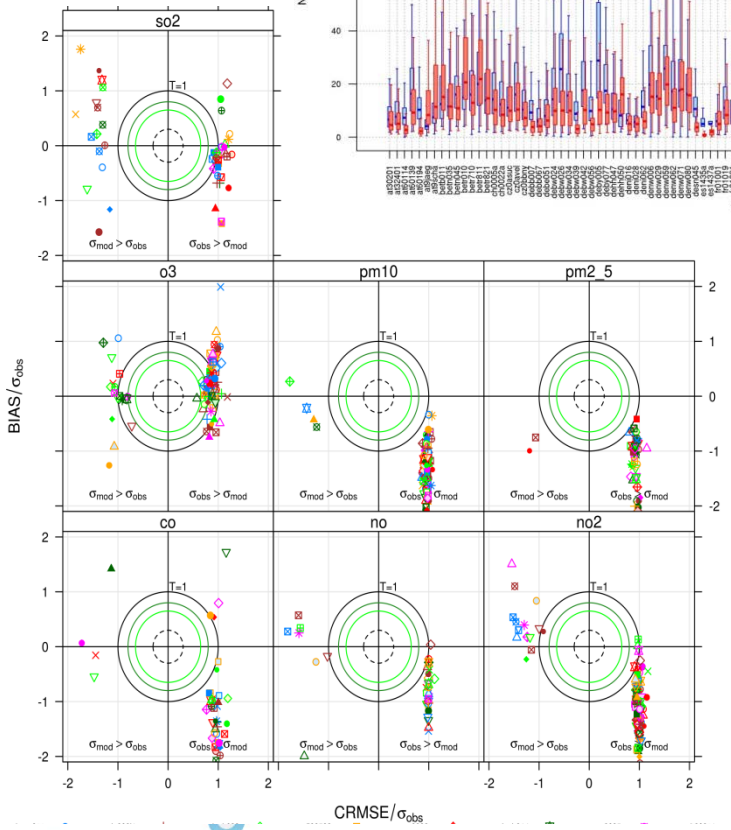
GMES

EUROPEAN COMMISSION

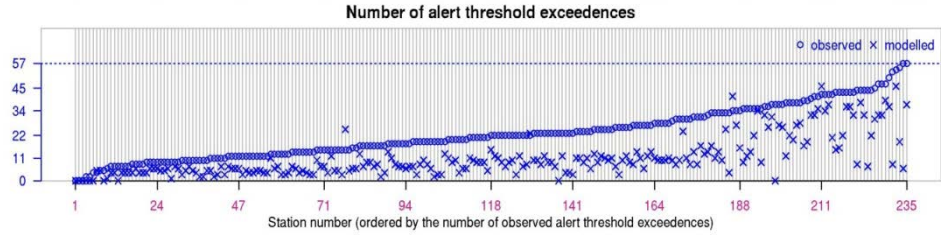
Toolkit: harmonised evaluation of airsheds







4 models/airsheds
7 components



Myair Pasodoble

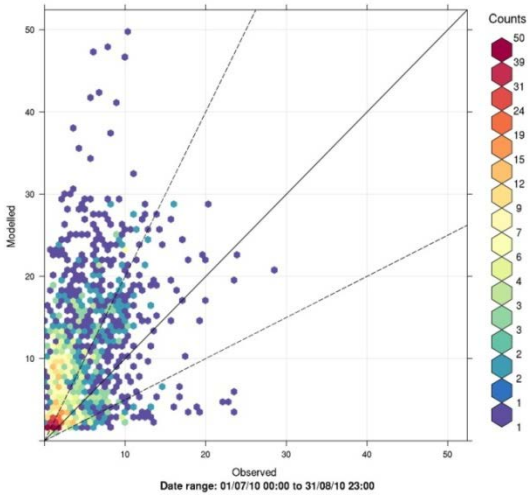
GMES

SEVENTH FRAMEWORK PROGRAMME

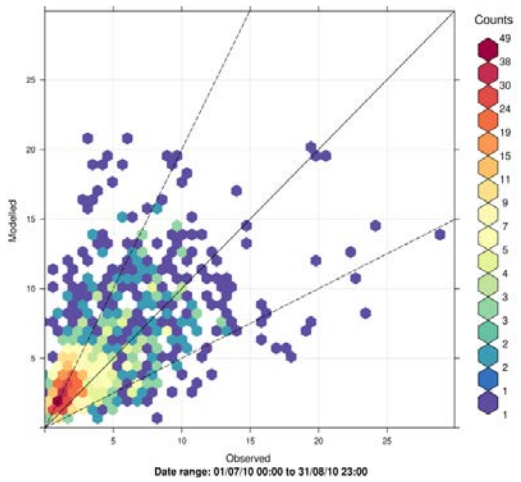
EUROPEAN COMMISSION

Comparison of overlapping airsheds

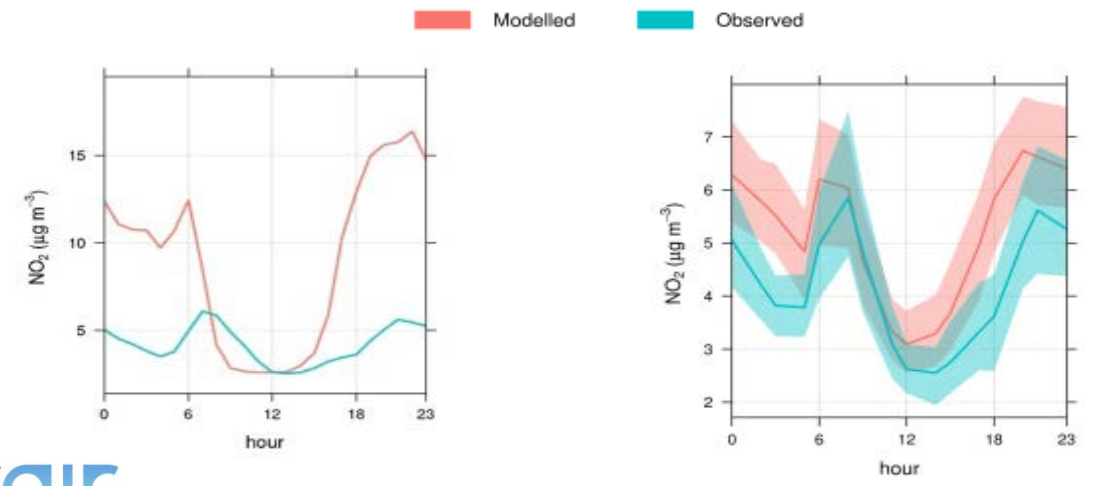
Model 1 (LOTOS-EUROS)



Model 2 (SILAM)



NO_2
Rural station
FI00208



Diurnal variation

Myair Pasodoble

Validation and Evaluation Tools

Two toolkits developed during PASODOBLE:

1. Myair Toolkit for Model Evaluation

   Developed by CERC

2. **AUTH Toolkit for retrieval and pre-processing of satellite and in situ data**

   Developed by AUTH

AUTH Toolkit

✓AUTH developed a prototype toolkit that collects observational data and post-processes model output data, in order to

- Provide data to the users (RCM and MINENV) for evaluation
- produce quick look validation plots of the AUTH Modeling system (WRF/CAMx).

✓The programming language for the software module is the Interface Data Language (IDL) and operates under WINDOWS environment (XP and 7 tested).

✓It does not require an IDL license, it uses IDL Virtual Machine available as free download

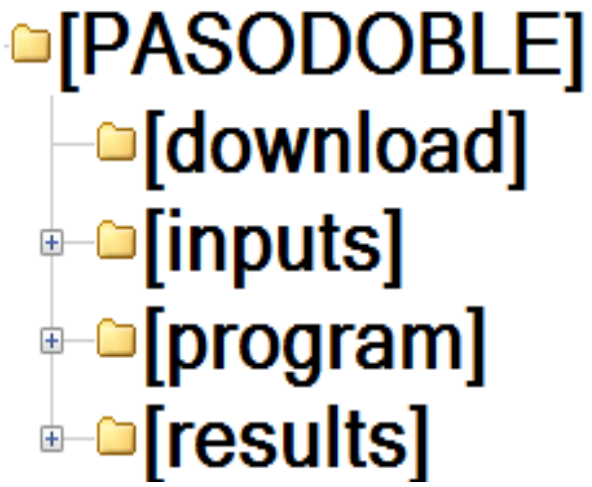
AUTH Toolkit

✓The basic functions of the AUTH Toolkit are:

1)Download the observational data

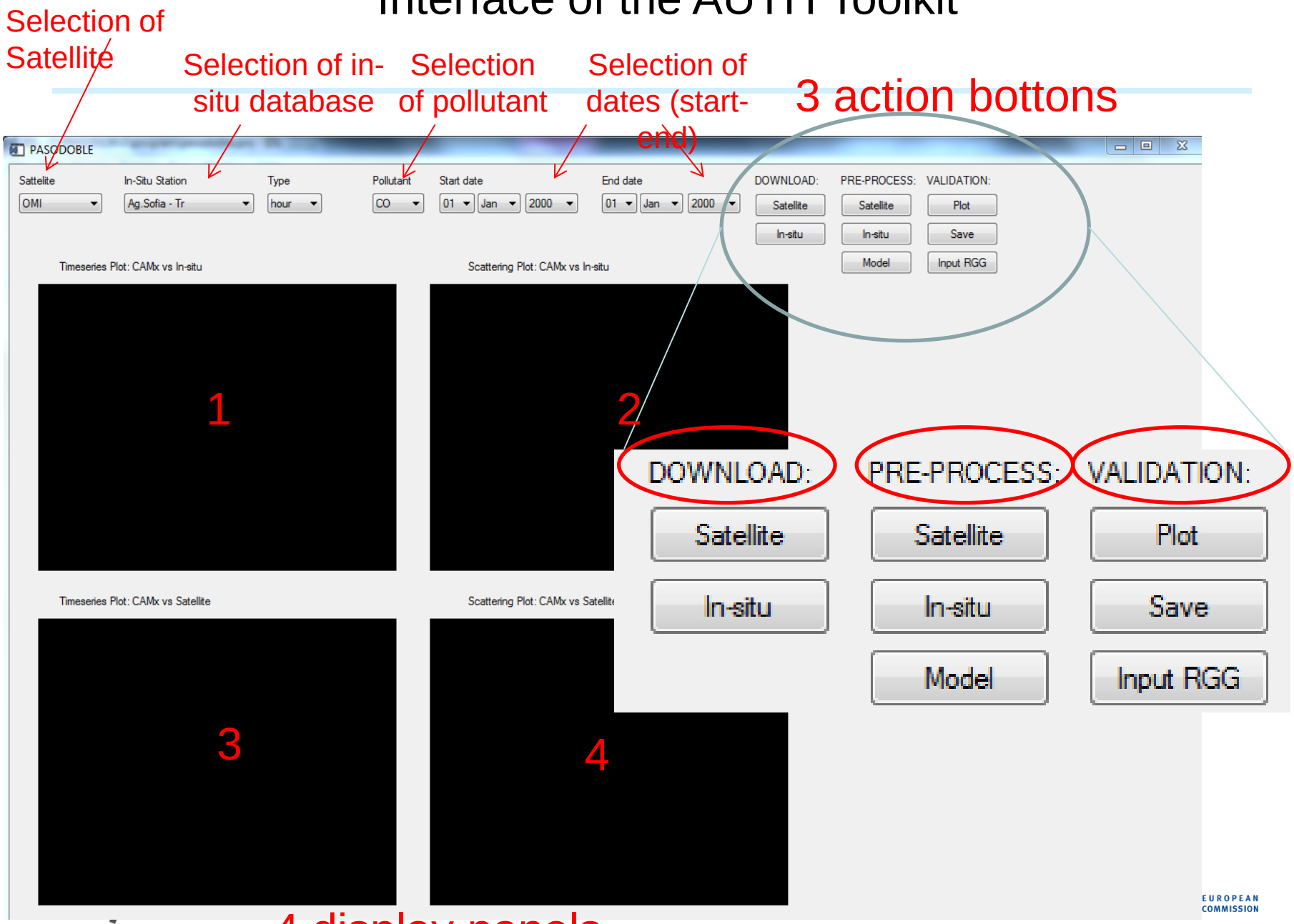
2) Pre-process model output and observations

3) Produce validation plots

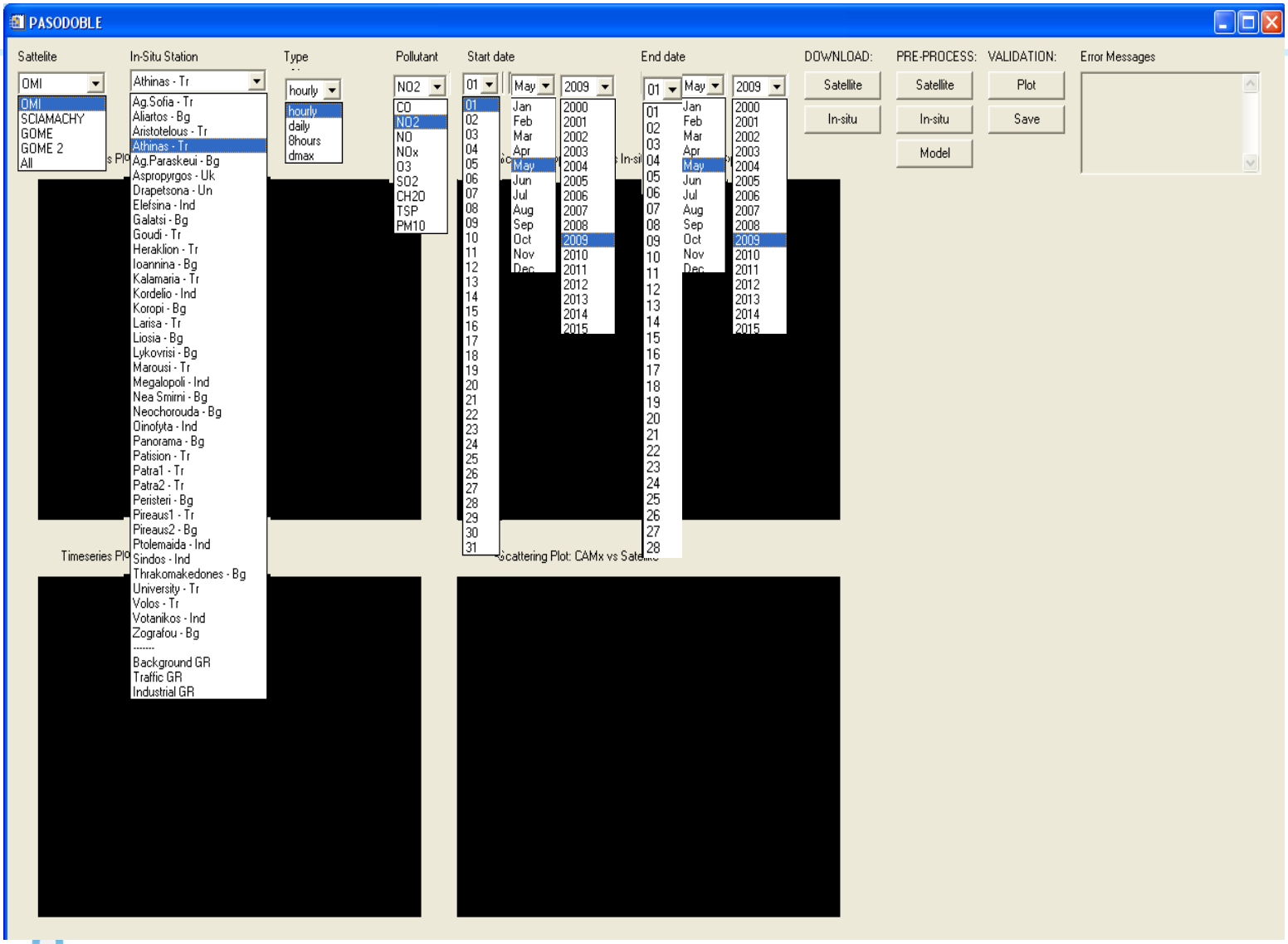


Directory listing created upon the AUTH Toolkit installation

Interface of the AUTH Toolkit



Interface of the AUTH Toolkit: drop down menus



Myair Pasodoble

GMES

EUROPEAN COMMISSION

AUTH Toolkit: The **input-RGG** action button

The AUTH Toolkit can be used as a pre-processing tool to prepare an observational data set to the Myair Model Evaluation Toolkit in the required CSV format.

✓Use the input-RGG action button if you want to prepare those datasets as input to the Myair Toolkit.

✓Currently there is the option of preparing the observations of AIRBASE and the ones received directly from the Hellenic Ministry of Environment and the Region of Central Macedonia to the Myair Toolkit-ready format.

✓The Myair Toolkit-ready format is a CSV file including information of date, hour and pollutant concentration.

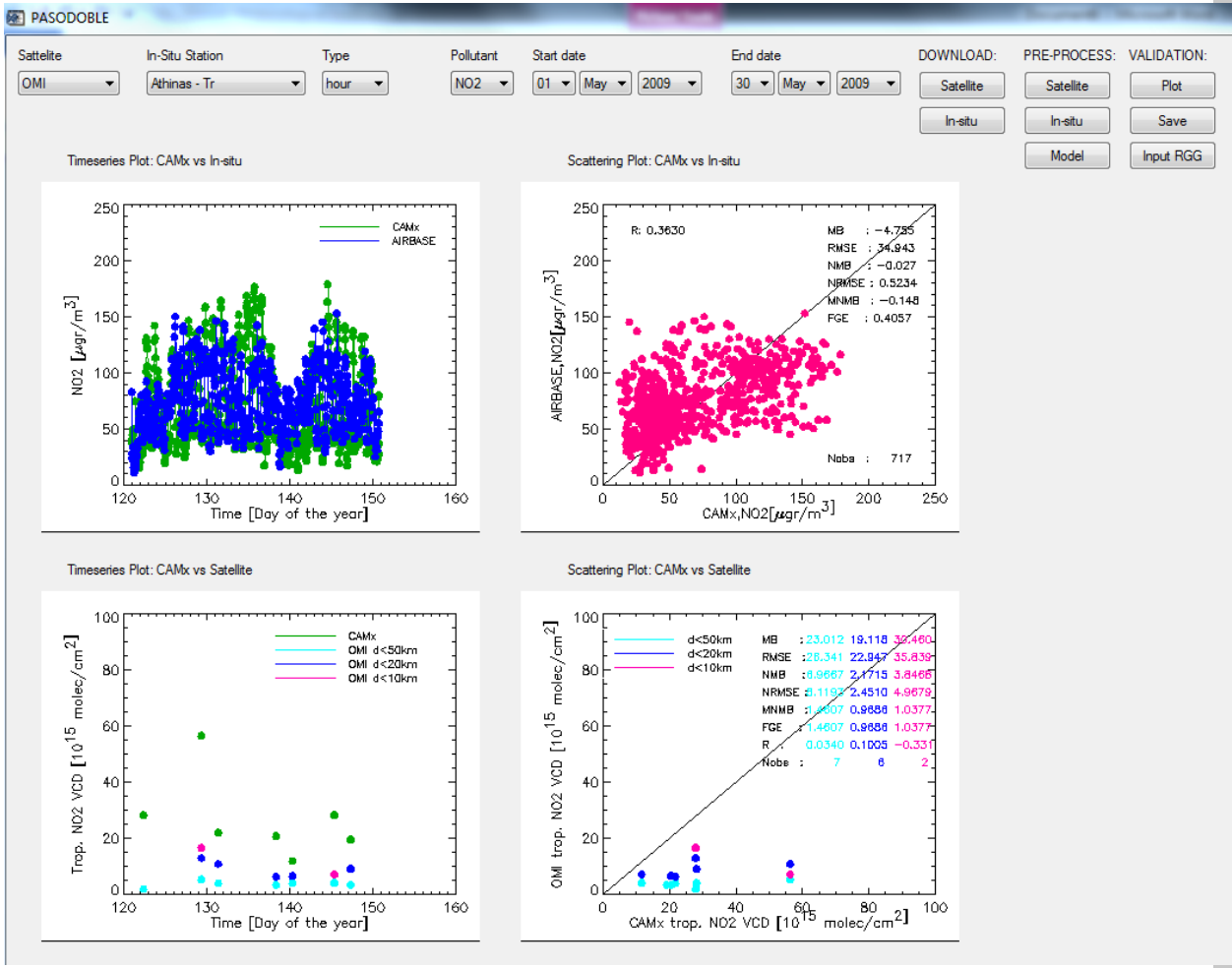
AUTH Toolkit: Results of a test case over Athens

Upper left: Time-series of surface NO2 concentration WRF/CAMx (green) and AIRBASE (blue) hourly values over an urban Athens station (Athinas station, May 2009)

Upper right: Scatter plot of measured vs. modeled surface NO2 and evaluation statistics (error, bias, rmse)

Bottom left: Time series of NO2 column for model (green) and satellite data

Bottom right: Scatter plot of measured vs. modeled NO2 column and evaluation statistics (error, bias, rmse)



Myair Pasodoble

GMES

SEVENTH FRAMEWORK PROGRAMME

EUROPEAN COMMISSION

Summary

- Demonstrated two toolkits developed during PASODOBLE for validation and evaluation:
  1. Myair Toolkit for Model Evaluation
  2. AUTH toolkit for download and pre-processing of satellite and in situ data
- For more information and to download the toolkits, go to http://www.myair.eu/products-services/local-model-evaluation/
- Many thanks to Renske Timmermans (TNO) for slides on the PASODOBLE AIRSHEDS evaluation, Dimitris Balis (AUTH) for slides on the AUTH toolkit, and Sam Royston (CERC) for graphs from the Myair Toolkit for Model Evaluation