

# Myair Toolkit for Model Evaluation

**Amy Stidworthy<sup>1</sup>, David Carruthers<sup>1</sup>, Jenny Stocker<sup>1</sup>, Sam Royston<sup>1</sup>, Dimitris Balis<sup>2</sup>, Eleni Katragkou<sup>2</sup>, Jaakko Kukkonen<sup>3</sup>**

<sup>1</sup>Cambridge Environmental Research Consultants (CERC)

<sup>2</sup>Aristotle University of Thessaloniki (AUTH)

<sup>3</sup>Finnish Meteorological Institute (FMI)

**15<sup>th</sup> International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes**

Madrid, Spain 6 – 9 May 2013



# Summary

---

- Background
- Existing tools and methodologies
- Myair Toolkit capabilities
- Practicalities
- Summary

# Background PASODOBLE project

- PASODOBLE is the Copernicus (GMES) downstream service project, producing **local-scale air quality services for Europe** under the name **Myair Pasodoble**

- Local forecast model evaluation support services

The screenshot shows the Myair Pasodoble website interface. At the top right, there is a language selector set to 'English' and a Twitter follow button for '@Myair-Pasodoble'. Below the navigation bar, the 'Products & Services' section is highlighted. A table lists various services, with one row circled in red.

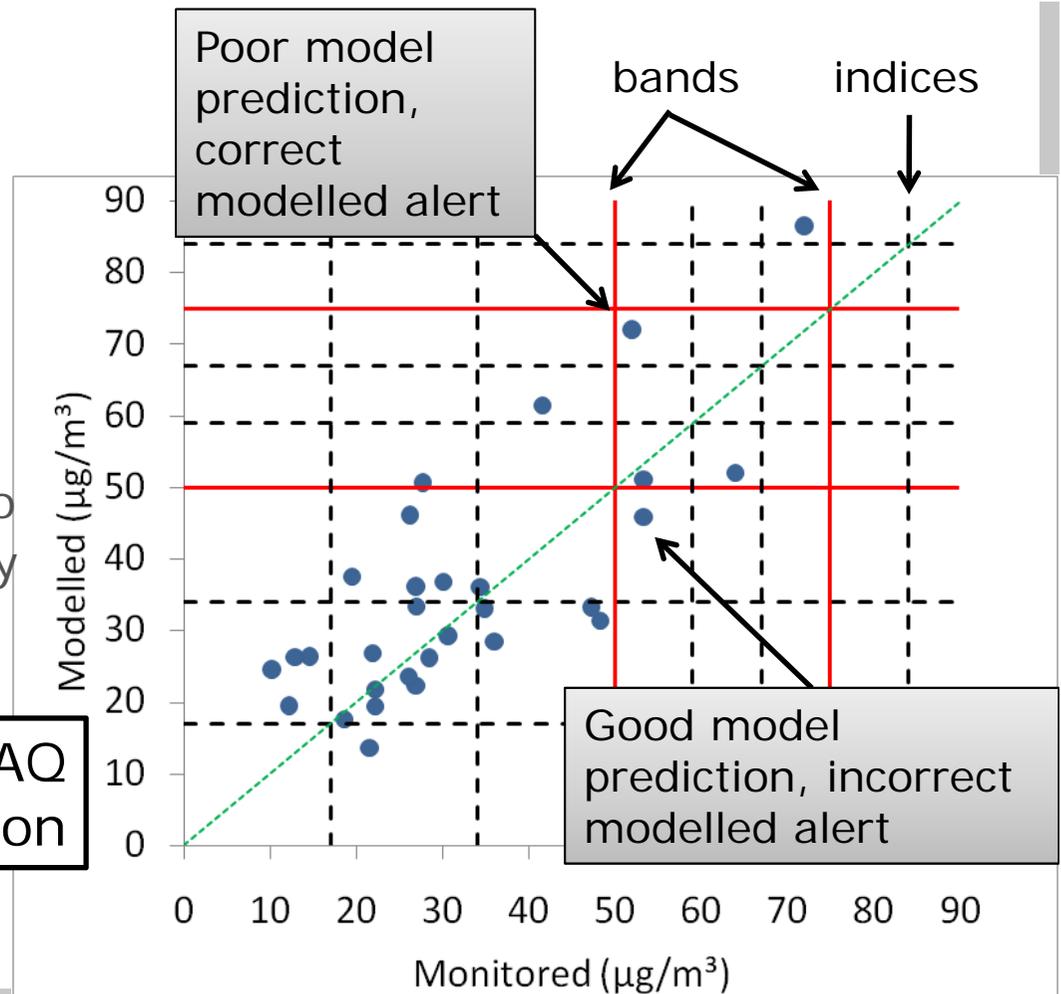
Service Line	Service	European region / city / cities covered	Service provider*	Products
Local forecast model evaluation support	Methodology and toolkit for local forecast model evaluation	All	CERC	Toolkit for local forecast model evaluation
	Tools for pre-processing of surface and satellite data that can be applied across Europe	All	AUTH	Tools for pre-processing of surface and satellite data

# Background

## Why AQ forecast models need special tools

- Air quality (AQ) forecasting systems predict air quality in terms of bandings.
- Forecasts aim to get the band correct (low, moderate etc).
- An alert is issued by the forecasting system if a moderate, high or very high band is forecast
- Therefore, validating a forecasting system is different to validating concentrations directly output from an AQ model.

### Daily Air Quality Index



Scatter plot for AQ forecast system validation

# Background

## Development procedure

---

- The Myair Toolkit validates air quality model output, focussing on requirements for **standardised evaluation of local air quality forecast** models.
- The Toolkit was designed following an extensive review (2010-11) of the state-of-the-art in air quality model evaluation.
- Toolkit builds on existing tools and initiatives.

# Existing tools and methodologies

---

- AQ model validation tools and methodologies:
  - Model Validation Kit
  - ASTM model evaluation methodology
  - FAIRMODE Delta tool
- Meteorological forecasting models:
  - Event-based statistics used at ECMWF for validating the accuracy of Numerical Weather Predictions can be used to validate AQ forecasting systems.
- openair data analysis tools:
  - UK project for the air pollution community
  - Free, open-source, innovative data analysis tools.

# Myair Toolkit capabilities

## What can Myair Toolkit do?

---

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Help you investigate model performance at individual stations using openair graphs
- 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
- Run in batch mode, for easy automation

# Myair Toolkit capabilities

## What can Myair Toolkit do?

---

- **Assess your model's forecast skill**
- Assess your model's concentration predictions
- Help you investigate model performance at individual stations using openair graphs
- 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
- Run in batch mode, for easy automation

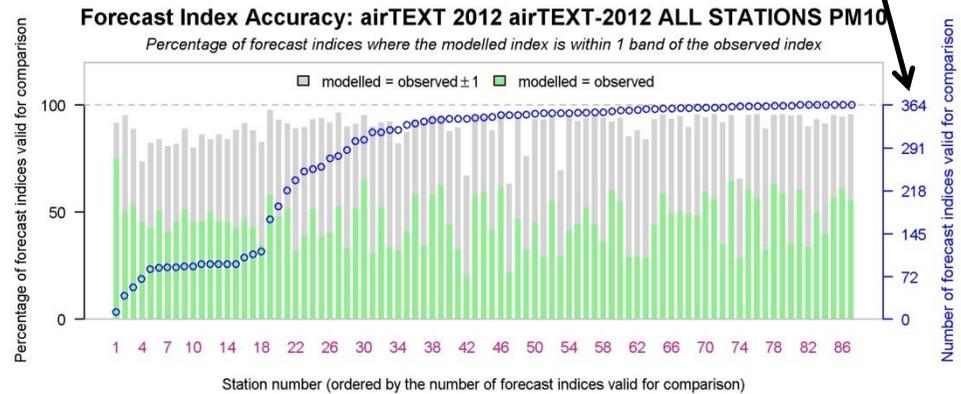
# Myair Toolkit capabilities

## Assess your model's forecast skill

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

- modelled = observed  $\pm 1$
- modelled = observed

### Number of forecast indices valid for comparison

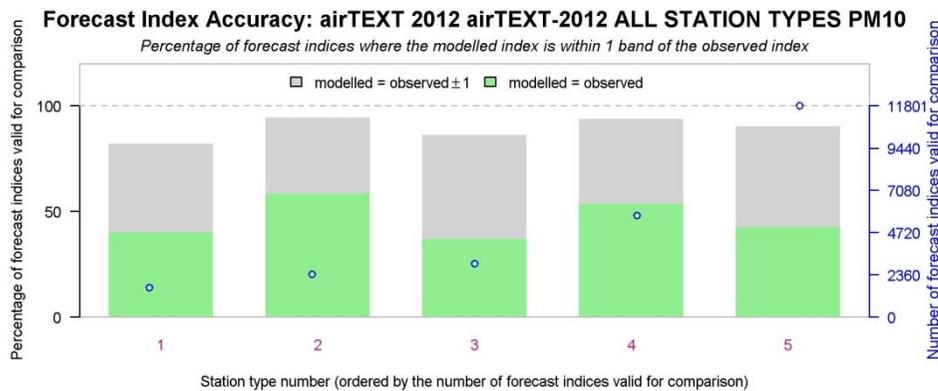


Key to numbers

1	EA2	11	HS2	21	ENS	31	HS8	41	B14	51	CT3	61	GR8	71	GR5	81	BL0
2	ST7	12	HS6	22	EI1	32	SK5	42	CT8	52	EI2	62	WA7	72	RB4	82	EA6
3	LT	13	HS9	23	GN4	33	CD3	43	WMO	53	MY1	63	KC5	73	GN3	83	GR0
4	LE5	14	LE5	24	LW2	34	MY7	44	CD1	54	ST4	64	TK8	74	BT6	84	GR7
5	ST5	15	ST5	25	KC2	35	ME2	45	GNO	55	HV3	65	BX2	75	BX0	85	RI1
6	BC8	16	BC8	26	NM2	36	ZB1	46	BQ7	56	LW3	66	GN2	76	KC1	86	RI2
7	BC9	17	BC9	27	VAG	37	EN4	47	HF4	57	HG1	67	CR4	77	ST6	87	RI3
8	HG2	18	HG2	28	WM6	38	BX1	48	IS6	58	HK6	68	TH4	78	GR9		
9	NM3	19	NM3	29	EI7	39	DR4	49	EA3	59	BQ8	69	BT6	79	HT1		
10	BC2	20	BC2	30	KC7	40	BQ5	50	HR2	60	EA7	70	CR3	80	IS2		

Station number

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



Key to numbers

1	INDUSTRIAL
2	SUBURBAN
3	KERBSIDE
4	URBAN BACKGROUND
5	ROADSIDE



# Myair Toolkit capabilities

## Assess your model's forecast skill

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

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

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

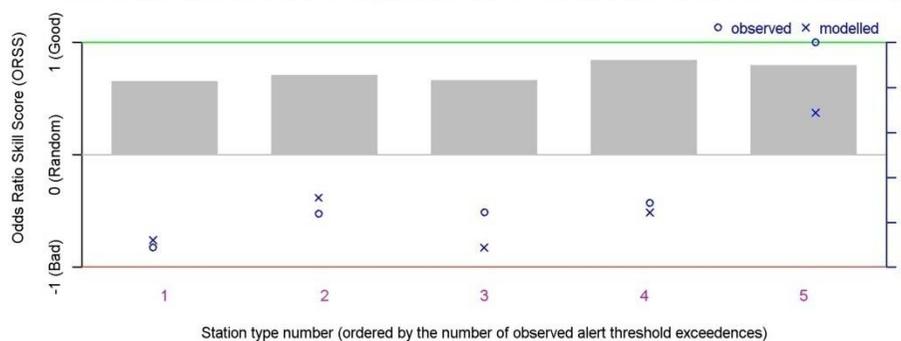
# Myair Toolkit capabilities

## Assess your model's forecast skill

ORSS grouped by station...

**Odds Ratio Skill Score (ORSS): airTEXT-2012 ALL STATION TYPES MODERATE PM10**

No bars are shown where ORSS is invalid (the number of observed or modelled alert threshold exceedences is zero)



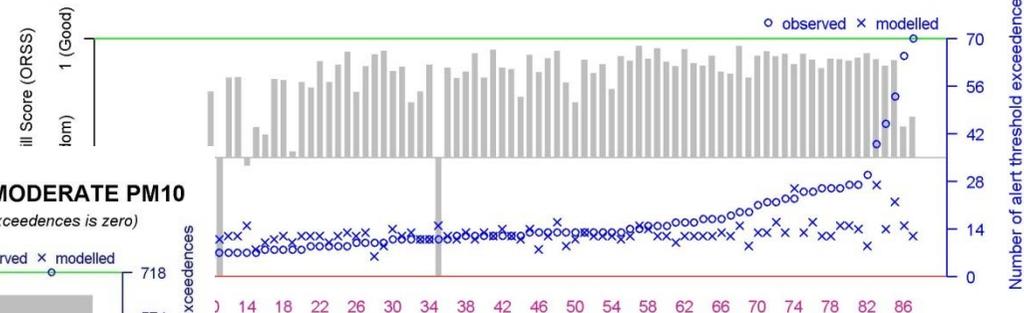
Key to numbers

1	SUBURBAN
2	URBAN BACKGROUND
3	INDUSTRIAL
4	KERBSIDE
5	ROADSIDE

Number of forecast indices valid for comparison

**Odds Ratio Skill Score (ORSS): airTEXT-2012 ALL STATIONS MODERATE PM10**

No bars are shown where ORSS is invalid (the number of observed or modelled alert threshold exceedences is zero)



Station number (ordered by the number of observed alert threshold exceedences)

Key to numbers

BQ8	21	CR3	31	BT6	41	FG1	51	IS2	61	LW2	71	GB0	81	TH4
BX2	22	EI2	32	CT3	42	KC2	52	IS6	62	ST4	72	LB4	82	BQ5
KC1	23	GR4	33	HS2	43	RI1	53	NW2	63	ST6	73	HS8	83	MY1
LB5	24	KO7	34	HV3	44	ST5	54	RB5	64	LW3	74	MY7	84	EA8
LB6	25	TH1	35	WA7	45	EN4	55	TK8	65	RB4	75	GR7	85	CT8
BN2	26	BX0	36	WM0	46	EN5	56	HR2	66	SK5	76	KC5	86	BT5
GN3	27	EAT	37	BN1	47	GR5	57	WM6	67	HS6	77	GN0	87	HF4
HS7	28	HS4	38	EA6	48	HK6	58	CD3	68	CD1	78	GN4		
RB3	29	WA9	39	EI1	49	HS5	59	CR4	69	ME2	79	GR8		
RI2	30	BL0	40	EI7	50	HS9	60	GR9	70	GN2	80	BT4		

... 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 Toolkit capabilities

## Assess your model's forecast skill

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...

Probability

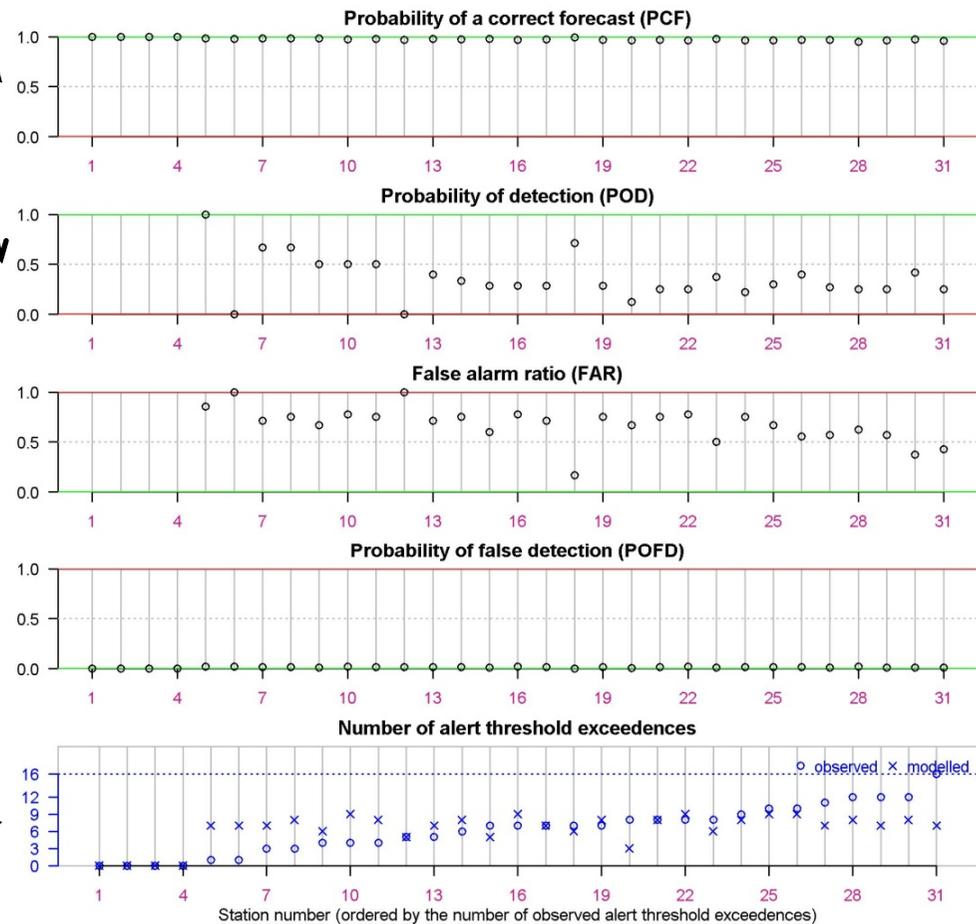


Number of alerts



Forecast Performance Metrics: airTEXT 2012

ALL STATIONS MODERATE O3



Key to numbers						
1	CT6	11	TH4	21	GB6	
2	EA1	12	CR3	22	RB1	
3	HS2	13	GR9	23	WM0	
4	MY1	14	WA2	24	GR4	
5	BT4	15	BL0	25	BQ7	
6	GN3	16	BX1	26	NM2	
7	HK6	17	EA7	27	NM3	
8	LW1	18	H10	28	HG2	
9	GR8	19	ST3	29	KC1	
10	TD0	20	CT1	30	RI2	
					31	TH1

# Myair Toolkit capabilities

## What can Myair Toolkit do?

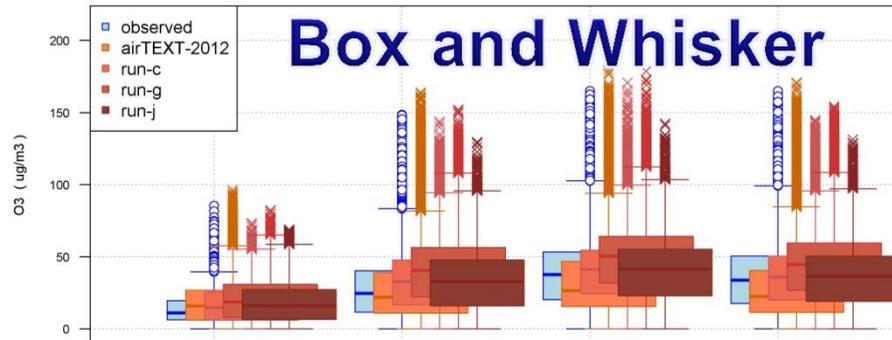
---

- Assess your model's forecast skill
- **Assess your model's concentration predictions**
- Help you investigate model performance at individual stations using openair graphs
- 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
- Run in batch mode, for easy automation

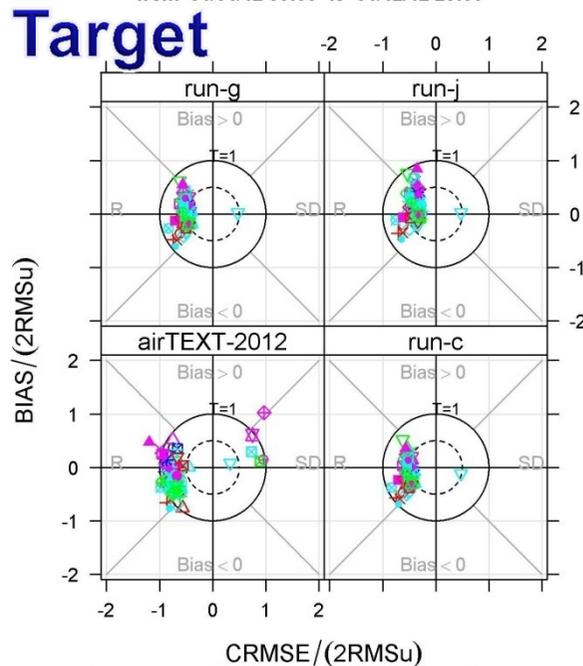
# Myair Toolkit capabilities

## Assess your model's concentration predictions

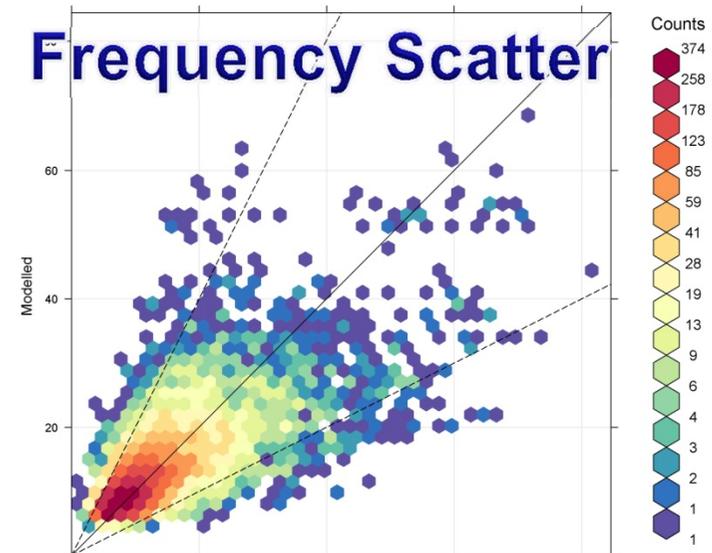
Box and Whisker plot: AIRTEXT 2012 VALIDATION  
ALL STATIONS, 8-HOUR ROLLING MEAN O3



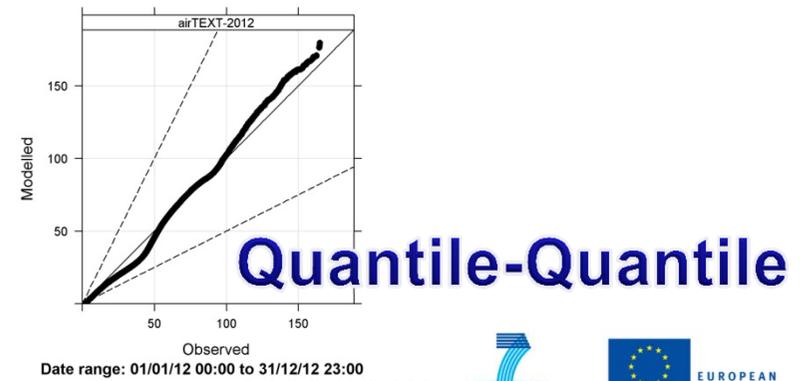
Target plot (DELTA 3.3): airTEXT 2012, ALL STATIONS, DAILY MEAN PM10  
from 01/01/12 00:00 to 31/12/12 23:00



Frequency Scatter Plot: AIRTEXT 2012 VALIDATION  
ALL STATIONS, DAILY MEAN PM<sub>2.5</sub> ( $\mu\text{g m}^{-3}$ )



Quantile-Quantile Plot: AIRTEXT 2012 VALIDATION  
ALL STATIONS, 8-HOUR ROLLING MEAN O<sub>3</sub> ( $\mu\text{g m}^{-3}$ )<sup>d</sup>  
00:00 to 31/12/12 23:00



# Myair Toolkit capabilities

## Assess your model's concentration predictions

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	Mean	$1/n \sum C$
mod.mean		
SDO	Standard Deviation	$\sqrt{1/n \sum (C - \bar{C})^2}$
SDM		
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	$\text{cov}(C_p, C_o) / \sigma_{C_p} \sigma_{C_o}$
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		max C
mod.max		
obs.RHC		$\chi(n) + (\chi - \chi(n)) \ln\left(\frac{3n-1}{2}\right)$ ,
mod.RHC		where $n$ is the number of values used to characterise the upper end of the concentration distribution, $\chi$ is the average of the $n - 1$ largest values, and $\chi(n)$ is the $n^{\text{th}}$ largest value; $n$ is taken to be 26.

Maximum Robust highest concentration

# Myair Toolkit capabilities

## What can the Myair Toolkit do?

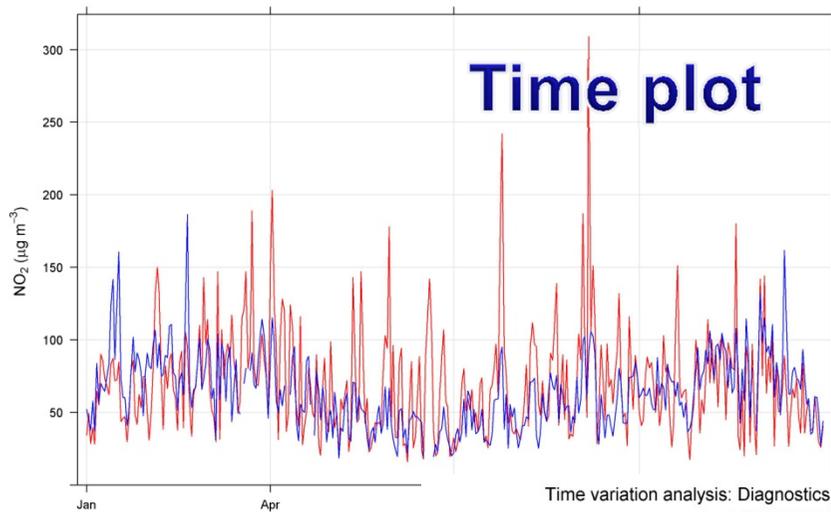
---

- Assess your model's forecast skill
- Assess your model's concentration predictions
- **Help you investigate model performance at individual stations using openair graphs**
- 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
- Run in batch mode, for easy automation

# Myair Toolkit capabilities

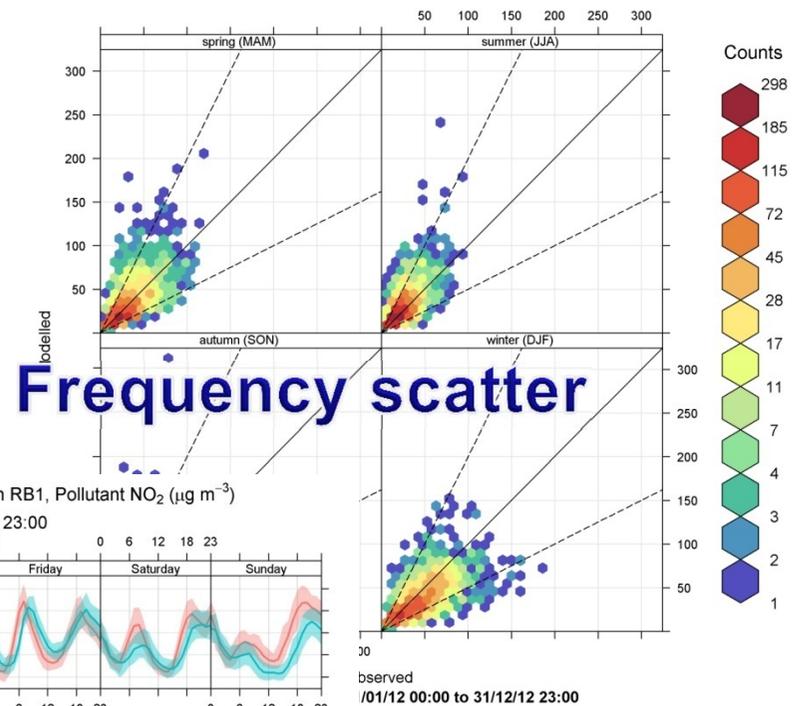
## Investigate model performance at individual stations

Time Plot: Diagnostics, run-g, Station RB1, Pollutant NO<sub>2</sub> (µg m<sup>-3</sup>)



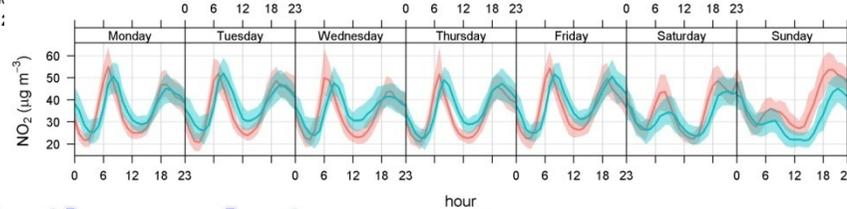
**Time plot**

Frequency Scatter Plot: Diagnostics, run-g, RB1  
Raw Data NO<sub>2</sub> (µg m<sup>-3</sup>), Filtered by Season

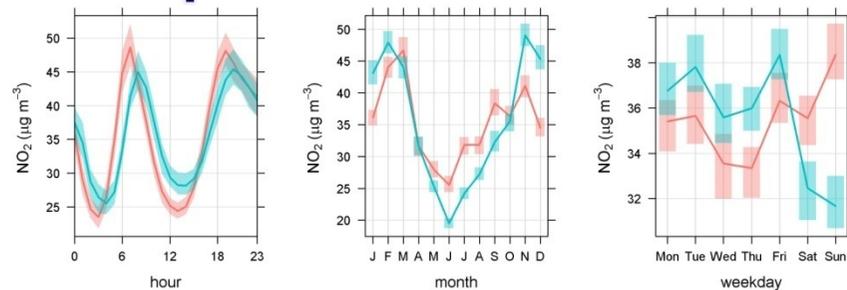


**Frequency scatter**

Time variation analysis: Diagnostics, run-g, Station RB1, Pollutant NO<sub>2</sub> (µg m<sup>-3</sup>)  
01/01/12 00:00 to 31/12/12 23:00



**Time variation plot**



# Myair Toolkit capabilities

## What can the Myair Toolkit do?

---

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Help you investigate model performance at individual stations using openair graphs
- **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
- Run in batch mode, for easy automation

# Myair Toolkit capabilities

## Import gridded and point modelled data formats

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

The screenshot shows the '2. Modelled data' section of the Myair Toolkit. It features a form with the following elements:

- Modelled data label:** A text input field containing 'airTEXT-2012'.
- Select modelled data format:** A section with three radio buttons: 'netCDF' (selected), 'ADMS PST', and 'CSV'.
- Choose format:** A dropdown menu currently showing 'AIRSHEDS', with a list of options including 'AIRSHEDS', 'MACC Ensemble', and 'CMAQ'.
- Select file:** A radio button option with a 'Browse' button.
- Select directory:** A radio button option with a 'Browse' button.
- Separator:** A dropdown menu set to 'comma'.
- Missing data indicator:** An empty text input field.

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

# Myair Toolkit capabilities

## What can the Myair Toolkit do?

---

- Assess your model's forecast skill
- Assess your model's concentration predictions
- Help you investigate model performance at individual stations using openair graphs
- 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
- Run in batch mode, for easy automation

# Myair Toolkit capabilities

## Download and import monitoring data

- 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

2. Observed data

CSV

London KCL

UK Automatic Urban and Rural Network (AURN)

Select file:

Select directory:

Separator:

Missing data indicator:

# Myair Toolkit capabilities

## What can the Myair Toolkit do?

---

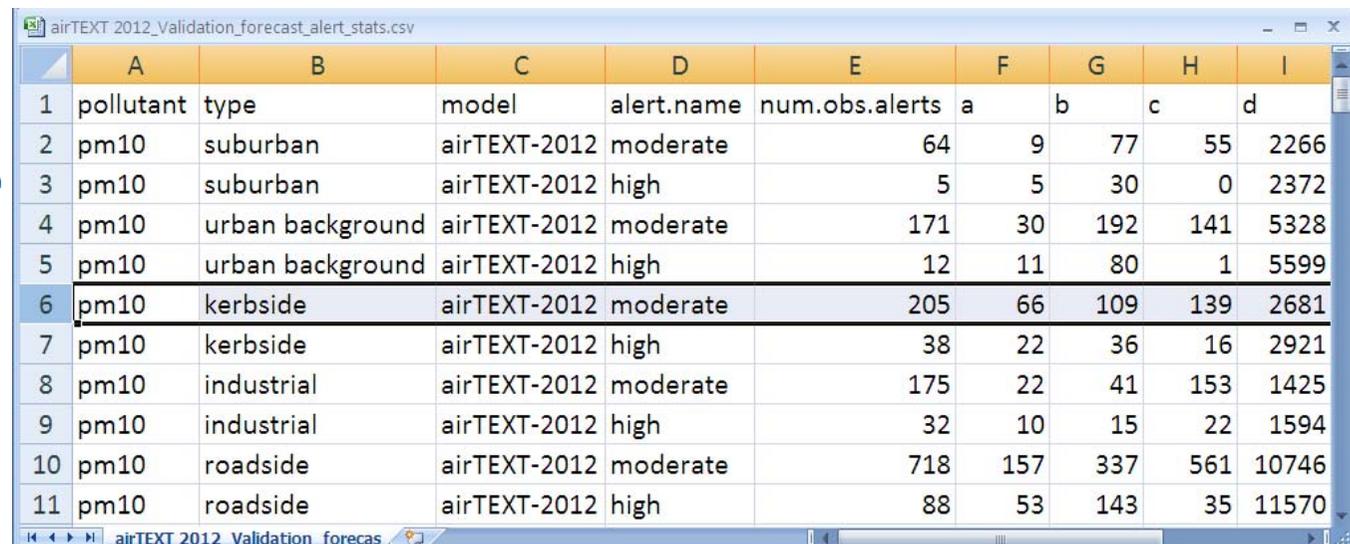
- Assess your model's forecast skill
- Assess your model's concentration predictions
- Help you investigate model performance at individual stations using openair graphs
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also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

# Myair Toolkit capabilities

## Save graphical and statistical output

- 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*



	A	B	C	D	E	F	G	H	I
1	pollutant	type	model	alert.name	num.obs.alerts	a	b	c	d
2	pm10	suburban	airTEXT-2012	moderate	64	9	77	55	2266
3	pm10	suburban	airTEXT-2012	high	5	5	30	0	2372
4	pm10	urban background	airTEXT-2012	moderate	171	30	192	141	5328
5	pm10	urban background	airTEXT-2012	high	12	11	80	1	5599
6	pm10	kerbside	airTEXT-2012	moderate	205	66	109	139	2681
7	pm10	kerbside	airTEXT-2012	high	38	22	36	16	2921
8	pm10	industrial	airTEXT-2012	moderate	175	22	41	153	1425
9	pm10	industrial	airTEXT-2012	high	32	10	15	22	1594
10	pm10	roadside	airTEXT-2012	moderate	718	157	337	561	10746
11	pm10	roadside	airTEXT-2012	high	88	53	143	35	11570

# Myair Toolkit capabilities

## What can the Myair Toolkit do?

---

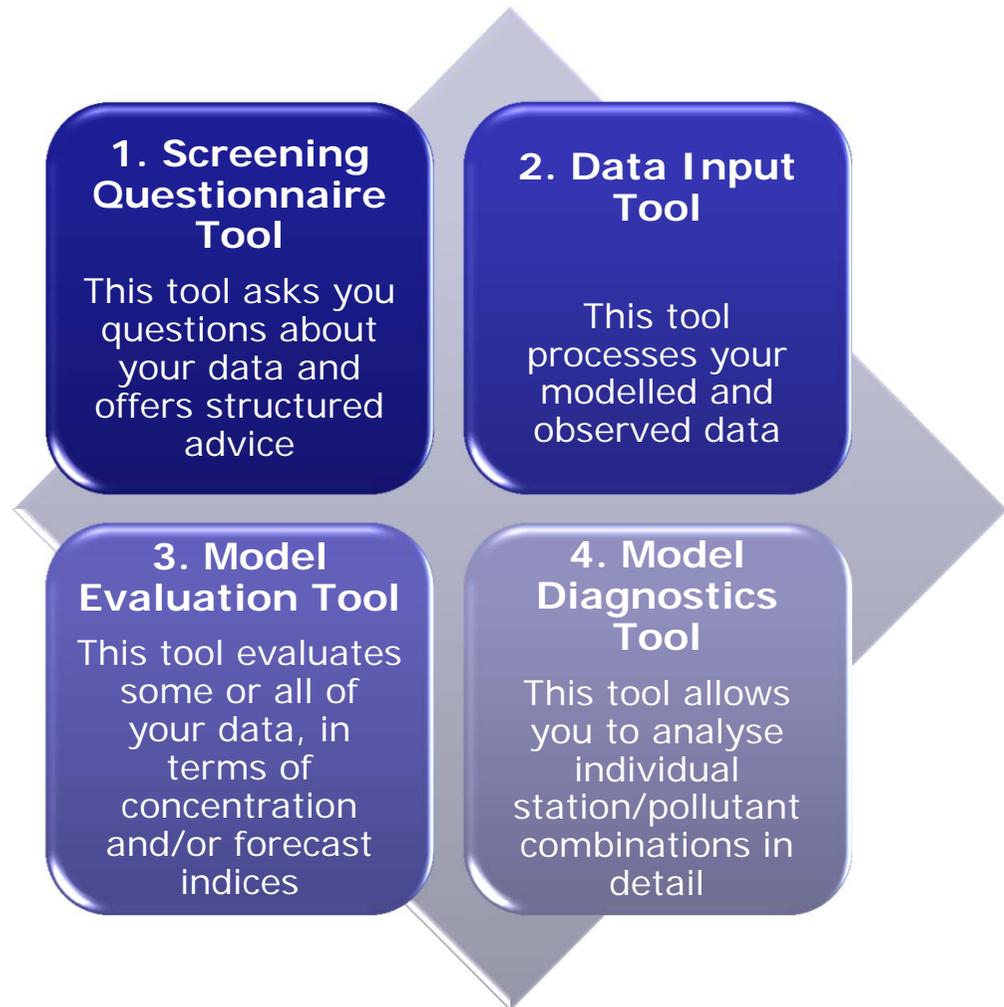
- Assess your model's forecast skill
- Assess your model's concentration predictions
- Easily import a wide range of gridded and point modelled data formats
- Download and import in situ monitoring data for the UK (also CSV files)
- Save graphical and statistical output to your computer
- Help you investigate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

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

# Practicalities

## 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



# Summary

---

- The Myair Toolkit for Model Evaluation is a powerful new tool for the evaluation of air quality forecasting models
- The Toolkit was developed building on existing tools and methodologies
- You can download the **free** toolkit from <http://www.cerc.co.uk/environmental-software/myair-toolkit.html>

# Acknowledgements

---

The authors wish to thank:

- PASODOBLE project partners, in particular the users in the 'Local forecast model evaluation support' work package and those involved in evaluating the AIRSHEDS products, for their valuable feedback during development and demonstration of the Myair Toolkit.
- EU FP7 programme for funding the 3-year PASODOBLE project (completed April 2013)