

Evaluation of DELTA forecast functionality

**Jenny Stocker,
David Carruthers
& Kate Johnson**

7th Plenary Meeting of FAIRMODE

April 2014

Kjeller

Norway

Contents

- *airTEXT* forecasting system for London
- Model performance according to DELTA version 3.6
 - Is the forecast better than persistence?
 - Is the forecasting target formulation robust?
- Why air quality forecast models need special tools
- Another forecasting evaluation tool: MyAir Toolkit for Model Evaluation
- Suggestions for additional forecasting parameters / criteria
- Summary

airTEXT forecasting system for London



MAYOR OF LONDON



Free air pollution, UV, pollen and temperature forecasts for Greater London

Currently providing free air quality alerts to more than 7000 subscribers

airTEXT forecasting system for London

airTEXT
Free air pollution forecasts

Forecasts

www.airtext.info

Daily Health Bulletin for Islington

Wednesday 25th July 2012

Air pollution MODERATE Action may be required. Health effects are unlikely to require action. If unwell, contact GP. <small>This is a daily air pollution forecast and may be LOW, MODERATE, HIGH or VERY HIGH.</small>	UV Index 7 (HIGH) Protection required. Seek shade during midday hours, cover up and wear sunscreen. <small>This is a forecast of maximum hourly cloud-adjusted solar UV index over a 24-hr period. 1 to 2 is LOW, 3 to 5 is MODERATE, 6 to 7 is HIGH or 8+ is VERY HIGH</small>
Pollen LOW <small>This is a daily grass pollen forecast and may be LOW, MODERATE, HIGH or VERY HIGH.</small>	Temperature Max. Day 29°C/84°F Min. Night 17°C/62°F <small>These are the minimum and maximum hourly temperatures predicted over a 24-hour period.</small>

Forecasts supported by funding from defra (www.defra.gov.uk) and EU FP7 PASODOBLE (www.myair-eu.org)

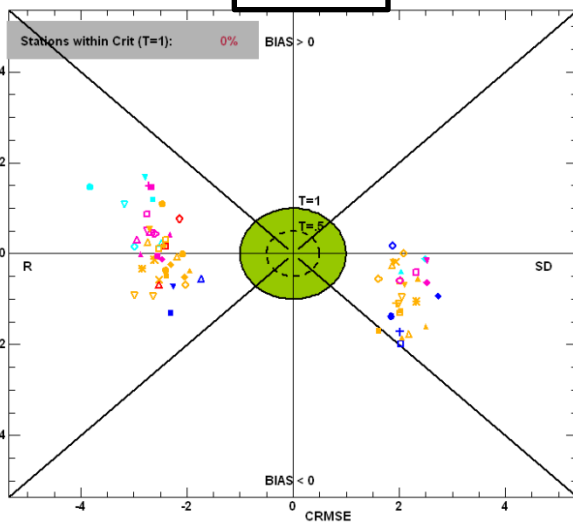
Nokia Phone Message:
airTEXT
10:48
MODERATE air pollution forecast for Monday. Health effects are unlikely to require action. If unwell, contact GP. Unsubscribe www.airtext.info
Options Reply Back

Smartphone App:
About airTEXT Info
Forecast for: Olympic Park
Today Tomorrow
Air pollution MODERATE (2)
UV Index MODERATE (3)
Pollen HIGH
Temperature 13C -20C

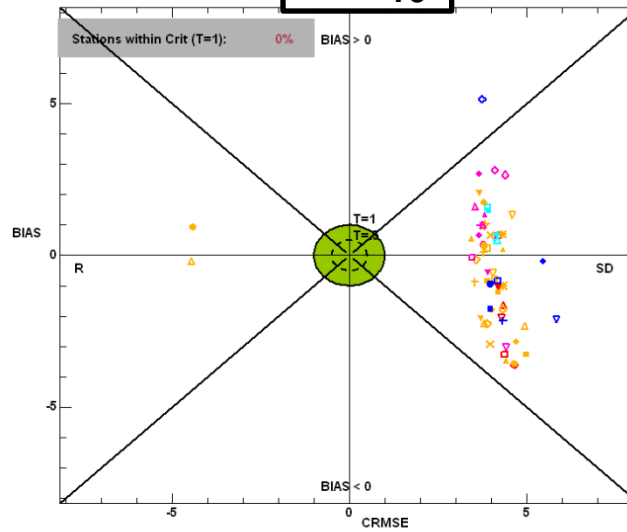
Model performance (DELTA version 3.6)

- How well is *airTEXT* performing according to DELTA, using the 2013 dataset?

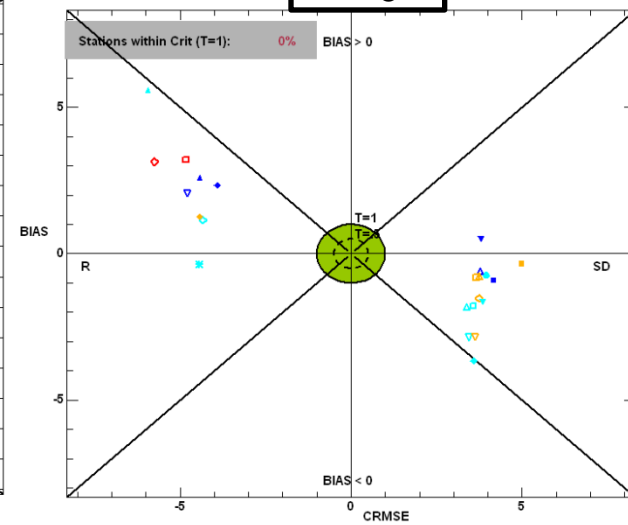
NO₂



PM₁₀



O₃



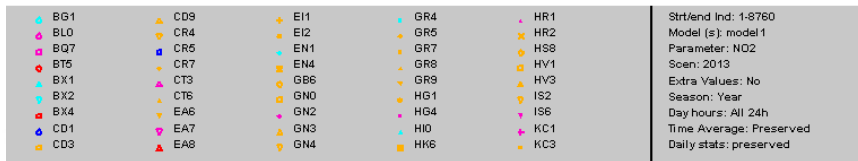
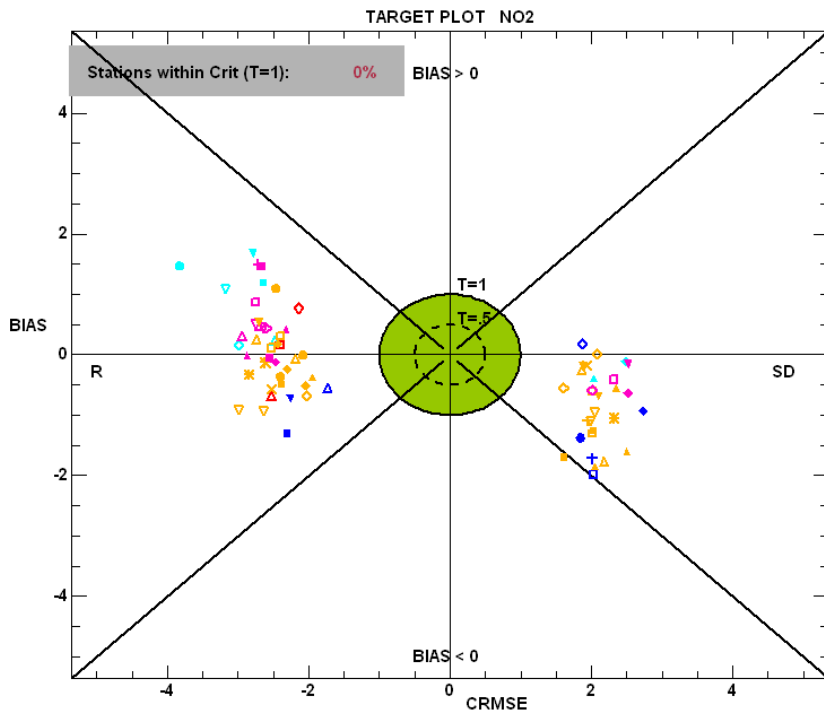
● BG1	● CD9	● E11	● GR4	● HR1	● BLO	● CD9	● E11	● GR4	● HR1	● BLO	● CD9	● E11	● GR4	● HR1
● BG2	● CR4	● E12	● GR5	● HR2	● BQ5	● CD9	● E11	● GR4	● HR1	● BQ5	● CD9	● E11	● GR4	● HR1
● BG7	● CR5	● E14	● GR7	● HR3	● BQ6	● CD9	● E11	● GR4	● HR1	● BQ6	● CD9	● E11	● GR4	● HR1
● BT5	● CR7	● E14	● GR8	● HV1	● BQ7	● CD9	● E11	● GR4	● HR1	● BQ7	● CD9	● E11	● GR4	● HR1
● BX1	● CT3	● GB6	● GR9	● HV3	● BQ8	● CD9	● E11	● GR4	● HR1	● BQ8	● CD9	● E11	● GR4	● HR1
● BX2	● CT6	● GB6	● GR9	● HV3	● BT4	● CD9	● E11	● GR4	● HR1	● BT4	● CD9	● E11	● GR4	● HR1
● BX4	● E46	● G12	● HG4	● IS6	● CT3	● CD9	● E11	● GR4	● HR1	● CT3	● CD9	● E11	● GR4	● HR1
● CD1	● E47	● G13	● HO	● KC1	● BT6	● CD9	● E11	● GR4	● HR1	● BT6	● CD9	● E11	● GR4	● HR1
● CD3	● E48	● G14	● HO6	● KC3	● BX0	● CD9	● E11	● GR4	● HR1	● BX0	● CD9	● E11	● GR4	● HR1
StratInd: 1-8760 Model (s): model f Parameter: NO2 Scen: 2013 Extra Values: No Season: Year Day hours: All 24h Time Average: Preserved Daily stats: preserved					StratInd: 1-8760 Model (s): ADM5 Parameter: PM10 Scen: 2013 Extra Values: No Season: Year Day hours: All 24h Time Average: Preserved Daily stats: Mean					StratInd: 1-8760 Model (s): ADM5 Parameter: O3 Scen: 2013 Extra Values: No Season: Year Day hours: All 24h Time Average: 8h Daily stats: Max				

- Terribly!!!

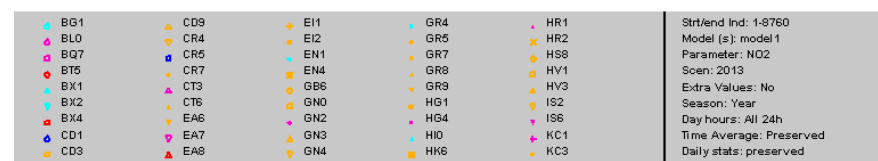
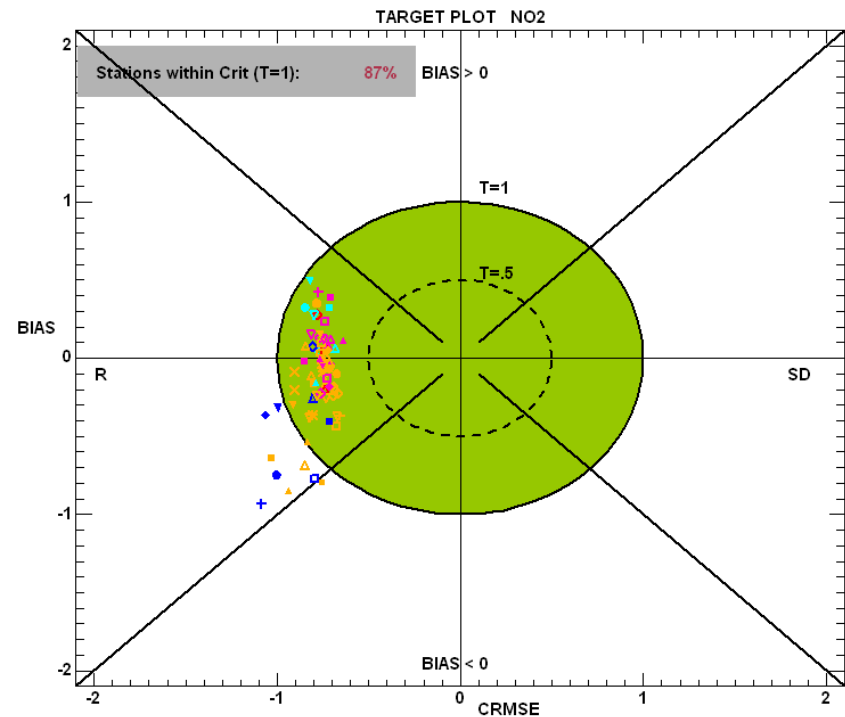
Model performance (DELTA version 3.6)

- Does this poor performance make sense when the model performs well in the standard Target plot (same dataset)?

NO₂ – Forecasting target



NO₂ – Standard target



Model performance according to DELTA version 3.6

Is the forecast better than persistence?

- Target for forecasting applications is related to the forecast being as good as a persistence model:

$$\text{Target} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (O_{i-1} - O_i)^2}}$$

where N is the number of observations, M_i is the modelled value and O_i is the observed value.

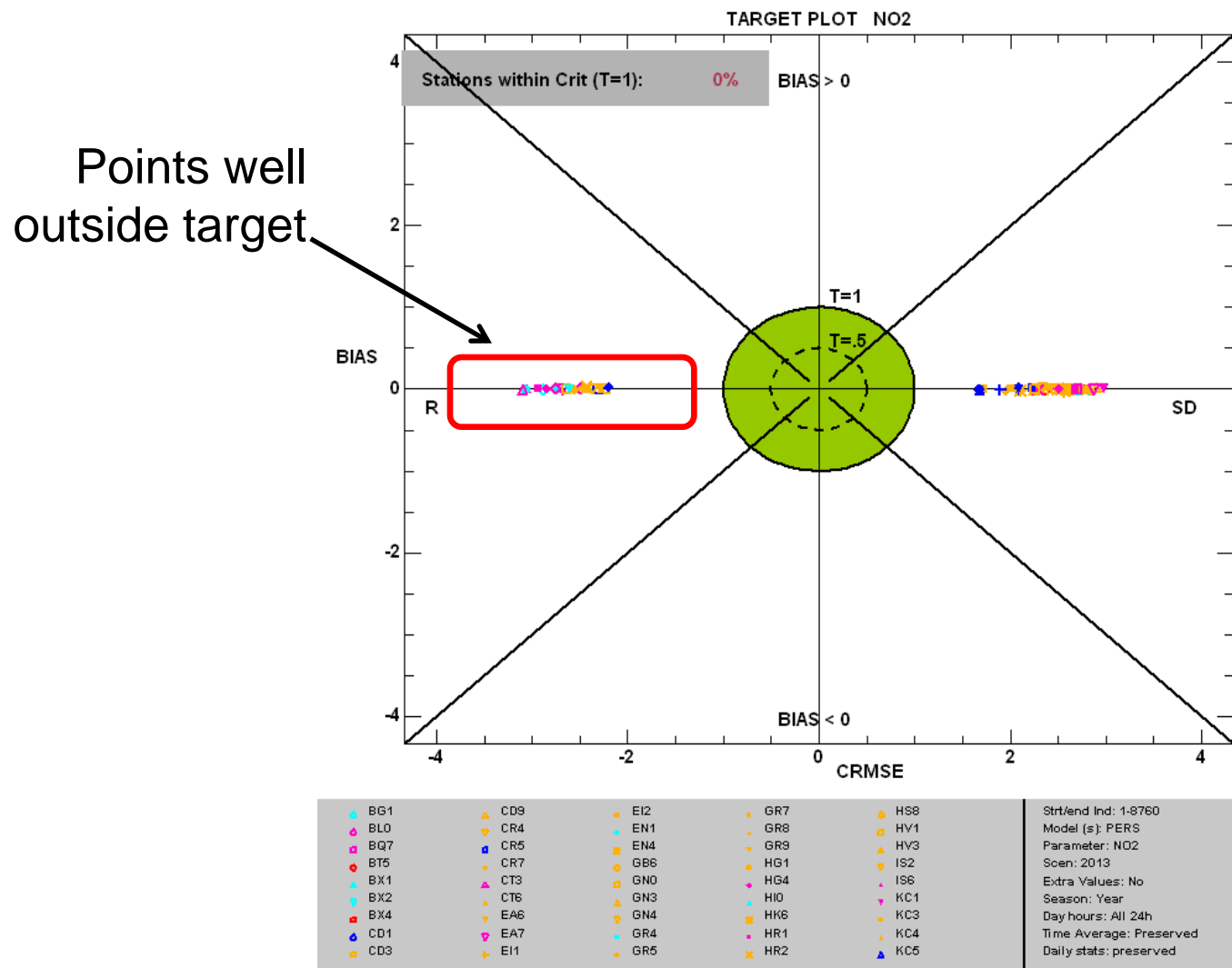
- So test the Forecasting plot with these values for London 2013 observations i.e. on a **day-by-day basis**:

$$M_i = O_{i-1}$$

Model performance according to DELTA version 3.6

Is the forecast better than persistence?

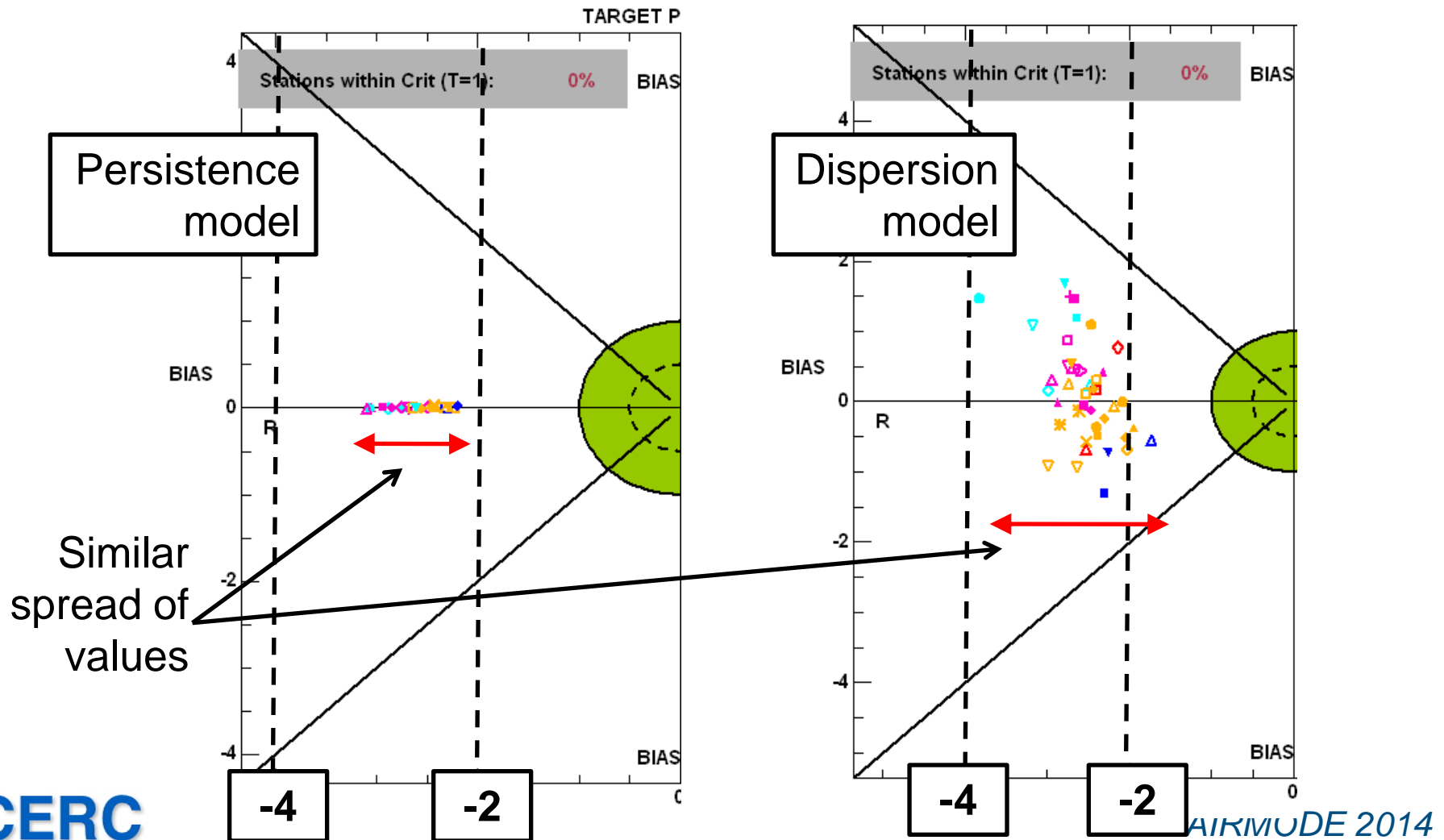
- Persistence plot for NO₂ (similar plot for other pollutants)



Model performance according to DELTA version 3.6

Is the forecast better than persistence?

- Persistence plot for NO₂ (similar plot for other pollutants)



Model performance according to DELTA version 3.6

Is the forecasting target formulation robust?

- Take:

$$\text{Target} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (O_{i-1} - O_i)^2}}$$

where N is the number of observations, M_i is the modelled value and O_i is the observed value.

- If you had a period where the levels of pollution remained the same on a **day by day basis** (either constant, or varying diurnally), then

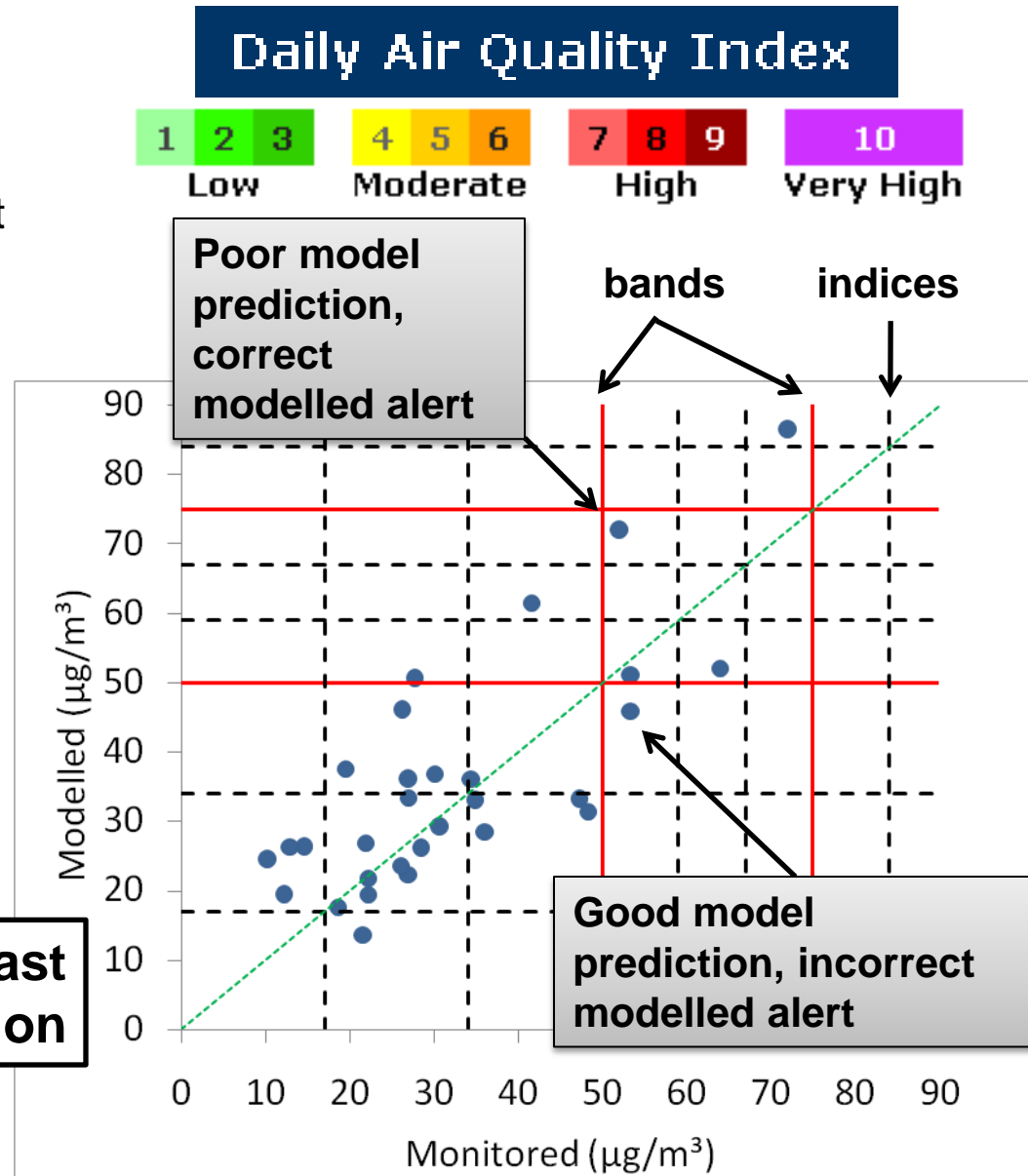
$$\frac{1}{N} \sum_{i=1}^N (O_{i-1} - O_i)^2 = 0$$

so the target \rightarrow infinity

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.
- Primarily interested in predicting high concentrations correctly

Scatter plot for AQ forecast system validation



Another forecasting evaluation tool

MyAir Toolkit for Model Evaluation

- PASODOBLE was the Copernicus (GMES) downstream service project, producing **local-scale air quality services for Europe** under the name 'Myair' (<http://www.myair.eu/>)

- Local forecast developed, local air quality **Evaluation**

- The **Myair** free downlo

English

Username Login

Follow @Myair-Pasodoble on Twitter

Myair Pasodoble

Copernicus The European Earth Observation Programme

Home Objectives **Products & Services** Airsheds User Interface Quality Documents and Results

Health community support services

Public forecasting and assessment

Compliance monitoring support services

Local forecast model evaluation

Products & Services

More details on the Products and Services can be viewed using the links in the menu on the left.

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

Service Line	Service	European region / city / cities	Service provider*	Products
--------------	---------	---------------------------------	-------------------	----------

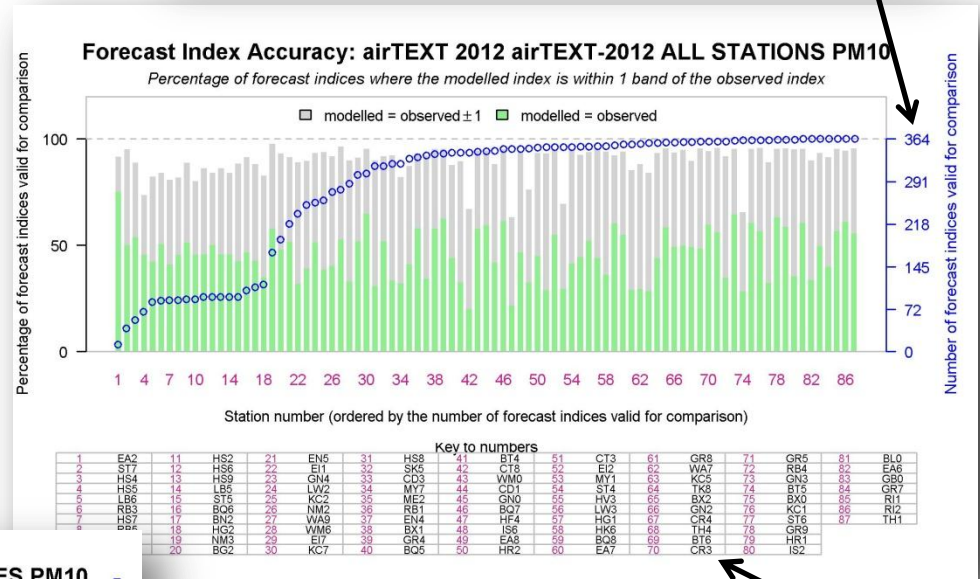
Suggestions for additional forecasting parameters/criteria (1 of 4)

Percentage of forecast indices ± 1 observations

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

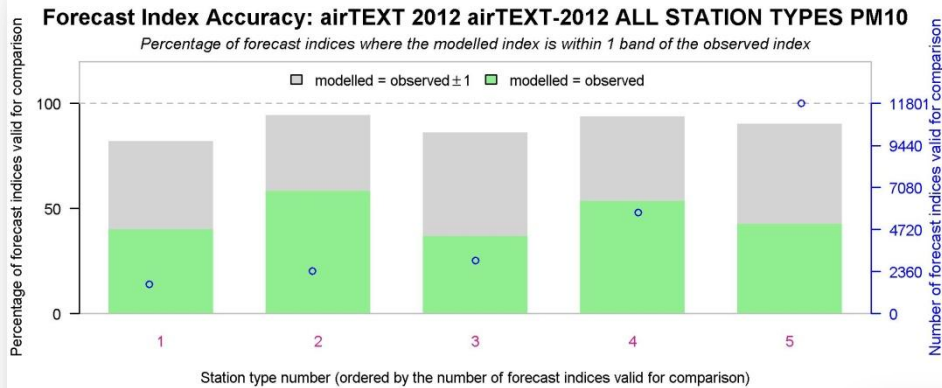
- modelled = observed ± 1
- modelled = observed

Number of forecast indices valid for comparison



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

INDUSTRIAL
SUBURBAN
KERBSIDE
URBAN BACKGROUND
ROADSIDE

Suggestions for additional forecasting parameters/criteria (2 of 4)

Model 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

$$\text{OddsRatioSkillScore(ORSS)} = \frac{ad - bc}{ad + bc}$$

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

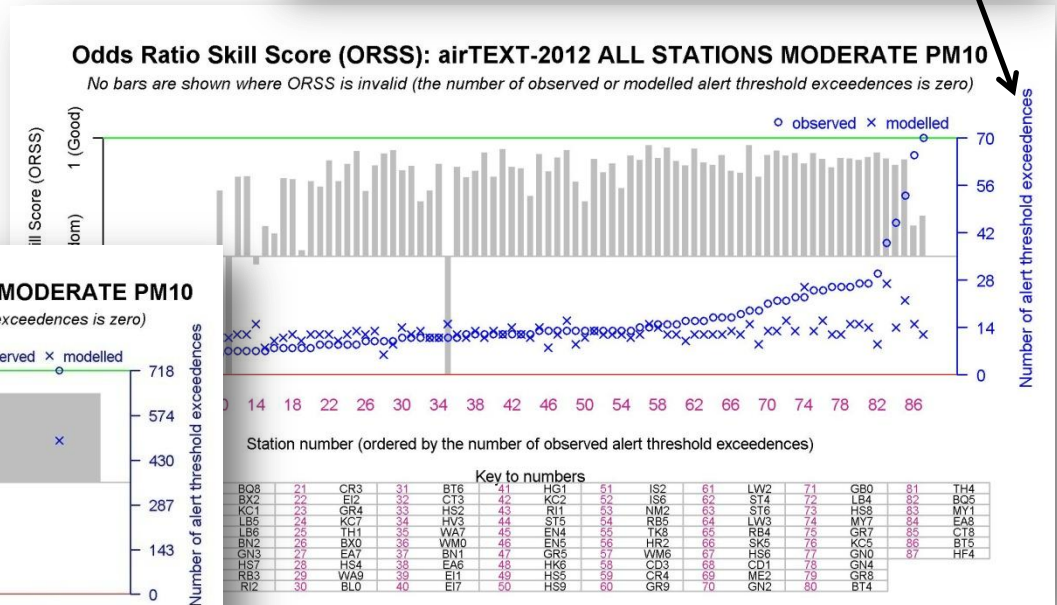
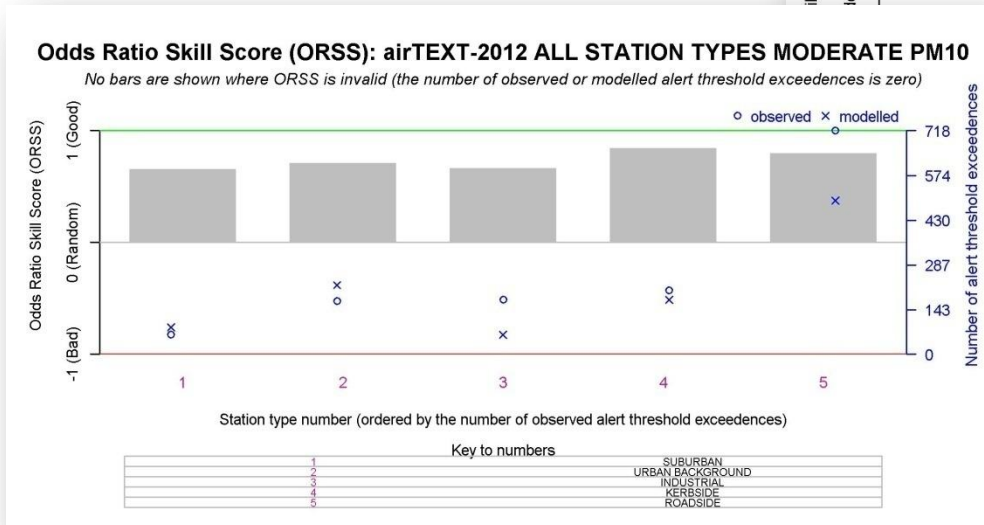
Perfect score:	b = c = 0	ORSS=1
Good score:	ad > bc	ORSS>0
Bad score:	bc > ad	ORSS<0
Fail score:	a = d = 0	ORSS=-1

Suggestions for additional forecasting parameters/criteria (3 of 4)

Model forecast skill

ORSS grouped by station...

Number of forecast indices valid for comparison



... 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

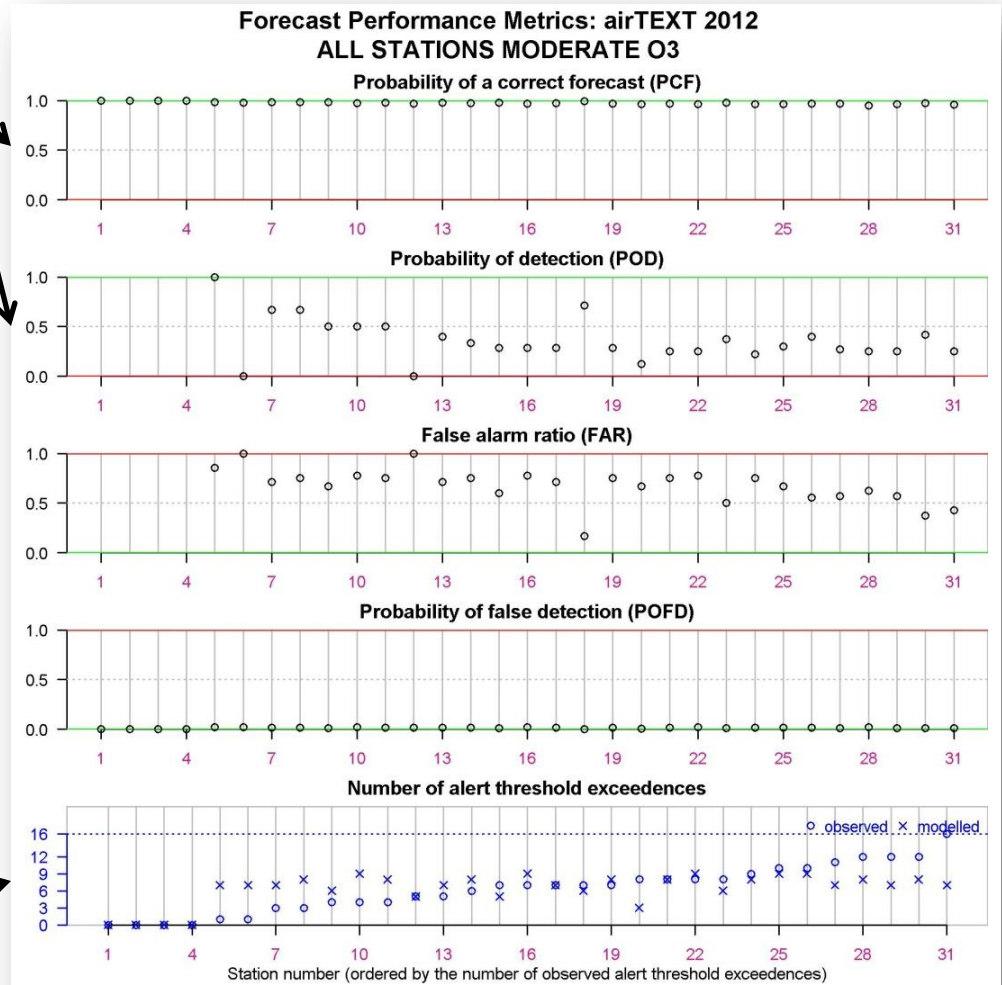
Suggestions for additional forecasting parameters/criteria (4 of 4)

Model 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



Key to numbers

1	CT6	11	TH4	21	GB6	31	TH1
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		

Summary

- There seem to be some issues with the formulation and/or the implementation of the forecasting Target plot
- There are forecasting-related statistics that could be calculated by DELTA that would help in the assessment of forecasting model output
- For additional information relating to the **MyAir Toolkit** functionality, refer to the Harmo presentation:
Stidworthy A, et al. 2013: Myair Toolkit for Model Evaluation. 15th International Conference on Harmonisation, Madrid, Spain, May 2013
To download the MyAir Toolkit:
<http://www.cerc.co.uk/environmental-software/myair-toolkit.html>