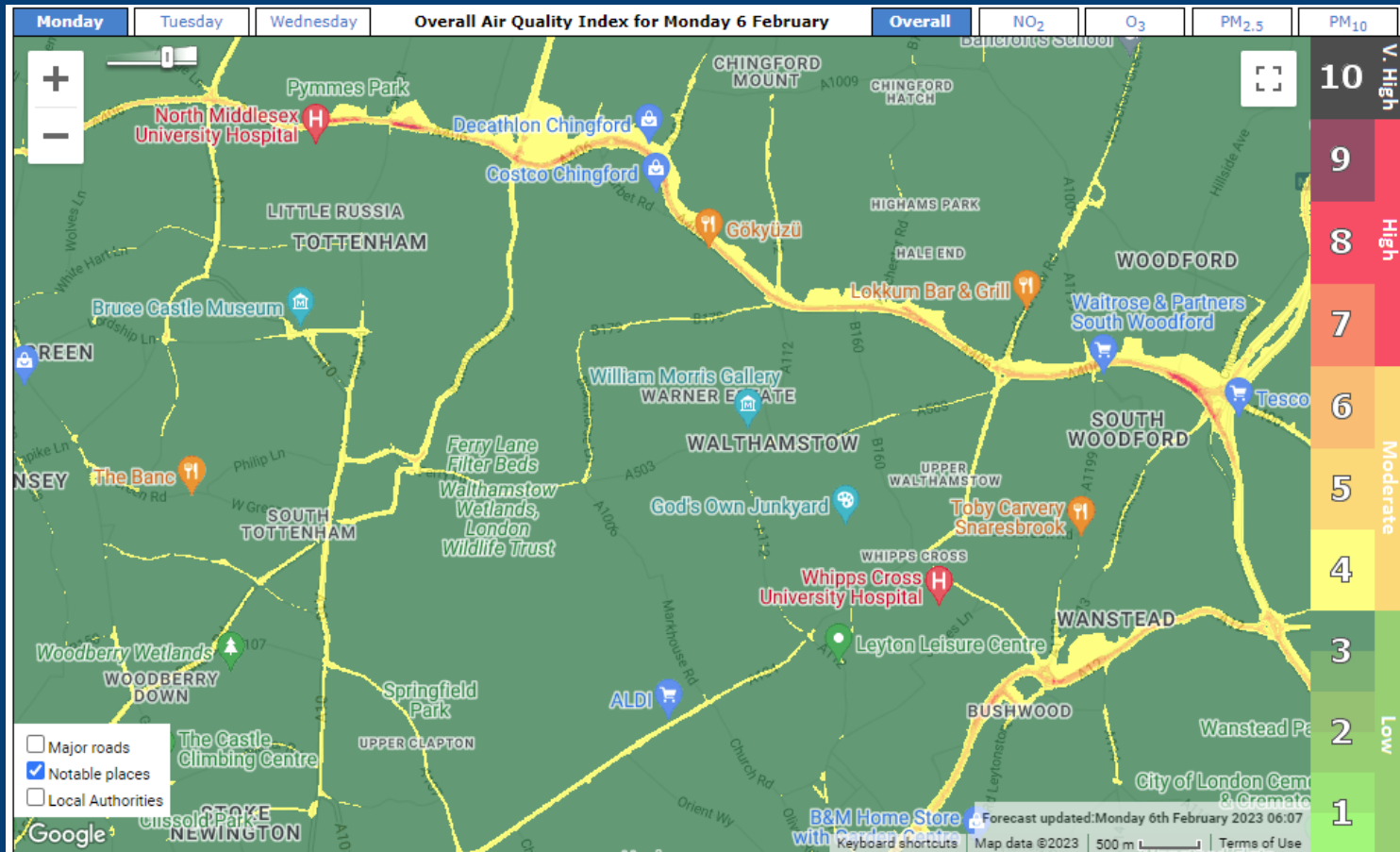# Delta v7.2 forecasting plots for London airTEXT

Kate Johnson, Amy Stidworthy, **Jenny Stocker**, Chris Johnson, Daniel Connolly

FAIRMODE

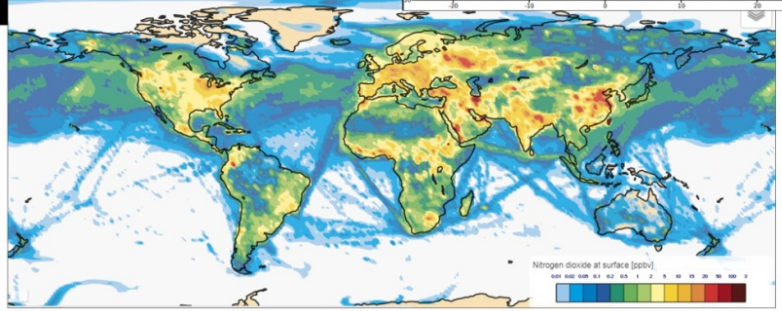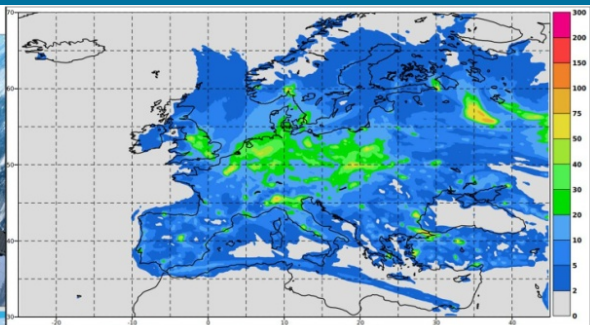Athens, 4-6th October 2023

CERC

Cambridge Environmental Research Consultants
Environmental Software and Services

CERC's London airTEXT System

ADMS-Urban
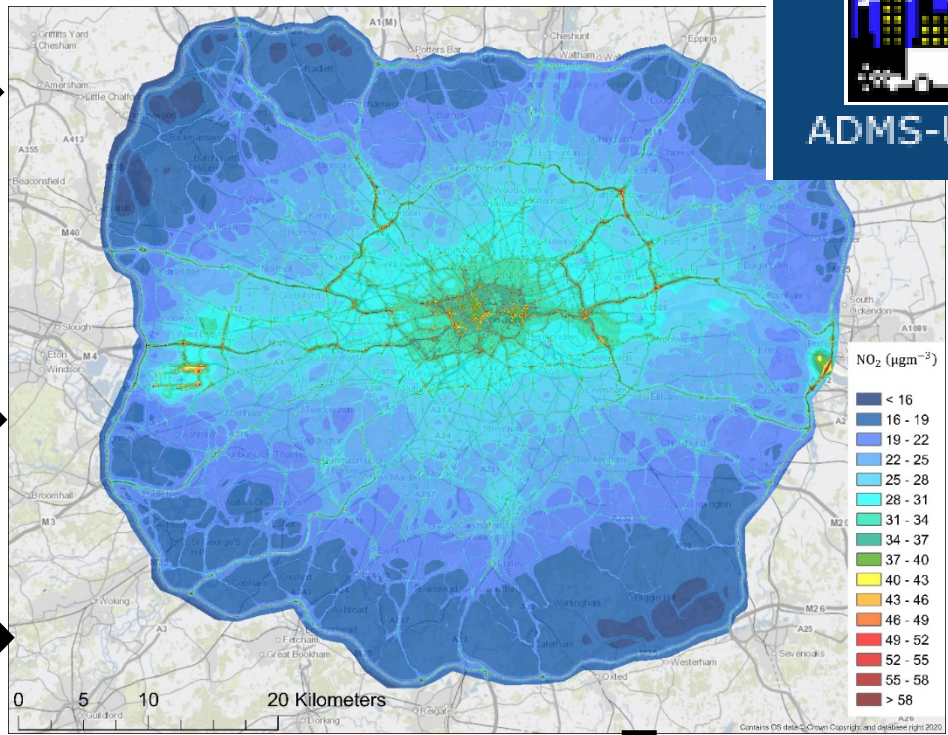
Copernicus Atmosphere Monitoring Service (CAMS)

Emissions Inventory (LAEI)

Measure-ments (AURN)

Weather Forecast (DTN)

NO$_2$ (μgm^{-3})

	< 16
	16 - 19
	19 - 22
	22 - 25
	25 - 28
	28 - 31
	31 - 34
	34 - 37
	37 - 40
	40 - 43
	43 - 46
	46 - 49
	49 - 52
	52 - 55
	55 - 58
	> 58

3-day pollution forecasts expressed using UK Daily Air Quality Index (DAQI)

NO$_2$ O$_3$ PM$_{10}$ PM$_{2.5}$

Recommended Actions and Health Advice

Air Pollution Banding	Value	Accompanying health messages for at-risk individuals*	Accompanying health messages for the general population
Low	1-3	**Enjoy** your usual outdoor activities.	**Enjoy** your usual outdoor activities.
Moderate	4-6	Adults and children with lung problems, and adults with heart problems, **who experience symptoms**, should **consider reducing** strenuous physical activity, particularly outdoors.	**Enjoy** your usual outdoor activities.
High	7-9	Adults and children with lung problems, and adults with heart problems, should **reduce** strenuous physical exertion, particularly outdoors, and particularly if they experience symptoms. People with asthma may find they need to use their reliever inhaler more often. Older people should also **reduce** physical exertion.	Anyone experiencing discomfort such as sore eyes, cough or sore throat should **consider reducing** activity, particularly outdoors.
Very High	10	Adults and children with lung problems, adults with heart problems, and older people, should **avoid** strenuous physical activity. People with asthma may find they need to use their reliever inhaler more often.	**Reduce** physical exertion, particularly outdoors, especially if you experience symptoms such as cough or sore throat.

airTEXT

Website | Email
Text message | Voicemail
Twitter | API

AirText London: evaluation for 2022 using Delta

Summary of sites & pollutants used to evaluate system performance:

Pollutants	Site types					
	Kerbside	Roadside	Industrial	Suburban	Urban background	**Total**
NO_2	7 (6)	28 (25)	3 (3)	8 (6)	10 (8)	**56 (48)**
O_3	1 (1)	7 (6)	0	4 (2)	3 (3)	**15 (12)**
PM_{10}	6 (5)	25 (17)	3 (2)	5 (2)	9 (7)	**48 (33)**
$PM_{2.5}$	2 (1)	10 (3)	1 (1)	4 (0)	5 (2)	**22 (7)**
Total	**16**	**70**	**7**	**21**	**27**	**141**

- Total number of sites (sites used in analysis*)

- Results presented for Day 1 forecast (previous analysis shown that Day 2 and 3 are better, relative to persistence)

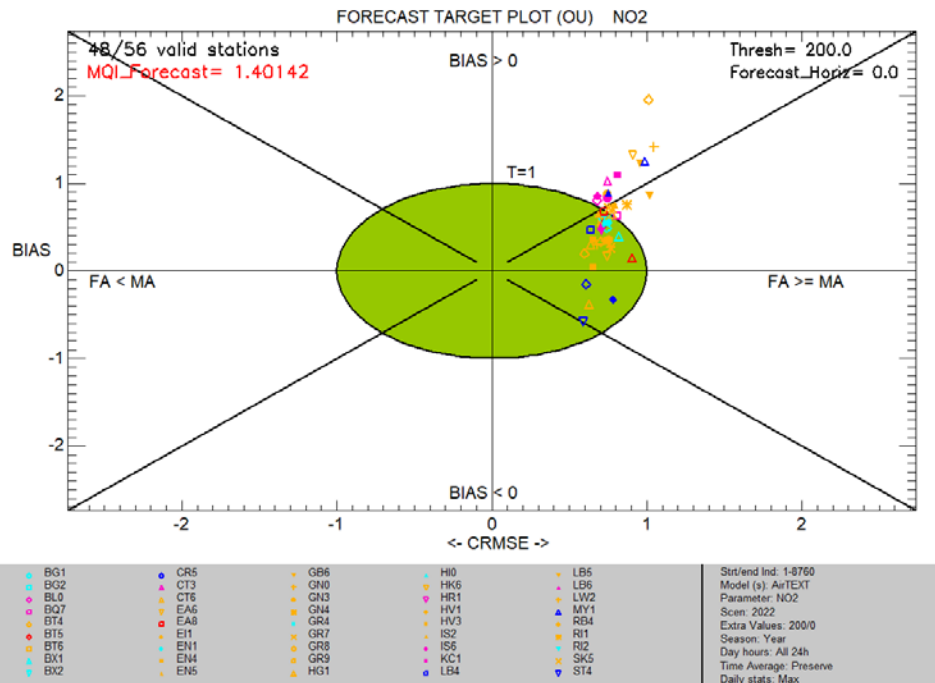*relates to to data capture requirement of > 75% for the relevant metric.

CERC

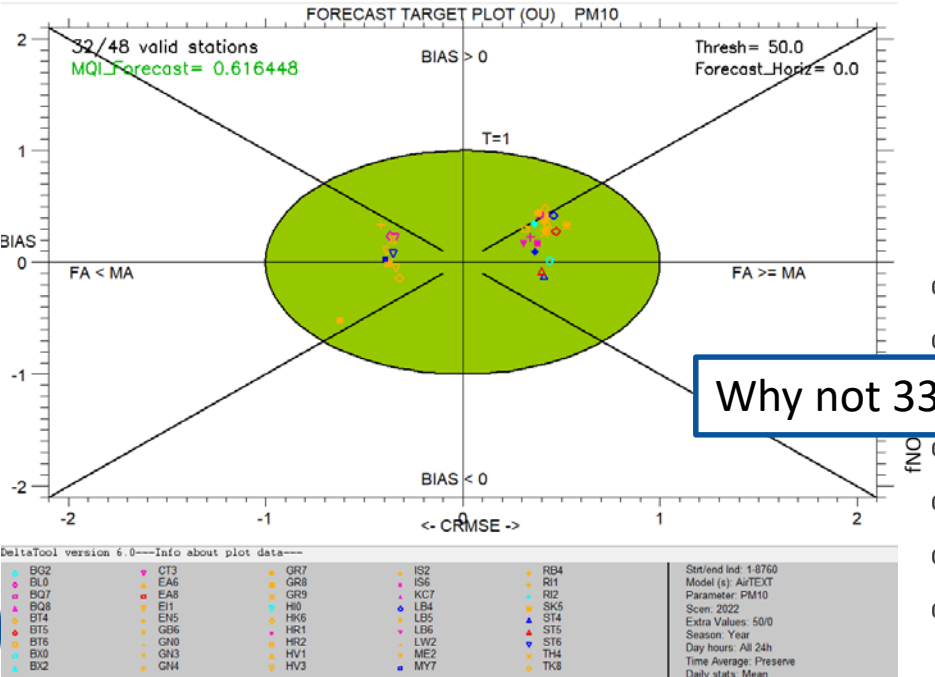# Forecast target plots

...are evaluating modelled concentrations

Description	In the Forecast target plot information is included on the following quantities (all normalized by the root mean squared error of the persistence model): • RMSE: distance from the origin to the point • BIAS: the bias can be either positive or negative and is represented along the vertical axis (Y) • CRMSE: The CRMSE is always positive and given by the distance from the origin to the point along the X axis. • False Alarm (FA) vs. Missed Alarm (MA): we use the FA/MA ratio to differentiate the right and left parts of the target diagram: $\frac{FA}{MA} < 1 \rightarrow \text{Left}$; $\frac{FA}{MA} \geq 1 \rightarrow \text{Right}$

MQO	Values lower than one (points within the green circle) indicate better capabilities than the persistent model whereas values larger than one indicate poorer performances. The MQI value corresponding to the 90th largest percentile is printed in the left upper corner and should be lower than 1. More details can be found in Concepts Section 6.

MQO are valid for the following parameters/ time statistic choices

Parameter	Time Avg	Daily Stats	Season	Day	Group
O3	8H	Max	Free	N/A	90%
NO2	Preserve	Max	Free	N/A	90%
PM10	Preserve	Mean	Free	N/A	90%
PM25	Preserve	Mean	Free	N/A	90%

	OBS	**MOD**	**PAR**	**SCEN**	Other		Single Mode	Group mode
Options	X	X			O-M		YES	Only 90% option

CERC

Forecast target plots

FORECAST TARGET PLOT (OU) NO2

48/56 valid stations
MQI_Forecast= 1.40142

FORECAST TARGET PLOT (OU) O3

11/15 valid stations
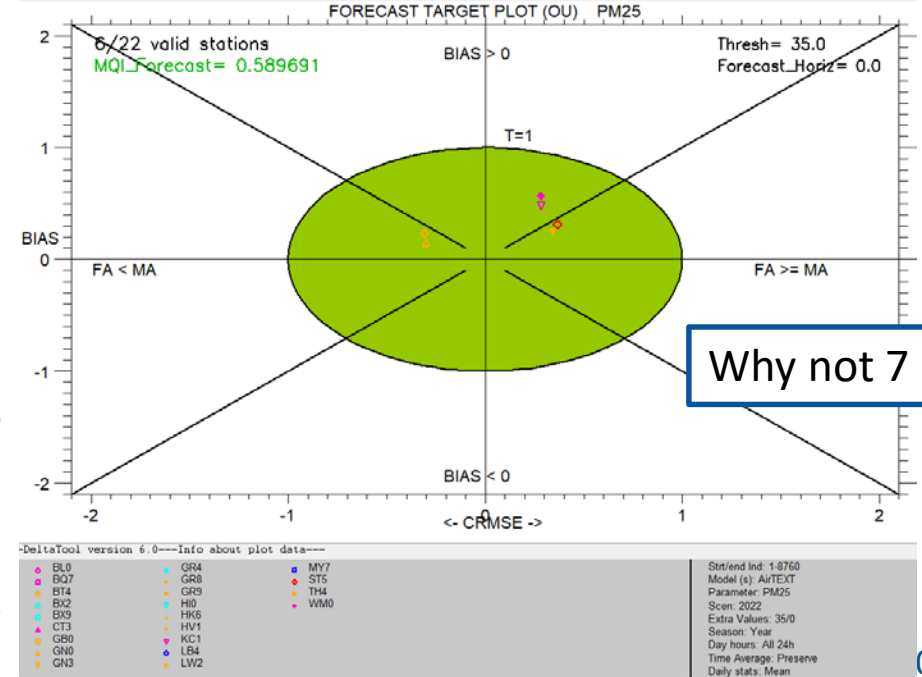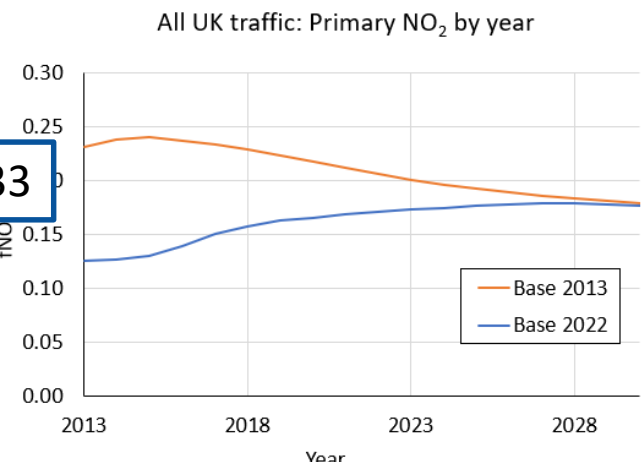MQI_Forecast= 0.585551

Why not 12

- This dataset: NO_2 target indicates over-prediction of hourly concentrations (emissions inventory currently being updated including revised primary NO_2), other pollutants fine
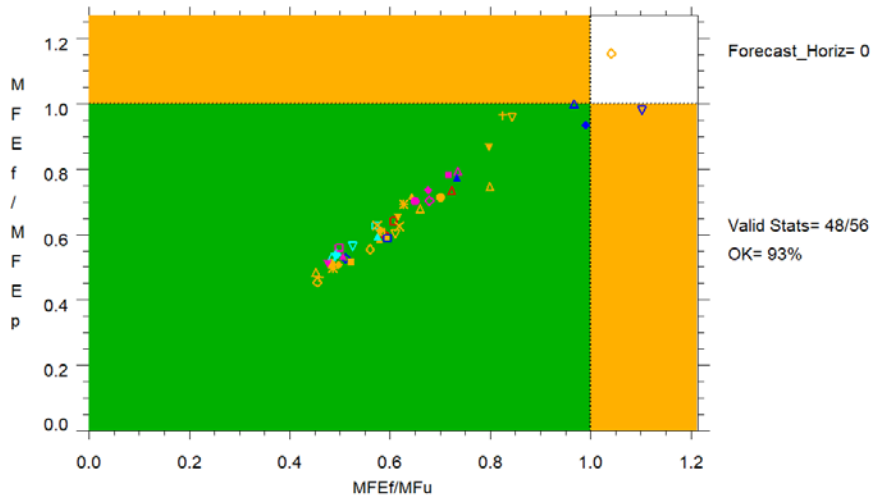

FORECAST TARGET PLOT (OU) PM10

32/48 valid stations
MQI_Forecast= 0.616448

Why not 33

All UK traffic: Primary NO_2 by year

Base 2013
Base 2022

FORECAST TARGET PLOT (OU) PM25

6/22 valid stations
MQI_Forecast= 0.589691

Why not 7

023
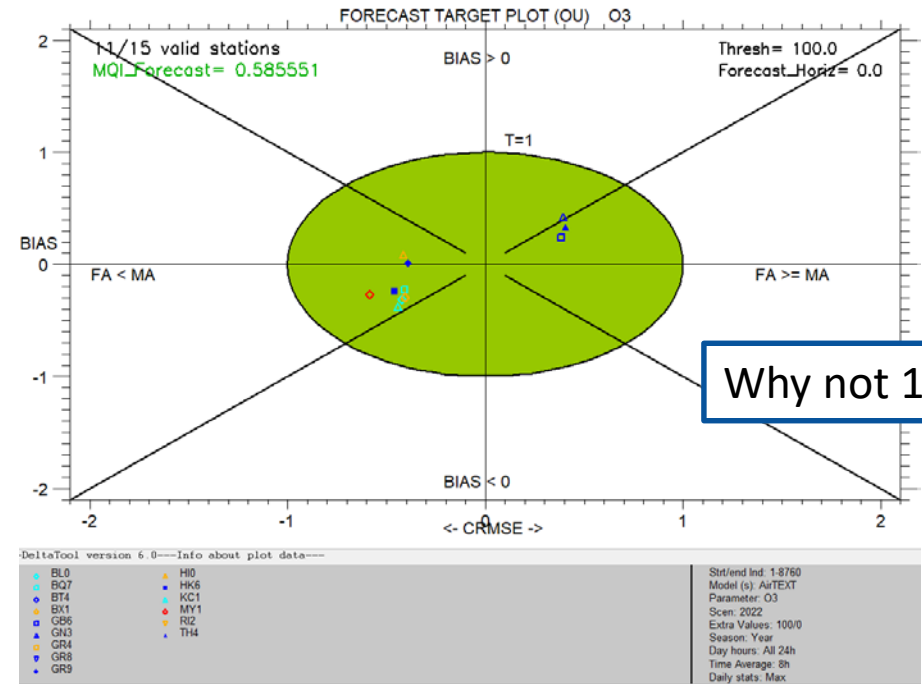
Forecast MPI plots

...are evaluating modelled concentrations

Description	Forecast MPI Plot shows the fulfilment of the MPCs defined in Section 6.2. Forecast performances (MFE_f) are compared to Mean Fractional Uncertainty (MF_U) along the X axis and to the persistence model performances (MFE_p) along Y axis.						

MQO	The green area identifies the area of fulfilment of both criteria. The orange areas indicate where only one of them is fulfilled.						

Analysis is valid for the following parameters/ time statistic choices

Parameter	Time Avg	Daily Stats	Season	Day	Group
O3	8H	Max	Free	N/A	90%
NO2	Preserve	Max	Free	N/A	90%
PM10	Preserve	Mean	Free	N/A	90%
PM25	Preserve	Mean	Free	N/A	90%

Options	OBS	MOD	PAR	SCEN	Other		Single Mode	Group mode
	X						YES	Only 90% option

CERC

Forecast MPI plots

Forecast_MPI Plot NO2

Forecast_Horiz= 0

Valid Stats= 48/56
OK= 93%

Forecast_MPI Plot O3

Forecast_Horiz= 0

Valid Stats= 11/15
OK= 100%

Why not 12

Forecast_MPI Plot PM10

Forecast_Horiz= 0

Valid Stats= 32/48
OK= 100%

Why not 33

Forecast_MPI Plot PM25

Forecast_Horiz= 0

Valid Stats= 6/22
OK= 100%

Why not 7

- Easy to understand that we want these metrics < 1. Plots should be square?

- Is it obvious why the majority of points are close to a straight line? i.e. that MFE_p scales linearly with Mf_u?

- For O_3 / $PM_{2.5}$, the error relative to persistence is lower / higher than relative to measurement uncertainty (gradient)

Forecast AQI plots

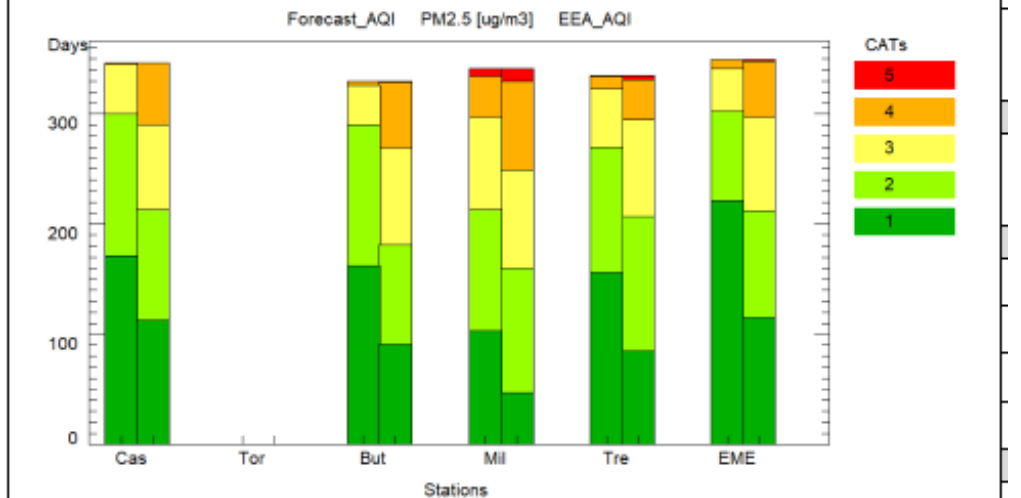*...are comparing threshold exceedances, all levels*

This plot is based on multiple thresholds as they appear in the Air Quality Categories and their Indices, like EEA, UK or US EPA indicators. In this diagram we compare the number of days that the forecast model (M) and the Measurements (O) have in common in each of the Air Quality Categories.
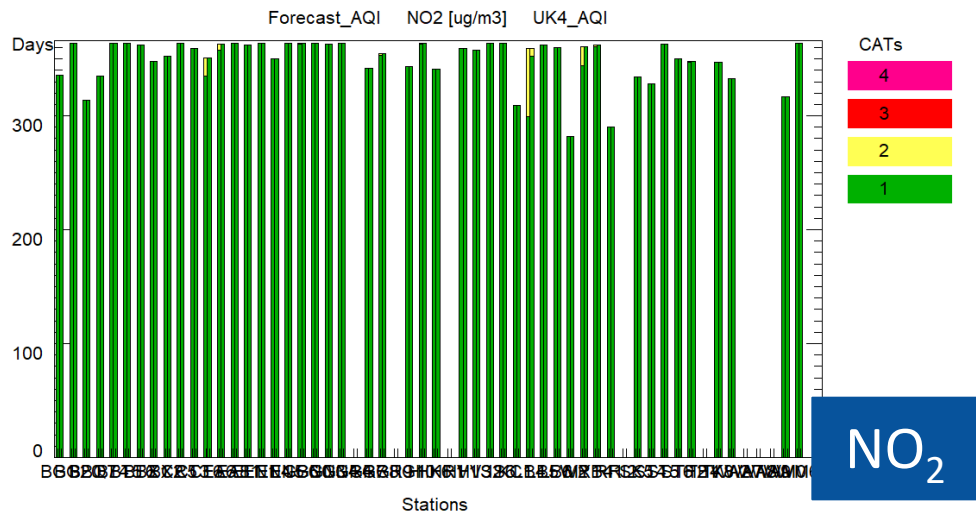
The Index table itself is shown in the grey area below, the corresponding colours on the right-hand side of the graphic.

In the current version of the DELTA Tool the following AQI tables are available: EEA (5 indices), UK4 (4 indices), UK10 (10 indices), USEPA (7 indices), and can be selected in the aqibounds.dat file in the DELTA tool configuration folder.

FORECAST DIAGRAMS USING AIR QUALITY INDICES

CERC

Forecast AQI plots (broad UK indices)

NO₂

O₃

PM₂.₅

- Some sites aren't appearing because of low data capture – do we need to include these? They are omitted from the Target and MPI plots (e.g. GR4, GR8 and RI2 for O₃)

- "SubBars M O" – please expand to say "SubBars Modelled Observed"

- Usual problem – too many sites for London means it is difficult to read plot (for London NO₂)

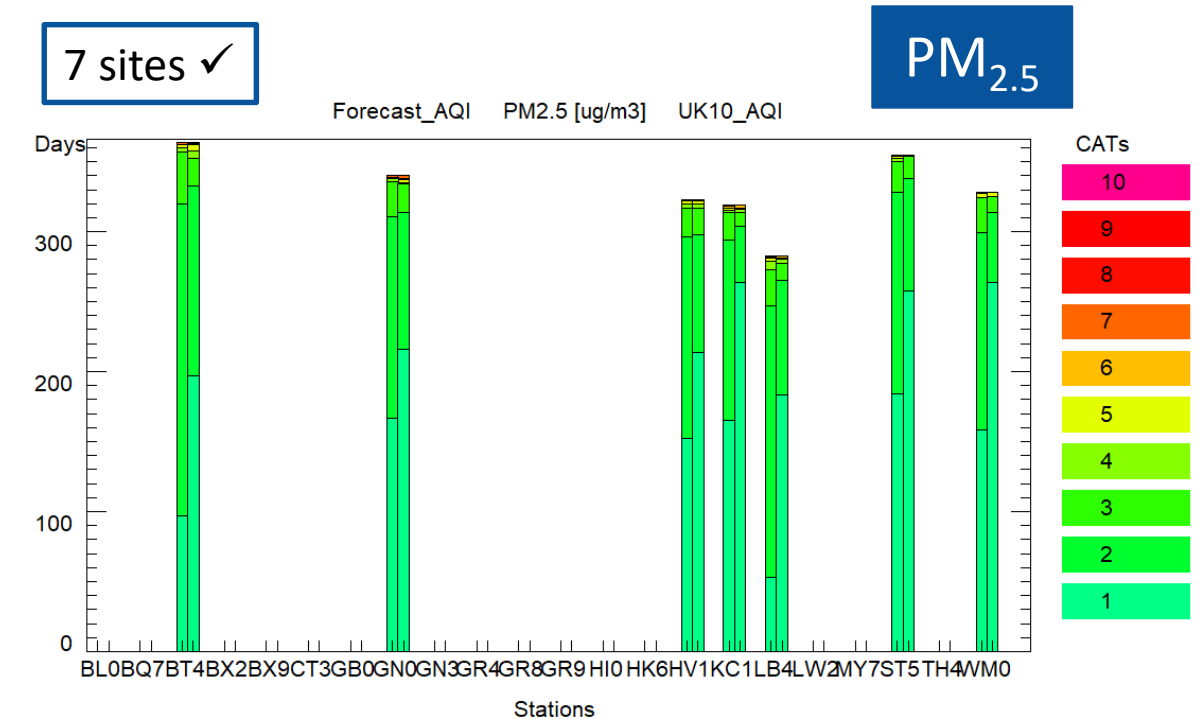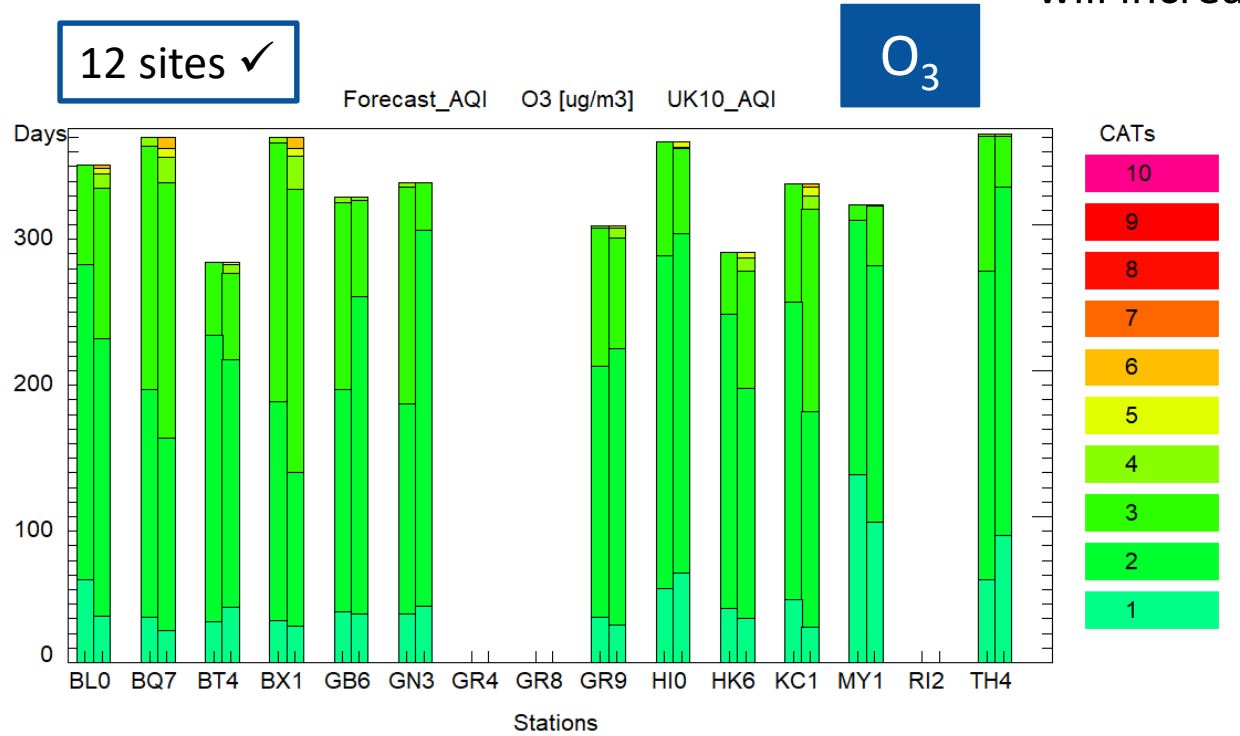- Is there any way of zooming in on the 'interesting bits'? Yes – can change to use more detail CATs….

CERC

2023

Forecast AQI plots (detailed UK indices)

- The user option to vary the number of thresholds is useful when thresholds are rarely exceeded (although the greens are quite similar!!)

- System misses the higher O_3 values – need to re-evaluate when revised traffic emissions are included (reducing NO_X emissions will increase O_3 concentrations)

12 sites ✓ **O_3**

7 sites ✓ **$PM_{2.5}$**

Bar plots for exceedance indicators

...are comparing exceedances of a specific threshold

Bar Plots show the values of the individual exceedances indicators (POD, SR, POD&SR, ACCURACY), together with their negative and positive sensitivities with respect to the threshold. The red bar indicates the change in the indicator when the threshold is reduced by 1 unit, the yellow bar the change when the threshold is increased by 1 unit.

- Probability of Detection: $POD = GA_+/(MA + GA_+)$

- Success Ratio: $SR = 1 -$ False Alarm Ratio: $1 - FAR = 1 - FA/(FA + GA_+) = GA_+/(FA + GA_+)$

- FBias score: $FBIAS = (GA_+ + FA) / (MA + GA_+)$

- Accuracy: $ACC = (GA_+ + GA_-) /$ Total

- Threat score: $TS = GA_+ / (MA + FA + GA_+) = GA_+/(FA + CA)$

- Gilbert Skill score: $GSS = (GA_+ - H_{random}) / (MA + FA + GA_+ - H_{random})$

  with $H_{random} = (GA_+ + MA)(GA_+ + FA) /$ Total

CERC

Bar plots

O₃, 100 µg/m³

Forecast_POD O3

Forecast_SR O3

12 sites ✓

Forecast_ACCURACY O3
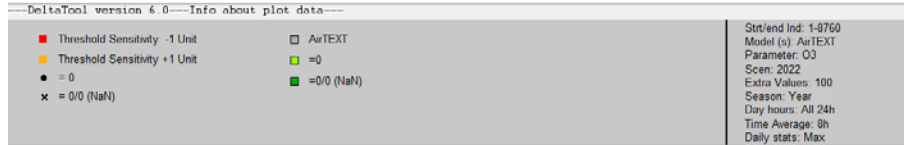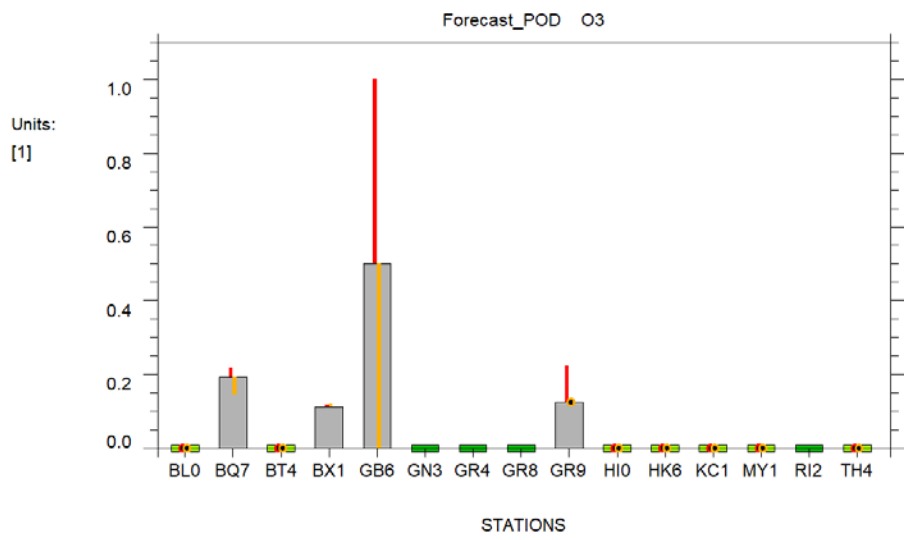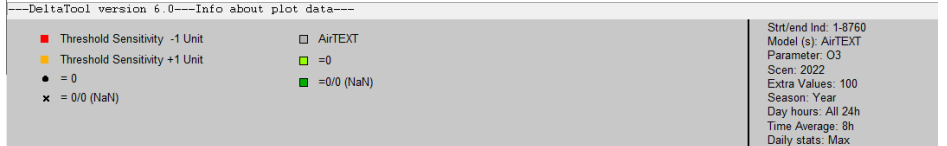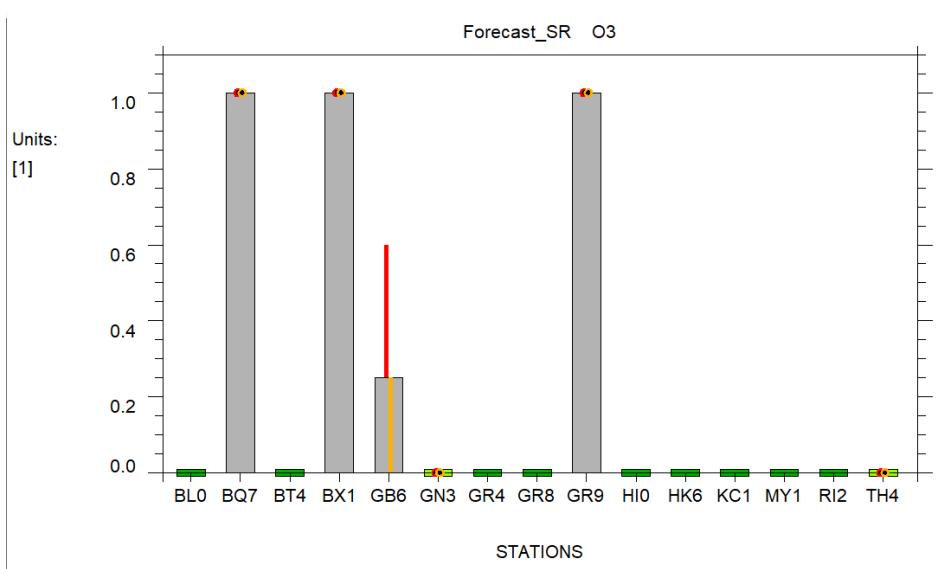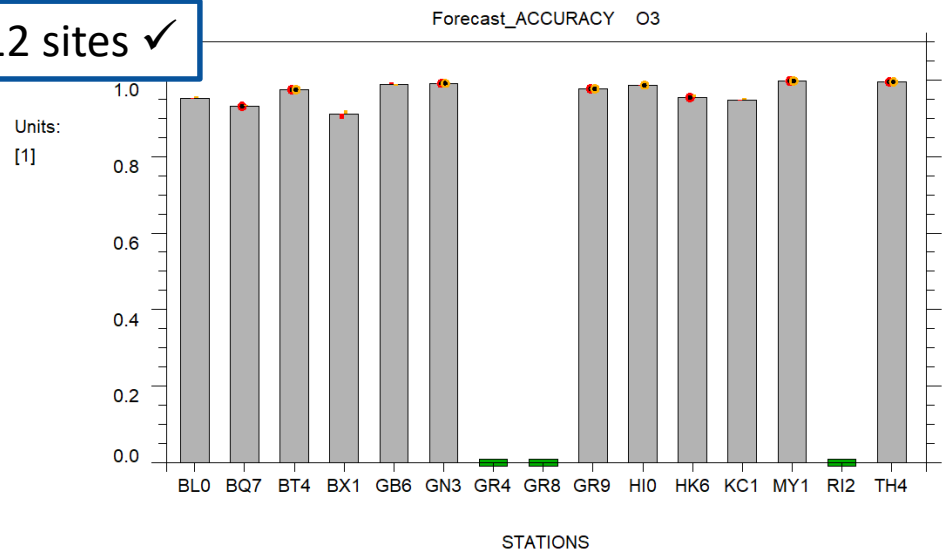
- Red/orange sensitivities:
  - Interesting how variable GA+, MA, FA for this small variation in threshold (given in .csv), but…
  - Should the step be related to measurement uncertainty ranges? Currently 1 µg/m³ - fairly arbitrary e.g. comparing across different pollutants

- Green always says 0/0 ~ NaN. Forecast AQI plots missing bars indicated low data capture, and is consistent with ACCURACY plot here. But POD has one extra, and SR many extra, 'green' sites. SR extras make sense (zero False Alarms). But no way of distinguishing between NaN sites and FA=0 sites

- Do we need to include the sites with poor data capture?

Forecast threshold performance

Description	The following indicators are included in this report: GA+ (counted exceedances), GA- (counted non-exceedances), FA (False Alarms), MA (Missed Alarms), ACC (accuracy), SR=1-FAR (Success Ratio), POD (Probability of Detection), FBIAS (FBias score), TS (Threat Score) and GSS (Gilbert Skill Score). See details in 6.3. A different graphical layout is applied depending on the number of stations taken into account in the analysis: if the number of stations is below 15, each of the dots represents a station for which the forecast indicators are evaluated (top of the Figure); if the number of stations is above or equal 15, boxplots are used to represent the statistical distribution of the indicators values (bottom of the Figure)

Description	The Forecast Threshold Performance Plot shows the four forecast indicators POD (Probability of Detection), SR (Success Ratio), FBIAS (FBias score) and TS (Threat score). It is based on the SR values on the X axis and POD values on the Y axis. Since FBIAS and TS are indicators related to POD and SR, they are represented by additional isolines. Good forecasts with a high POD and SR are situated in the upper right corner.

Forecast threshold performance – p normalised

MQO	In order to evaluate whether a forecast model is "good enough", indicator values obtained with the persistence model are used as a benchmark. Within the normalized version of the Forecast Summary Report the "good enough" zone is shaded in green and indicates that the model performs better than the persistence model for this particular indicator.

...are comparing exceedances of a specific threshold

CERC

Description	In order to indicate whether a forecast model is "good enough", in the Forecast Threshold Normalized Performance Plot the POD and SR values obtained with the persistence model are used as a benchmark (i.e. POD and SR of the forecast model are normalized with POD and SR of the persistence model).
MQO	In this plot, the green area represents forecasts with better POD and SR threshold indicators than the persistence model. In the white zone, the model performs worse than the persistence model on both indicators. In the orange zone, one of the two indicators is better than the benchmark. The normalized POD and SR values (i.e. POD/PODp and SR/SRp) are also given as indicative Modelling Performance Indicators. These MPIs correspond to the 10th largest percentile value and should be larger than one for a "good enough" forecast.

PM$_{2.5}$ forecast threshold performance

| Name	Obscode	Region	Type	lon	lat	alt	MO	MM	SO	SM	1-R	RMSE	GA+	GA-	FA	MA
BT4	LDN0006	roadside	Default	-999	-999	-999	12.797	15.425	7.713	7.135	0.891	4.38	7	351	0

Threshold	GAplus	GAminus	FA	MA	CA	SR	POD	ACC
34.0000	8.0000	349.0000	1.0000	5.0000	13.0000	0.8889	0.6154	0.9835
34.5000	8.0000	350.0000	0.0000	5.0000	13.0000	1.0000	0.6154	0.9862
35.0000	7.0000	351.0000	0.0000	5.0000	12.0000	1.0000	0.5833	0.9862
35.5000	7.0000	351.0000	0.0000	5.0000	12.0000	1.0000	0.5833	0.9862
36.0000	7.0000	352.0000	0.0000	4.0000	11.0000	1.0000	0.6364	0.9890

PM$_{2.5}$, 35 µg/m³
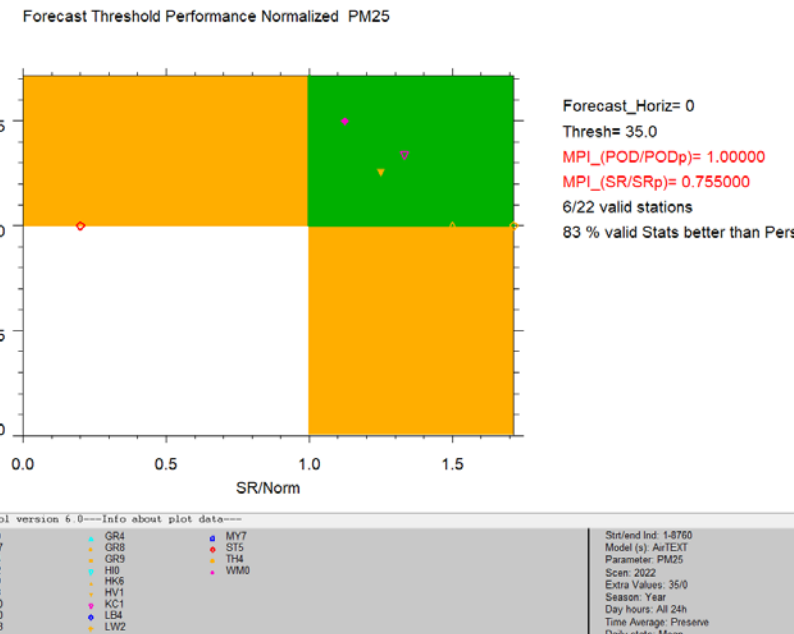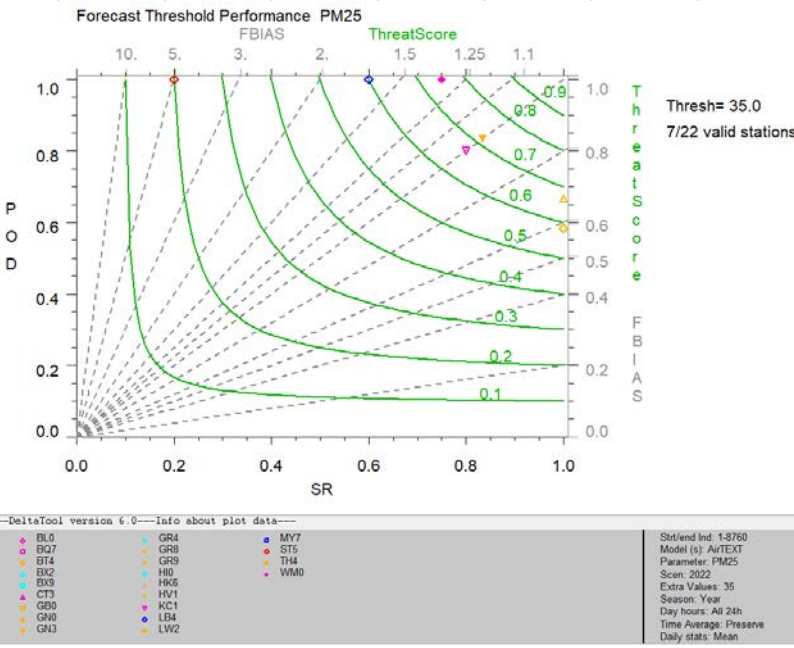
- Why are there a different number of valid stations for Summary stats & "Forecast Threshold Performance" compared to "Forecast Normalised" 6/22 vs 7/22 here (similar discrepancy for Target, 6, and MQI, 7)

- Rather than "valid", could we say "low data capture" or "usable"?

- Forecast threshold performance plots should be square? (they are in the User Guide)

Why not 7
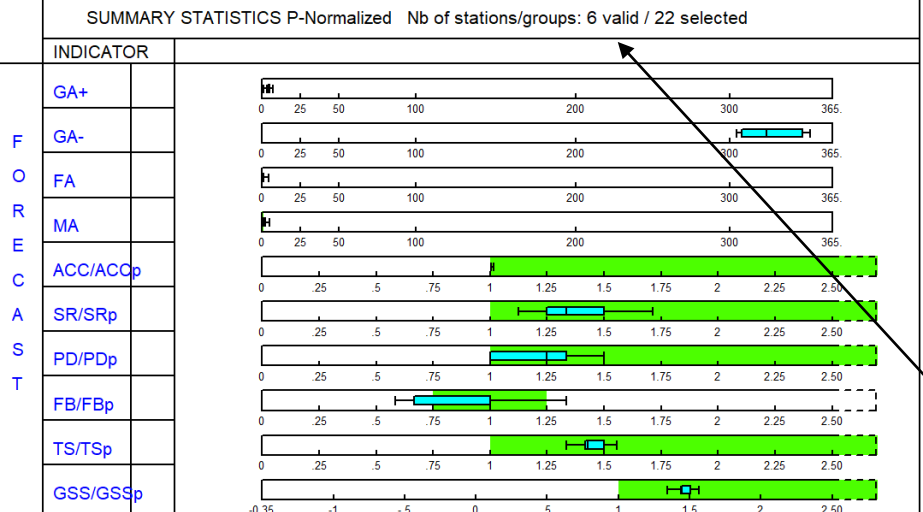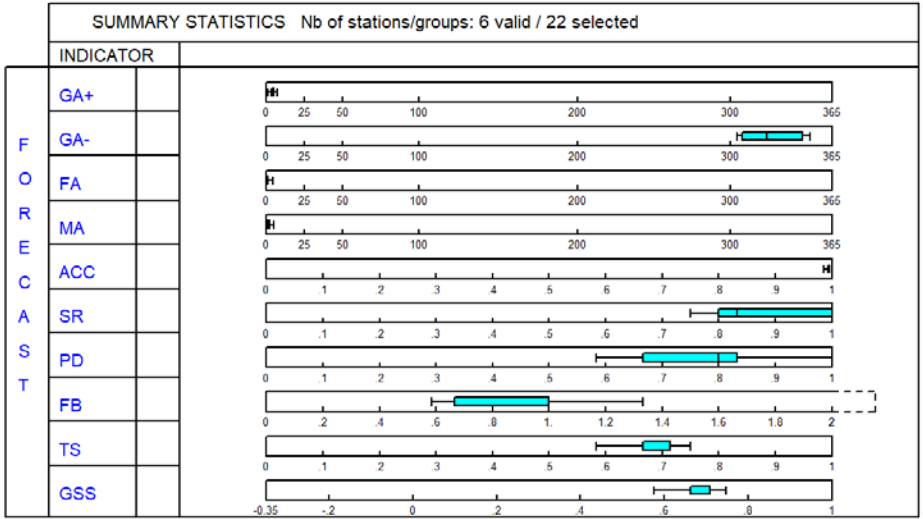
Summary

- Forecast tool generates a useful set of plots and statistics that allow analysis of forecasting system performance

- Plots can be categorised:

  - Over all evaluation (full range of concentrations, similar analyses to Delta Assessment)

  - Evaluation of threshold exceedances (all thresholds)

  - Detailed assessment of threshold exceedances (single thresholds)

Concept queries:

- In the MPI plots, Is it obvious why the majority of points are close to a straight line? i.e. that MFEp scales linearly with Mfu?

- Should we be excluding low data capture sites from all plots?

- Is there an inconsistency between the number of sites included on the different plots – highlighted throughout the presentation

- Should a the sensitivity range on the bar plots relate to observation uncertainty rather than the arbitrary 1 μg/m³ currently used

Presentation comments:

- Some plots should be square e.g. forecast MPI, forecast threshold performance

- More helpful to state "low data capture" (if space) or "usable" rather than "valid"

- On forecast MQI plots– please expand "SubBars M O" to say "SubBars Modelled Observed" if room

CERC

Any questions?

Jenny.Stocker@cerc.co.uk

CERC