



Regional air quality forecast verification at the Met Office

Paul Agnew

Oxford 30th September



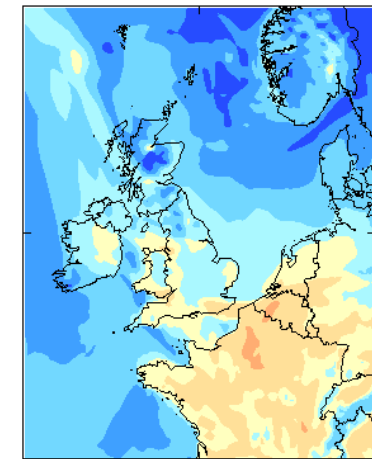
Verification: Some questions to begin with....

- What is your modelling system trying to do?
 - Forecast hourly values, evaluate annual means, exhibit correct spatial variations....etc...?
- What constitutes a good forecast/modelling outcome?
 - E.g. predict exceedance of an air quality threshold
- How reliable are the measured values?
 - Quality of measurements, representivity
- How to achieve like-for-like comparison?

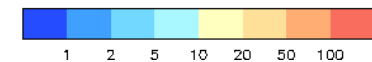
AQUM

- AQUM is an off-line, regional atmospheric composition model
- Uses UKCA gas phase chemistry
- Currently using CLASSIC aerosol scheme (will migrate to UKCA-MODE)
- Run once per day
- Hourly forecasts to T+48
- MACC chem/aer LBCs
- Daily verification

PM10 ($\mu\text{g}/\text{m}^3$) 28-09-2010 01:00
From 00Z on 20100928

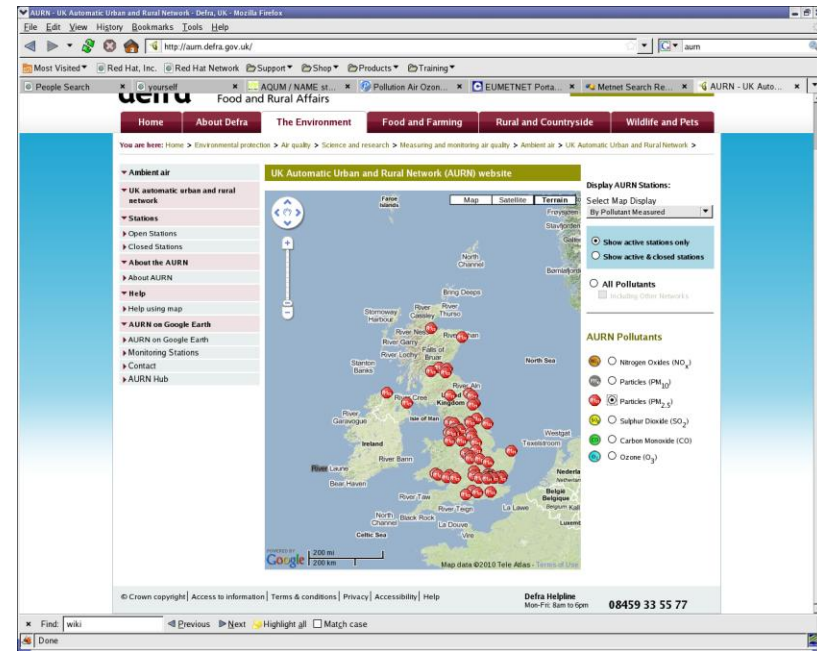


© Crown Copyright 2010. Source: Met Office



Air Quality Measurements: PM

- Good news:
 - A substantial surface network of hourly measurements across much of Europe
 - Many measurements available in 'near-real-time' (thanks to GEMS)
 - Site locations classified (rural, urban b/g, kerbside etc.)
- Not so good news.....
 - NRT data not quality controlled
 - Low precision
 - PM not speciated





Air Quality Metrics: address two aspects of model 'skill'

- Verification of forecasts c.f. observations
 - (N)RMS error, bias and correlation take into account all forecasts and observations, across the range of values
 - Sensitive to model resolution: 'smoother' (low-res) models can have better scores than hi-res models overall
 - Susceptible to 'double penalty' problem
 - but may under-forecast exceedance events
- Skill scores focussed on threshold exceedance events



Metrics

Traditional measures

- Normalised mean bias

$$B_n = \frac{1}{N} \sum_i \frac{(f_i - o_i)}{o_i}$$

- Normalised rmse

- Asymmetry problem:
limited to -1 for under-
prediction; unlimited for
over prediction

$$E_n = \left(\frac{1}{N} \sum_i \left(\frac{f_i - o_i}{o_i} \right)^2 \right)^{1/2}$$



Metrics : employed for operational system at MetO

Employ measures which are fair to both under/over prediction

$$B'_n = \frac{2}{N} \sum_i \left(\frac{f_i - o_i}{f_i + o_i} \right)$$

- Modified mean bias
- Fractional gross error

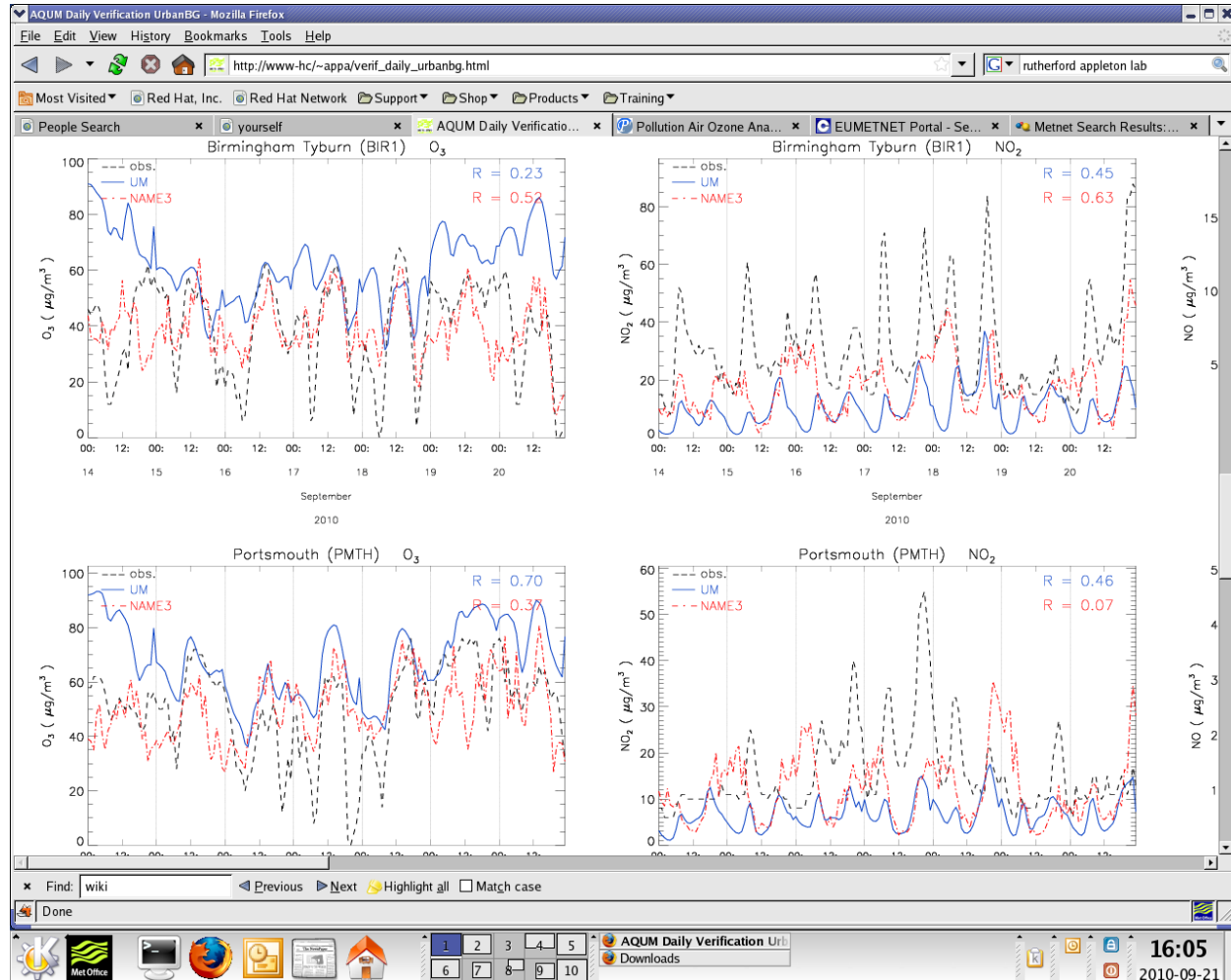
$$E_f = \frac{2}{N} \sum_i \left| \frac{f_i - o_i}{f_i + o_i} \right|$$

- Correlation used to indicate match between obs/forecast patterns

$$R = \frac{\frac{1}{N} \sum_i (f_i - \bar{f})(o_i - \bar{o})}{\sigma_f \sigma_o}$$

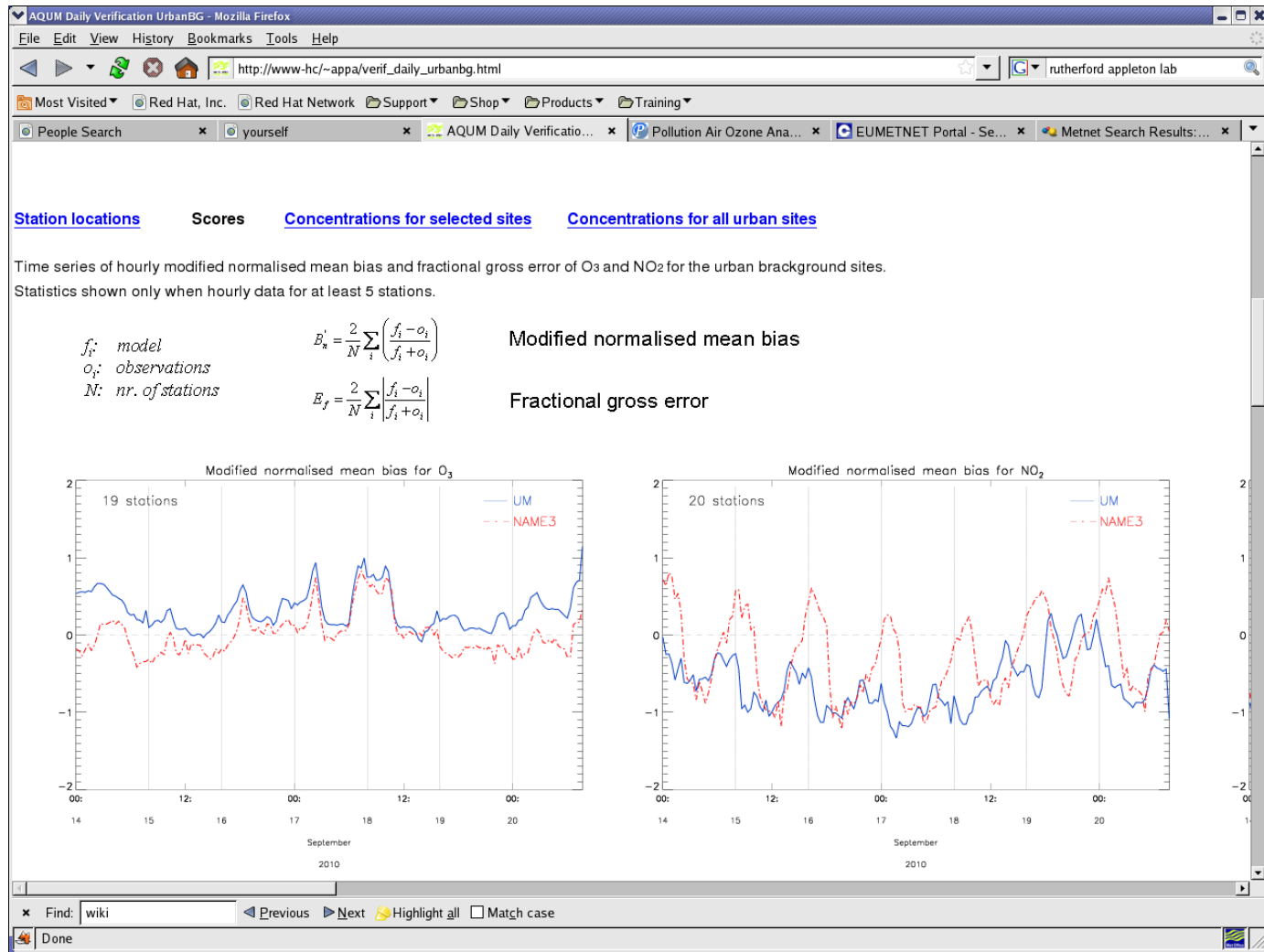


Met Office operational system: site time series



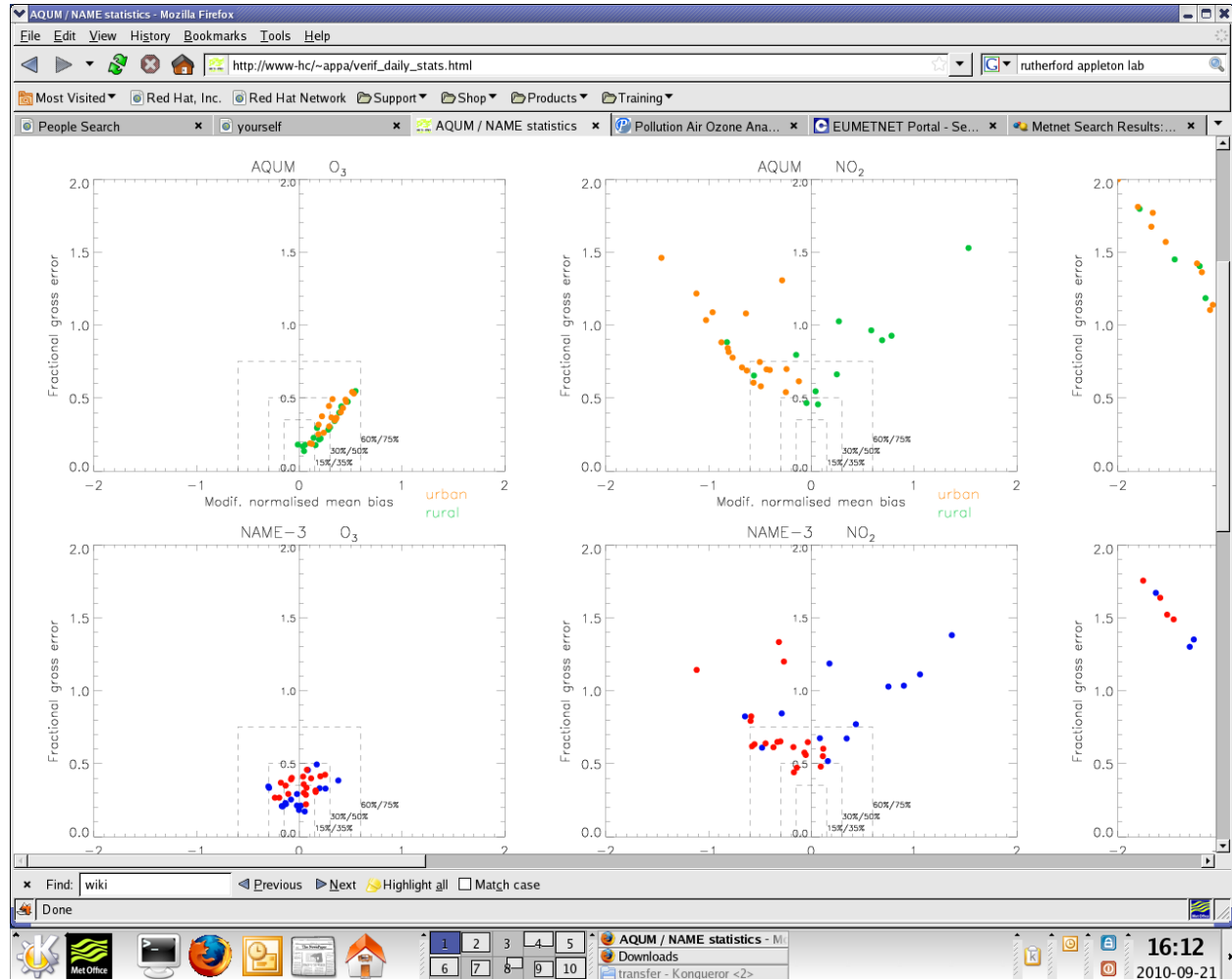


Met Office operational system: summary statistics



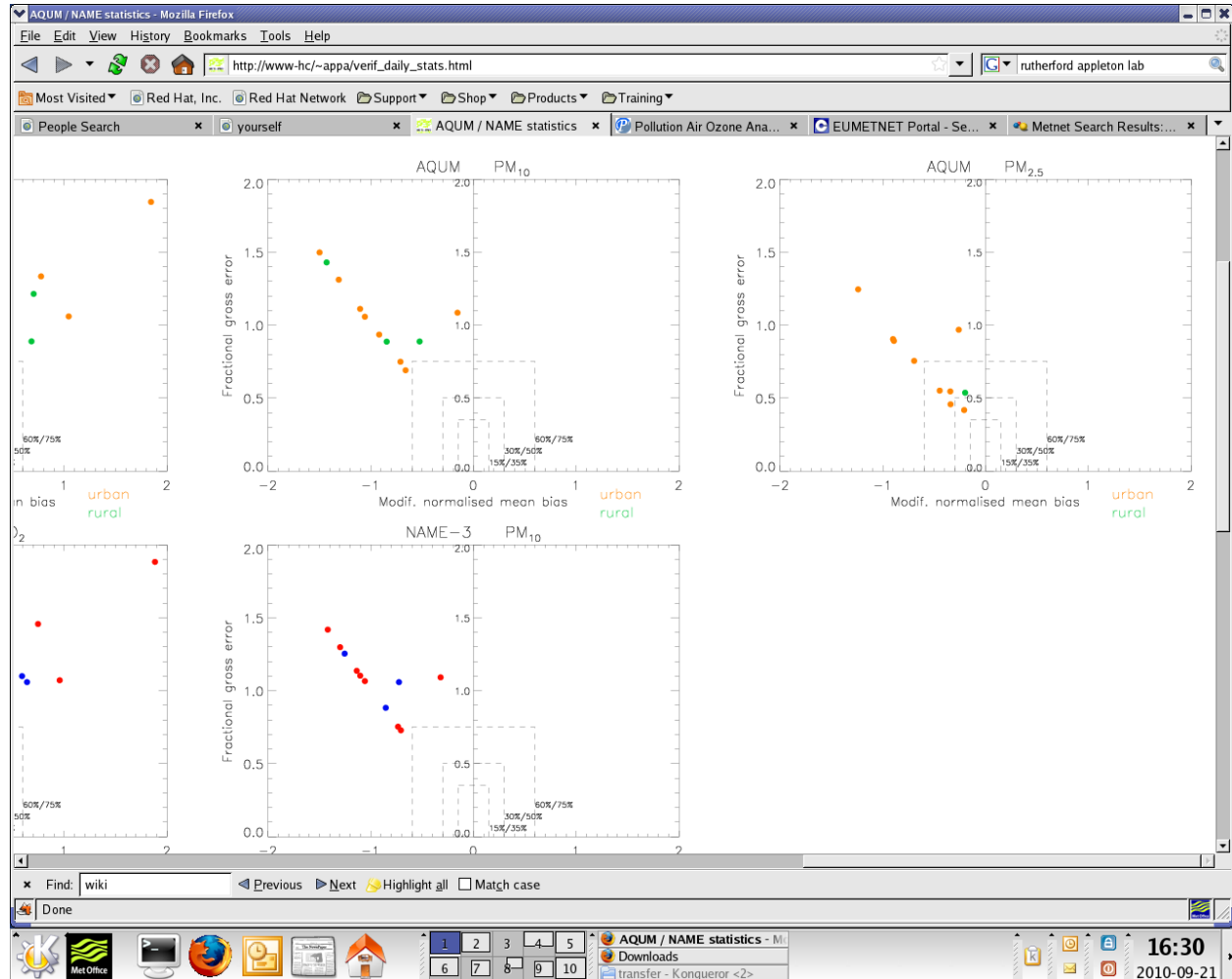


Summary statistics: 'soccer' plots





Soccer plots: PM





Exceedance forecasting: AQ 'Objectives'

- SO₂
 - 1hr mean of 350 μ g/m³ not exceeded more than 24 times per year
 - 24hr mean of 125 μ g/m³ not more than 3 times/yr
- NO₂
 - 1hr mean of 200 μ g/m³ not exceeded more than 18 times per year
- PM₁₀
 - 24hr mean of 50 μ g/m³ not exceeded more than 35 times per year
- Ozone
 - 24hr max of 8hr running mean not to exceed 100 μ g/m³ more than 10 times per year



Exceedance events

- Requirement: a single statistic indicating the relative skill of each model in forecasting threshold exceedences
- Basis: 2x2 contingency table
 - a – Hit
 - b – False alarm
 - c – Miss
 - d – Correct rejection
 - $n=a+b+c+d$ total no. events

		Events	Observed
		Yes	No
Events	Yes	a	b
Forecast	No	c	d



Properties of a useful skill score:

- A range of indicators traditionally developed for meteorological forecasts:
 - Proportion Correct, Heidke Skill Score, Gilbert SS, Peirce (Kuipers) SS etc.
- Require a Skill Score which is:
 - Simple to calculate and interpret
 - Not sensitive to the thresholds chosen
 - Not sensitive to the 'base rate'
 - Robust – not easily 'hedged'
 - Can be tested for significance if required
- The 'Odds Ratio' meets these requirements



Odds ratio skill score

- Odds ratio: In terms of contingency table, $OR=ad/bc$
- A skill score can be derived by a simple transformation:
 - $ORSS=(OR-1)/(OR+1)$
 - This mapping produces a skill score in the range -1 to +1
- When $ORSS=-1$ forecasts and observations are independent
- *Providing number of forecasts is statistically significant, ORSS approaching +1 indicates a skillful forecast*
- A test exists for the statistical significance

Exceedance forecast skill

Verification statistics - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.nwp/~apdg/AirQuality/ObsDailyStats.html

Most Visited Red Hat, Inc. Red Hat Network Support Shop Products Training

People Search [Met-jobs] PD... AQUM Daily V... yourself Oxford Physic... Capita Busines... Login Verification sta...

Verification statistics

Modified: 23 March 2010 by Lucy Davis

For statistics for just the previous week, see [here](#).

For statistics for the period 23/06/2009-22/03/2010, see [here](#).

Overall Pollution Index:

	NetCen	With AEA	Without AEA	Coarse AQ run	airq-d - Without AEA	airq-d - Coarse AQ run	AQUM**	Persisted Obs 24hr	Persisted Obs 48hr
Number of correct predictions:	2374	2129	2519	599	2542	1794	1961	2994	2655
Number of under predictions:	353	389	1021	39	835	251	177	735	880
Number of over predictions:	1789	2046	1024	3926	1187	2519	2095	722	879
Percentage of correct predictions:	52.6	46.6	55.2	13.1	55.7	39.3	46.3	67.3	60.1
Percentage of under predictions:	7.8	8.5	22.4	0.9	18.3	5.5	4.2	16.5	19.9
Percentage of over predictions:	39.6	44.8	22.4	86.0	26.0	55.2	49.5	16.2	19.9
Bias:	0.3470	0.4191	0.0177	1.4380	0.0819	0.6122	0.5200	-0.0036	0.0000
RMSE:	0.7898	0.8781	0.7752	1.7669	0.7519	1.0092	0.8828	0.6416	0.7514
ORSS* (exceedance index= 4) :	0.7116	0.7253	0.7668	0.7218	0.5405	0.5046	0.8061	0.8503	0.5597
Hit rate* (index= 4) :	0.3517	0.4828	0.3172	0.9241	0.1552	0.4897	0.6275	0.3382	0.1730
False alarm rate* (index= 4) :	0.0838	0.1294	0.0578	0.6631	0.0519	0.2401	0.1531	0.0397	0.0558

Comparison dates: 20100305 - 20100927 inclusive

* For contingency tables used to generate this statistic can be found [here](#)

** Until 26/04/2010: based on forecast on the previous day. (Except 20100303,20100304,20100305,20100329,20100330 which are based on the forecast for the present day). From 27/04/2010: based on forecast for the present day.

O3 Pollution Index:

Done

© Crown copyright

14:52
2010-09-28



Key elements of air quality forecast verification

- Metrics which evaluate standard field stats AND exceedance skill
- Stratify evaluations according to site type
- Baseline comparison provided by persistence forecast
- Value of innovative visualisation (Taylor, Soccer plots etc....)
- PM verification should ideally examine speciated components (to be done.....)