ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

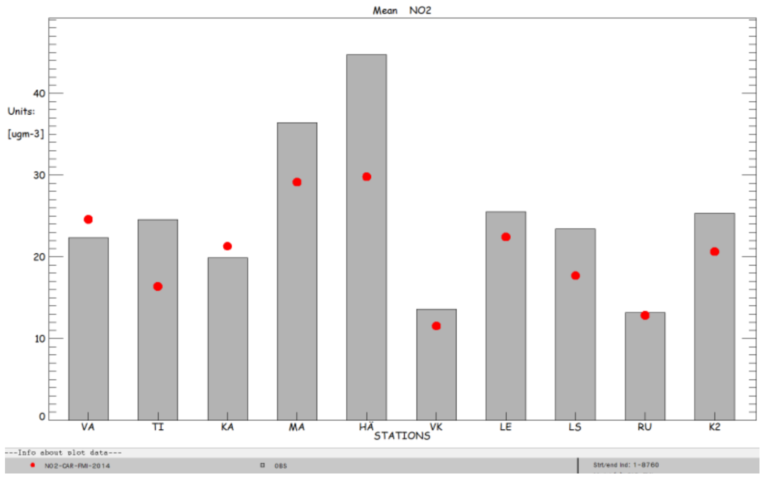
Some notes on Delta forecast benchmarking

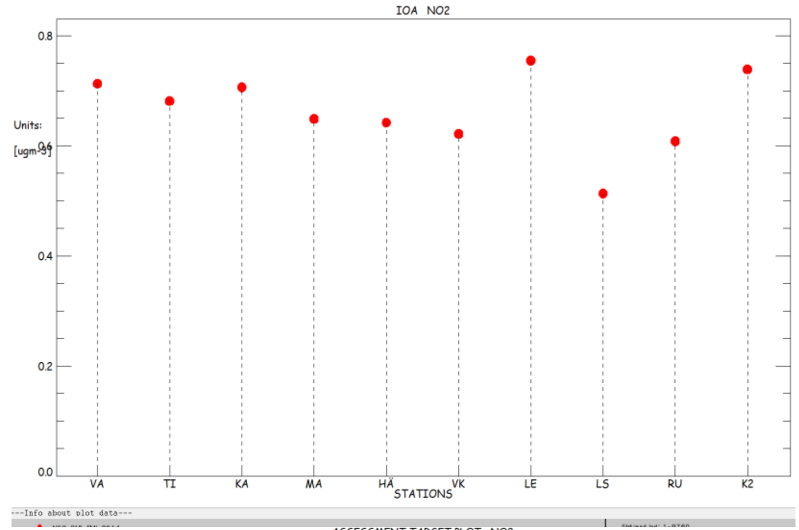
Ari Karppinen

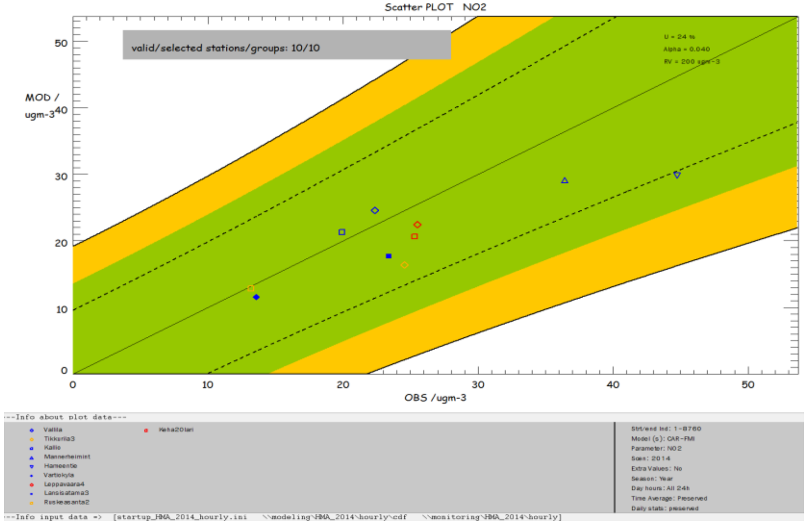
Research Manager

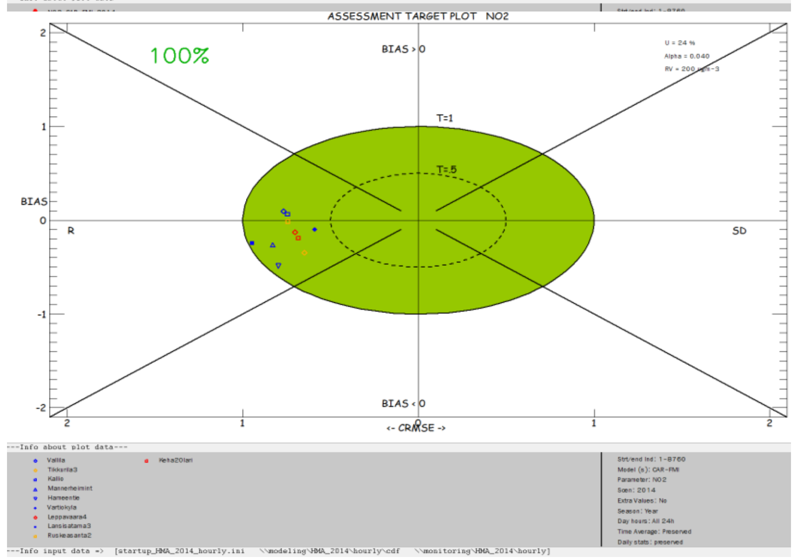
Finnish Meteorological Institute

Helsinki area: ~10 stations / ~10 years









Normalisation

$$\text{Target}_{\text{forecast}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(M_i^* - O_i\right)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(O_{i-j} - O_i\right)^2}} \qquad (1)$$

Can this ever been comparabale without constant/fixed (i-j) ?

- Do e.g. longer or ”unsyncronized-length” forecasts (<->reference day j) automatically produce better target values

Observation uncertainty

$$\text{if } M_t < O_t \text{ then } M_t^* = \min(M_t + OU * O_t, O_t)$$

$$\text{if } M_t \geq O_t \text{ then } M_t^* = \max(M_t - OU * O_t, O_t)$$

This transfrom seems to have an interesting effect in forcing *substantial (?)* amount of model results to be exactly the observed value

..is this a problem for the delta –statistics calculations.?
. The statistical distribution of M* certainly completly different than the original distribution of M?

Some comments on the diagrams

$$\text{If} \quad \frac{FA}{MA} \leq 1 \Rightarrow \text{Left}$$

$$\text{If} \quad \frac{FA}{MA} > 1 \Rightarrow \text{Right}$$

this is effective use of the diagram-space, but I am slightly worried about the discontinuity(?) it creates at FA=MA

Furthermore: I am not so convinced if the FA<>MA relation gives any strong indication on the model skill, although it may make sense to ”hope” more False alarms than missed alarms

Just speculation: giving strong priority to this requirement, might force the modellers to add some ”random” coefficients to model just to make sure that FA>>MA..

Diagrams cntd…

$$\frac{GA_+}{FA + MA + GA_+} < 0.2 \Rightarrow \text{Red}$$

$$0.2 \leq \frac{GA_+}{FA + MA + GA_+} < 0.4 \Rightarrow \text{Orange}$$

$$0.4 \leq \frac{GA_+}{FA + MA + GA_+} < 0.6 \Rightarrow \text{Yellow}$$

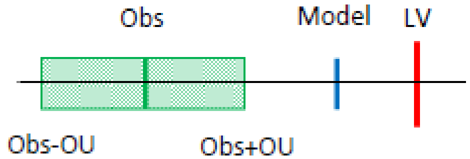
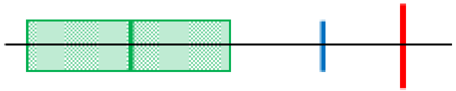
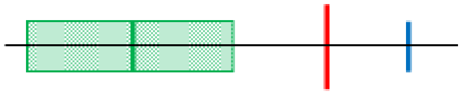
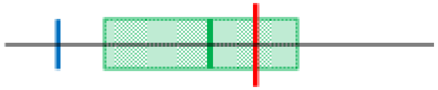
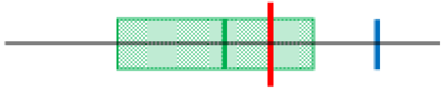
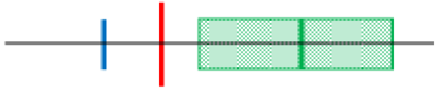
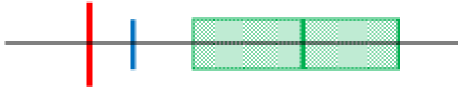
$$0.6 \leq \frac{GA_+}{FA + MA + GA_+} < 0.8 \Rightarrow \text{Light green}$$

$$0.8 \leq \frac{GA_+}{FA + MA + GA_+} \Rightarrow \text{Dark green}$$

Colors nice ..but.. maybe still some additional justification why GA- is dropped from the diagram-metrics completely.
Would it make sense to add metrics where GA+ is replaced with (GA+) + (GA-) ?

Finland is simply too clean country for properly assessing this with our own data (with ”correct” limit values)

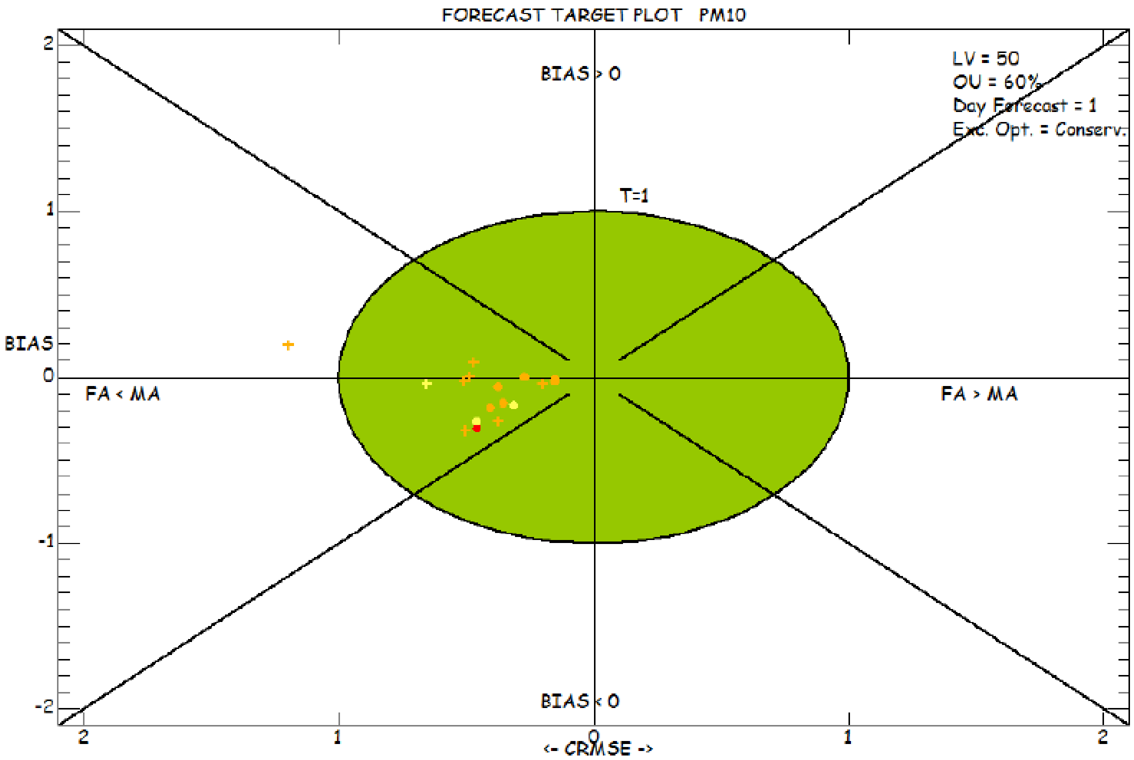
Diagrams cntd..

	Observations		Model		DELTA
	relation to LV	Alarm?	relation to LV	Alarm?	
	$O_+<LV$	No	M<LV	No	GA-
	$O_+<LV$	No	M≥LV	Yes	FA
	$O_-<LV$ $O_+\geq LV$	1: Yes, conservative 2: No, cautious 3: Same as model	M<LV	No	MA GA- GA-
	$O_-<LV$ $O_+\geq LV$	1: Yes, conservative 2: No, cautious 3: Same as model	M≥LV	Yes	GA+ FA GA+
	$O_-\geq LV$	Yes	M<LV	No	MA
	$O_-\geq LV$	Yes	M≥LV	Yes	GA+

Good one to explain the whole story!

A (stupid?) question: *would it make sense to somehow indicate also model uncertainty in the figures + add it to the “alarm”-logic ?*
(~easy way of getting FA>> MA ?)

Diagrams contd..

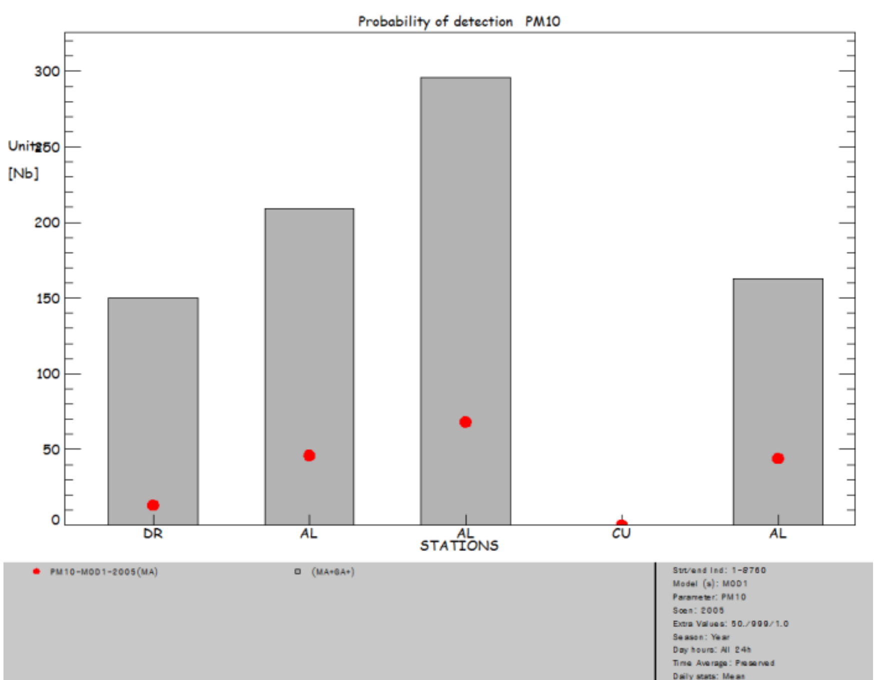


Quite ok for me/us. (although not sure if FA<MA "separation is really needed?) Maybe some additional real-data case studies will help to decide/understand this minor(?) issue better…

DP bar-plot

Based on the following definitions:
* Probability of detection: DP=GA+/(MA+GA+) and

the probability of detection plots GA+ as red dots and (MA+GA+) as grey column for each station. A good model capability would see all red dots on top of the column.

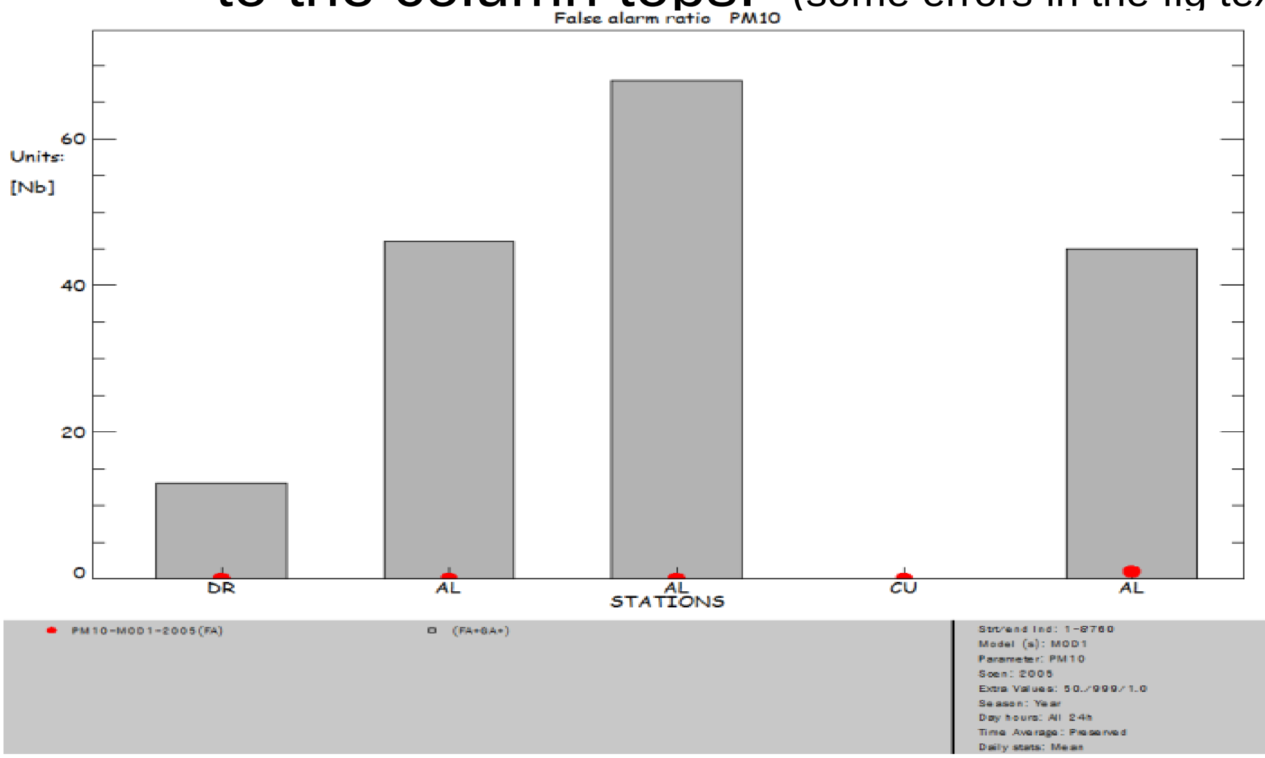


Ok.

False alarms

* False alarm ratio: FAR=FA/(FA+GA+), ->

1-FAR=GA+/(FA+GA+) : the red dots are for GA+ and the grey column for (FA+GA+). A good model again would see red dots close to the column tops. (some errors in the fig text)



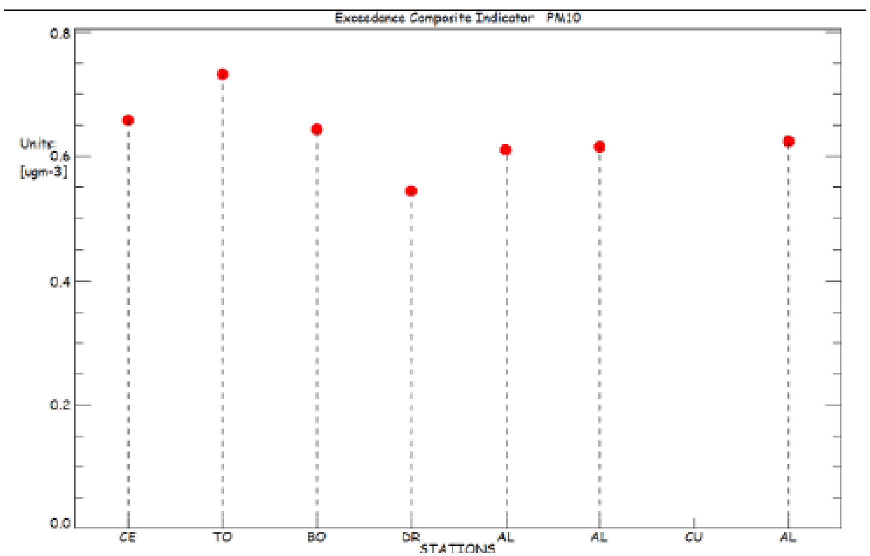
? Strange data.. Where did those GA+ cases vanish <-> previous figure

Combined 1

ratio of the previous two indicators to create a “composite exceedances ratio” as:

A value of CEI=1 would represent the optimal value although it does allow for compensating effects between FA and MA. A value above one is indicative of a dominance of modeled false alarm while a ratio value above 1 indicates dominance of missed alarms in the results. This indicator is bound to a value of 2 and therefore varies between 0 and 2

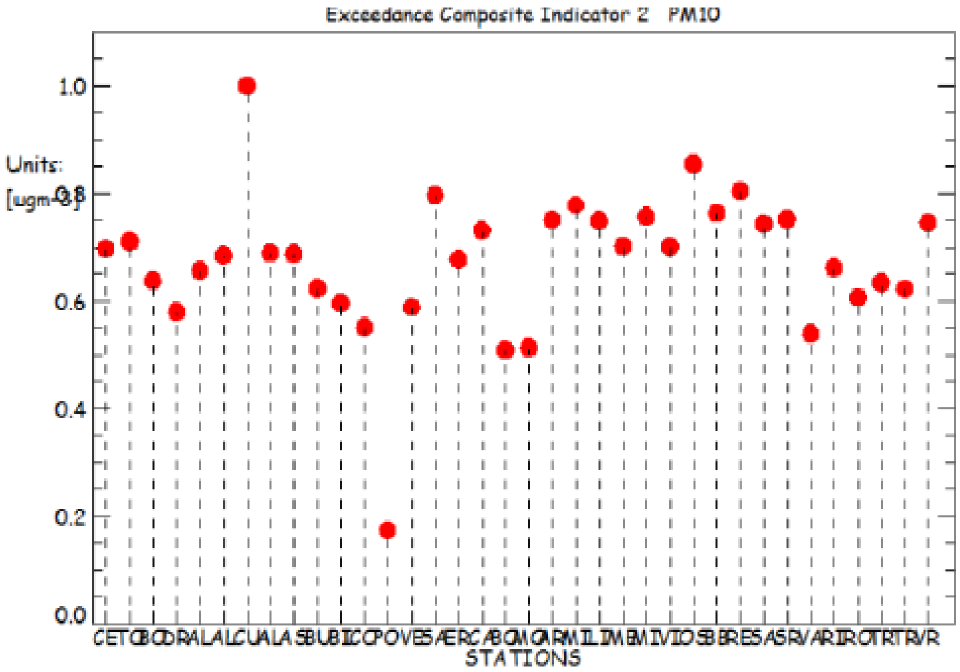
$$CEI_1 = \frac{DP}{1 - FAR} = \frac{FA + GA_+}{MA + GA_+} = \frac{\text{Modelled exceedance s}}{\text{Observed exceedance s}}.$$



Very unclear what this plot/parameter can tell ?
”optimal” value 1 can be reached with extremly poor model if just FA ~= MA ??

Combined 2

$$CEI_2 = 0.5(DP + 1 - FAR) = 0.5\left[\frac{GA_+}{MA + GA_+} + \frac{GA_+}{FA + GA_+}\right].$$



- Much easier to understand/accept !

Conclusions

- In general the added parameters and diagrams make sense
  - some definitions still ”drafty”
  - Justification for the choices made should be clearer
  - Some parameters/diagrams seem to be not so useful or at least very hard to interpret
  - Some additional thought to the importance (or unimportance) of FA/MA ratio should be given -> some changes in the parameters/diagrams ?
- Evaluation with real Finnish data :
  - Hard to find stations/time-series with statistically relevant amount of ”alarming” situations..