

Forecast & Exceedances Indicators

Δ tool – Version 5.4

Based on contributions from J. Stocker, M.T. Pay, A. Monteiro, B. Mahieu, S. Janssen, F. Meleux, M. Stortini and Ph. Thunis (May 2016)

1. Introduction

In this note, indicators and diagrams are proposed for the evaluation of model results in forecast mode. The main objective is to offer a common standardized template to facilitate the screening and comparison of model results.

2. Brief overview of the indicators

As we are mostly interested in forecast mode to check the model ability to accurately reproduce daily forecasts, the Target indicator is modified to consider these aspects. In place of using a normalization based on the standard deviation of the observation uncertainty, we normalize by a quantity representative of the “intra-day” variations, i.e.

$$\text{Target}_{\text{forecast}} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (M_i^* - o_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (o_{i-j} - o_i)^2}} \quad (1)$$

where N is the number of days included in the time series. In such case the Target indicator becomes one when the model forecast is as good as a persistent model (i.e. a model, which for day “i” forecasts the day “i-j” value). The indice “j” represents the forecast time length and is expressed in hours or days . M^* represent the modelled forecast values after accounting for uncertainty (see next section). Values lower than one indicates better capabilities than the persistent model whereas values larger than one indicate poorer performances.

3. Observation uncertainty

The observation uncertainty (OU) is introduced to allow for some margin of tolerance. In DELTA it is now a user input parameter which can be used to test the sensitivity of the results to uncertainty. To account for this uncertainty each model results M_t at time t is transformed into a new model result M_t^* as follows :

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

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

This uncertainty will affect the calculation of the target indicator (1) but also the counting of:

- False Alarms (FA): Model values are above the limit value (LV) but not the observations.
- Missed Alarms (MA): Model values are below the LV but observed values are above it.
- Good values below LV (GA-): both model and observation are below the LV.
- Good values above LV (GA+): both model and observations are above the LV.

Three options to include the observation uncertainty into the calculations are proposed: (1) Conservative, (2) Cautious and (3) as model (Note point 2). These are described in the below table.

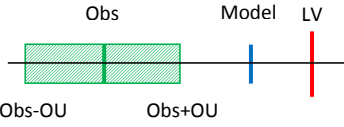
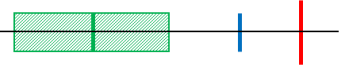
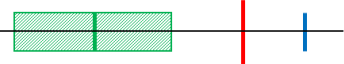
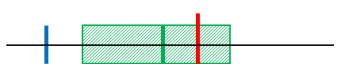

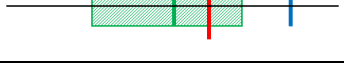
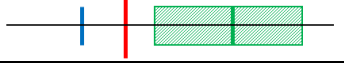
|  | Observations | | Model | | DELTA |
|---|-----------------------------|---|----------------|--------|------------------|
| | relation to LV | Alarm? | relation to LV | Alarm? | |
|  | $O_+ < LV$ | No | $M < LV$ | No | GA- |
|  | $O_+ < LV$ | No | $M \geq LV$ | Yes | FA |
|  | $O_- < LV$ $O_+ \geq LV$ | 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 | $M \geq LV$ | Yes | GA+ FA GA+ |
|  | $O \geq LV$ | Yes | $M < LV$ | No | MA |
|  | $O \geq LV$ | Yes | $M \geq LV$ | Yes | GA+ |

Table 1: Possible cases with respect with model, observation and associated uncertainty. Please note that some “<” or “>” signs from the Note table have been changed to “≤” or “≥” to make sure all situations are included (please check). The DELTA column indicates how DELTA considers the specific cases here described.

The observation uncertainty can either be introduced as a percentage (from 1 to 100%) constant over the whole range of concentration or be as proposed in the frame of the assessment target approach (in this case the uncertainty varies with the concentration). This variable uncertainty can be introduced in the target by selecting “999” for the uncertainty field (see user’s guidance below).

4. Diagrams in the delta tool

a. Target diagram

The **Assessment target diagram** provides information on the following quantities (all normalised by the root mean squared observation uncertainty):

- RMSE: distance from the origin to the point (if distance inferior to one, differences between the model and measurements are not significant)
- 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.
- R and SD: Since the CRMSE is always positive only the right hand side ($X > 0$) of the diagram is used so far. We therefore use this opportunity to add information on the diagram: if the CRMSE is dominated by Standard deviation (SD) the point is plotted on the right whereas if the CRMSE is dominated by correlation error the point is plotted on the left hand side.

In the **Forecast target diagram** information on the following quantities (all normalised by the root mean squared intra-day variations):

- Target forecast (RMSE): distance from the origin to the point (if distance inferior to one, the model behaves better than the persistent approach)
- 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)): CRMSE is still the X axis but we use the ratio FA/MA ratio to differentiate the negative and positive portions of the X axis. This ratio is used to differentiate the right and left parts of the target diagram:

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

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

Points are also given a characteristic colour which depends on the following ratio:

$$\begin{aligned} \frac{FA}{FA + GA_+} < 0.2 &\Rightarrow \text{Red} \\ 0.2 \leq \frac{FA}{FA + GA_+} < 0.4 &\Rightarrow \text{Orange} \\ 0.4 \leq \frac{FA}{FA + GA_+} < 0.6 &\Rightarrow \text{Yellow} \\ 0.6 \leq \frac{FA}{FA + GA_+} < 0.8 &\Rightarrow \text{Light green} \\ 0.8 \leq \frac{FA}{FA + GA_+} &\Rightarrow \text{Dark green} \end{aligned}$$

If more than one model forecast is used (e.g. D+1, D+2...) different symbols will be used to differentiate the forecasts (see Figure 2).

Values lower than one (within the green circle) indicates better capabilities than the persistent model whereas values larger than one indicate poorer performances. Points should ideally be located within the right hand side of the circle ($FA > MA$) with symbols in green. An example of the forecast diagram is provided below for the case of a single model forecast (Figure 1) and in the case of two model forecasts compared to each other (Figure 2).

Note: Only stations characterized by at least one exceedance over the time period selected are considered in the analysis. Currently this filter on stations does not account for observation uncertainty (this could however be changed if necessary).

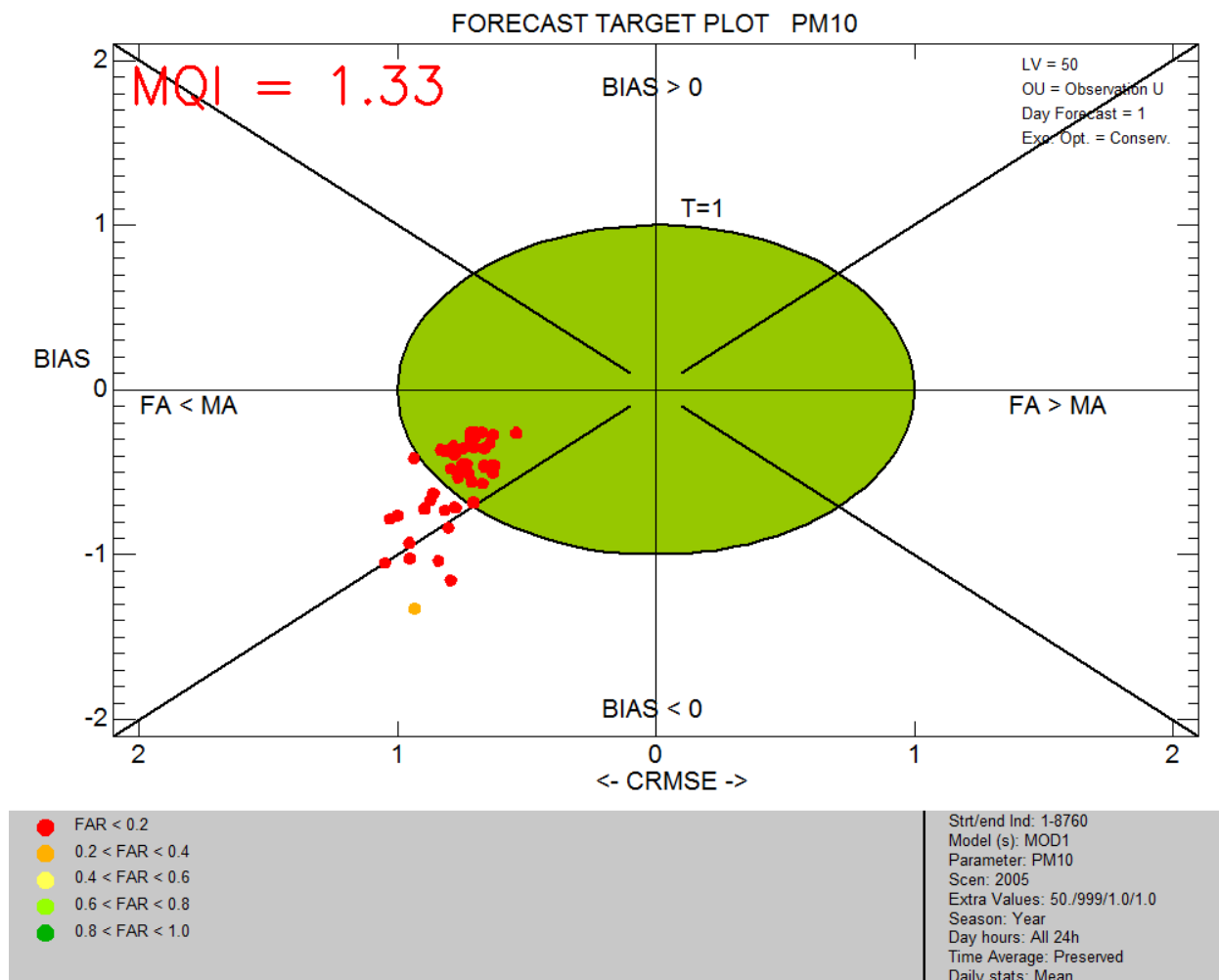


Figure 1: Target forecast for one model. The options selected (Limit Value, observation uncertainty, uncertainty flexibility option and forecast horizon) are reported in the right hand top corner of the figure. All symbols are similar and colors correspond to the value obtained for the FAR indicator.

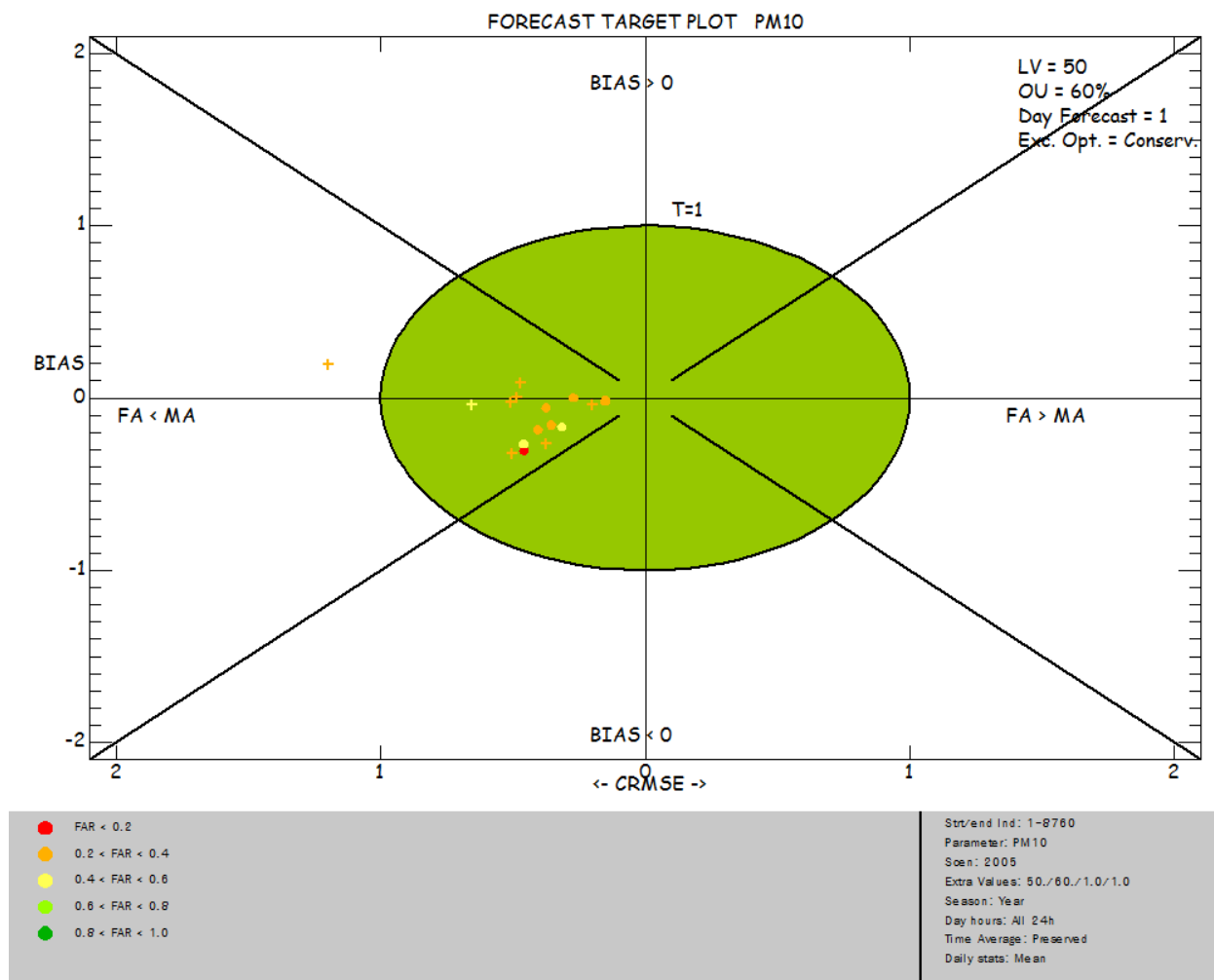


Figure 2: Same as Figure 1 but for two models. Different symbols (yet to be added in legend) are used to differentiate model results but color still correspond to the value obtained for the FAR indicator

User's guidance: The target plot requires 4 values to be inserted as user's options under the threshold field (Analysis window interface). They will be introduced in the following formats:

Val1#Val2#Val3#Val4. Some of these parameters will be hardcoded after the testing phase.

- Val1: Limit value
- Val2: Observation uncertainty – can be set as a fixed percentage (from 0 to 100%) or as the variable uncertainty used in the assessment target (use 999 as value for this option)
- Val3: Flexibility – Options to calculate the uncertainty (see table above). 1 for Conservative, 2 for cautious, 3 for same as model.
- Val4: Forecast horizon – time lag used for the comparison with the persistent model (see formula 1). The value is set in terms of a number of hours or days. Note that days will be selected only if the "Daily Stat" is not set to "preserve" in the Analysis window interface (i.e. either set to "Max", "Mean" or "Min"). Note that this value will be reset to "preserve" automatically if the diagram selection or the variable is changed.

Notes: The four user values will be reset each time a new diagram or a new variable is selected. Use CTRL-C & CTRL-V to copy/paste these values to avoid re-introducing them each time.

b. “Probability of detection” and “False alarm ratio” plots

Based on the following definitions:

- Probability of detection: $DP = GA+ / (MA + GA+)$ and
- False alarm ratio: $FAR = FA / (FA + GA+)$,

two bar plots have been created:

- The first (Figure 3) for 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.
- The second (Figure 4) for false alarms is based on $1 - FAR = GA+ / (FA + GA+)$ in which again 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.

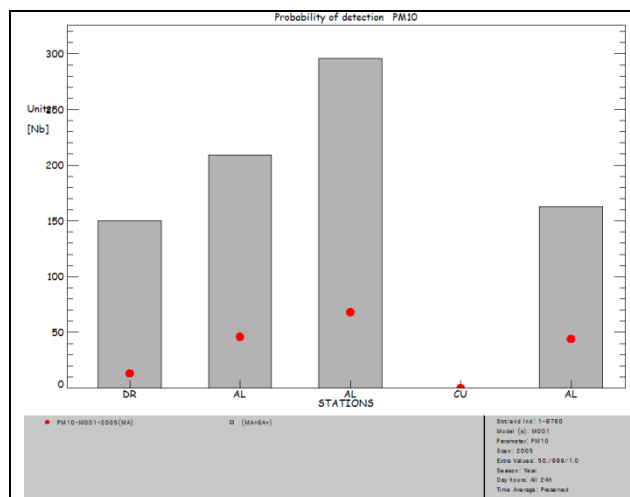


Figure 3: Probability of detection diagram: the dots ($GA+$) should be at top of grey columns (representing the total of observed alarms) for good model predictions (i.e. $MA \approx 0$). The choices of limit value, uncertainty and flexibility options (conservative, cautious...) are listed in the lower right corner.

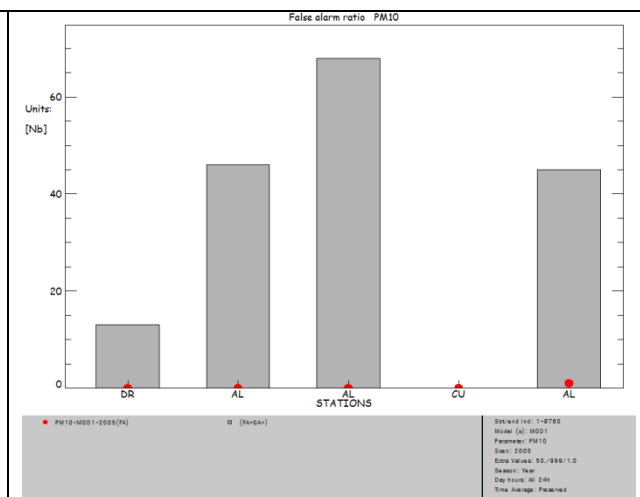


Figure 4: False alarm ratio diagram: the dots (FA) should be as close as possible from 0 (i.e. $FA \approx 0$), with respect to the total of alarms predicted by the model (grey columns). The choices of limit value, uncertainty and flexibility options (conservative, cautious...) are listed in the lower right corner.

The values of FA , $GA+$, $FA + GA+$ and $MA + GA+$ are automatically saved in a “csv” file in the “dump” directory. The file will be over-written at each new diagram production.

c. “Exceedances indicator” bar plots

Indicator 0:

A barplot based on the delta user’s guide high percentile MQO:

$$MPI_{perc} = \frac{|M_{perc} - O_{perc}|}{\beta U_{95}(O_{perc})} \quad \text{and} \quad MPC: MPI_{perc} \leq 1$$

is included in DELTA (named as “percentile indic PM, NO₂, O₃” in the barplot menu). One threshold value corresponding to the value of β is requested. This diagram is available for PM₁₀, PM_{2.5}, NO₂ and O₃.

Indicator 2:

It is defined as the ratio combination of the previous forecast indicators to create a “composite exceedances ratio” as:

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

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. This ratio is available in the DELTA barplots choices under the label “exceedance ratio 1” (Figure 5)

Indicator 3:

It is defined as the summation combination of the previous two indicators to create a “composite exceedances ratio” as:

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

This indicator varies between 0 and 1 and does not allo for compensating effects between false and missed alarms. A perfect model forecast would lead to a value of 1. This ratio is available in the DELTA barplots choices under the label “exceedance ratio 2” (Figure 6).

The CEI indicators 2 & 3 defined as above will vary as follows:

| Case | Obs Exceed. | Mod. Exceed | GA+ | FA | MA | CEI ₁ | CEI ₂ |
|------|-------------|-------------|-----|----|----|-------------------|------------------|
| A | 0 | 0 | 0 | 0 | 0 | Set to 1 | Set to 1 |
| B | 0 | 10 | 0 | 10 | 0 | Bound to 2 | 0 |
| C | 10 | 0 | 0 | 0 | 10 | 0 | 0 |
| D | 10 | 10 | 10 | 0 | 0 | 1 | 1 |
| E | 10 | 10 | 5 | 5 | 5 | 1 | 0.5 |
| F | 10 | 5 | 5 | 0 | 5 | 0.5 | 0.75 |

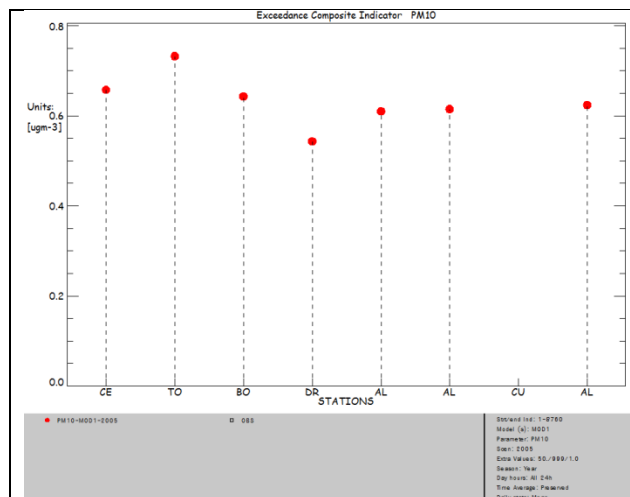


Figure 5: Composite exceedances indicator n1. The choices of limit value, uncertainty and flexibility options (conservative, cautious...) are listed in the lower right corner.

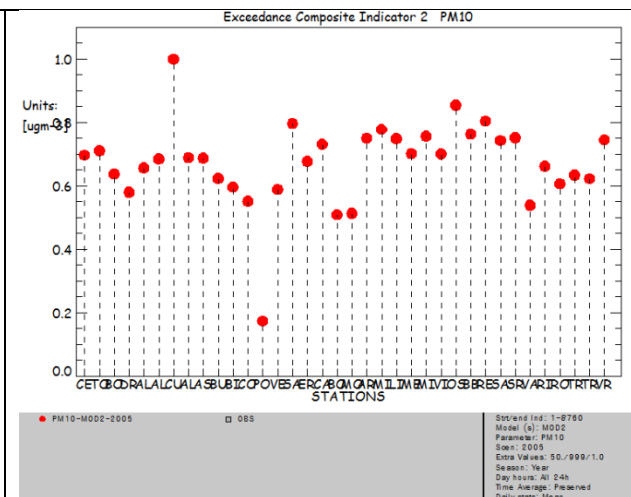


Figure 6: Composite exceedances indicator n2. The choices of limit value, uncertainty and flexibility options (conservative, cautious...) are listed in the lower right corner.

The values of FA, GA+, FA+GA+ and MA+GA+ are automatically saved in a “csv” file in the “dump” directory. The file will be over-written at each new diagram production.

d. User's guidance:

All bar-plots require 3 values to be inserted as user's options under the threshold field (Analysis window interface). They will be introduced in the following formats: Val1#Val2#Val3. These three values are the three first described in the Target Forecast diagram (see above). Some of these parameters will be hardcoded after the testing phase.

Note that the forecast target and barplots diagram are only available in the advanced DELTA version as this diagram is yet under development and intended for testing purposes only. To activate this functionality, go in the resource/init.ini file and uncomment the following line:
ELAB_FILTER_TYPE=ADVANCED and comment this one: ELAB_FILTER_TYPE=STANDARD

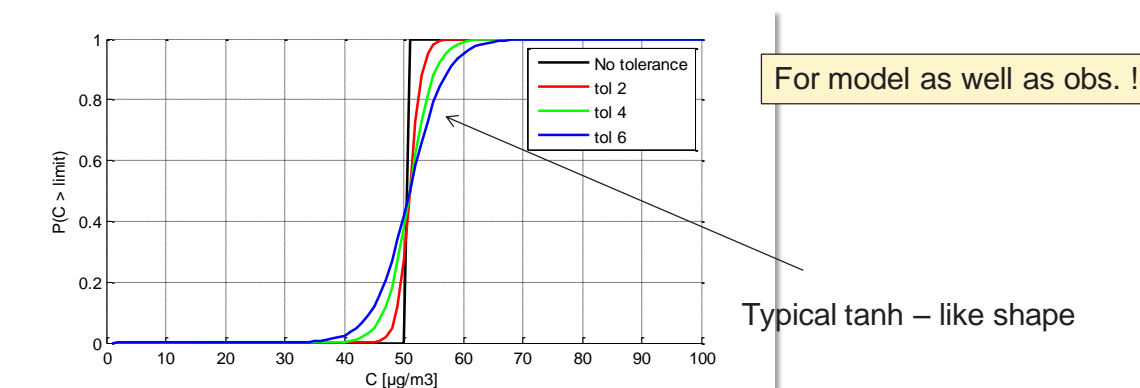
5. Remaining issues

These are the points mentioned in the note for which NO IMPLEMENTATION HAS BEEN DONE in DELTA so far.

a. Probabilities (Note point 4)

Currently, if we have a large amount of missing alerts $FA/MA < 1$, it means that the model is underestimating the observations near the threshold, but introducing the OU will increase the model values, sometime to values above the threshold leading, so the ratio FA/MA would increase. Also, the opposite holds. So in the target diagram the points can switch from left to right, but in reality, you don't really know since you are still within in the observation uncertainty. This behavior makes it somewhat difficult to interpret the left/right labelling of the points in the target plot and relates back to the discussion above

What about introducing tolerance on the threshold? Here we don't count hard exceedances, count probabilities i.e.:



The normal way of counting exceedances is you add 1 when the C value is e.g. above 50 µg/m³, if not you add 0. In the logic presented above, for a concentration of say 45 µg/m³, we would already add say 0.1 to the number of exceedances as there is a probability, given uncertainties that there were exceedances. This way of counting exceedances could be applied both for observations & for model values (representing observation & model uncertainty) and might be a more natural way for taking into account uncertainties in the exceedances indicators. To illustrate, say that theoretically both model & observations are 50 µg/m³ and the threshold is at 50 µg/m³ as well, there is for both model & observation a 50 % change that we have an exceedances. This would mean for FA,MA,GA+,GA- we would add in each of the classes $0.5 \times 0.5 = 0.25$ instead of picking one (which remains uncertain whether it's correct)

It is however unclear how this method would affect the target value/diagram, or how an increase in tolerance would affect the exceedances indicators such as FCF & FFA & this option would have to be further explored & perhaps presented by the next FAIRMODE meeting

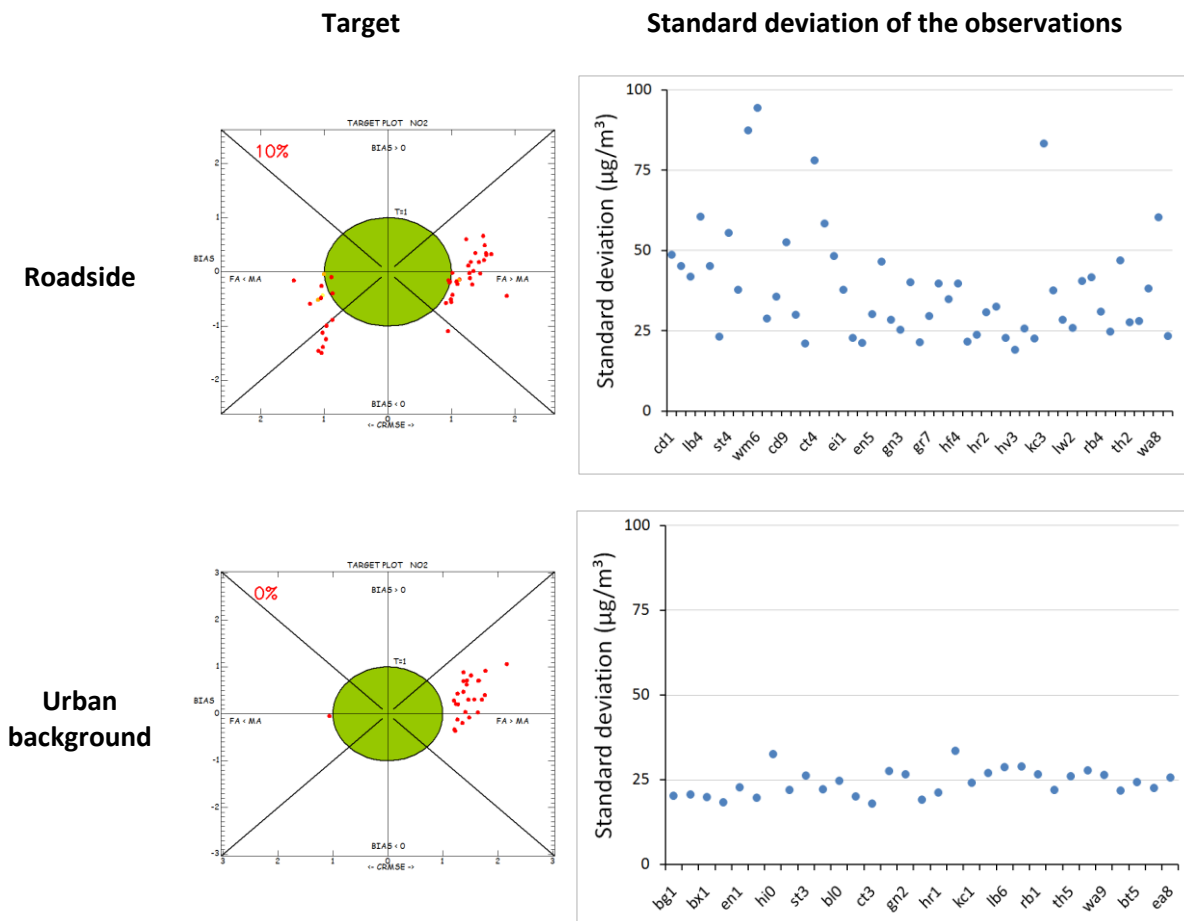
b. Standard deviation (note point 6):

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. This means that at background sites where the standard deviation of the observations is relatively low compared to roadside sites, the target is harder to achieve. In fact what we are saying is that persistence is a better model at urban background sites than it is for roadside sites. Besides which, the persistence model fails mainly when there are peaks and exceedances to forecast, i.e. the main purpose of a forecasting system.

Examples of target and SD plots for roadside and background sites for London are shown below (note the different scales on the target plot):



Is it robust for the Target formulation to be independent on the sd of the observations? The assessment target includes normalization by the sd of the observations. One suggestion is to include:

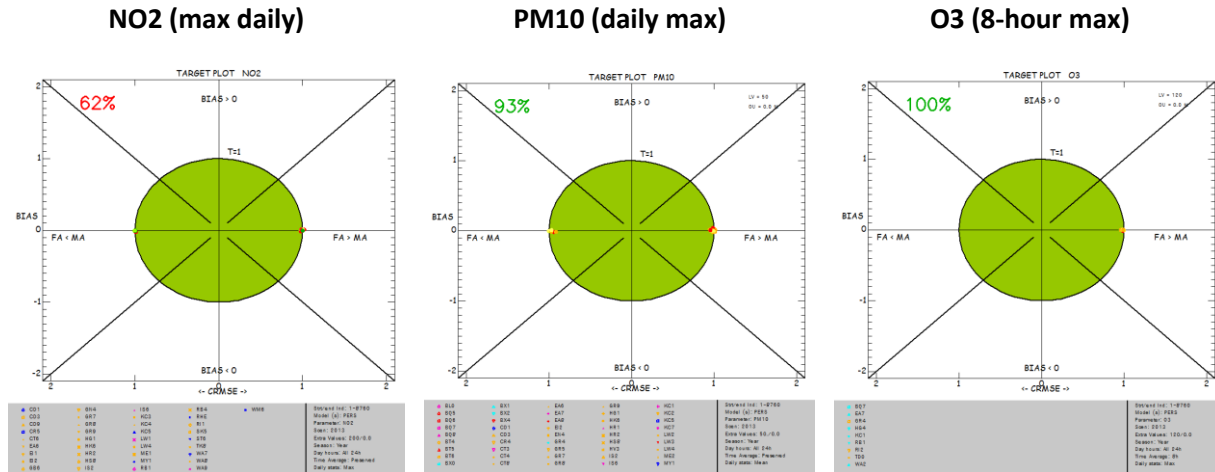
$$\sigma_o \sqrt{2(1 - \alpha)}$$

where α is the autocorrelation in the observation time-series, which could be then adjusted for day+1, day+2 etc...¹

¹ BM: I however don't think this would solve the problem mentioned above as it is simply a reformulation of the formula using $O_{i-1} - O_i$. But perhaps if one should replace the sigma_O by the OU (wasn't this the strategy that was taking with the original target plot ?), this would solve the problem. Not sure how to interpret the target value then though...

c. Persistence plots (Note point 10):

Running the tool with a persistence dataset and OU=0.0 gives approximately what you would expect – but it seems rounding errors are causing misleading stats, for example NO2 has 62% within the Target, PM10 has 93% yet O3 has 100% - as you would expect:



Comment: To be tested with version 5.2. Note also that it is only implemented for 8h max O3, hourly NO2 and daily PM2.5 and PM10. The reason for that is the use of the observation uncertainty only available for these parameter-time average statistics.

d. Summary report:

Currently there are no forecast indicators in the summary report. Some suggestions for inclusion are the following:

| Participant | Index agreement | Threshold agreement | GA+ | FA | FCF | FFA |
|-------------|-----------------|---------------------|-----|----|-----|-----|
| MP | | ✓ | | | | |
| JS | ✓ | ✓ | | | ✓ | ✓ |
| AM | ✓ | ✓ | | ✓ | | |
| BM | | | | | ✓ | ✓ |

Comment: To be implemented later on. Definition of proposed indicators needed. Does not seem to be consensus for all indicators

e. X: Axis (Point 7)

replace the ratio FA/MA along the x-axis by a new indicator combining the probability of good detection of threshold ($FCF = GA+ / (GA++ MA)$)

Comment: Does not seem to represent a consensus.

