FAIRMODE
Forum for air quality modelling in Europe

FORECAST DELTA TOOL

Pawel Durka
Stijn Janssen
Kees Cuvelier

FORECAST EVALUATION

General concept for the evaluation metrics

» Forecast Modelling Quality Objective comes on top of FAIRMODE's assessment MQO

» The forecast MQO should test two different features of a forecast model:

1. Detection of the start / end of an episodes (sudden changes in the concentration)
2. Threshold exceedances (as trigger for short term action plans)

» For 1. we use the "persistence model" as a benchmark

» For 2. we use standard threshold indicators

MQI FOR FORECAST

Target is to do better than the "persistence" model

» Definition: MQI

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

$$MQO_{\text{forecast}} : \quad MQI_{\text{forecast}} < 1$$

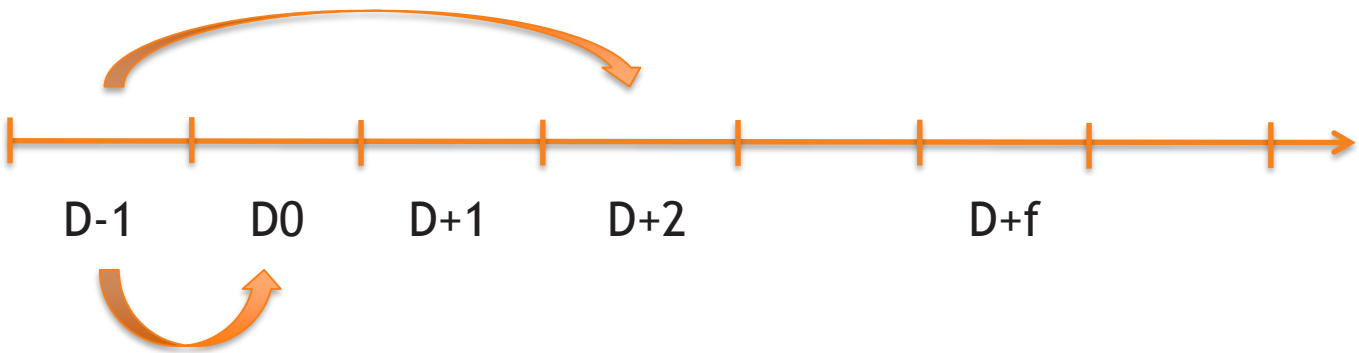
PERSISTENCE MODEL

» At D0 only data for D-1 is available
→ Persistence model for D+f = D-1 (f = horizon)
→ j = f+1

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

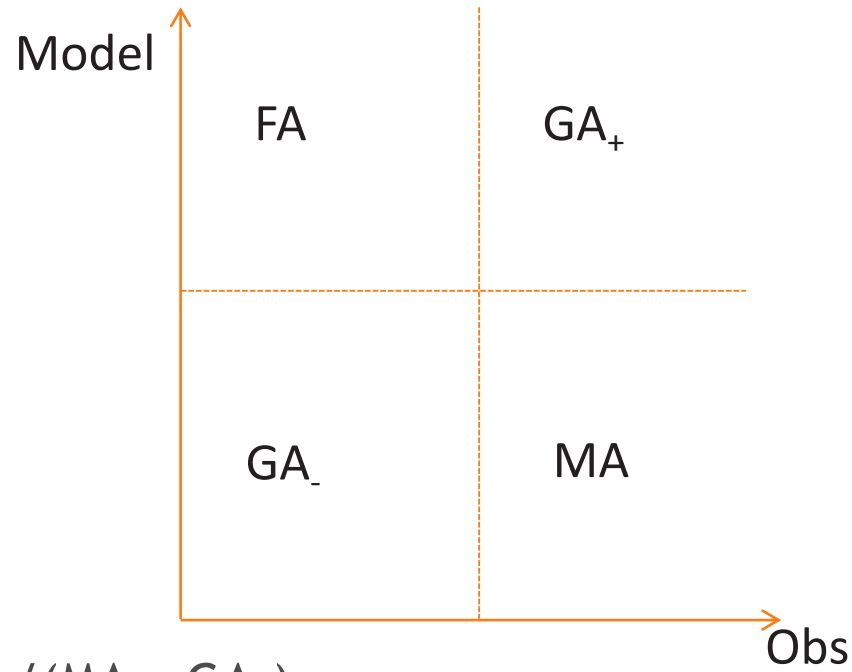
» How to extent to hourly data?

$$MQO_{\text{forecast}}: \quad MQI_{\text{forecast}} < 1$$



THRESHOLD EXCEEDANCES

Definition of threshold exceedance indicators



- » False Alarms (FA)
- » Missed Alarms (MA)
- » Good values below thr (GA_-)
- » Good values above thr (GA_+)

- » Probability of detection: PoD = $GA_+/(MA + GA_+)$
- » 1 - False alarm ratio: 1 - FAR = 1 - $FA/(FA + GA_+) = GA_+/(FA + GA_+)$
  - » better name? → valid/true alarm ratio / VAR?

THRESHOLD EXCEEDANCES

FA OBS<=LV & MOD>LV 38	**GA^+** OBS>LV & MOD>LV 82	Forecast YES events 120
GA^- OBS<=LV & MOD<=LV 222	**MA** OBS>LV & MOD<=LV 23	Forecast NO events 245
Observed NO events 260	Observed YES events 105	**Total** 365

- LV = Limit Value or Threshold
- GA^+ = Mod and Obs > LV
- GA^- = Mod and Obs <= LV
- FA = False Alarm
- MA = Missed Alarm

MAIN FORECAST INDICATORS

» What fraction of the observed Y events were correctly forecast?
» **POD** = GA^+ / (GA^+ + Mi) = 0.78
» Probability of Detection.

» What fraction of the forecast Y events were correctly observed?
» **1-FAR** = GA^+ / (GA^+ + FA) = 0.68
» False Alarm Ratio, Success ratio (SR).

» $0 \leq POD \leq 1$ => POD=1
» $0 \leq 1\text{-}FAR \leq 1$ => 1-FAR=1

» If GA^+=0 and MA=0 the POD=1, If GA^+=0 and MA>0 the POD=0
» If GA^+=0 and FA=0 the 1-FAR=1, If GA^+=0 and FA>0 the 1-FAR=0

» Health approach: No MA, i.e. POD=1
» Fin/Econ approach: No FA, i.e. 1-FAR=1

FORECAST BARPLOT

Description of indicators

Extra Values: Threshold value and sensitivity range %:

» LV#SensRange#, Example 50#10#

No criteria needed

POD and 1-FAR are calculated at LV

Sensitivity of POD (cq 1-FAR) is calculated using 5 threshold values

» in the range LV – SensRange % → LV + SensRange %

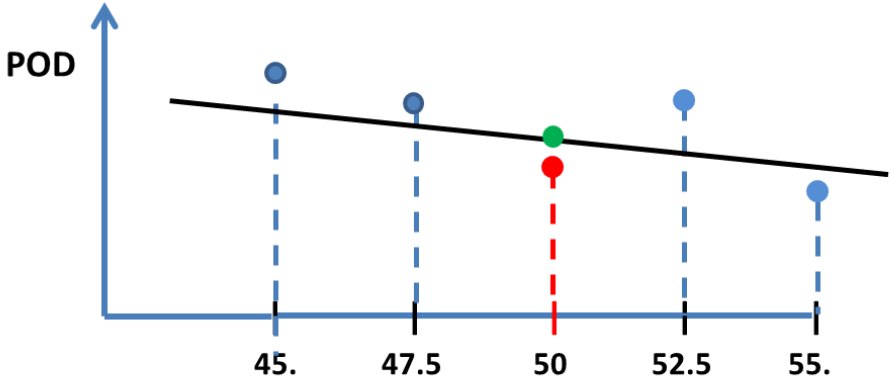
» Example PM10, POD: Extra values 50#10#.

» The 5 threshold values are 50-10% → 50+10%,

» i.e. 45., 47.5, 50., 52.2, 55.

FORECAST BARPLOT – SENSITIVITY

DeltaTool Demo data: Threshold=50, SensRange=10.



Slope is sensitivity of POD for a change of 1 unit of the threshold value

Threshold=50.0000 SensRange=10.0000

Threshold	GA+	GA-	FA	MA	CA	1-FAR	POD	
45.000	18.000	241.000	8.000	64.000	82.000	0.692	0.220	
47.500	17.000	248.000	5.000	61.000	78.000	0.773	0.218	
50.000	16.000	250.000	6.000	59.000	75.000	0.727	0.213	●
52.500	11.000	258.000	7.000	55.000	66.000	0.611	0.167	
55.000	10.000	264.000	5.000	52.000	62.000	0.667	0.161	

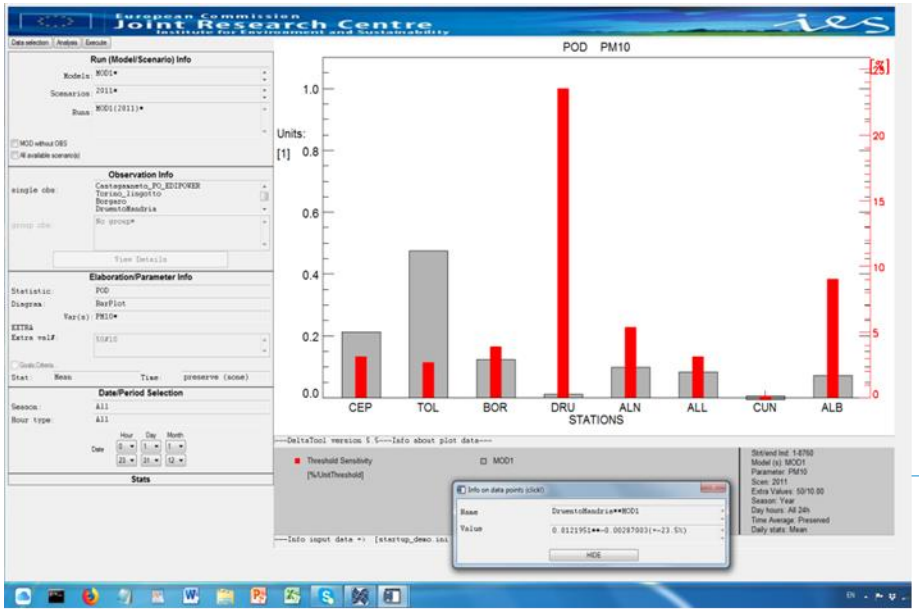
Interc ●	14.400	252.200	6.200	58.200	72.600	0.694	0.196	●
Slope[%]	-5.500	0.896	-2.667	-2.034	-2.773	-1.171	-3.145	wrt ●
Slope[abs]	-0.880	2.240	-0.160	-1.200	-2.080	-0.009	-0.007	

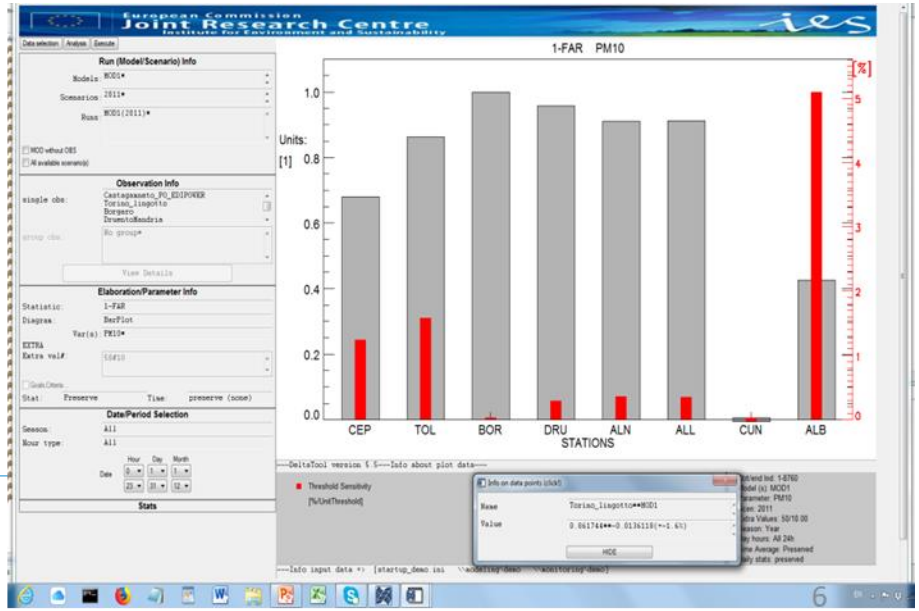
FORECAST BARPLOT - EXAMPLE

» In the DeltaTool forecast barplot, the grey bars represent POD (c.q. 1-FAR), red bars the sensitivity of POD, which is defined as the change in % of the POD value (grey bar) when the threshold value is changed by 1 unit.

» Input 'extra values' to the Tool: Threshold value (example 50 for PM10), and a sensitivity range (example 10): 50#10#. POD is calculated for 5 values within a range of 2*10% of the threshold, i.e. for 45, 47.5, 50, 52.5 and 55. The red bar is the slope (in %) of the regression line through the 5 corresponding values of POD.

» In the next diagram, the recognize feature for station DRU tells us that changing the threshold value from 50 to 51 will change the POD value of 0.012195 by -23.5%





FORECAST TARGET DIAGRAM

Normalization by the intra-day variations

» 3 extra values: LV#Cut-off# DayForecast #.

» Only hours/days with Obs >= Cut-off are taken into account for

» the calculation of the resilience (i.e. the normalization factor)

» Needs criteria, but is set automatically (auto-correction)

» **Resilience Normalisation:**

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

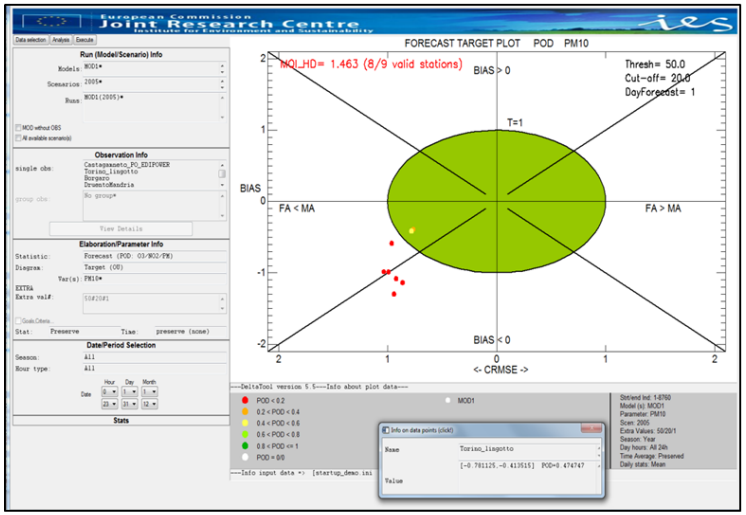
» Persistence model predicts for day i the Obs of day i-1.

» For day i: Compare Model(day i) with Persistence model (day i), i.e.

» Compare Model (day i) with Obs (day i-1)

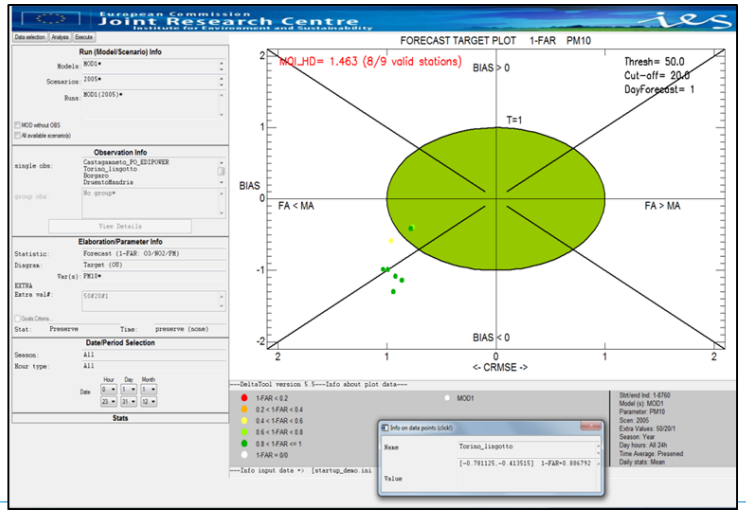
» i-1 => i-j, j = DayForecast = 1,2,3,…

FORECAST TARGET DIAGRAM - EXAMPLE



Example: PM10, POD
Threshold=50, Cut-off=20, DayForecast=1

Example: PM10, 1-FAR,
Threshold=50, Cut-off=20, DayForecast=1

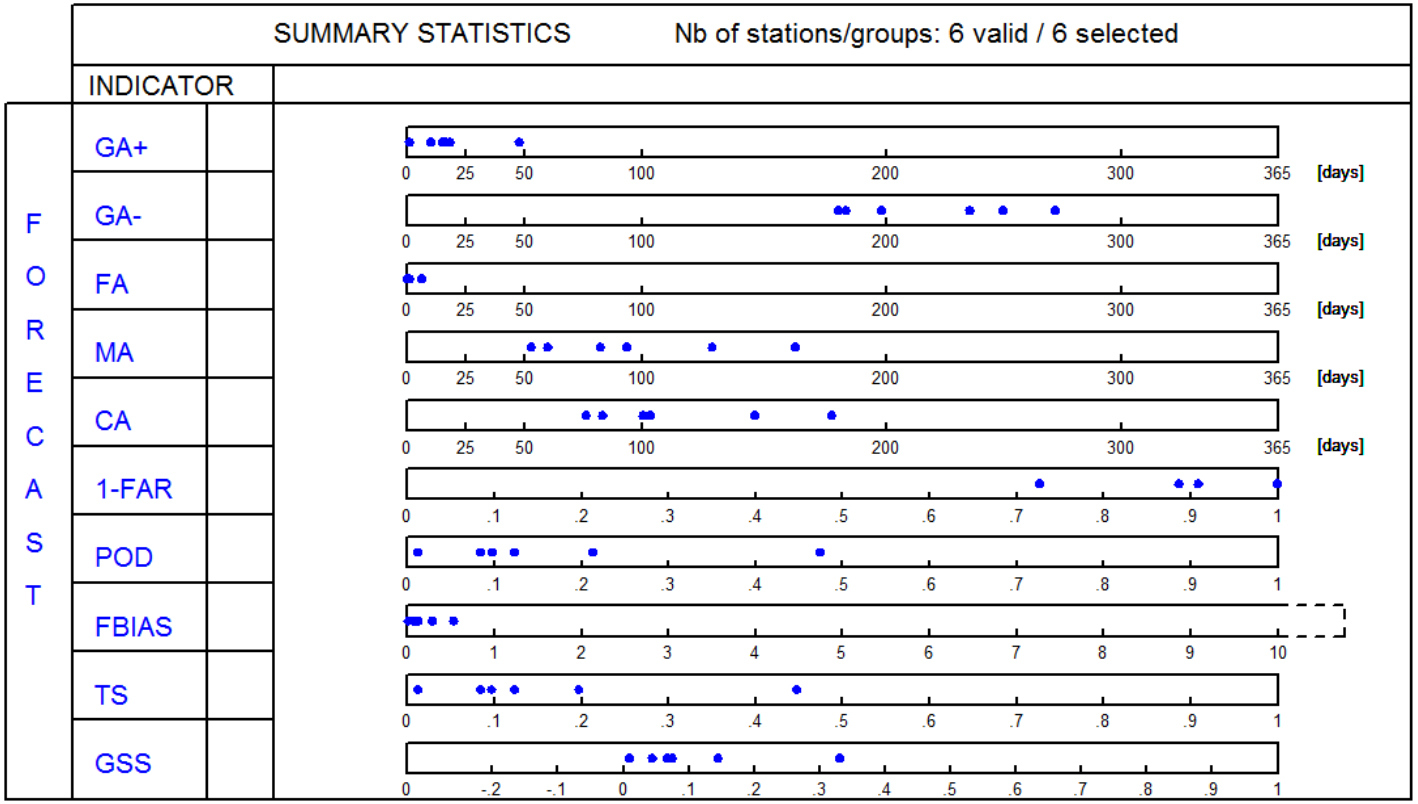


SUMMARY REPORT: aditional indicators

Indicator	Optimal value
GA+	→ Counted exceedances
GA-	→ Counted non-exceedance
FA	→ 0
MA	→ 0
CA	→ Counted alerts
1-FAR	→ 1
POD	→ 1
FBAIS	→ 1
TS	→ 1
GSS	→ 1

- FBias score: FBIAS= (GA+ + FA) / (MA + GA+)
- Threat score: TS = GA+ / (MA + FA + GA+) = GA+/(FA + GA+)
- Gilbert Skill score: (GA+ - H_{random}) / (MA + FA + GA+ - H_{random})
  - with H_{random} = (GA+ + MA)(GA+ + FA) / Total

SUMMARY REPORT



FORECAST PERFORMANCE DIAGRAM

FA OBS<=LV & MOD>LV 38	**GA⁺** OBS>LV & MOD>LV 82	Forecast YES events 120
GA⁻ OBS<=LV & MOD<=LV 222	**MA** OBS>LV & MOD<=LV 23	Forecast NO events 245
Observed NO events 260	Observed YES events 105	**Total** 365

FA	GA^+
GA^-	MA

$= \frac{\square}{\square} =$ POD

FA	GA^+
GA^-	MA

$= \frac{\square}{\square} =$ 1-FAR

FA	GA^+
GA^-	MA

$= \frac{\square}{\square} =$ FBIAS

FA	GA^+
GA^-	MA

$= \frac{\square}{\square} =$ TS

FORECAST PERFORMANCE DIAGRAM - INDICATORS

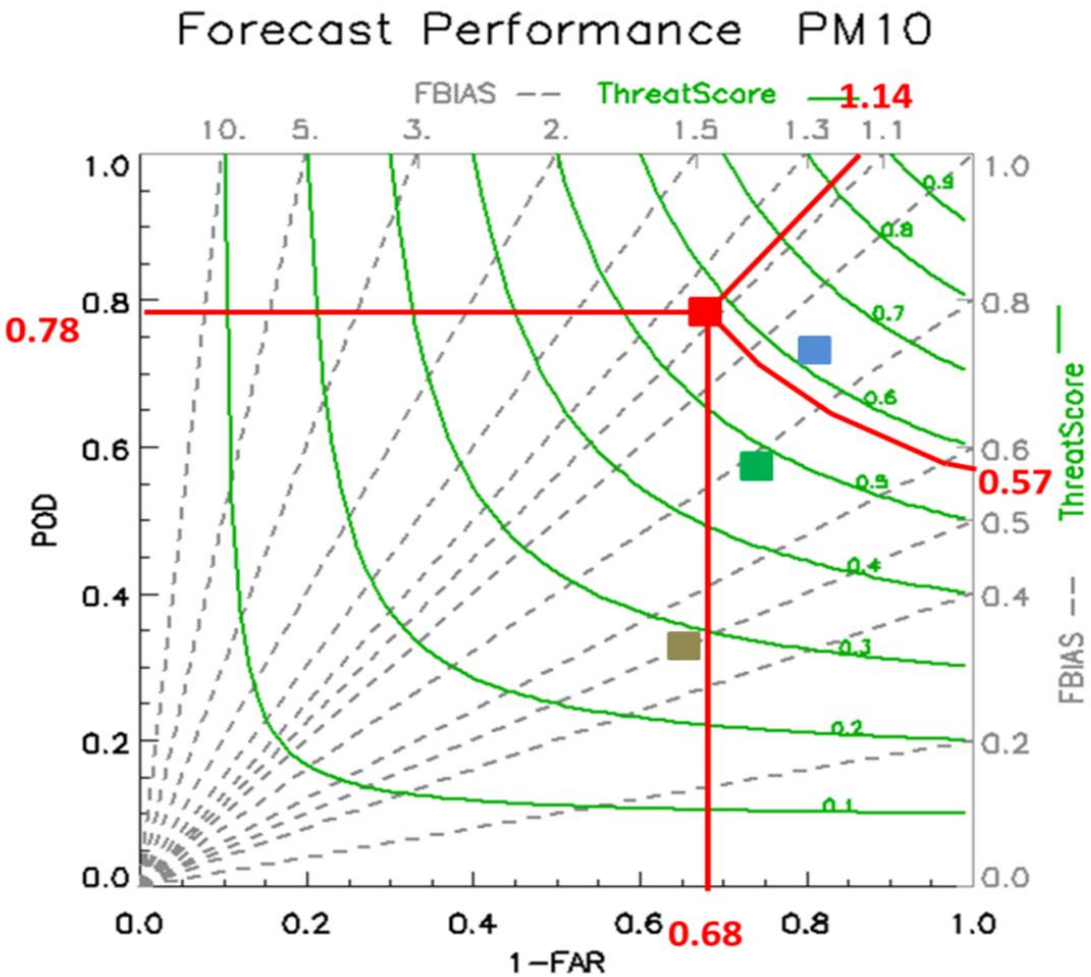
» What fraction of the observed Y events were correctly forecast?
» **POD** = GA^+ / (GA^+ + MA) = 82/(82+23) = 0.78
» Probability of Detection.
» 78%

» What fraction of the forecast Y events were correctly observed?
» **1-FAR** = GA^+ / (GA^+ + FA) = 82/(82+38) = 0.68
» Success ratio (SR).
» 68%

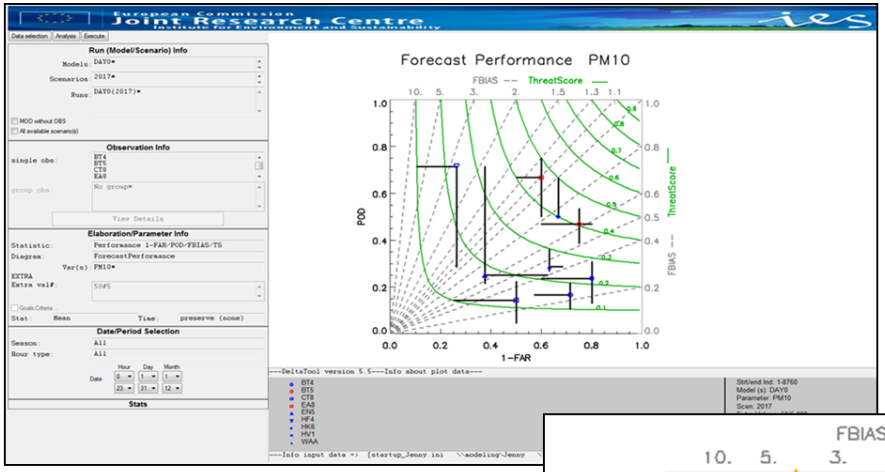
» How did the forecast Y events compare to the observed Y events?
» **FBIAS** = (GA^+ +FA) / (GA^+ +MA) = 120 / 105 = 1.14
» FBias score, FBias < 1 underforecasting, FBias > 1 overforecasting
» Slight overforecasting (14%)

» How well did the YY events correspond to all Y events (Obs+Forecast)?
» **TS** = GA^+ / (GA^+ +MA + FA) = 0.57
» Threat score, CSI – Critical success index.
» 57%

FORECAST PERFORMANCE DIAGRAM - EXMAPLE

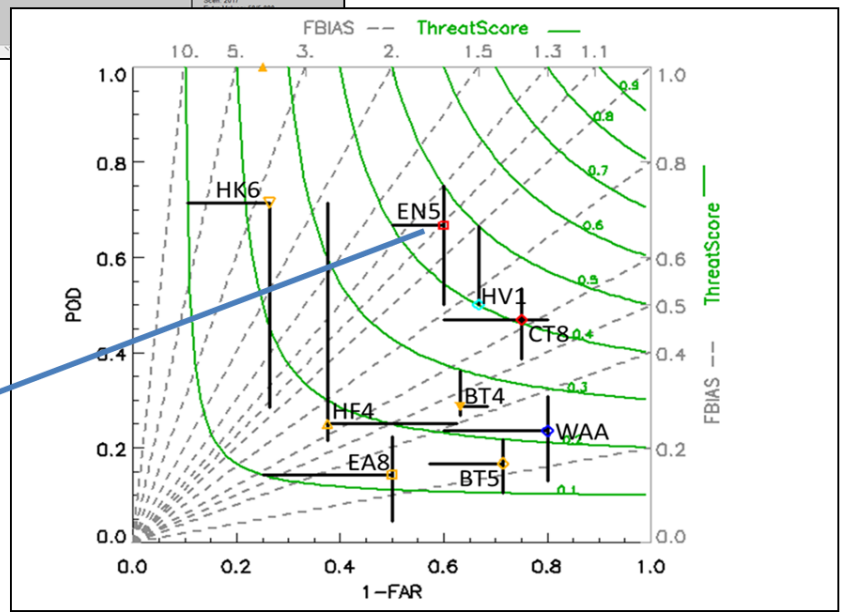


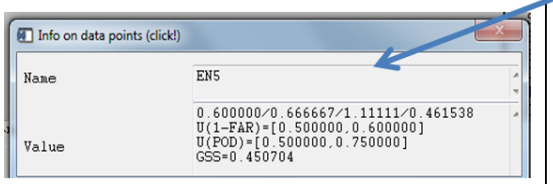
FORECAST PERFORMANCE DIAGRAM – OBSERVATION UNCERTAINTY



Example Data Jenny Stocker, PM10, extraValues 50#5#

Zoom





AGENDA HACKATHON

- Tuesday 17/9 (13u30 – 17u30):
  - Introduction to the Forecast Benchmarking methodology (Stijn & Pawel)
  - Introduction to the Forecast Benchmarking DELTA tool (Kees)
  - Participant feedback on the Forecast Benchmarking methodology
    - Poland (Pawel)
    - Portugal (Alexandra)
    - Italy ENEA (Antonio)
    - Italy ARPAE (Michele)
    - Netherlands (Sjoerd)
    - France (Adrien)
    - Germany (*Florian*/Stijn)

- Wednesday 18/9 (9u30 – 12h30):
  - Cont. participant feedback
  - Discussion & Conclusions
  - Update the forecast chapter in the Guidance Document

OPEN ISSUES & DISCUSSION

- Definition of the Persistence Model:
  - j = forecast horizon / j = forecast horizon + 1
  - j = 1 (or 24h) → no input parameter (forecast horizon) anymore!
  - Only focus on one statistic per day (max1h NO2, max8h O3, dayavg PM10) → predefined options per pollutant in DELTA tool
- Is Target_forecast diagram the right indicator/message to bring
  - Other layout than assessment target plot? → proposal ARPAE
  - We keep Target diagram
    - MQI_forecast P90
    - POD_norm P90
    - SR_norm P90
    - Add FBAIS on x-axis
    - Color code based on TS as additional option
- Forecast barplot
  - Change sensitivity (red bar) from percentage to absolute values
- Normalized version of Performance Diagram → Cuvelier-plot:
  - Based on persistence model as normalization factor
- What about the cut-off in the Target diagram → sensitivity? Fix value needed!? 20µg/m^3
  - Free parameter for time being but over time fix it to a fraction (e.g. 50%) of the threshold value
  - Check how cut-off is applied (model and obs) → modify in DELTA: preprocessing of the time series
- Exclude stations with no threshold exceedances from performance plot?
  - White dots are solution for time being → POD = 0/0 → 1 needs explanation
- 1 – FAR → Fraction of Correct Alarms (FCA) / Success ratio (SR)
- Do we recommend fixed thresholds per pollutant/statistics? → open for time being. Maybe over time fix it according to AQD

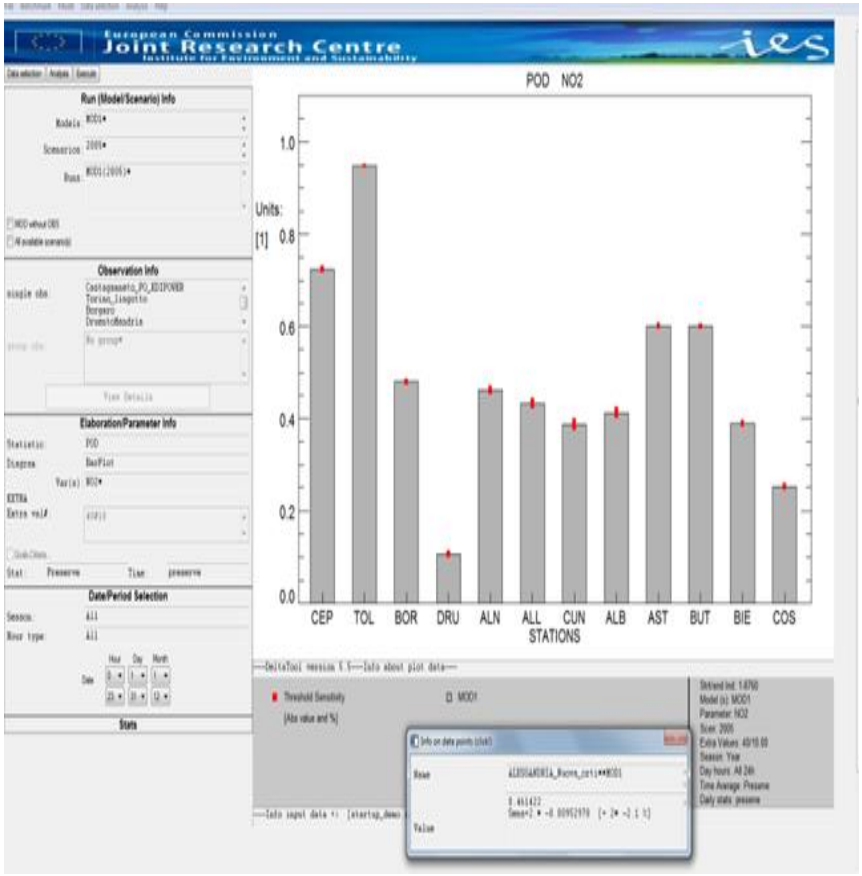
- Publication on MQI_forecast → who takes the lead? → too early now? Sjoerd??
- Update Guidance Document → we need updated graphs, before Madrid meeting

NEW AND FRESH!

FORECAST DELTA TOOL FIRST RESULTS AFTER CHANGES DISCUSSED AND ESTABLISHED AT HACKATON!!!

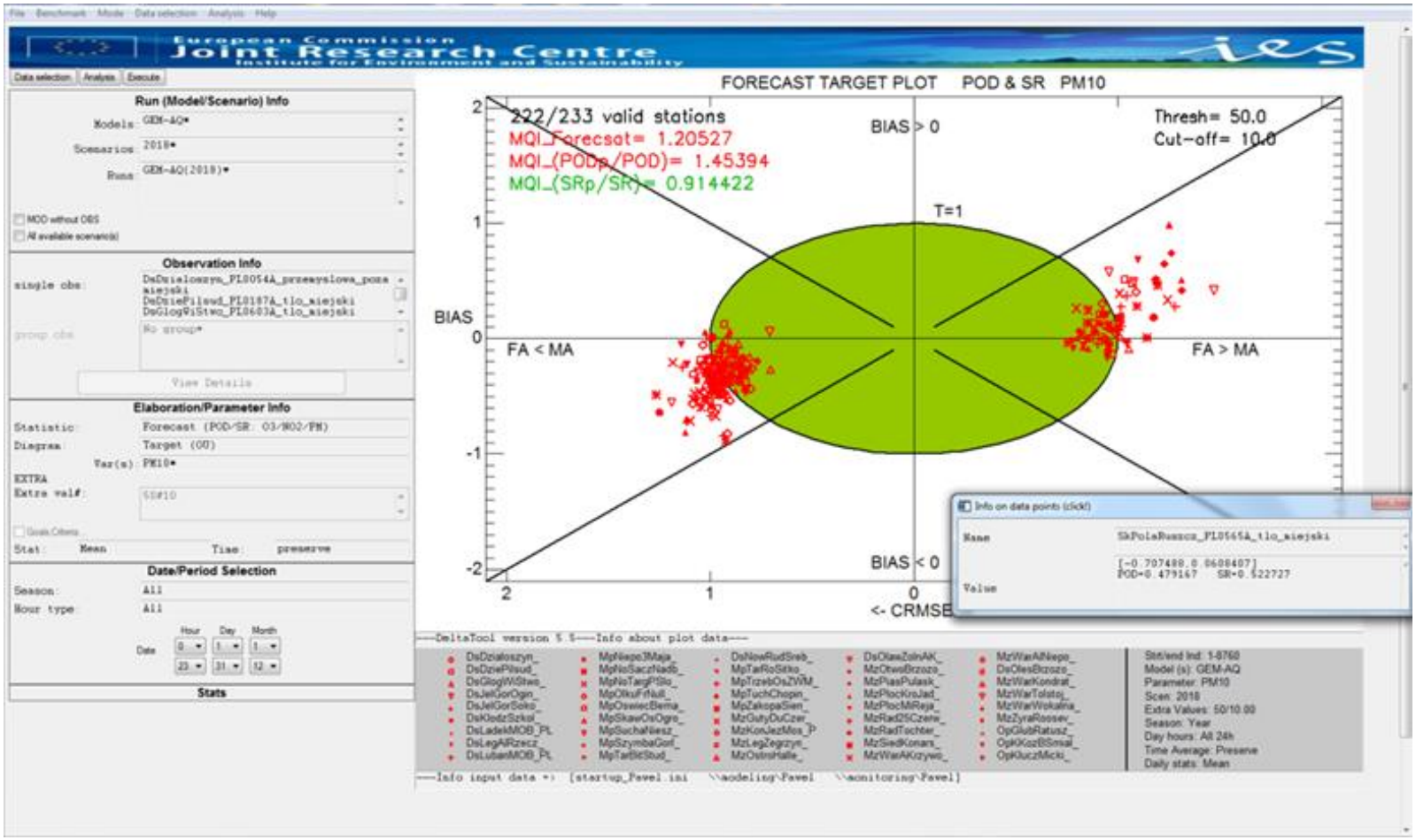
UPDATED FORECAST DELTA TOOL

Barplots with sensitivty bar as absolute value



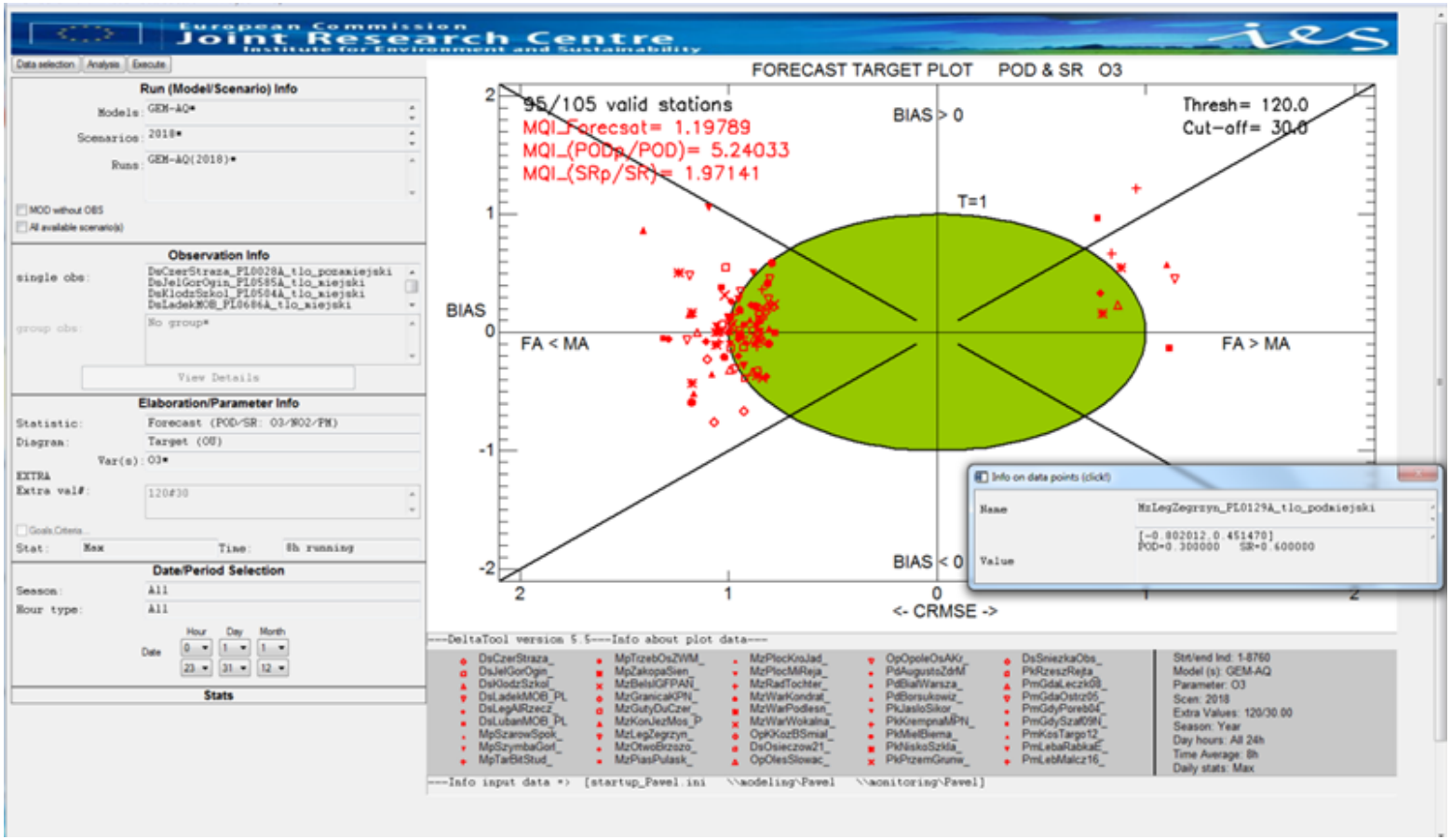
FORECAST TARGET PLOT – PM10

With normalised PODp/POD and SRp/SR approach



FORECAST TARGET PLOT – O3

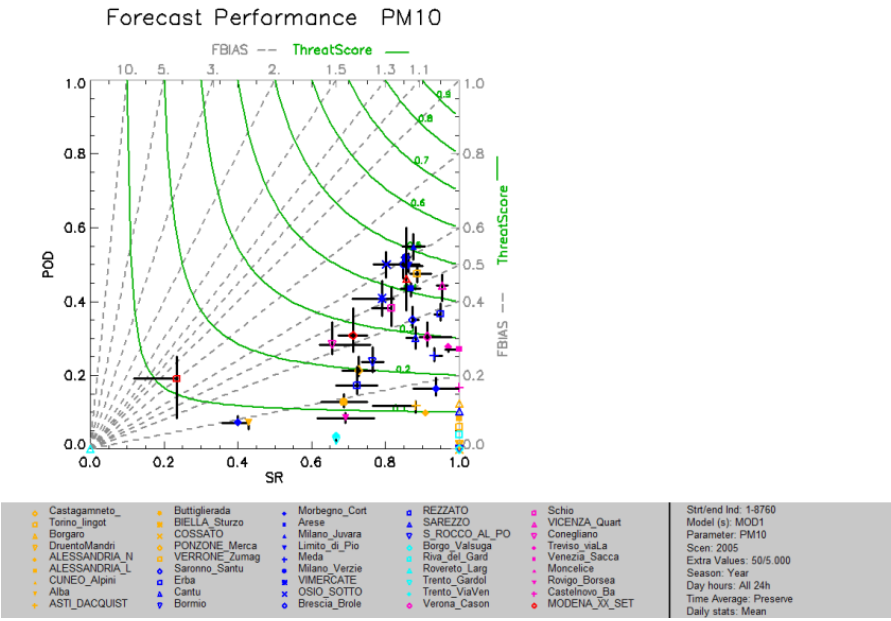
With normalised PODp/POD and SRp/SR approach

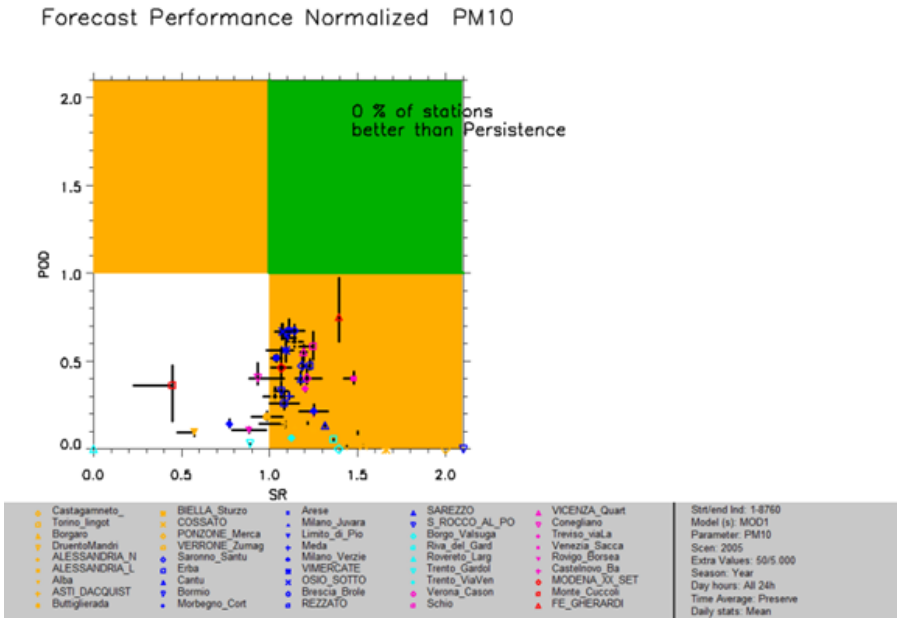


NORMALISED FORECAST PERFORMANCE DIAGRAM (CUVELIER PLOT?)

The idea is to have a yellow zone for stations with POD/PODp or SR/SRp >= 1., Green if both are >= 1., and white if both are <1

» Demo all stations

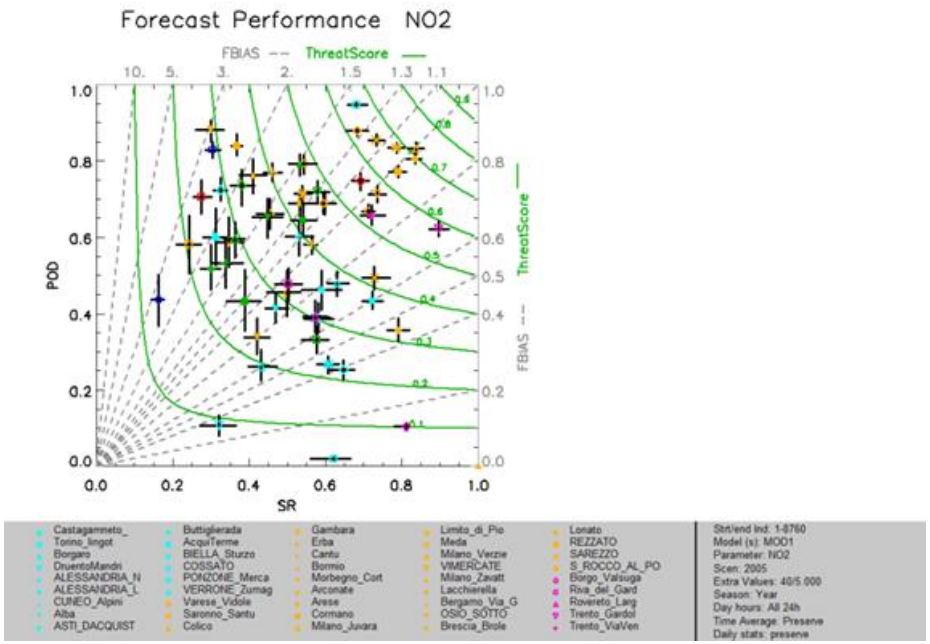


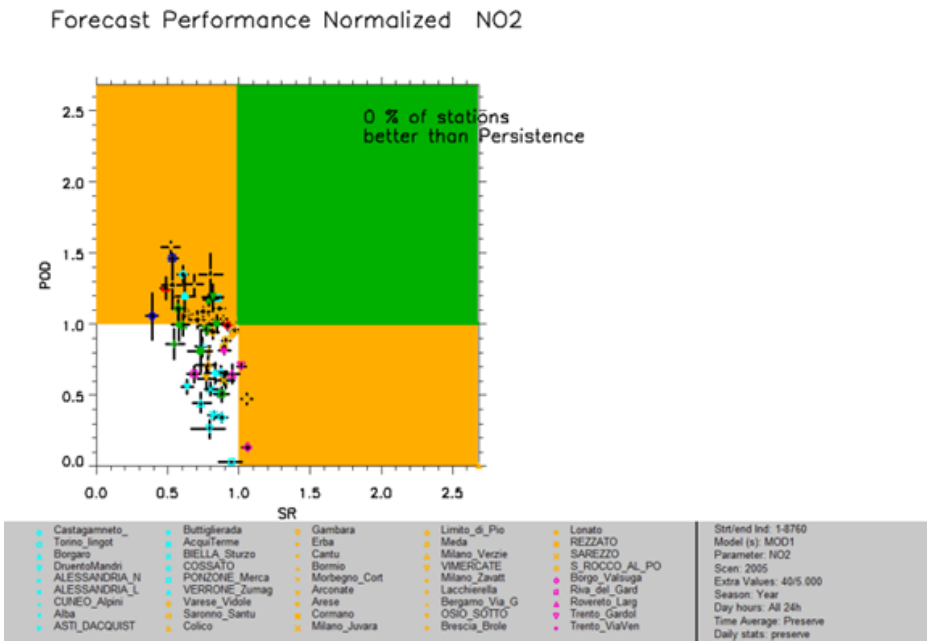


NORMALISED FORECAST PERFORMANCE DIAGRAM (CUVELIER PLOT?)

The idea is to have a yellow zone for stations with POD/PODp or SR/SRp >= 1., Green if both are >= 1., and white if both are <1

» Demo all stations

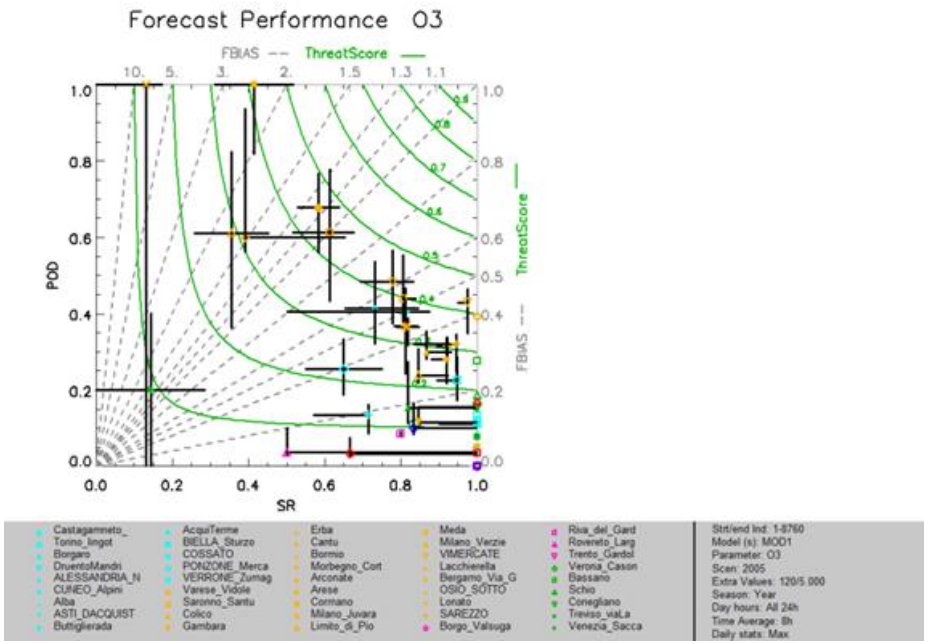


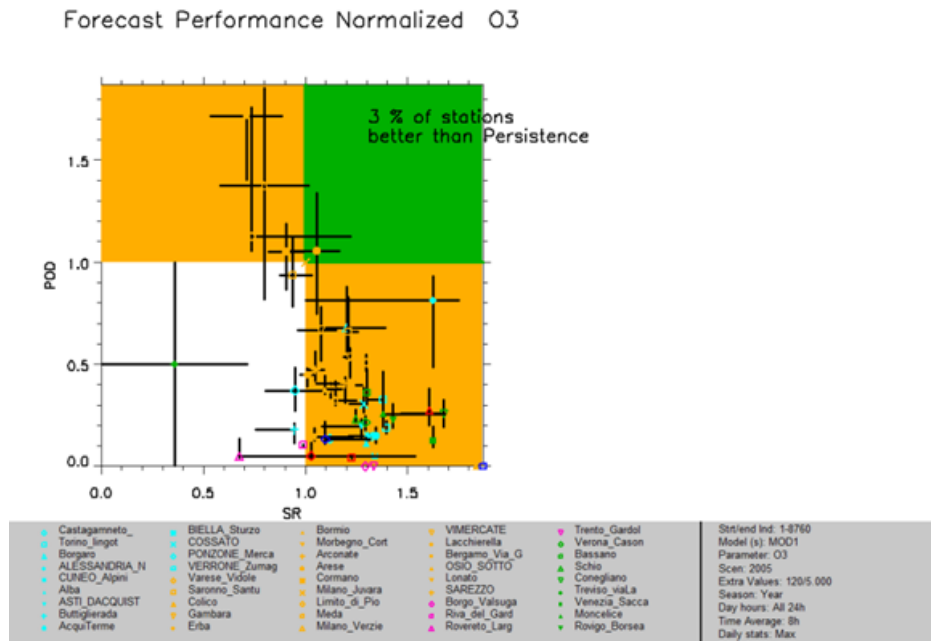


NORMALISED FORECAST PERFORMANCE DIAGRAM (CUVELIER PLOT?)

The idea is to have a yellow zone for stations with POD/PODp or SR/SRp >= 1., Green if both are >= 1., and white if both are <1

» Demo all stations

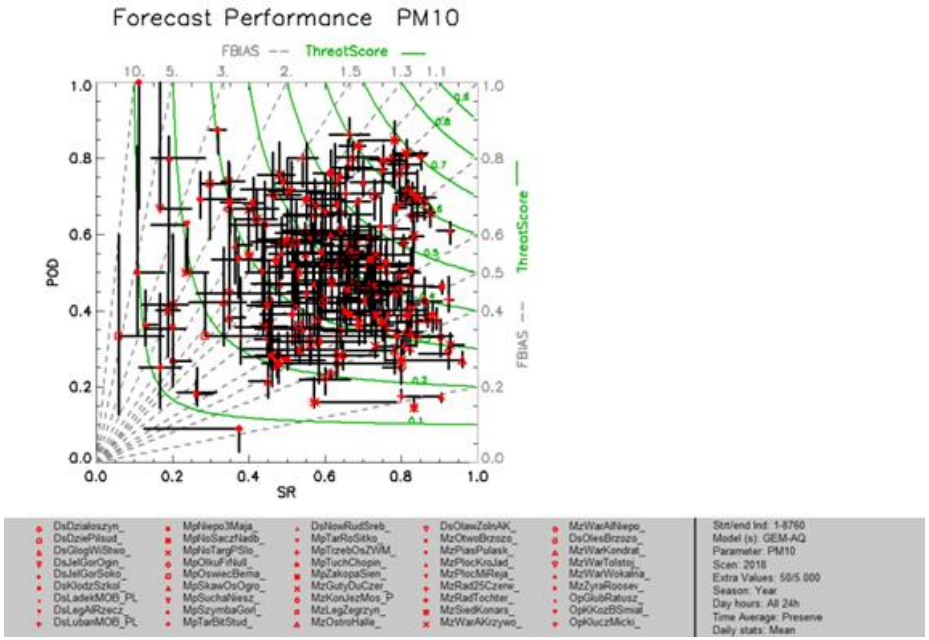


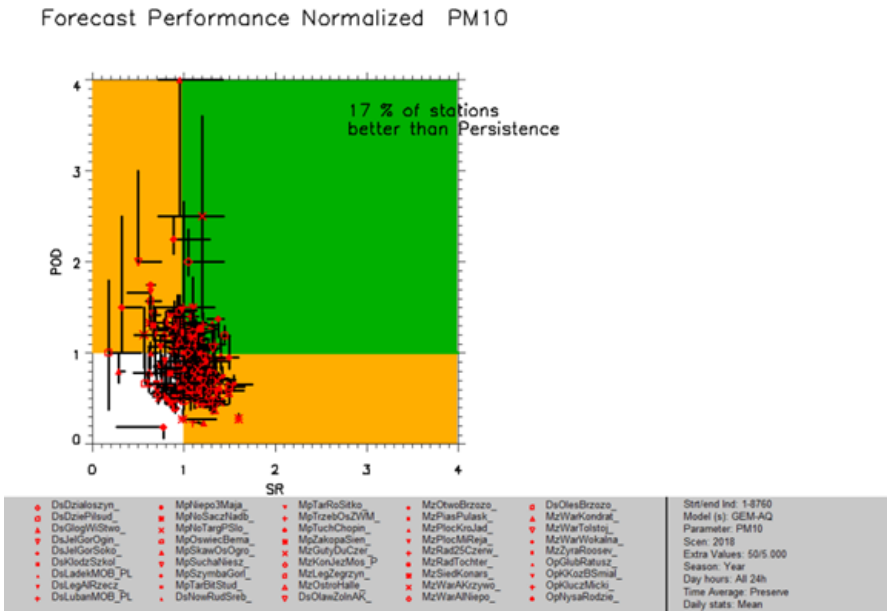


NORMALISED FORECAST PERFORMANCE DIAGRAM (CUVELIER PLOT?)

The idea is to have a yellow zone for stations with POD/PODp or SR/SRp >= 1., Green if both are >= 1., and white if both are <1

» Poland – all stations

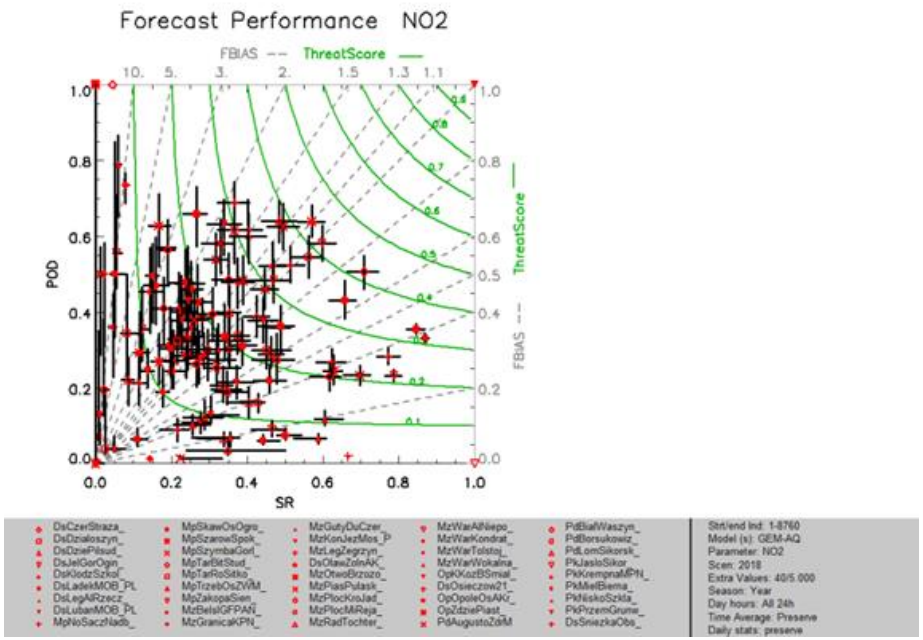


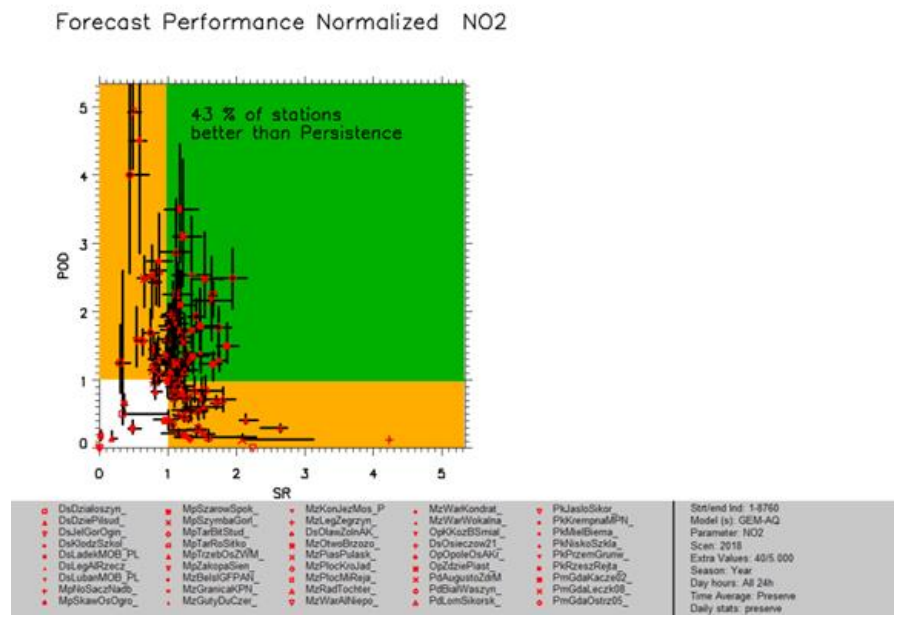


NORMALISED FORECAST PERFORMANCE DIAGRAM (CUVELIER PLOT?)

The idea is to have a yellow zone for stations with POD/PODp or SR/SRp >= 1., Green if both are >= 1., and white if both are <1

» Poland – all stations

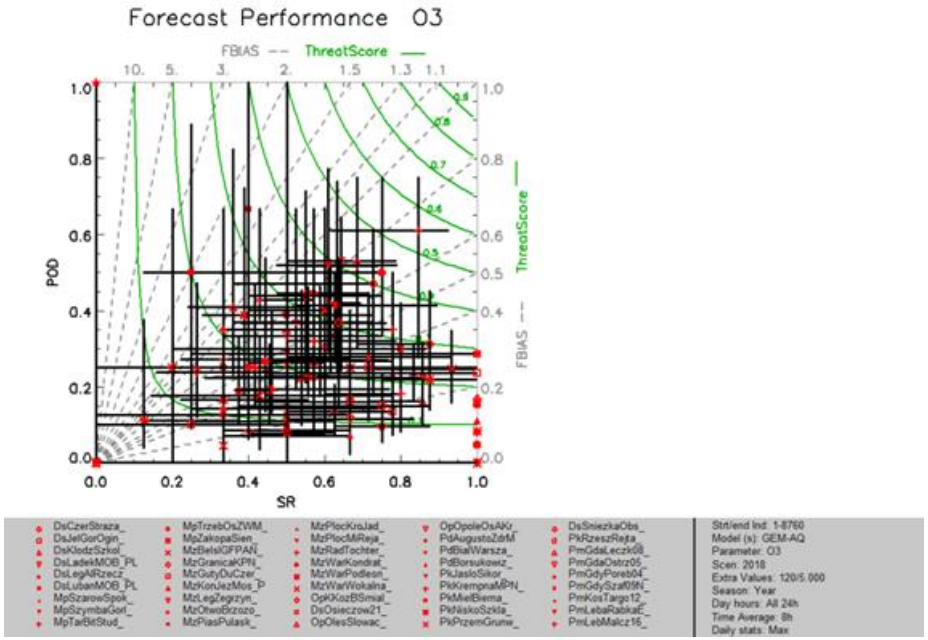


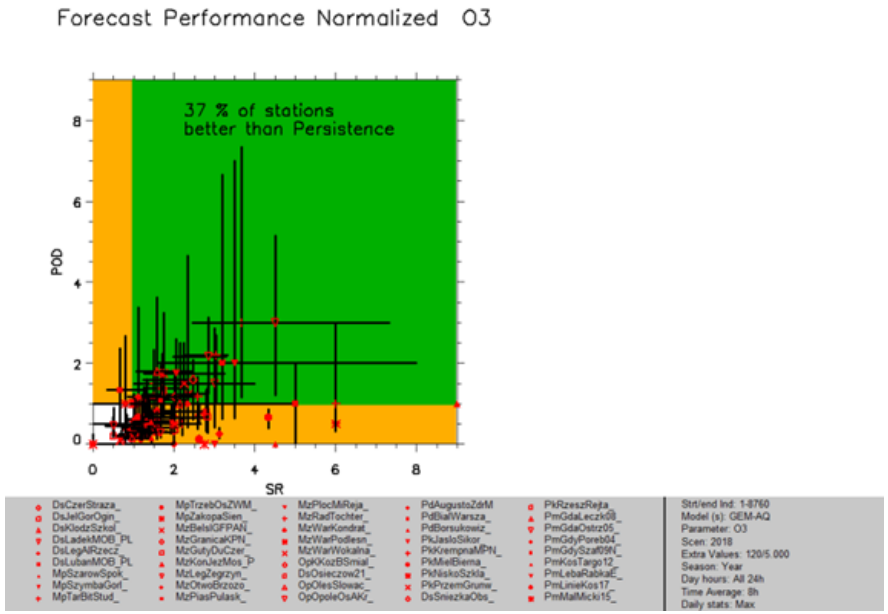


NORMALISED FORECAST PERFORMANCE DIAGRAM (CUVELIER PLOT?)

The idea is to have a yellow zone for stations with POD/PODp or SR/SRp >= 1., Green if both are >= 1., and white if both are <1

» Poland – all stations





CONCLUSIONS

- Indicators and plots:
  - Are they clear and informative?
  - Are they sufficient / fit for purpose of forecast evaluation?

- Persistance model
  - Is it realy this good?
  - Check the methodology and implementation – starting with results for one station

- Time is important!
  - Duration of an episode
  - Check forecast performance (persistance vs evaluated model) on shorter and longer episodes