

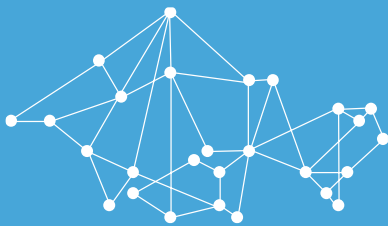


FAIRMODE

Forum for air quality modelling in Europe

STATUS OF THE FORECAST MODEL QUALITY OBJECTIVES AND NEXT STEPS

**PAWEL DURKA ,STIJN JANSSEN, KEES CUVELIER
& PHILIPPE THUNIS**

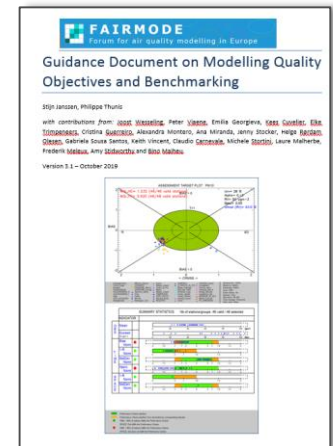


CT3 - Quality control indicators for AQ forecasts



Guidance

- ✓ Elaborate a guidance document on the use of forecast indicators
- Test the proposed indicators (national and CAMS data) – looking for volunteers
- DELTA tool vs 6.0 available from 18.09 (vs 6.01 available from 22.09 – minor bug fix)



FORECAST EVALUATION

General concept for the evaluation metrics

- » 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)

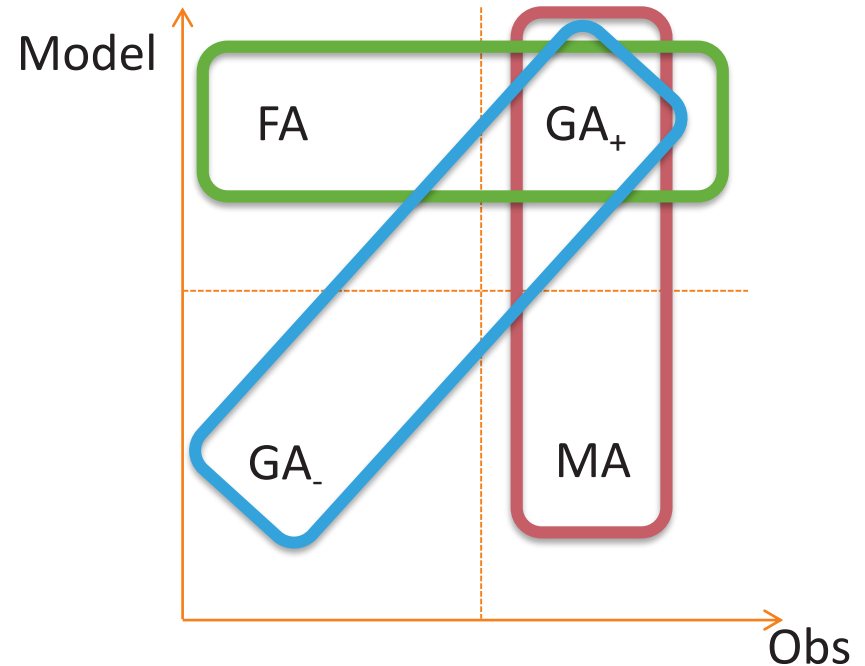
- » For 1. we use the “persistence model” as a benchmark
- » For 2. we use standard threshold indicators

- » Whole methodology in guidance document under QA/QC Protocol for assessment (CT2) - CHAPTER 8 FORECASTING & EXCEEDANCES INDICATORS
https://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/MQO_Guidance_V3.2_online.pdf

THRESHOLD EXCEEDANCES

Definition of threshold exceedance indicators

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



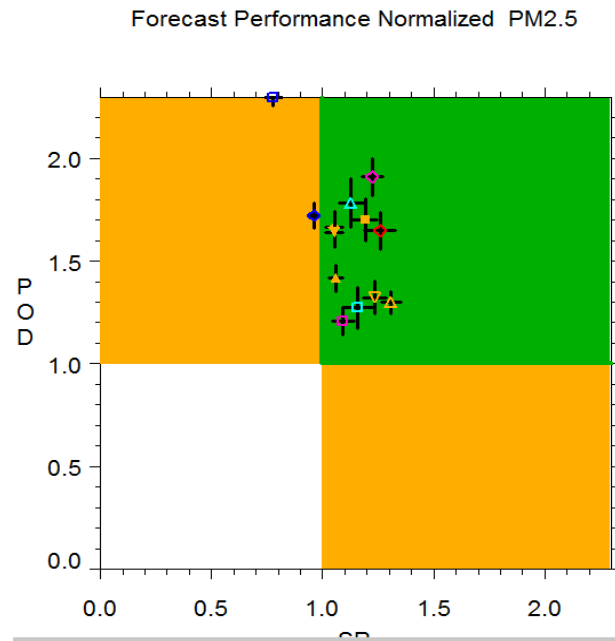
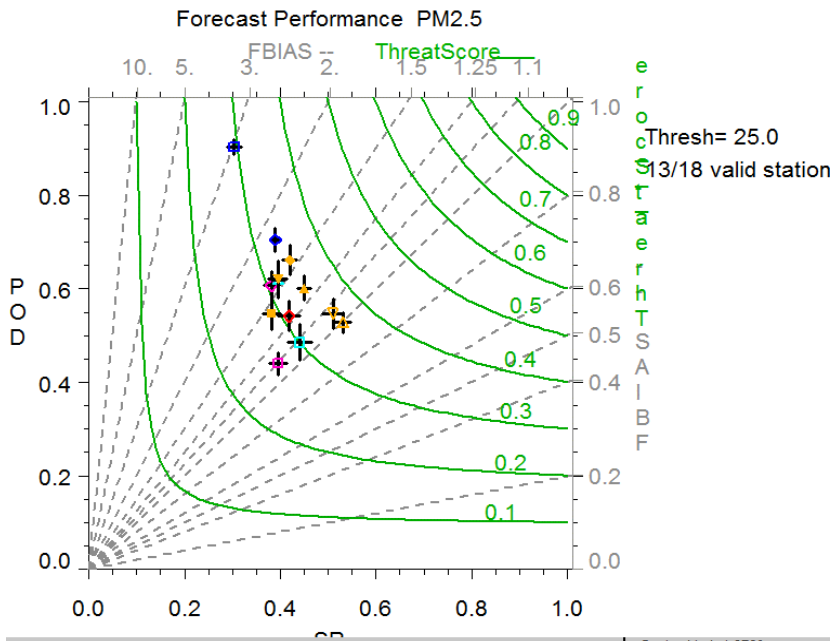
- » **Probability of detection:** $PoD = GA_+ / (MA + GA_+)$
- » **Success ratio:** $SR = 1 - FAR = 1 - FA / (FA + GA_+) = GA_+ / (FA + GA_+)$
- » **Accuracy:** $ACC = (GA_+ + GA_-) / (GA_+ + GA_- + MA + FA)$

FORECAST PERFORMANCE DIAGRAM & CUVELIER PLOT

Performance: combination of POD & SR

Cuvelier Plot: Normalised Performance → do better than Persistence model

“Good enough”:
dots in the green area.



DayForecast= 1
Thresh= 25.0
MQI_(POD/PODp)
MQI_(SR/SRp)= 0
13/18 valid station
84 % of valid station

<ul style="list-style-type: none"> ● BL0 ● BX1 ● BX9 ● CD1 ● CD9 ● CT3 ● GB0 ● GN0 ● GN3 ● GR4 ● GR9 ● HV1 ● KC1 ● LW2 ● MY7 ● ST5 ● TH4 ● VVM0 	<p>Strt/end Ind: 1-8760 Model (s): Day0 Parameter: PM2.5 Scen: 2018 Extra Values: 25 Season: Year Day hours: All 24h Time Average: Preserve Daily stats: preserve</p>
---	---

<ul style="list-style-type: none"> ● BL0 ● BX1 ● BX9 ● CD1 ● CD9 ● CT3 ● GB0 ● GN0 ● GN3 ● GR4 ● GR9 ● HV1 ● KC1 ● LW2 ● MY7 ● ST5 ● TH4 ● VVM0 	<p>Strt/end Ind: 1-8760 Model (s): Day0 Parameter: PM2.5 Scen: 2018 Extra Values: 25/1.000 Season: Year Day hours: All 24h Time Average: Preserve Daily stats: preserve</p>
---	---

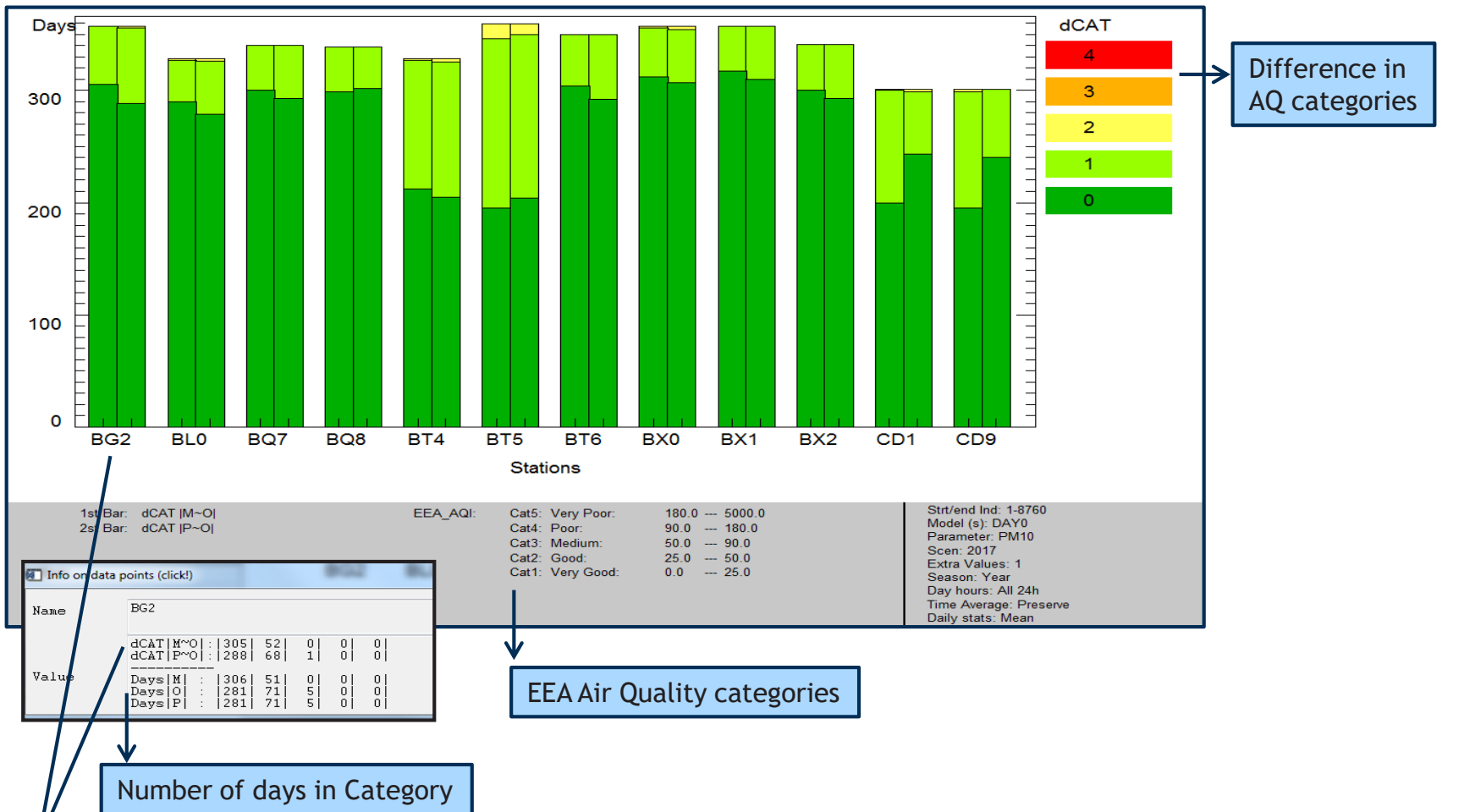
EEA Air Quality Categories

Forecast_AQI

PM10 [$\mu\text{g}/\text{m}^3$]

ForecastDay=1

M=Model, O=Observations, P=Persistence

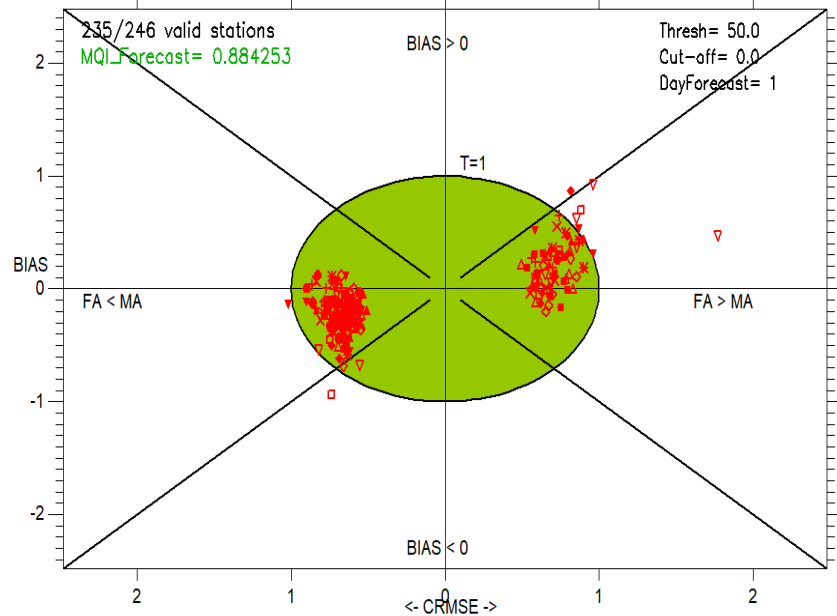


1st Bar: For 305 days M and O forecast the same AQ category (difference dCAT is 0)
 For 52 days the difference in category dCAT between M and O is 1

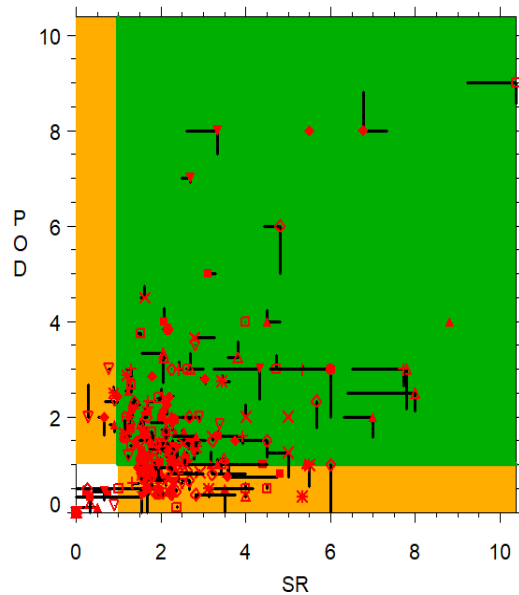
2nd Bar: For 288 days P and O forecast the same AQ category (difference dCAT is 0)
 For 68 days the difference in category dCAT between P and O is 1
 For 1 day the difference in category dCAT between P and O is 2

PM10 FORECAST - YEAR 2019

FORECAST TARGET PLOT PM10



Forecast Performance Normalized PM10



FUTURE PLANS

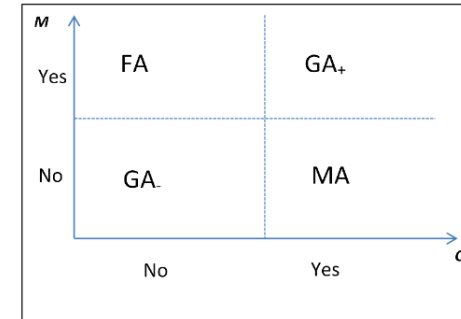
- » Merge of WG1-CT3 mailing list
- » E-mail with on-line hackaton proposition and doodle with possible dates
 - » Results of forecast evaluation on users data sets - countries/regions/data sets with different air quality would be ideal!
 - » Discussion on methodology and possible improvements
 - » Further testing
- » Joint publication- next slide
- » New guidance document (V 3.2) in CT3 activity section on FAIRMODE web



(8 experts, some are missing on the picture)

PUBLICATION OF FORECAST EVALUATION IN DELTA TOOL

- » Guidance document as „core” of the publication
- » User tests with DELTA tool 6.01 - available now to download!
- » Experience, feedback, possible issues or improvements - on hackaton
- » Agreement on final shape of the methodology, as well indicators and diagrams - after hackaton
- » What could be added?



As a consequence, the counted alarms $CA = GA_+ + MA$ and includes all cases where $O >$ threshold.

For a good forecast both FA and MA are small compared to GA_+ and GA_- . Based on these quantities the following indicators can be calculated:

- Probability of Detection: $POD = GA_+ / (MA + GA_+)$
- Success Ratio: $SR = 1 - \text{False Alarm Ratio} = 1 - FAR = 1 - FA / (FA + GA_+) = GA_+ / (FA + GA_+)$

The POD indicator is comparing the correct modelled alerts with the **observed** alerts whereas the SR indicator is comparing the correct modelled alerts with all alerts **issued** by the model.

We also define four additional indicators as:

- FBias score: $FBIAS = (GA_+ + FA) / (MA + GA_+)$
- Accuracy: $ACC = (GA_+ + GA_-) / \text{Total}$
- Threat score: $TS = GA_+ / (MA + FA + GA_+) = GA_+ / (FA + GA_+)$
- Gilbert Skill score: $GSS = (GA_+ - H_{\text{random}}) / (MA + FA + GA_+ - H_{\text{random}})$
with $H_{\text{random}} = (GA_+ + MA)(GA_+ + FA) / \text{Total}$

- » Feedback from countries with different „air quality problems” - in terms of pollutant and/or number of exceedances
- » Is this a useful set of evaluation tools?
- » Is the persistence model a realistic benchmark?
- » What about cut-off value?
- » What could be added to planned publication?