European Commission

FAIRMODE
Forum for air quality modelling in Europe

# WG3 session
# Forecast

*FAIRMODE Plenary Meeting*

*Paris - France, February 26-27 2024*

## Status, next steps & discussion

# AGENDA

- **Status and future steps**: points for technical discussion in 2024 and proposal for hackathon in spring

- **Discussion** - how to establish the fitness for purpose of a forecast model application? Is fulfilling the $MQO_f$ sufficient?

# STATUS

*All the valuable contributions and input received during the last FAIRMODE Technical Meeting (Athens, October 4-6, 2023) were collected and considered*

- [x] Some questions/doubts on the methodology have been already addressed and a questions & answers document was distributed to WG3 participants (20/12/2023)

- [x] Most of the technical issues and graphical improvements suggested by WG3ers were considered → a new beta version of the DELTA Tool is available including the suggested developments (*Thanks @Kees*!)

- [ ] Concept issues and ideas for developments were considered and summarized according to the three main WG3 topics → ***today's discussion***

# 1. Comparison with the Persistence Model

Main issue:

*Should we plan to make MQI criterion stricter?*

The new AQUILA-based parameters are now available for measurement uncertainty estimates, which will make the criterion stricter

A beta version of the Delta Tool is now available enabling the choice of two different set of parameters for measurement uncertainty estimates: the current parameters and the new ones based on AQUILA

# 2. CAPABILITY IN PREDICTING EXCEEDANCES

Main issue:

*What could be the target for predictability of episodes?*
*Should the model have to be better than persistence for all stations?*

WG3ers' outcomes showed that Persistence Model is very difficult to beat in predicting exceedances

→ **Proposal**: by default, model performances skills in predicting Exceedances are presented without the comparison with Persistence Model ones.

The comparison with the Persistence Model can be included by the user to better understand the results but it is not mandatory

**In perspective, WG3 should work on defining objective criteria for threshold exceedance's indicators (POD, SR, ACC, FBIAS, TS, GSS), also considering the new daily limit for PM2.5**

# 3.CAPABILITY IN PREDICTING AIR QUALITY INDICES

Two main proposals arise from WG3ers feedback

*A) Including the correct timing of the forecasted AQI levels*

*B) Multi-pollutant approach*

We propose starting from point A, since point B needs more effort, due to the single-pollutant approach currently adopted within the Delta Tool

# 3.CAPABILITY IN PREDICTING AIR QUALITY INDICES

## PROPOSAL FOR INCLUDING THE CORRECT TIMING EVALUATION

ANALYSIS BASED ON **MULTI-CATEGORY CONTINGENCY TABLE** (example for 5 Classes)

| | | Forecast | | | | | |
|---|---|---|---|---|---|---|---|
| | | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | $TOT_{obs}$ ↓ |
| Observations | Class 1 | $n_{1,1}$ | $n_{1,2}$ | $n_{1,3}$ | $n_{1,4}$ | $n_{1,5}$ | $O_1$ |
| | Class 2 | $n_{2,1}$ | $n_{2,2}$ | $n_{2,3}$ | $n_{2,4}$ | $n_{2,5}$ | $O_2$ |
| | Class 3 | $n_{3,1}$ | $n_{3,2}$ | $n_{3,3}$ | $n_{3,4}$ | $n_{3,5}$ | $O_3$ |
| | Class 4 | $n_{4,1}$ | $n_{4,2}$ | $n_{4,3}$ | $n_{4,4}$ | $n_{4,5}$ | $O_4$ |
| | Class 5 | $n_{5,1}$ | $n_{5,2}$ | $n_{5,3}$ | $n_{5,4}$ | $n_{5,5}$ | $O_5$ |
| | $TOT_{forecast}$ → | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ | |

$n_{i,i}$ Number of Correct occurrences in Class i

$O_i$ Number of Total Observed occurrences in Class i

$F_i$ Number of Total Forecast occurrences in Class i

A **Multi-category Contingency Table** can be created for each station

A **"cumulative" Multi-category Contingency Table** can be created where $n_{i,j}$ are obtained as the sum of the corresponding $n_{i,j}$ values in the Contingency Tables of all the stations

**Multi-category skill scores** can be computed and plotted (*)

$$POD_{AQI} = \left(\frac{n_{i,i}}{O_i}\right)$$

$POD_{AQI}$ represent the percentage of correct events in each class $i$ with respect to the total events based on measurements

$$SR_{AQI} = \left(\frac{n_{i,i}}{F_i}\right)$$

$SR_{AQI}$ represent the percentage of correct events in each class $i$ with respect to the total events based on forecast

(*) e.g. EPA, 2003: *Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program*; Stanski, H.R., L.J. Wilson, and W.R. Burrows, 1989: *Survey of common verification methods in meteorology*. World Weather Watch Tech. Rept. No.8, WMO/TD No.358, WMO, Geneva

# FUTURE STEPS

**Proposal of an exercise to be launched**

*The new beta version of the Delta Tool will be distributed to participants in the exercise in order to test the effect of changing the current parameters for measurement uncertainty estimates (goals_criteria_oc_OLD.dat) with the new AQUILA-based parameters (goals_criteria_oc_NEW.dat), which will make the criterion stricter*

**An hackathon is planned 7 or 8 May 2024**

- *Results and feedback of the exercise will be shared and discussed*
- *We will continue today's discussions (next slide)*

*Who would be potentially interested in joining the exercise and the following hackathon?*

# FUTURE STEPS

**Short-time perspective (2024 - AAQD guidance - Fairmode guidance):**
**1 criterion for forecast evaluation**

**Beating persistence:** $MQO_f < 1$ = **fitness for purpose**

# FUTURE STEPS

## Medium-time perspective (202x?): tiered approach for forecast evaluation

**Tier 1 - beating persistence:** $MQO_f$ <1 = **necessary but not sufficient for fitness for purpose**

**Tier 2 - beating persistence & catching episodes:** Tier 1 + criteria/goals satisfied for threshold exceedance's indicators (POD? SR? GSS?) = **fitness for purpose**

**Tier 3 - beating persistence, catching episodes & catching AQI levels:** Tier 2 + criteria/goals satisfied for AQI indicators ($POD_{AQI}$? $SR_{AQI}$?) = **fitness for purpose**

# DISCUSSION

How to establish the fitness for purpose of a forecast model application?

- Is fulfilling the $MQO_f$ sufficient?
- Objective criteria for threshold exceedance's indicators (POD, SR, ACC, FBIAS, TS, GSS): a medium/long-term perspective? Do you agree with the proposed approach?
- Including the correct timing within AQI evaluation: comments on our proposal? Ideas?