



Slides based on:

**8th meeting of
CEN/TC 264/WG 43 “Ambient air – Modelling quality objectives”**

**18 and 19 September 2018
Warsaw (Poland)**

Ari Karppinen



The project proposal/submitted longer WP descriptions (homework)



1.	Work package	Description	Estimated budget (kEur)
	1	Project coordination	40
	2	Open source R-tool for the statistic calculations + UI (Philippe)	50
	3	Statistics : assessment of measurement uncertainty parameters and proportionality factor: robustness (Philippe)	80
	4	Statistics : assessment of measurement uncertainty parameters and proportionality factor: stringency (Stefan)	80
	5	Measurements requirements for modelling validation:, requirements for measurement time series and requirements for measurement campaigns (Joost+Stijn)	70
	6	Requirements for data driven modelling systems (data – assimilation , data fusion, bias correction) (French team+Heinke+Helen)	80



Station representativeness

- * In the current approach only the uncertainty related to the measurement device is accounted for but another source of divergence between model results and measurements is linked to the lack of spatial representativeness of a given measurement station (or to the mismatch between the model grid resolution and the station representativeness). Although objectives regarding the spatial representativeness of monitoring stations are set in the AQD these are not always fulfilled in real world conditions. The formulation proposed for the MQI and MPI could be extended to account for the lack of spatial representativeness if quantitative information on the effect of station (type) representativeness on measurement uncertainty becomes available.



Station representativeness



(may be not the right approach at all...)

- * Current status: In order to clarify ambiguities in the definition, interpretation and assessment methodologies of spatial representativeness, an inter-comparison exercise was conducted within FAIRMODE. Different teams were encouraged to apply their preferred methodology and interpretation framework in a common case study. **A report summarizing the inter-comparison exercise** and putting forward a set of **recommendation** is now available at the FAIRMODE website under the WG1 documents section.



Performance criteria for high percentile values (important! For forecasting & episodes)



The MQI and MPI described in this document provide insight on the quality of the model average performances but **do not inform on the model capability to reproduce extreme events** (e.g. exceedances). For this purpose, a specific *MPI* indicator is proposed as: $MPI_{perc} = |M_{perc} - O_{perc}| \beta U_{95}(O_{perc})$ and *MPC*: $MPI_{perc} \leq 1$ (35)

- * **Current status:** Apart from the extension of the MQI for percentiles as described above, a specific evaluation of model performance for episodes is now implemented in the test/expert version of DELTA vs5.3 and higher. The threshold evaluation criteria as implemented for forecast models (see further in Section 10) can also be used for the evaluation of episodes.



Hourly/daily versus annual MQI (minor issue)



- * Background: Some model applications seem to pass the hourly/daily objective but apparently fail to meet the criteria when annual averaged values of the time series are used in the annual MQI procedure.
- * **Current status:** As a pragmatic solution to the above mentioned problem, it is suggested that model applications with hourly/daily output should also comply with the annual MQO. In DELTA vs5.3 and higher both criteria are implemented in one graph.



Data availability

(no absolutely correct number existing)



Currently a value of 75% is required in the benchmarking both for the period considered as a whole and when time averaging operations are performed for all pollutants.

Current status: The 75% criteria was a pragmatic choice when the methodology was elaborated. It can be questioned if this is still a valid choice.

A higher value for this criterion will limit the number of stations available for the evaluation whereas a smaller criteria value leads to truncated comparisons (only a small fraction of the year is indeed evaluated in these cases).



Data assimilation



(important! /not really resolved)

The AQD suggests the integrated use of modelling techniques and measurements to provide suitable information about the spatial and temporal distribution of pollutant concentrations. When it comes to validating these integrated data, different approaches can be found in literature that are based on dividing the set of measurement data into two groups, one for the data assimilation or data fusion and one for the evaluation of the integrated fields. The challenge is how to select the set of validation stations.

For the time being, FAIRMODE recommends the “**leaving one out**” validation strategy as a methodology for the evaluation of data assimilation or data fusion results. It has to be noticed that such an approach (might) not be appropriate for on-line data assimilation methodologies (4D VAR, Ensemble Kalman Filter...) due to computational constraints. In such cases, an a priori selection of assimilation and validation stations has to be made. However, the modeler should be aware of the fact that this a priori selection of validation stations will have an impact on the final result of the evaluation of the model application.



Model benchmarking and evaluation for zones with few monitoring data

During the 2016 Plenary Meeting it was put forward by a number of participants that the FAIRMODE Modelling Quality Objective might not be fully applicable for urban scale applications. In many cases, only a limited set of monitoring stations is available in a single city or town. When less than 10 stations are at hand, the 90% criteria for the Target value requires that a model application fulfil the MQO criteria in all available stations, reducing the level of tolerance which is available for regional applications. In the new formulation of the MQO which implicitly takes into account the 90% principle (§7.3.3) accounts for this shortcoming to a certain level. **It still remains questionable what a minimum number of stations should be to evaluate a modelling application for a specific (urban) region. (..or – does it ??)**

In addition, it is noticed that at the urban scale additional auxiliary monitoring data sets might be available (e.g. passive sampling data, mobile or temporary campaigns). Those monitoring data might **(or simply: are!)** be very valuable to check the quality of the urban applications but at present the FAIRMODE Benchmarking procedure is not capable to deal with those observations.



Application of the procedure to other parameters



- * Currently only PM, O₃ and NO₂ have been considered but the methodology could be extended to other pollutants such as heavy metals and polyaromatic hydrocarbons which are considered in the Ambient Air Quality Directive 2004/107/EC.
- * The focus is clearly on applications related to the AQD and thus those pollutants and temporal scales relevant to the AQD. However the procedure can of course be extended to other variables including meteorological data as proposed in Pernigotti et al. (2014)



Conclusions



- * Project proposal submitted to CEN – but – it may take quite a substantial time to get the project started
- * Meanwhile- the group is updating the technical document: still several important points to consider (like the ones presented here) – but- we firmly believe that the standard will be reality .. in 3-4 years
- * Other standards may follow : e.g.French initiative on European model standard (vs. just evaluation standard)