Proposal for a QA/QC protocol to support modelled assessments of air quality

Authors: P. Thunis, L. Tarrasón

Contributors:

Version 1.0 - 4/15/2020

FAIRMODE
Forum for air quality modelling in Europe

Contents

1. Introduction

The assessment of the air quality situation in different areas over Europe is increasingly relying on a widespread use of air quality dispersion models. The ability of these modelling assessments to capture both the spatial and temporal variability of the air quality situation becomes therefore an important question, especially when such modelling results are to be used to support environmental policies. Systems for the quality assessment and control (QA/QC) of the results from modelling applications are normally based on a quantitative comparison of the model output to observations. In particular, the Modelling Quality Objective (MQO) methodology developed by FAIRMODE, and currently discussed within CEN TC264/43, is generally used to evaluate the quality of the results of a given modelling application. Such type of quality assessment provides a good measure of the accuracy of the modelling results.

For air quality management purposes, however, the overall accuracy of the modelling results is not the only condition to determine the quality of a modelling application. The MQO procedure is recognised as a necessary step by the modelling community but is however not sufficient to guarantee that the complete modelling chain used to produce the results is error-free. Nor does it guarantee that the spatial and temporal variability of the air quality situation is accurately considered. Compensation of errors either in the model itself but also in its input data can occur. This document proposes a QA/QC protocol to (1) ensure that sufficient information (metadata) is associated to the modelling results for their interpretation and to (2) control the quality of the modelling results with different QA/QC tests to account for the variability of the air quality modelled situation.

In conclusion, the MQO is not sufficient to guarantee high quality modelling results. We need additional metadata information and QA/QC tests. These are proposed in this document.

2. Scope

Following the Commission Implementing Decision of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament (AAQDs) as regards the reciprocal exchange of information and reporting on ambient air quality (IPR decision, 2011/850/EU), modelling results can be used for assessment purposes. When modelling results are used for assessment of exceedances to limit values under the AAQDs, Article 9, Paragraph 5 requests that at least the following information is reported (a) the description of the modelling system and its inputs; (b) the model validation through measurements; (c) the coverage area; (d) the documentation of data quality.

The aim of this document is to propose a QA/QC protocol to ensure the quality of modelling results for assessment applications in the context of the AAQDs, to respond to some of the requirements in the IPR decision. It fulfils this scope by:

- **Providing recommendations on the preferred information on the modelling application** and its input data to be collected to document the application (metadata information). This metadata information could be compiled and made available as part of the FAIRMODE composite mapping platform in a first stage to test the capabilities of the QA/QC system.
- **Proposing a specific suite of QA/QC tests** to help understanding the capability of the modelling application to reproduce the existing air quality situation, both in terms of its

magnitude as well as in terms of its spatial and temporal variability. This information is to be derived in comparison with observations and could be summarized via specific diagrams and summary tables, following a harmonised template.

3. Which models, which pollutants?

The proposed QA/QC protocol is to be applied to establish the quality of all modelling applications used for assessment purposes in the context of the AAQDs.

This includes all types of modelling approaches, whether these are statistical, Gaussian, Lagrangian or Eulerian. The protocol is designed so that it can also be applied to computation fluid dynamic models and models using artificial intelligence. Here it is important to mention that it is not the models that are evaluated but the modelling application that uses such models.

The pollutants considered in the proposed QA/QC protocol are in first instance particulate matter, nitrogen dioxide and ozone: **$PM_{2.5}$, PM_{10}, NO_2 and O_3.**

4. Requisites and assumptions

Guidance on fitness for purpose is already provided in the FAIRMODE documents (EEA technical guide, PM and NO2 FAIRMODE guidance). A few extracts are reproduced in the Annex 1 of this document for illustration. **In the context of this QA/QC protocol, we assume that the modelling results originate from fit-for-purpose models**.

The FAIRMODE guidance document on "MQO and benchmarking"[1] provides practical recommendations on how to perform the evaluation process. Such recommendations are currently under consideration by the work of CEN TC264/43. We assume in the context of this QA/QC protocol that all recommendations contained within these documents apply (e.g. which measurements to use? how frequent should the assessment be? what to do when sufficient measurements are not available?)

Finally, we also assume that the **station classification and station types currently proposed by the IPR apply** in the context of this QA/QC protocol. When various (sub)models in a modelling chain need validation, this IPR concept of "station type/classification" will be efficient to select appropriate stations for partial evaluations (one sub-model) of the modelling chain. The same classification will be very useful to structure the QA/QC tests within the protocol.

5. QA/QC Protocol: documentation of modelling application

Harmonised documentation on the modelling application should include all information necessary to understand model results and compare them to those obtained from other modelling applications. If we consider that **the general 'fit-for-purpose' criteria proposed in EEA (2011**) should apply, the information compiled should indicate whether:

- the model has the appropriate spatial and temporal resolution for the intended application;
- the model is adequately validated for the particular application, and is well documented;

[1] https://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/Guidance_MQO_Bench_vs3.1.1.pdf

- the model contains the relevant physical and chemical processes suitable for the type of application, the scale and the pollutant for which it is applied;
- the relevant emission sources for the application are adequately represented;
- suitable meteorological data are available

The IPR Decision already indicates a series of harmonised information to be reported concerning modelling applications used for air quality assessment. The actual requirements from the IPR Decision on the modelling information are those listed in Table 1 below.

Table 1: IPR Decision documentation requirements for assessment methods using modelling

IPR Decision documentation requirements for assessment methods using modelling **IPR dataflow (D) Information on the assessment methods (Articles 8 and 9)** (iv) *Modelling Information*	
(1)	Modelling ID
(2)	Environmental objective type (data type 'Environmental Objective')
(3)	Modelling method: name
(4)	Modelling method: description
(5)	Modelling method: documentation (web link)
(6)	Modelling method: validation by measurement
(7)	Modelling method: validation by measurement at sites not reported under the AQD
(8)	Modelling period
(9)	Area for modelling (data type 'Spatial Extent')
(10)	Spatial resolution
(11)	Assessment method for winter sanding and salting (where Article 21 of Directive 2008/50/EC applies)
(12)	Assessment method for natural contribution (where Article 20 of Directive 2008/50/EC applies)
(13)	Data quality objectives: uncertainty estimation
(14)	Data quality objectives: documentation of QA/QC (Web Link)

We propose to complement the IPR requirements on the basis of existing sources of information such as

1. **The Model Documentation System** (suggestion of key fields proposed in green in Table 2).
2. **Existing national procedures**: see for example the French AASQA audit (see Annex 2)

Table 2: Model documentation grid (as extracted from the EEA technical guide[2]

MDS descriptive topics	**Sub-topics**	
Basic information	Model name	
	Model version and status	
	Latest date of revision	
	Contact information	
	Level of knowledge needed to operate model	
Intended field of application		
Model type and dimension		
Model description summary		

[2] The application of models under the European Union's Air Quality Directive: A technical reference guide, EEA, 2011

Model limitations/approximations		
Resolution	Temporal resolution	
	Horizontal resolution	
	Vertical resolution	
Schemes	Advection & Convection	
	Turbulence	
	Deposition	
	Chemistry	
Solution technique		
Input	Availability and Validation of Input data	
	Emissions	
	Meteorology	
	Topography	
	Initial condition	
	Boundary conditions	
	Data assimilation options	
Output quantities		
User interface availability		
User community		
Previous applications		
Documentation status	Ranking levels 1 – 5	
Validation and evaluation	Ranking levels 1 – 5	
	Model inter-comparison	
Frequently asked questions		

At this stage, other national procedures, or other sources of information exists and are welcome. Exchanges of views will be organised during the next technical meetings to identify the key fields to pick from these information source to build a harmonised reporting of metadata.

The IPR requirements also call for specific QA/QC tests (fields 13 and 14 in Table 1). The next sections propose a way forward to fulfil this task.

6. QA/QC Protocol: Benchmarking tests

The QA/QC benchmarking tests below are a first proposal for a possible protocol to determine how a modelling application actually is able to establish the spatial and temporal variability of air concentrations in the modelled domain to support air quality assessment applications. The proposed protocol is based on the outcome of the discussions during the past meetings as well on information provided by some modelling groups on their QA/QC procedure (see Annex C with QA/QC protocol for CHIMERE, applying to their regional scale modelling applications).

Additional QA/QC protocols for assessment applications are needed to reinforce this section! Please feel free to add information about any existing protocol in your country and/or complement the current proposal.

STEP 1. Current MQO

For this first step, we follow the formulation proposed and discussed within FAIRMODE and CEN TC264/43. The Modelling Quality Indicator (MQI) is a statistical indicator of the accuracy of a specific modelling application calculated based on measurements and modelling results. It is defined as the ratio between the model-measured bias at a fixed time (i) and a quantity proportional to the measurement uncertainty as:

$$\mathrm{MQI(i)} = \frac{|\mathrm{O_i} - \mathrm{M_i}|}{\beta U(\mathrm{O_i})} \quad (1)$$

Where U(Oi) is the measurement uncertainty and β a coefficient of proportionality. The normalisation of the bias by the measurement uncertainty is motivated by the fact that both model and measurements are uncertain. We want to account for the fact that when measurement uncertainty is large, some flexibility on the model performance can be accepted, translating in accepting larger model-observed errors. With a current value of 2 proposed for β, the quality of a modelling application is said to be sufficient when the model-observation bias is less than twice the measurement uncertainty.

Applied to a complete time series, Equation (1) can be generalized to:

$$MQI = \frac{\mathrm{RMSE}}{\beta RMS_U} \quad (2)$$

With this formulation, the RMSE between observed and modelled values (numerator) is compared to a value (RMS_U) representative of the maximum allowed measurement uncertainty (denominator).

For yearly averaged pollutant concentrations, the MQI formula is adapted so that the mean bias between modelled and measured concentrations is normalised by the uncertainty of the mean measured concentration:

$$MQI = \frac{|\bar{O} - \bar{M}|}{\beta \mathrm{U}(\bar{\mathrm{O}})} \quad (3)$$

More details on formulation (1), (2) and (3) can be found in the MQO guidance document.

For the evaluation of the model results against measurements, we apply the following two rules:

- Following the FAIRMODE recommendation, both equations (2) where relevant and (3) must be fulfilled, i.e. $MQI \leq 1$.
- Following the requirements prescribed by the EU Ambient Air Quality Directives (AAQD), equations (2) where relevant and (3) must be fulfilled ($MQI \leq 1$) for at least 90% of the available measurement stations.

Dedicated diagrams have been proposed to detail the information contained in the MQI (*Figure 1*). They provide a common set of key statistical indicators, bias, standard deviation and correlation as well as a harmonised way of visualising them.

The main drawback of the MQOs is that they provide a single summary pass/fail information for a modelling application. This simple test does not prevent a modelling application to pass for the

wrong reason, under certain circumstances. In addition, it does not provide any information on the capability of the model to reproduce hot spot areas (spatial variability) or the timing of the pollution peaks (temporal variability). This information is key in the context of the AAQD and is only partially addressed in the current MQO proposal. This is why complementary tests are required and proposed below for a more exhaustive QA/QC process.

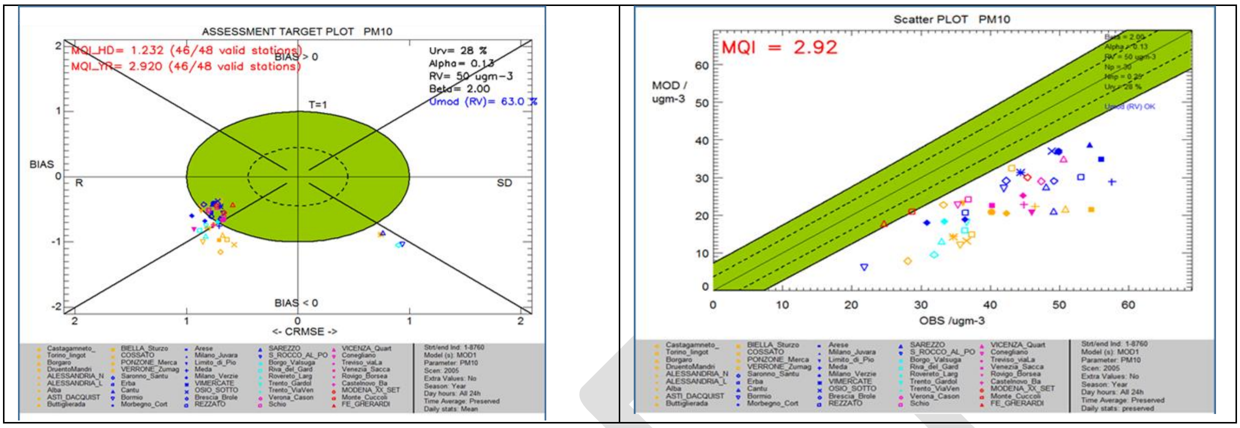


Figure 1: MQO diagrams for daily/hourly (left) and yearly values (right). Each point represents the model results for one measurement station. The Target diagram (Figure 5) is designed to visualize the MQI and its components. The MQI represents the distance between the origin and a given station (represented by a point on the diagram). This distance should be less than unity, i.e. fall within the green circle, for at least 90% of the available stations. The MQI for the yearly or seasonal averaged results are generally more challenging to fulfill. Equation (3) for yearly or season averaged results (i.e.) is based on bias only and is used as the main model quality indicator. In a scatter plot, the MQI is used to represent the distance from the 1:1 line.

STEP 2. Additional indicators for spatial and incremental variability

Within the benchmarking reporting template discussed in the MQO guidance, specific indicators, based on spatial correlation and spatial standard deviations, normalised by the measurement uncertainty, are proposed to assess the capacity of models to capture the spatial variability of the measurements. These indicators are constructed as follows.

For hourly frequency model output, values are first averaged yearly at each station. A spatial correlation and a spatial standard deviation indicator are then calculated for this set of values. The two indicators are normalised by the measurement uncertainty of the average concentrations: $\mathrm{RMS}_{\bar{U}} = \sqrt{\frac{1}{N}\sum \mathrm{U}(\bar{O})^2}$. The same approach holds for yearly frequency output.

These indicators are defined as:

	Model Performance Indicator (MPI) (currently existing)	Model Performance Criteria (MPC)
Correlation	$\mathrm{MPI} = \dfrac{1-\mathrm{R}}{0.5\beta^2 \dfrac{RMS_{\bar{U}}^{\,2}}{\sigma_O \sigma_M}}$	MPC: MPI $\leq$ 1
Standard deviation	$MPI = \dfrac{\lvert\sigma_M - \sigma_O\rvert}{\beta RMS_{\bar{U}}}$	

Where the Model performance criteria is the criteria to be fulfilled in order to reach the quality objective of the modelling application.

On top of these already agreed indicators included in FAIRMODE MQI system approach, we propose to complement them with incremental indicators, where relevant[3], to assess how concentration gradients between rural and urban or between traffic and urban stations are reproduced by the model. This is relevant in the context of the AAQDs because the design of the monitoring network aims to capture existing gradients and differences occurring as a result of different dominant pollution sources and dispersion situations. These additional spatial indicators can be constructed similarly to other MQIs, i.e. normalised by the measurement uncertainty. The implementation of these indicators in the DELTA tool is discussed in Annex 4.

For example, the modelled incremental change between rural background (RB) and urban background (UB) locations is defined as

$$INC_{UB-RB}^{model} = \bar{M}_{UB} - \bar{M}_{RB}$$

with M is the model value and similarly for the measured increment:

$$INC_{UB-RB}^{observed} = \bar{O}_{UB} - \bar{O}_{RB}$$

These indicators are then normalised by the measurement uncertainty.

	Model Performance Indicator (MPI) (to be implemented)	Model Performance Criteria (MPC)
UB - RB	$\text{MPI} = \frac{INC_{UB-RB}^{model} - INC_{UB-RB}^{observed}}{\beta * 0.5 * (RMS_{\overline{U(UB)}} + RMS_{\overline{U(RB)}})}$	$\text{MPC: MPI} \leq 1$
UT - UB	$\text{MPI} = \frac{INC_{UB-UT}^{model} - INC_{UB-UT}^{observed}}{\beta * 0.5 * (RMS_{\overline{U(UB)}} + RMS_{\overline{U(UT)}})}$	

where UT stands for "urban traffic". No harmonised graphical template yet exists for these new indicators, but the existing ones that cover the spatial correlation and standard deviation could be generalised.

The proposal is to work within FAIRMODE to include these additional spatial indicators to the current graphical MQO system.

STEP 3. Additional indicators for temporal variability

The MQO generally applies to the average of a specific period, currently, one year. Consequently, it provides no information whether the modelling application manages to capture the actual temporal variability of the air quality situation. Since the AAQDs include also in the assessment the evaluation of exceedances for specific temporal indicators, the capability of the modelling application to

[3] Indicators can only be applied with models that are designed to simulate the station types that are used in the indicators (e.g. urban-traffic incremental indicators cannot be applied to models that only simulate background levels).

reproduce the temporal variations becomes highly relevant in the context of air quality management.

Although challenging, the current MQO and benchmarking proposal already includes a specific indicator to see how well the model capture high concentrations. This indicator is currently formulated as follows.

	Model Performance Indicator (MPI)	Model Performance Criteria (MPC)
	$MPI_{perc} = \frac{\lvert M_{perc} - O_{perc} \rvert}{\beta U(O_{perc})}$	$\mathrm{MPC{:}\ MPI} \leq 1$

where “perc” is a selected percentile value and M_{perc} and O_{perc} are the modelled and observed values corresponding to this selected percentile. The denominator, $U(O_{perc})$ is directly given as a function of the measurement uncertainty characterizing the O_{perc} value. An alternative to the formulation of the MQI for percentiles as described above, the evaluation criteria as implemented for forecast models can also be used for the evaluation of high percentiles and episodes.

We propose to complement the existing “high concentration indicator” with additional indicators to assess the temporal coherence of model results, at different frequencies: seasonal, week/week-end or day/night. Measurement and modelling results are then aggregated (all stations belonging to a certain type (urban – rural –traffic – industrial) together and checks are made through the following indicators:

		Model Performance Indicator (MPI) (to be implemented)	Model Perf. Criteria (MPC)
Seasonal	Industry	$\mathrm{MPI} = \frac{SeasDiff_{Ind}^{mod} - SeasDiff_{Ind}^{obs}}{\beta RMS_{\bar{U}}}$	$\mathrm{MPC{:}\ MPI} \leq 1$
	Traffic	$\mathrm{MPI} = \frac{SeasDiff_{traffic}^{mod} - SeasDiff_{traffic}^{obs}}{\beta RMS_{\bar{U}}}$	
	Background	$\mathrm{MPI} = \frac{SeasDiff_{bg}^{mod} - SeasDiff_{bg}^{obs}}{\beta RMS_{\bar{U}}}$	
Week / weekend	Industry	$\mathrm{MPI} = \frac{WeekDiff_{Ind}^{mod} - WeekDiff_{Ind}^{obs}}{\beta RMS_{\bar{U}}}$	
	Traffic	$\mathrm{MPI} = \frac{WeekDiff_{traffic}^{mod} - WeekDiff_{traffic}^{obs}}{\beta RMS_{\bar{U}}}$	
	Background	$\mathrm{MPI} = \frac{WeekDiff_{bg}^{mod} - WeekDiff_{bg}^{obs}}{\beta RMS_{\bar{U}}}$	
Day/night	Industry	$\mathrm{MPI} = \frac{DayDiff_{Ind}^{mod} - DayDiff_{Ind}^{obs}}{\beta RMS_{\bar{U}}}$	

	Traffic	$\text{MPI} = \frac{DayDiff_{traffic}^{mod} - DayDiff_{traffic}^{obs}}{\beta RMS_{\bar{U}}}$	
	Background	$\text{MPI} = \frac{DayDiff_{bg}^{mod} - DayDiff_{bg}^{obs}}{\beta RMS_{\bar{U}}}$	

The additional steps suggested below could complement the QA/QC protocol described so far (steps 1 to 3). They are much more challenging and demanding. We welcome any suggestion/comments on their relevance and usefulness.

Input consistency (emission, meteorology…)

Specific input to the air quality models (i.e. emissions) can be tested for consistency. Overall consistency/plausibility of the model input can be checked through the use of specific indicators (e.g. max-min values compared to references...). Inconsistencies can be flagged out to the user. As this type of tests does not exist yet, this step is very challenging.

Ex-Post Assessment/Multi-annual variability

The QA/QC analysis (STEPs 1 to 3) can be repeated tentatively every 3 or 5 years, including different air quality situations. The objective of this step is to check that good MQO performances occur not only for a given year but for at least two years, that differ significantly in terms of emissions, meteorology…. In addition to their individual yearly evaluation, the difference between the two simulations (modelled delta) could be compared with measurements (observed delta), possibly normalised by measurement uncertainty.

Multi pollutants checks

We can probably learn from looking at pollutant ratios. Specific indicators can be developed to tests these aspects. Again, this is a very challenging step.

Annex 1: Fit for purpose modelling

References to EEA technical guide (2011)[4]

Table 3.1 List of typical model characteristics, formulations and processes, for the various scales and pollutants needed for air quality assessment

	Area of assessment		
Description	**Local/hotspot (1–1 000 m)**	**Urban/agglomerate (1–300 km)**	**Regional (25–10 000 km)**
Model type	Gaussian and non-Gaussian parameterised models Statistical models Obstacle-resolving fluid dynamical models Lagrangian particle models	Gaussian and non-Gaussian parameterised models Eulerian chemical transport models Lagrangian particle models	Eulerian chemical transport models Lagrangian chemical models
Meteorology	Local meteorological measurements Obstacle-resolving fluid dynamical models Diagnostic wind field models	Mesoscale meteorological models Localised meteorological measurements Diagnostic wind field models	Synoptic/mesoscale meteorological models
Chemistry	Parameterised or none	Ranging from none to comprehensive, depending on application	Comprehensive
Emission modelling	Bottom-up traffic emissions Source-specific emissions	Bottom-up and/or top-down emission modelling Emission process models	Top-down emission modelling Emission process models
Compound	**Local/hotspot**	**Urban/agglomerate**	**Regional/continental**
PM_{10}	No chemical processes	Deposition Secondary inorganic particle formation	Deposition Primary (combustion) particles Secondary inorganic and organic particle formation Suspended dust Sea salt
$PM_{2.5}$	No chemical processes	Deposition Secondary inorganic particle formation	Deposition Secondary inorganic and organic particle formation
NO_2	Simple photo-oxidant chemistry Statistical/empirical relations	Limited photo-oxidant chemistry Photo-stationary scheme Statistical/empirical relations Deposition	Deposition Full photo-oxidant chemistry
NO_x	No chemical processes	No chemical processes Full photo-oxidant chemistry for larger scales	Full photo-oxidant chemistry
O_3	As in NO_2	As in NO_2	As in NO_2
SO_2	No chemical processes	Deposition Secondary inorganic particle formation	Deposition Secondary inorganic particle formation Full photo-oxidant chemistry
Pb	No chemical processes	Deposition No chemical processes	Deposition Specialised chemical schemes
Benzene	No chemical processes	n/a	Deposition Full photo-oxidant chemistry
CO	No chemical processes	No chemical processes	Full photo-oxidant chemistry
Heavy metals and B(a)P	No chemical processes	Deposition Specialised chemical schemes	Deposition Specialised chemical schemes

[4] The application of models under the European Union's Air Quality Directive: A technical reference guide, EEA, 2011

References to guide on modelling NO2 (2011)[5]

Table 1. Fitness for purpose matrix for dispersion models. Shown are the four major model types (columns) and applications (rows). Fitness for purpose is indicated by colour and appropriate comments. Green = 'fit for purpose'; Orange = 'conditionally applicable'; Purple = 'not fit for purpose'.

Model types and applications	Gaussian models	Lagrangian particle models	Obstacle resolving Eulerian models (CFD)	Terrain resolving Eulerian models
Open roads			No obstacles, computationally expensive	Unresolved
Street canyon	In combination with parameterised wind field model	In combination with parameterised wind field model	Computationally expensive	Unresolved
Urban scale	Requires homogenous meteorology	Computationally expensive	Not computationally feasible	
Regional scale	Requires homogenous meteorology	Computationally expensive	Not computationally feasible	

Table 9. Fitness for purpose matrix for NO_2 chemical schemes. Shown are the major chemical scheme types (columns) and spatial scales (rows) for both assessment and planning applications. Fitness for purpose is indicated by colour and appropriate comments. Green = 'fit for purpose'; Orange = 'conditionally applicable'; Purple = 'not fit for purpose'.

Chemical schemes and applications	Empirical schemes	Photo-stationary and ozone limitting schemes	Distance from source and mixing schemes	Reduced photochemical schemes	Full photochemical schemes
Assessment					
Street level	Given sufficient observations	Overestimates NO_2 in the presence of ozone		Difficult to apply at this scale. CFD only	Only reduced schemes necessary
Urban scale	Given sufficient observations	Suitable for winter or low hydrocarbons			
Regional scale				Missing significant chemistry	
Planning					
Street level	Only if scheme includes ozone and NO_2 primary emissions	Sensitivity to ozone and NO_2 emissions represented		Difficult to apply at this scale. CFD only	Only reduced schemes necessary
Urban scale	Only if scheme includes ozone and NO_2 primary emissions	Suitable with low light or hydrocarbons	Suitable with low light or hydrocarbons		
Regional scale				Missing significant chemistry	

[5] https://www.eionet.europa.eu/etcs/etc-atni/products/etc-atni-reports/etcacm_tp_2011_15_fairmode_guide_modelling_no2

Annex 2: AASAQ Audit grid (extracts)[6]

Emissions, modelling, forecast, mapping and regulatory statistics

1. Methods and tools for regulatory modelling, forecasting, mapping and statistics
    1.1. Tools and methods
        1.1.1. Is there an up-to-date list of modelling tools, emission inventories and data processing?
        1.1.2. For each model, give its applications. How are the software used chosen? Which versions are implemented?
        1.1.3. What are the calculation methods associated with the different tools? (Modelling, emission inventories, statistical and geostatistical processing, ...)
        1.1.4. How do you carry out technical monitoring on the calculation tools? How do you carry out scientific monitoring on modelling and mapping methods? What sources of information do you use?
    1.2. Staff skills
        1.2.1. Are there personnel identified as a duplication of competence on key positions related to modelling? How is staff training provided?
    1.3. Tool quality assurance
        1.3.1. How and under what time do you integrate the latest versions of the tools? Have you implemented a model evaluation procedure?
        1.3.2. What types of data are stored (inputs, outputs, settings)? Do you have a data storage and archiving procedure? Are the retention periods of the modelling data defined?
        1.3.3. How do you control the correct performance of calculations and data transfers? Do you have a procedure to implement corrective actions in the event of a calculation error or any undesirable situation detected?
2. Emissions
    2.1. Spatial regional inventories
        2.1.1. What year is your most up-to-date inventory based on? How often is it updated?
        2.1.2. Does the inventory comply with the PCIT guide? How do you assess the adequacy of the methods used with those recommended in the guide?
        2.1.3. What tools are used to calculate the emissions of the different sectors?
        2.1.4. Do you make an inventory of indirect emissions?
        2.1.5. How do you ensure the validity of the input data used? calculations? Final inventory data? In the case of non-compliance of the input data (GEREP declaration in particular), do you forward the information? What year is your most up-to-date inventory based on? How often is it updated?
3. Inventory for modelling
        3.1.1. What version of the inventory is used to develop the cadastre?
        3.1.2. How is spatialization carried out? What are the specifics of the emission temporal distributions?
        3.1.3. How are the constituent data of the cadastre validated?
4. Modelling chain on a regional scale
        4.1.1. Specify the components of the chain or of each of the modelling chains
    4.2. Conditions to the limits

[6] Personal communication

        4.2.1. Where do the bottom boundary conditions come from? What knowledge do you have? What follow-up is done with the production centres to guarantee optimal use of this data?
    4.3. Meteorology
        4.3.1. What weather data do you use? Is the adequacy between resolution of the weather model and transport chemistry in relation to local problems verified?
        4.3.2. When using a high resolution weather model, have you configured the model to take into account urban specificities?
    4.4. Pollutants
        4.4.1. What knowledge do you have of the physico-chemical parametrizations activated in regional production?
    4.5. Calculation parameters
        4.5.1. Do you know the sensitive parameters of the models you use? Do you perform sensitivity tests on these parameters? How do you stop a configuration? in relation to which issues?
        4.5.2. How do you check the consistency and quality of the input data? How do you check the spatial coherence of the different data sources between them?
    4.6. Output grid
        4.6.1. What output grids do you use? According to what spatial and hourly resolutions? How high? If the AASQA has worked on the definition of the mesh, ask for details
    4.7. Output variables
        4.7.1. Describe the output
        4.7.2. Do you make analysed maps?
        4.7.3. How do you calculate the regulatory indicators from the modelling results?
5. Local scale modelling chain
    5.1. Specify the components of the modelling tool or each of the modelling chains
    5.2. Domain
        5.2.1. How do you justify the size of the modelling domain?
        5.2.2. How do you take into account land use?
    5.3. Conditions to the limits
        5.3.1. How is background pollution taken into account?
    5.4. Meteorology
        5.4.1. Where does the weather data you use come from?
        5.4.2. What weather variables do you take into account?
        5.4.3. Do you use parameters specific to urban meteorology?
        5.4.4. How does topography intervene in modelling?
        5.4.5. Are particular elements of land use taken into account in the treatment of the weather?
    5.5. pollutants
        5.5.1. What pollutants do you model?
        5.5.2. How is chemistry (gas + particulate) modelled?
        5.5.3. What values do you use for particle density and diameter?
    5.6. Calculation parameters
        5.6.1. What are the parameters used for the calculation of the deposit (e.g. deposition rate, leaching coefficient)?
        5.6.2. Do you know the sensitive parameters of the models you use? Do you perform sensitivity tests on these parameters?
        5.6.3. How do you check the consistency and quality of the input data?

- 5.7. Output data
  - 5.7.1. Which outlet grid do you use? According to what spatial resolution and what hourly resolution? How high?
  - 5.7.2. Do you take into account particular receptor points?
- 5.8. Output variables
  - 5.8.1. Describe the output
  - 5.8.2. How do you map the modelling results?
  - 5.8.3. What indicators and statistics do you calculate from the modelling results?

6. Evaluation of the results of modelling in diagnostic mode
   - 6.1. Do you make model-measure comparisons?
     - 6.1.1. What comparison indicators are used? Do you use the Delta Tool made available by LCSQA? Have you set yourself performance targets?
     - 6.1.2. What types of metrics are used for these comparisons? Which measurement sites? Over what periods?
     - 6.1.3. What actions do you take following the modelling assessment?
7. Air quality forecasts
   - 7.1. Production
     - 7.1.1. What models do you use to produce air quality forecasts? Does AASQA integrate Prev'air data into its forecasting model? What adaptations of your systems do you make for forecasting purposes? What adaptations do you make on the input data for forecasting?
     - 7.1.2. Does AASQA exchange with Prev'air?
     - 7.1.3. Do you have a time limit for availability of forecast data? Are the forecasts refreshed during the day? If so, under what conditions?
     - 7.1.4. When are the forecast deadlines produced? For which pollutants? What are the time constraints for producing the forecast system?
     - 7.1.5. Do you use statistical methods to predict concentrations? According to what methodology? How are the parameters selected? What is the learning period? Identical for all pollutants?
     - 7.1.6. Are other treatments performed on the forecasts from the models? Expert adjustment of the spatialization of concentration fields? Use of forecasts in relation to the "emergency measures" decree: which forecasts are used to assess the criteria?
     - 7.1.7. How are the forecasts distributed? In what forms (maps, data)?
8. Evaluation of forecasts
   - 8.1. Daily evaluation.
     - 8.1.1. Do you assess the performance of your forecasting system daily? What criteria do you use to estimate the quality of the forecasts? By pollutants? By indices?
     - 8.1.2. What is the impact of the evaluation? How is it integrated into forecasting expertise? In the event of an anomaly detected, what actions can be taken?
   - 8.2. Annual evaluation.
     - 8.2.1. Do you evaluate the performance of your forecasting system each year for the past year? With what tools? Is this evaluation the subject of an annual report? Is it sent to DREAL and LCSQA before June 30?
     - 8.2.2. How are the statistical methods used to predict or explain concentrations evaluated?
9. Exploitation of model outputs and mapping
   - 9.1. Do you estimate the zones and populations exposed to exceeding thresholds?
   - 9.2. In what form are the maps distributed? Are they accompanied by opinions, interpretations? How often are they updated?

10. Statistics
    10.1. What statistics do you calculate from the modelling data? With what tools and according to what rules?
    10.2. In addition to calculating basic statistics, do you use statistical methods to analyse and interpret measurement data?
11. Objective estimate
    11.1. Do you use objective estimation methods?
    11.2. In the context of a measurement campaign, how do you ensure the temporal representativeness of the data?
    11.3. Do you make maps from measurement campaigns? If yes, by what methods?
    11.4. How do you assess the quality of the results of the objective estimation? Do you have a method to assess the uncertainty?
12. Outlook
    12.1. Have you identified areas for improvement in modelling and mapping?
    12.2. What developments do you envisage in the context of regulatory oversight?

Annex 3: CHIMERE protocol followed before a new version is released[7]

- The CHIMERE developments are performed under the git[8] environment
- Any new version needs to be systematically compared to the previous one and compared with observations before the new version can be released
- The QA/QC protocol demands one year simulation at 50 km resolution encompassing a domain covering Europe and North Africa (to include desert dust sources) is performed
- In some cases, additional tests are performed to assess the improvements due to new parameterisations (meteorology, numerics, biogenic emissions, etc…) are required
- The capability of the new version is tested using a series of score indicators (Bias, RMSE, Cor.) on background stations (EEA sites) for criteria pollutants (O3, O3max, PM10, PM2.5, SO2, NO2)
- Averaged time series and focussed at specific sites are analysed
- Before the official release of the version, the results are discussed during the CHIMERE weekly meetings with the core modelling group at LMD
- The results from the new version need to show improved scores for the one year simulation with respect to the old version.

Annex 4: Spatial and temporal indicators: DELTA tool

[7] Personal communication
[8] Git is a distributed version-control system for tracking changes in source code during software development

The additional indicators are based on station characteristics defined in the DELTA tool “startup.ini” file (see extract below), in particular the station type (column 9) and station area (column 10). The station types (industry, traffic, background) and station areas (urban, rural) are combined to create

```
[MODEL]
;Year
;frequency
;Scale
2005
hour
urban
[PARAMETERS]
;Specie*type*measure unit
NO2;GAS;ugm-3
O3;GAS;ugm-3
PM10;PM;ugm-3
PM25;PM;ugm-3
[MONITORING]
Station Code;Station Name;Station abbreviation;Altitude;Lon;Lat;GMTLag;Region;Station Type;Area Type;Siting;listOfvariables
803620;   MODENA   MXXS;   30;    10.9292;  44.6431;  GMT+1;E  MR;    traffic;      urban;  plane;  NO2
803711;   Monte;   CUCC;   260;   11.3336;  44.4714;  GMT+1;   EMR;   background;   rural;  hilly;  NO2
803805;   GHERARDI;GHER;   0;     11.9611;  44.8417;  GMT+1;   EMR;   background;   rural;  plane;  NO
603010;   SGiovanni; SGIO; 6;     13.3934;  45.9752;  GMT+1;   FVG;   industrial;   rural;  hilly;  NO2
```

the spatial indicators: UB, UT, and RB, whereas the station types (industry, traffic , and background) are used for the temporal indicators: I, T, and B.

Figure 2 shows the diagram as it appears in the DELTA tool. Each row corresponds to a specific indicator and each dot to a specific station (or pair of stations). A dot will be in the green zone (sufficient quality) when the absolute value of the indicator is less than one (indicator normalised by the measurement uncertainty). The circle on the right-hand side indicates if sufficient quality is reached for that indicator, i.e. when at least 90% of the available dots are within the green area.

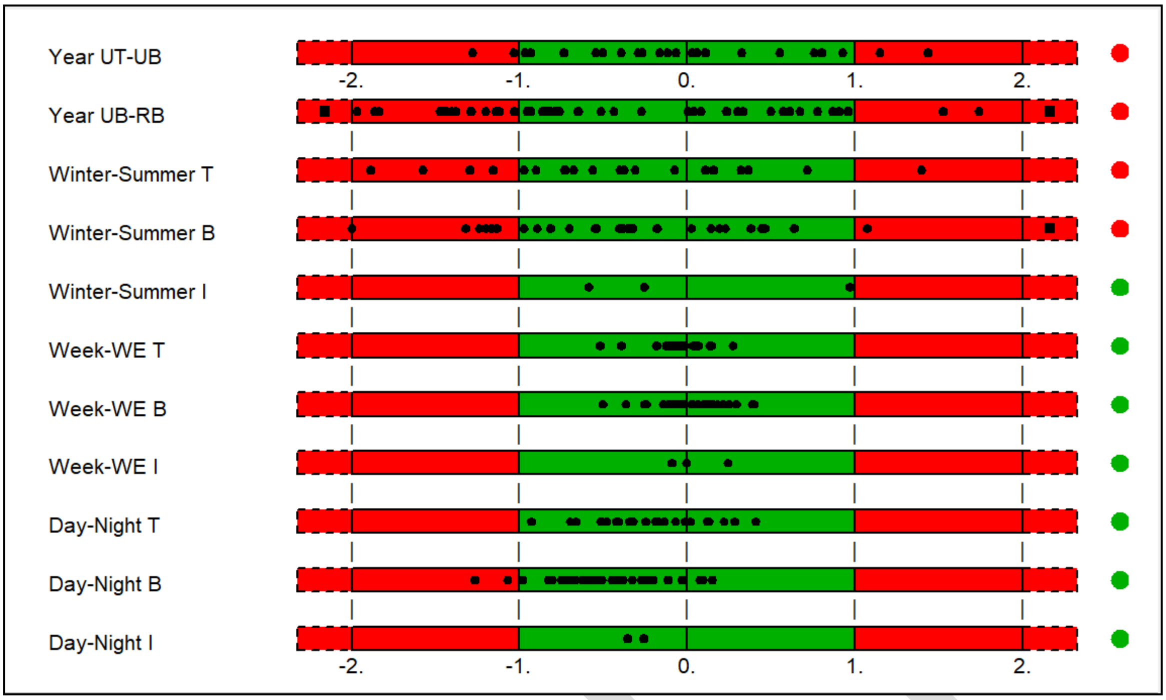


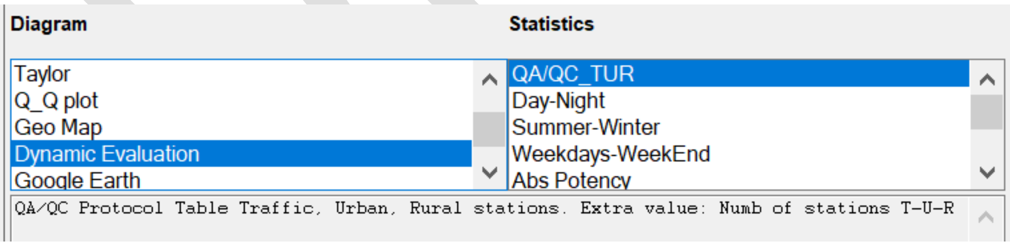
Figure 2: Summary report with the QA/QC spatial indicators (bars1 and 2) and temporal indicators (bars 3-11), normalised by the measurement uncertainty.

Information on each station can be obtained by clicking on the desired dot in the diagram.

<u>How to run DELTA: technical setting and options</u>

1) Make sure that your startup file correctly includes the desired station characteristics (station types and station areas)

2) Set the Diagram to "Dynamic Evaluation" and statistics to "QAQC_MQI" as shown in the window below.



3) Choose the desired threshold. This threshold applies for the spatial indicators and sets the number of stations that will be used to calculate the spatial gradients. For example in bar 1, if a choice of 3 is made, three urban background (UB) stations will be used to calculate three gradients with each available urban traffic (UT) station. The larger the threshold, the more points will be shown and the better the estimation of the capacity of the model to reproduce spatial gradients is.

4) For each diagram, a data file (QAQCdump.dat) is produced in the help folder) that includes all detailed results.