Institute for Environment and Sustainability

Procedure for Air Quality Models
Benchmarking

P. Thunis, E. Georgieva, S. Galmarini

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/
Outline

• Objectives & Background
• Key elements of the procedure
• The Benchmarking service
• Usage of the procedure
• Work Plan
• Contributions & links to other SG
Objectives

- Develop **procedure** for the benchmarking of AQ models in order to evaluate their performances.

- **Support** both model users & model developers in the implementation of the AQD (Assessment & plans)

- Provide technical/scientific **support to policy** on the analysis of the quality of model results

- Identify a **common scale** for model evaluation

- Identify a **common** (= to all member states), **permanent** (= through directives) and **periodic** (= every x years) practice to assess model quality improvement
Many tools & methodologies already existing…

- **BOOT software** (Chang and Hanna, 2005)
- **Model validation Kit** (Olesen)
- **USA-EPA AMET package** (Appel and Gilliam, 2008)
- CityDelta and EuroDelta projects
- **ENSEMBLE** platform (Galmarini S. et al. 2001, 2004).

- PM model performance metrics (Boylan and Russell 2006)
- Summary diagrams (Jolliff et al. 2009)
- SEMIP project
- **AIR4EU conclusions** (Borrego et al. 2008)
- **ASTM Guidance** (ASTM, 2000)
- Mesoscale Model Evaluation – COST728 (Schluenzen & Sokhi, 2008)
- Proposal for a Benchmarking Tool: FAIRMODE WG2 Meeting November 2009

- Document “Procedure for AQ models Benchmarking” sent out to SG4 participants April 2010 (uploaded on FAIRMODE web page)

  - **Application types:** AQ assessment and Planning
  
  - **Models included:** Regional, urban and local scales
  
  - **Focus:** pollutants considered in the AQ Directive (NO2, PM and O3) depending on the spatial scale addressed.
| **DELTA:** | Evaluation tool based on City- & Euro- Delta inter-comparison exercises |
| **ENSEMBLE:** | Multi-model evaluation and inter-comparison platform used by several modeling communities (e.g. Galmarini S. et al. 2001, 2004a and b). |
| **Data Extraction** | Extraction of Monitoring data, Emissions, BC… links to other projects data (GMES, EC4MACS…). |
| **Benchmarking Service** | Performance indicators, criteria and goals, summary reports. |
The benchmarking service

PURPOSE:

produce summary performance reports for a given model application in the frame of the AQD.

FEATURES:

• Reports follow a pre-defined template structured around a core set of indicators and diagrams.

• Definition of some bounds for specific indicators, called hereafter goals and criteria (regularly revised based on future joint modelling exercises).

• Decomposition of the evaluation in temporal and spatial segments on a reduced dataset but for an entire year.

• Reports are obtained through an automatic procedure
The benchmarking service
Core set of statistical indicators

- R  Correlation
- B  Bias
- SD Standard deviation
- FAC2 Factor 2
- RMSE Root Mean Square Error
- RMSEs Systematic RMSE
- RMSEu Unsystematic RMSE
- CRMSE Centered RMSE
- IOA Index of Agreement
- MFB Mean Fractional Bias
- MFE Mean Fractional Error
- RDE Relative Directive Error
- RPE Relative Percentile Error
Criteria: Acceptable performance for a given type of application (e.g. PM: MFE=75%, MFB=+/-60%)

Goal: Best performance a model should aim to reach given its current capabilities (e.g. PM: MFE=50%, MFB=+/-30%)

Observation Uncertainty
The benchmarking service: Performance summary report

Performance summary report (pollutant/scale specific)

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>Z1</th>
<th>Z2</th>
<th>Z3</th>
<th>Z4</th>
<th>Z5</th>
<th>Z6</th>
<th>Z7</th>
<th>Z8</th>
<th>Z9</th>
<th>Z10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias (Yearly mean)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias (Summer max)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (Yearly mean)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>RMSE (Summer max)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>HMODs (Yearly mean)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>RMSE (Summer max)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>IOA (Yearly mean)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>IOA (Summer max)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>MPRF (Hourly)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>MRDE (3h-average)</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
</tbody>
</table>
Usage of the procedure

USER

Model results

JRC

Data Extraction Facility

DELTA

BENCHMARKING service

Unofficial Working Report

Official Reports

Unofficial Working Report
Usage of the procedure: testing levels

Model Results

Input Consistency Indicators (ICI)  
(model vs. input data & min-max tests)

Model Observation Indicators (MOI)  
(model vs. measurements)

Multi-Model Indicators (MMI)  
(model vs model: type, version, user…)

Model Response Indicators (MRI)  
(model vs model: type, version, user…)
Work Plan

- Discussion and consensus on overall methodology (FAIRMODE meeting 09/2010)

- Development of the DELTA and benchmarking service prototypes (Dec 2010)

- Testing of the prototypes on existing datasets (2011)

- Development of the JRC Web facilities (data extraction, links ENSEMBLE-Benchmarking service…)

- Set-up of a joint exercise for testing of the whole system (2012)
Contributions needed

• Discussion and definition of the benchmarking service elements (species, statistics, goals and criterias…) for model performance reporting per pollutant/scale (especially local scale).

• Links to other SGs (station representativeness, emissions…)

• Definition of and participation to the joint activities