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/

Agenda

Objective
Key elements of the proposed procedure
Usage of the procedure
Discussion
Lunch Break
The Benchmarking service
Discussion
Work Plan
Contributions & links to other SG
Discussion

Develop a **procedure** for the benchmarking of AQ models **to evaluate and keep track** of their performances:

- **based on a** common **and** permanent **evaluation "scale"**
- **with** periodic **joint exercises to assess and compare model quality.**

Constraints:

- **Make use of available tools and methodologies**
- **Based on consensus**
- **Application specific (assessment & planning)**

Many tools & methodologies already existing…

- USA-EPA *AMET package* (Appel and Gilliam, 2008)
- Tools from CityDelta and EuroDelta (Cuvelier et al. 2007)
- E*NSEMBLE* platform (Galmarini S. et al. 2001, 2004).
- *BOOT software* (Chang and Hanna, 2005)
- *Model validation Kit* (Olesen, 2005)

- EPA Guidance (2007, 2009)
- AIR4EU conclusions (Borrego et al. 2008)
- Mesoscale Model Evaluation – COST728 (Schluenzen & Sokhi, 2008)
- Quality assurance of microscale models – COST732 (2007)
- SEMIP project (Smoke & emissions model inter-comparison, 2009)
- Evaluating the Performance of Air Quality Models, AEA (2009)
- *ASTM Guidance* (ASTM, 2000)

- PM model performance metrics (Boylan and Russell 2006)
- Summary diagrams (Jolliff et al. 2009)

Key elements of the procedure

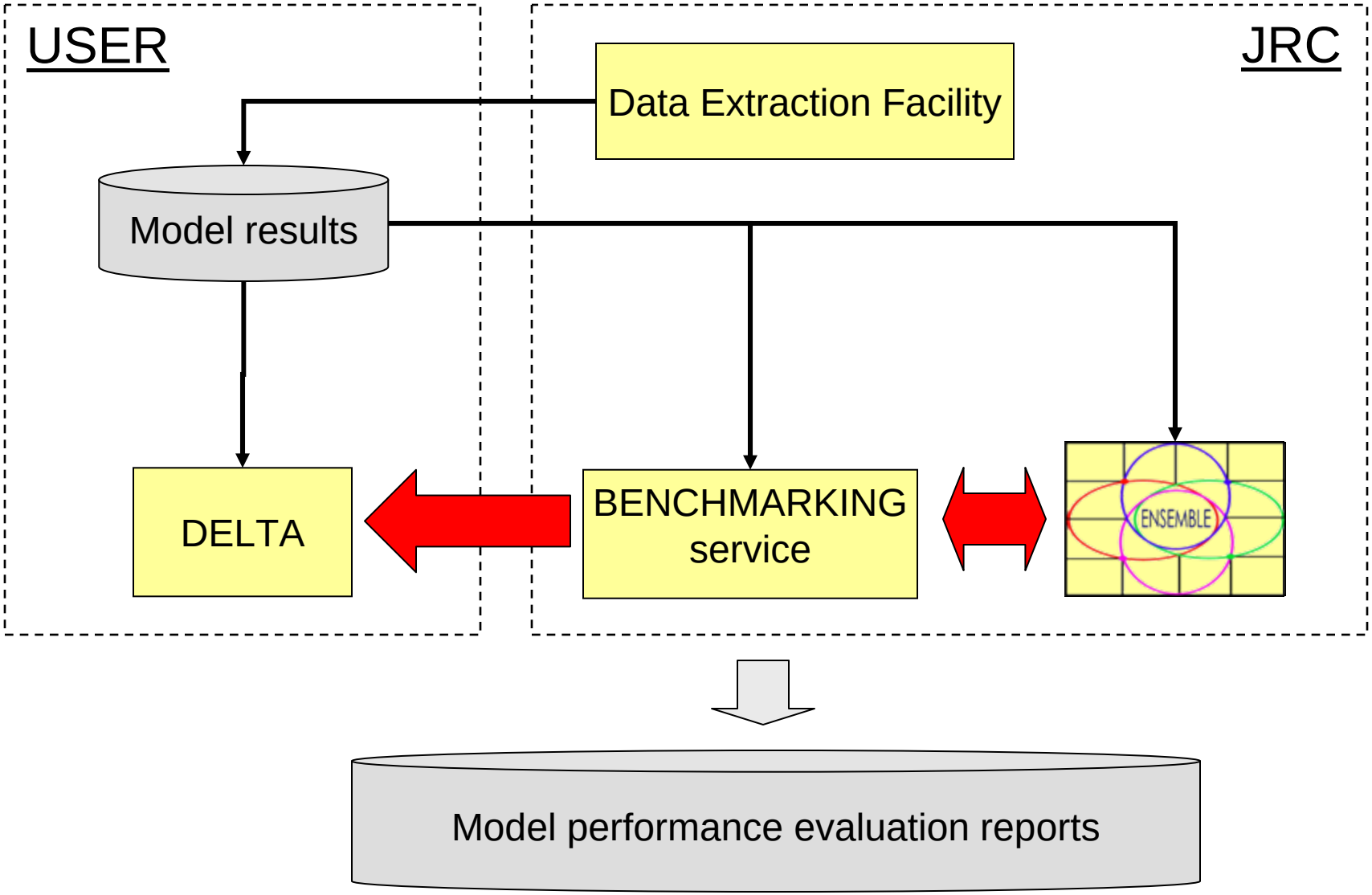
DELTA: → Evaluation tool based on City- & Euro-Delta, POMI and HTAP inter-comparison exercises

ENSEMBLE → Multi-model evaluation and inter-comparison platform used by several modeling communities

Benchmarking Service → Statistical indicators and diagrams, criteria and goals, automatic reporting.

Data Extraction → Extraction of Monitoring data, Emissions, BC…

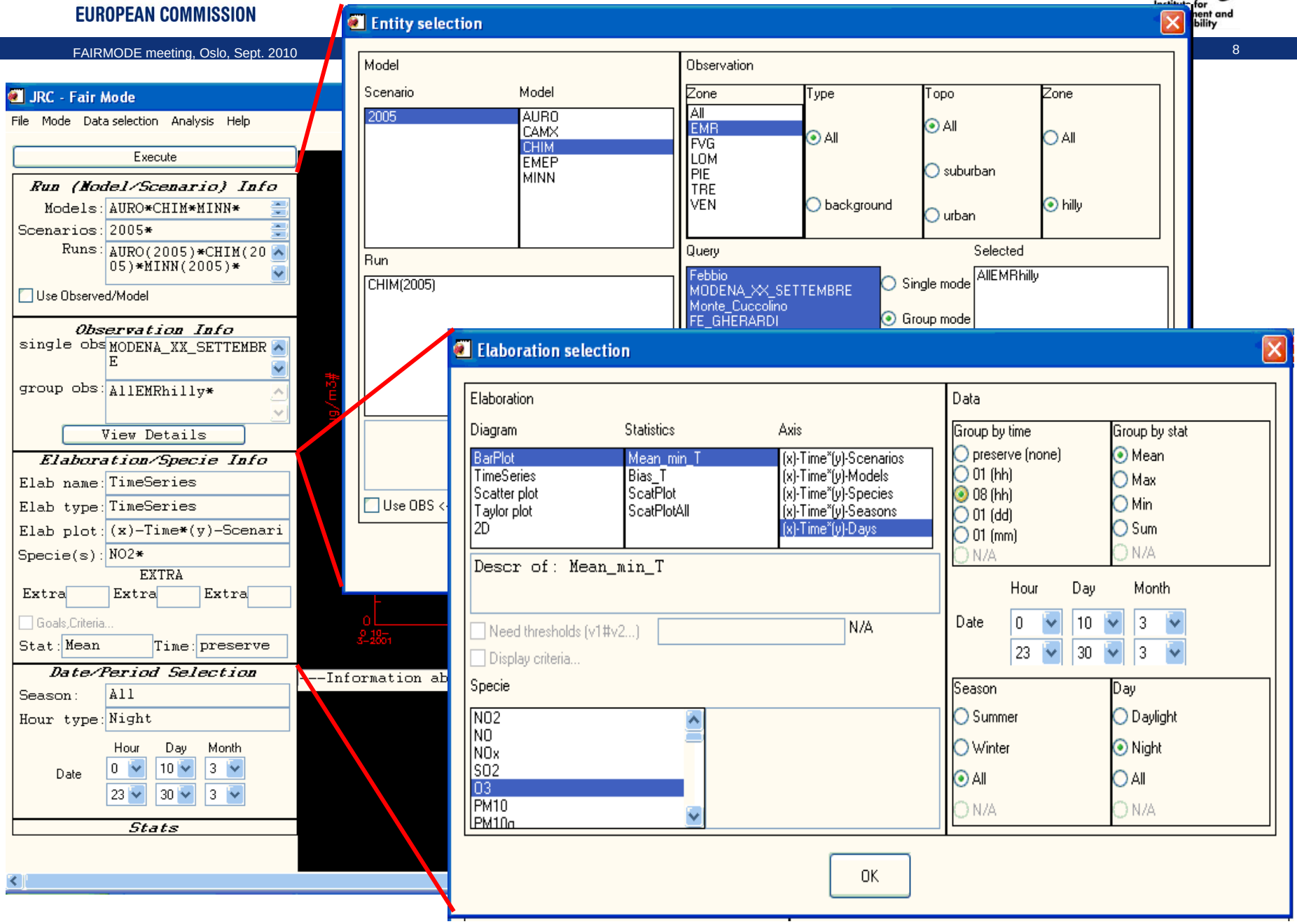
Benchmarking procedure: Key elements



The DELTA tool

- **Intended for rapid diagnostics by single users (at home)**
- **Focus mostly on surface measurement-model pairs (reduced set) → "independence" of scale**
- **Focus on AQD related pollutants on a yearly period (but AQ related input data also checked)**
- **Exploration and benchmarking modes**
- **Includes a set of statistical indices and diagrams (agreed)**
- **Flexibility in terms of:**
  - Addition of new statistical indicators & diagrams
  - Choice of monitoring stations, models, scenarios…

The DELTA Tool



The ENSEMBLE platform

- **JRC Web based platform**
- **All variables AQ and Meteo (4D fields) may be considered (full set)**
- **Exploration and benchmarking modes**
- **Used for multi-model analysis & evaluation**
- **Includes a set of statistical indices and diagrams (agreed)**
- **Acts as a model results depository**
- **Flexibility in terms of:**
  - Model vs model comparison, model vs obs,model vs. groups of models
  - Choice of monitoring stations, models, scenarios…

PURPOSE:

- Selection of a core set of statistical indicators and diagrams for a given model application in the frame of the AQD

- Production of summary performance reports based on a common scale

FEATURES:

- Based on different testing levels (obs., mod. vs. mod., responses to emission scenarios, input data, BC)
- Decomposition of the evaluation in temporal and spatial segments on a reduced dataset but for an entire year.
- Structured around an agreed core set of indicators and diagrams specific for each AQD related application
- Definition of bounds for specific indicators, called hereafter goals and criteria (regularly revised based on future joint modelling exercises).
- Reports are obtained through an automatic procedure and follow a pre-defined template
- JRC based service but with replica included in the DELTA tool, i.e. one unique “scale” used in ENSEMBLE and DELTA to evaluate models

The EXTRACTION facility

Single usage

- Observations (AIRBASE,…)
- Reference model data (EU)
- Boundary conditions

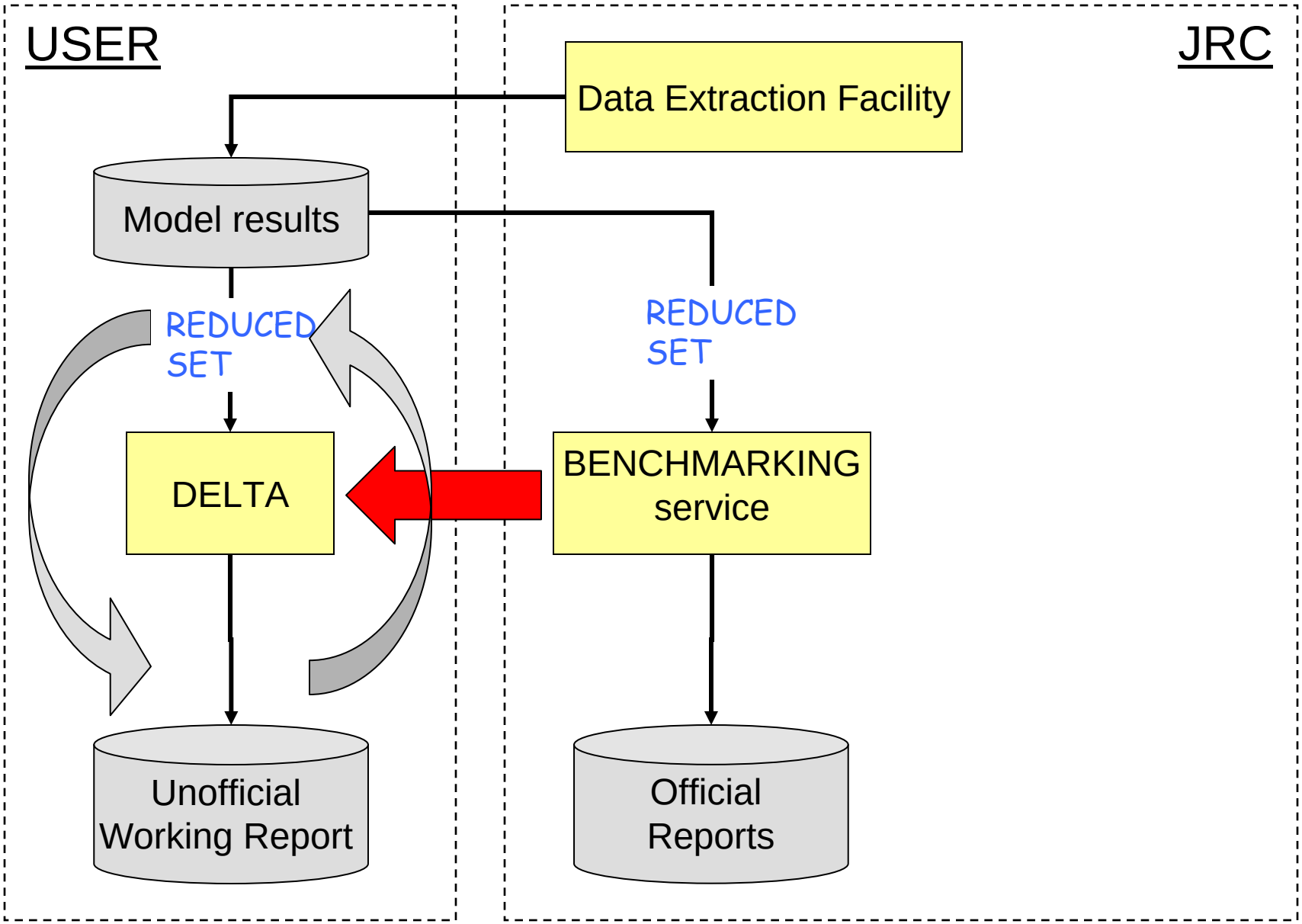
Joint exercise

- All required input data

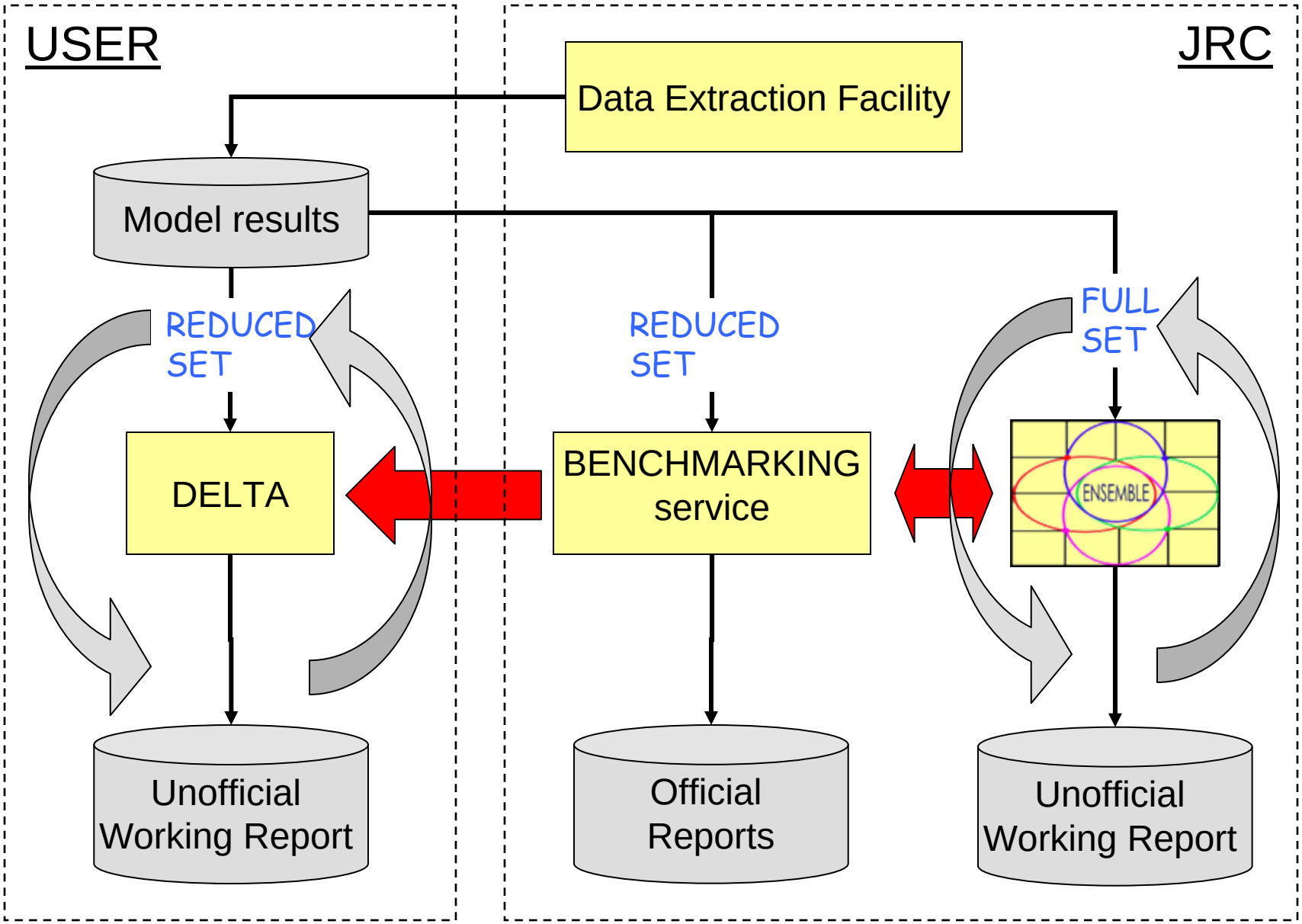
Usage of the procedure

- Usage 1: Individual model / MS
- Usage 2: Periodical Joint Activities

Usage 1: Individual Model/MS



Usage 2: Joint activities



Expected benefits

- **Same single evaluation tool**
- **Common (JRC based) place for evaluation & inter-comparison and acquisition of data**
- **Tracking of the historic evolution of model quality relevant for policy decisions**
- **Evolving reporting tool**
- **Data depository**
- **Quantification of uncertainty in model results**

Conclusions & discussion

- Common and general frame for model evaluation
- Application-specific benchmarking service
- User and JRC based components
- Updating process via expert-judgment bounds
- Common joint exercises

PURPOSE:

- Selection of a core set of statistical indicators and diagrams for a given model application in the frame of the AQD

- Production of summary performance reports based on a common scale and pre-defined template

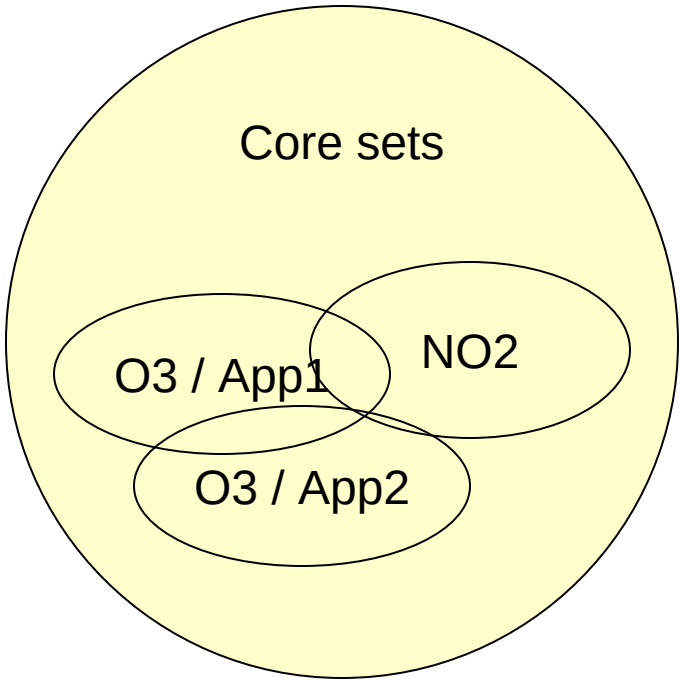
  - ✓**Reduced vs. full model datasets**
  - ✓**Organized around different testing levels**
  - ✓**Updating process: bounds (goals and criteria)**
  - ✓**Breakdown of the analysis into temporal and spatial segments**
  - ✓**Summary and annexes**

Testing levels:

- Input data ICI Model vs. Input data
- Observations MOI Model vs. Observations
- Multi-model MMI Model vs. model (base-case)
- Scenarios MRI Model vs. model (scenarios)

- **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

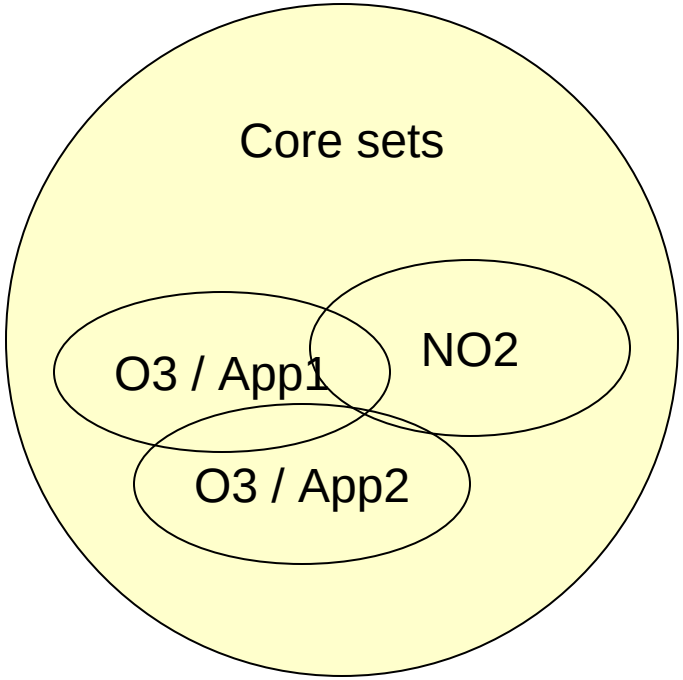


Set and core-sets of diagrams

- Scatter plots
- Q-Q plots
- Bar-plot
- Time series

- Taylor diagrams
- Target diagrams
- Soccer plots
- Bugle plots

- Conditional plots
- Multi-model diagram
- …



Bounds

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%)

Dev. ENS: Deviation from ensemble mean. Flagged when model results are deviating from fixed bounds around the ensemble mean and no observation is available.

Obs. Unc: Best performance a model should aim to reach given the observation uncertainty

Updating of bounds based on outcome of joint exercises

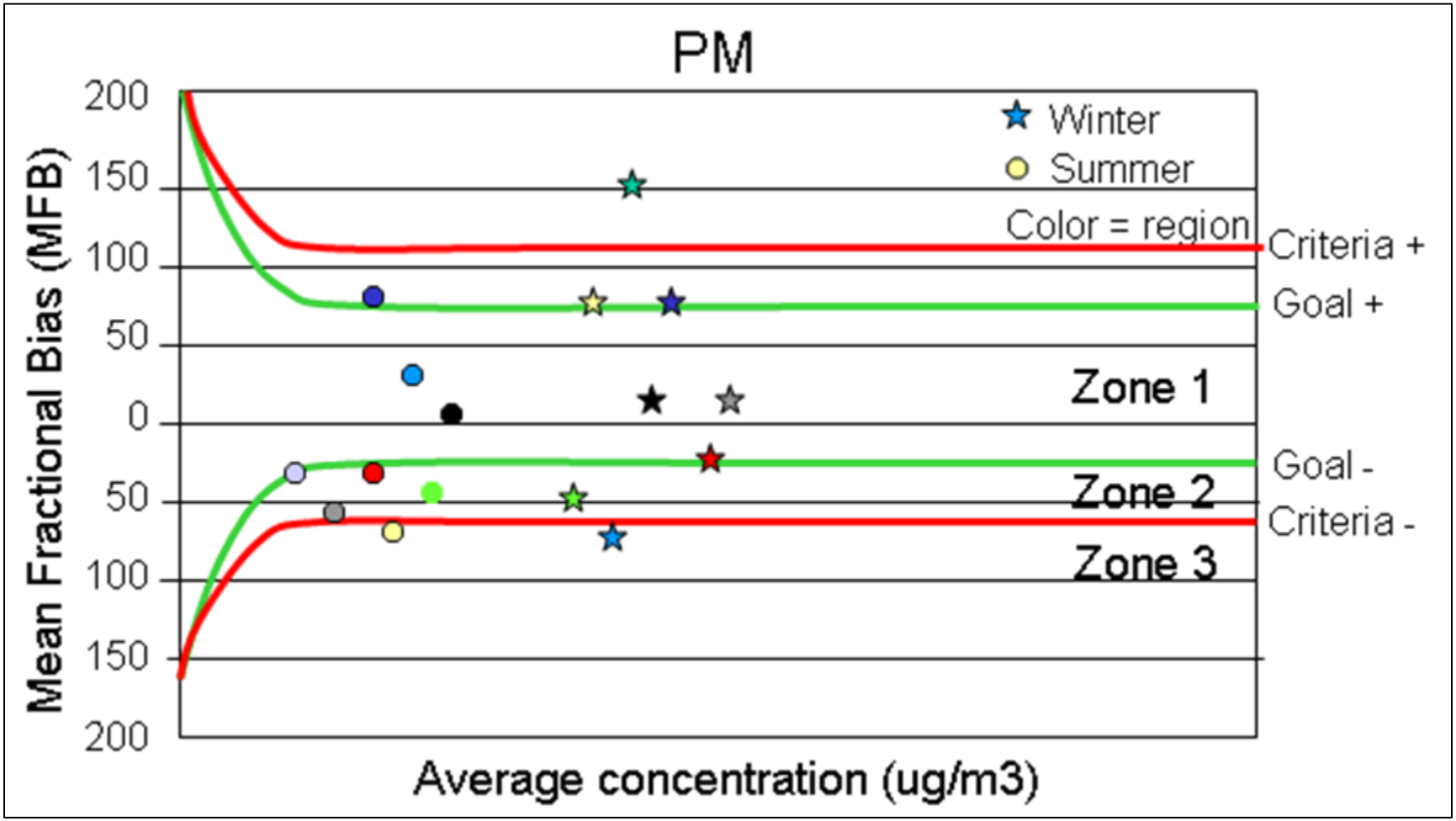
Meteorology- regional scale (Emery et al., 2001)

Parameter	Metric	Criteria
Wind speed	RMSE Bias IOA	$\leq$ 2 m/s $\leq \pm$ 0.5 m/s $\geq$ 0.6
Wind direction	Gross error Bias	$\leq$ 30 deg $\leq \pm$ 10 deg
Temperature	Gross error Bias IOA	$\leq$ 2K $\leq \pm$ 0.5 K $\geq$ 0.8
Humidity	Gross error Bias IOA	$\leq$ 2 g/kg $\leq \pm$ 1 g/kg $\geq$ 0.6

Air Quality (Regional scale modelling)

Species	Metric	Criteria	Goal
Boylan and Russel, 2005, EPA report 2007			
Main PM constituents (> 30% total mass), PM2.5	MFE MFB	75% ±60%	50% ±30%
Minor PM constituents (< 30% total mass)		Exp variations to reach 100% / 200% at 0 concentrations	
Ozone	MFE MFB	35% 15%	
Evaluating the Performance of Air Quality Models, AEA (2009)			
Any pollutant	FAC2 NMB	Half points within $-0.2 < MFB < 0.2$	
Air quality model performances evaluation, Chang et Hanna (2004)			
NOx, CO, PM10	FAC2 FB NMSE	Half points within $-0.3 < FB < 0.3$ $NMSE < 4$	

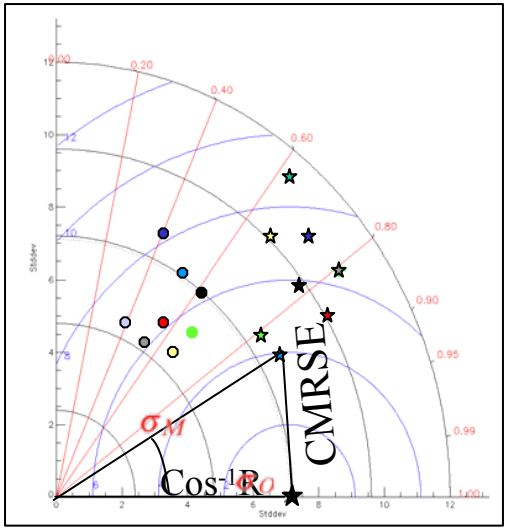
Bugle plot (Boylan 2005)

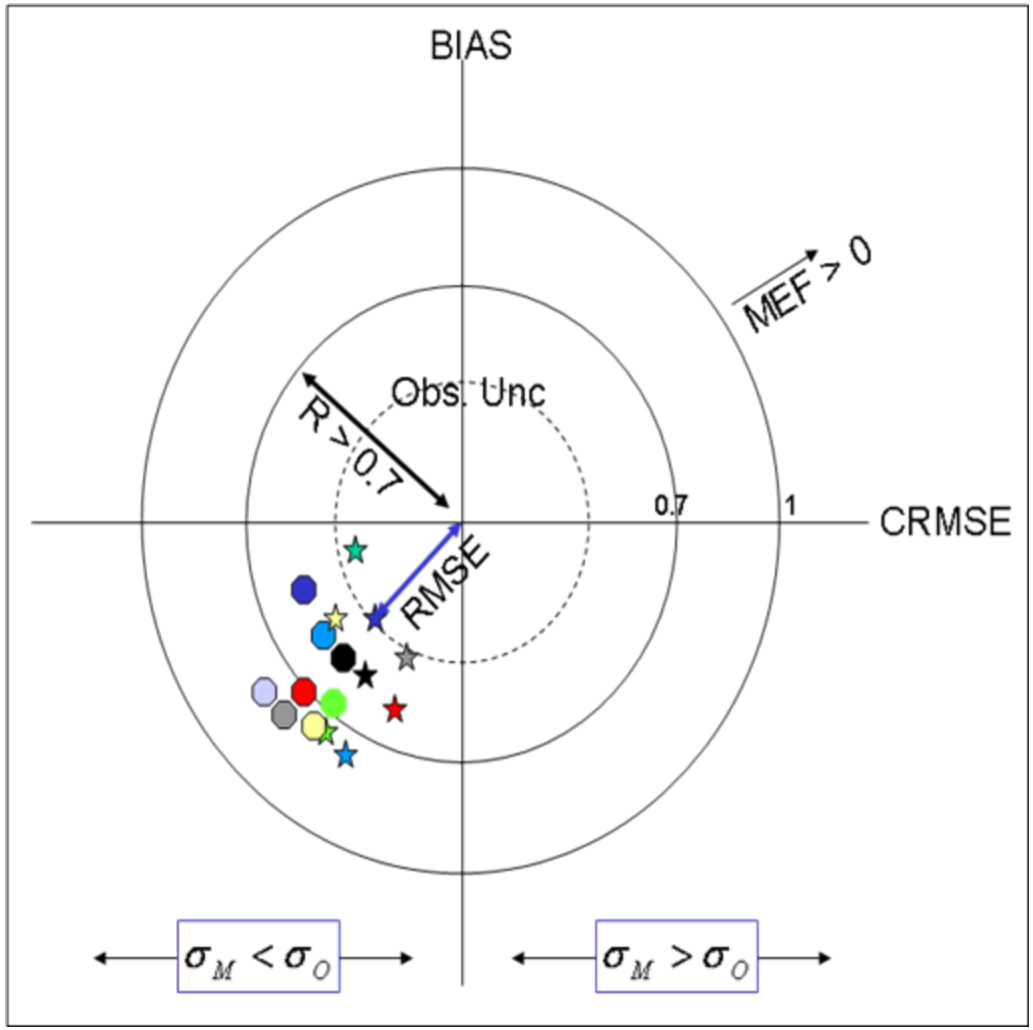


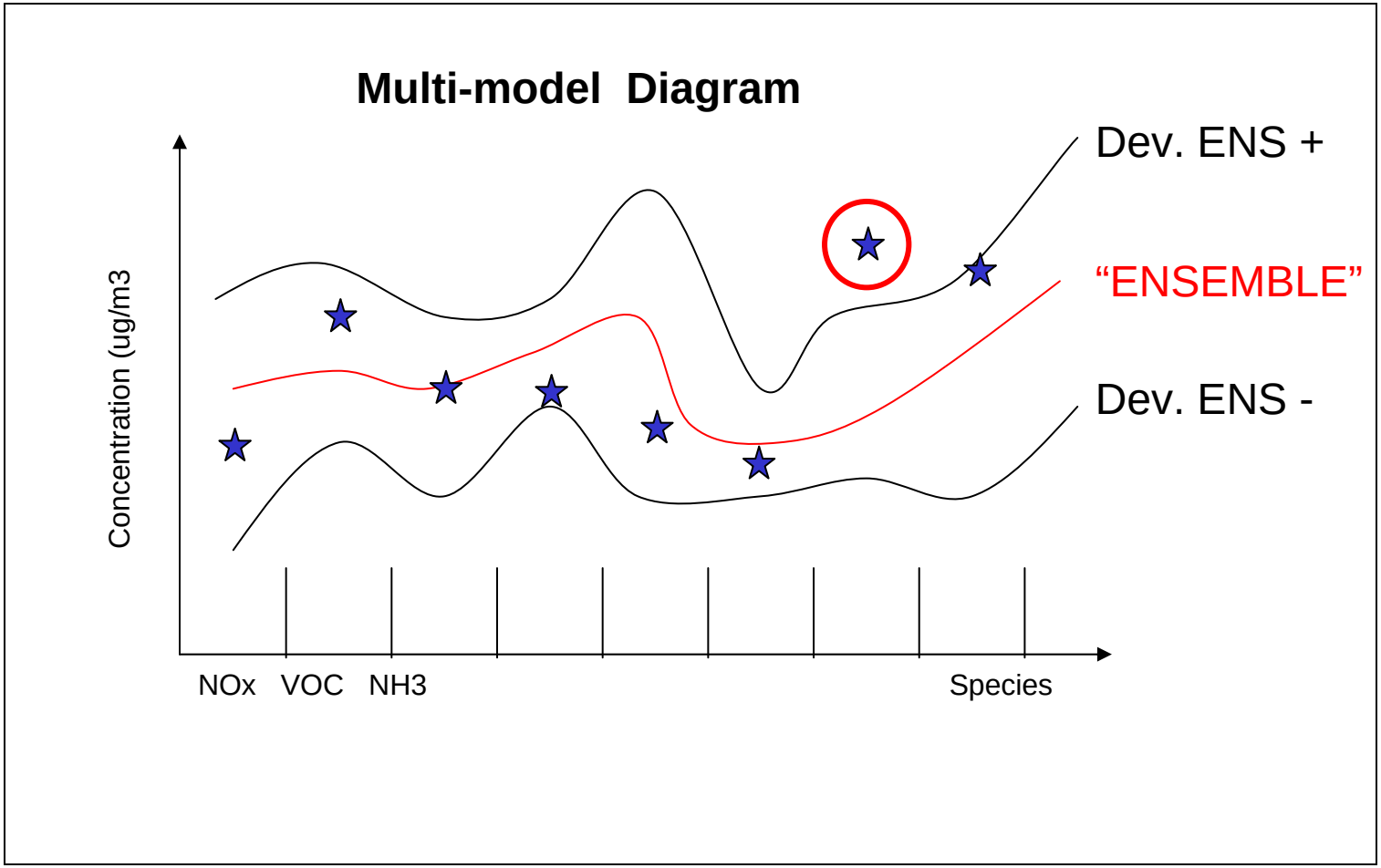
Summary diagrams

From Taylor diagram to

Target plot (Jolliff 2009)

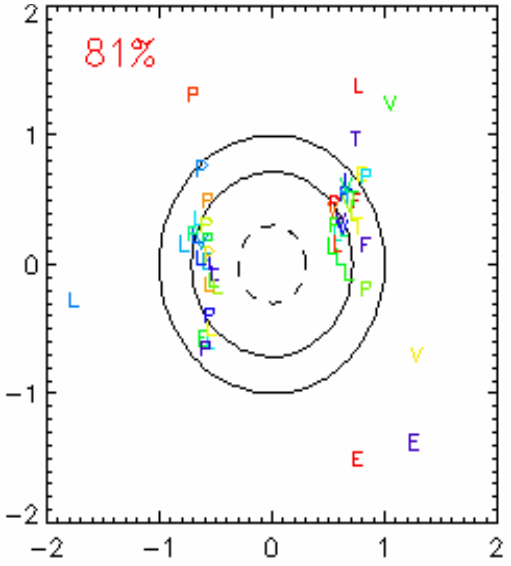





Multi-model Diagram
Dev. ENS +
"ENSEMBLE"
Dev. ENS -
Concentration (ug/m3
NOx
VOC
NH3
Species

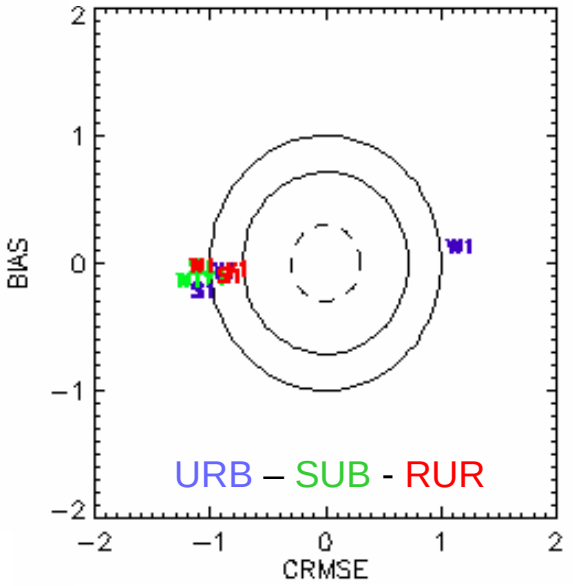
Summary diagrams

Target plot: all stations

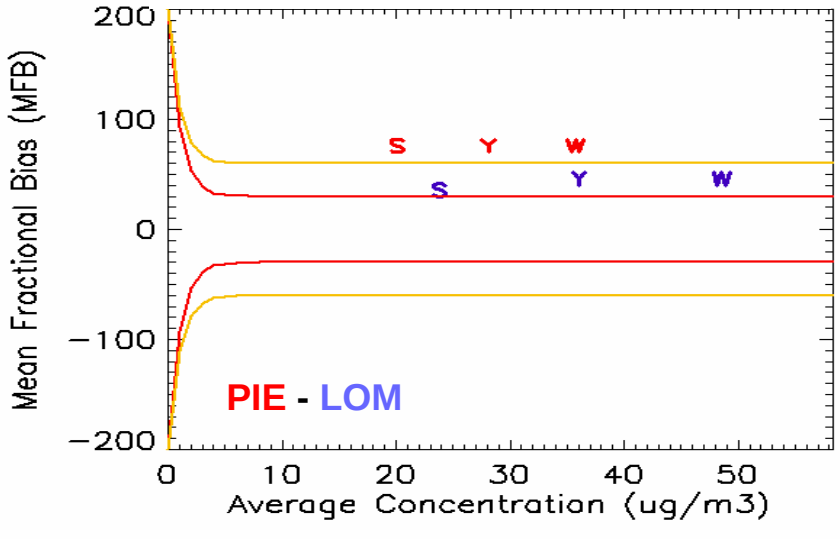


An example: POMI data

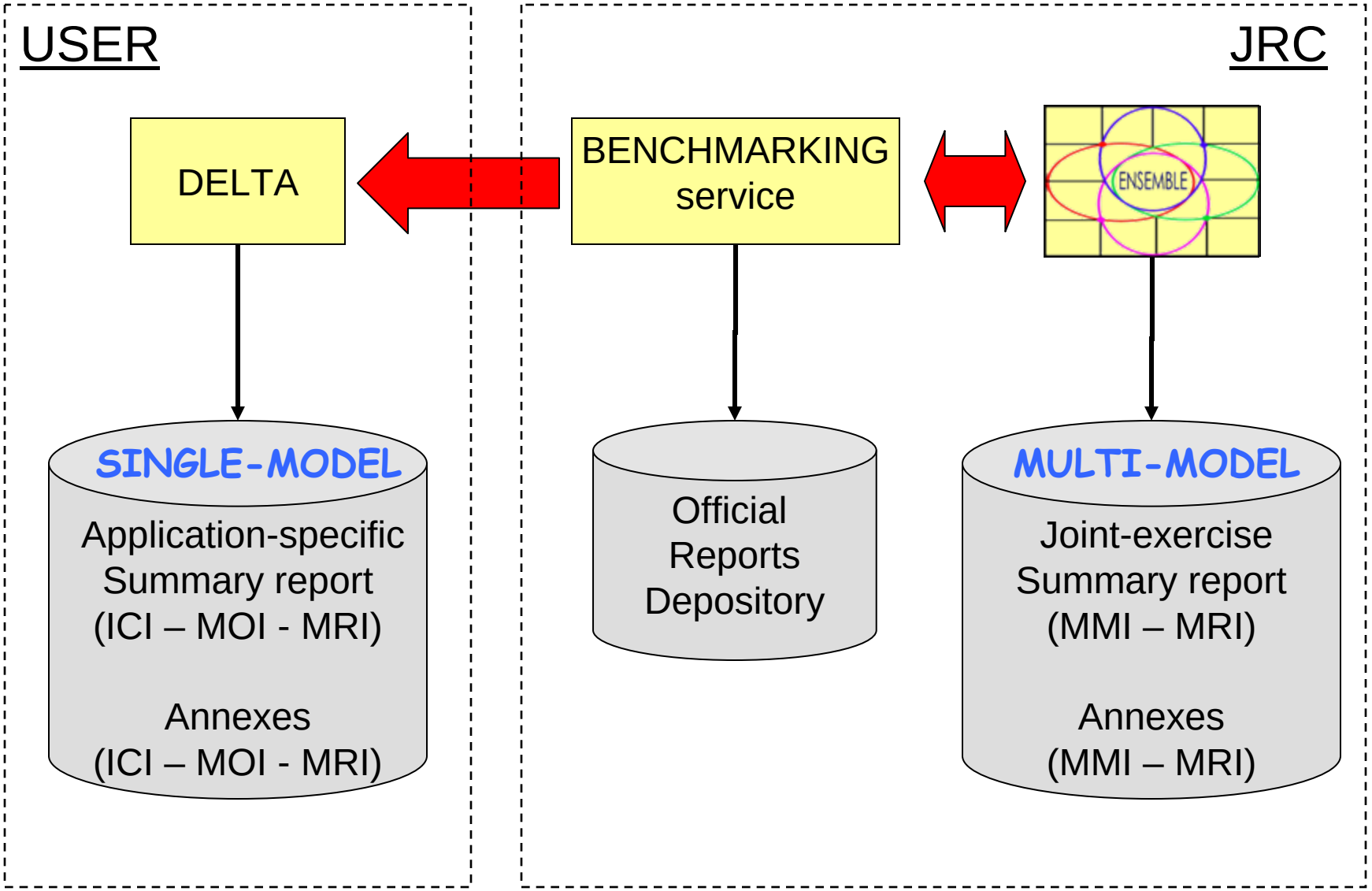
Target plot: Groups

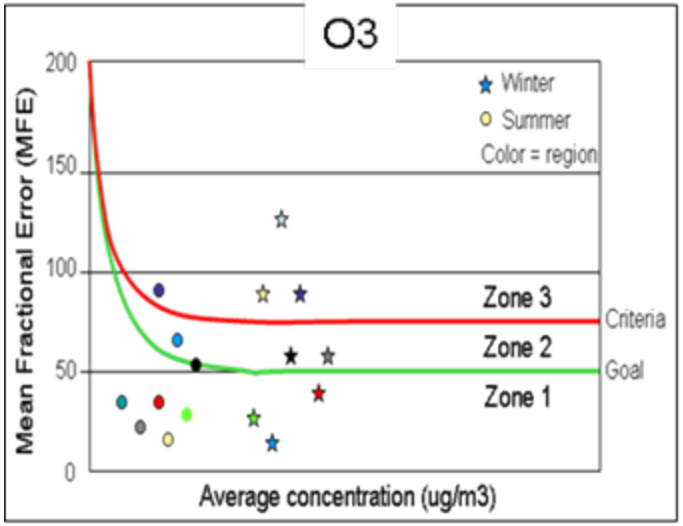


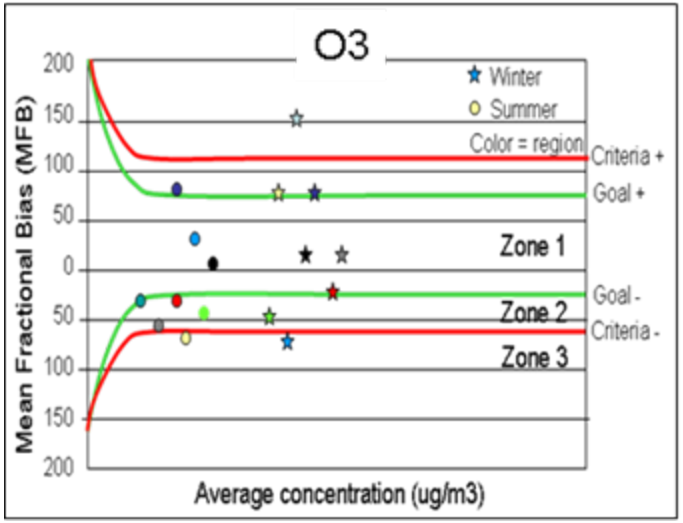
Bugle plot: Groups

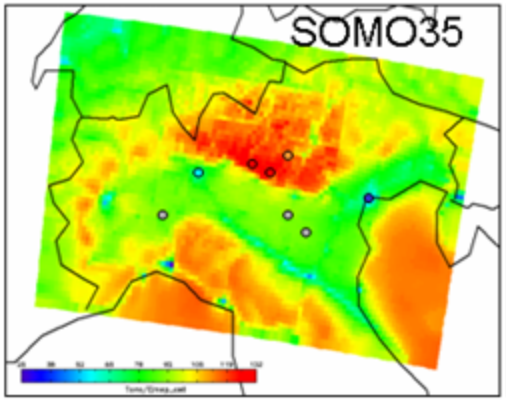


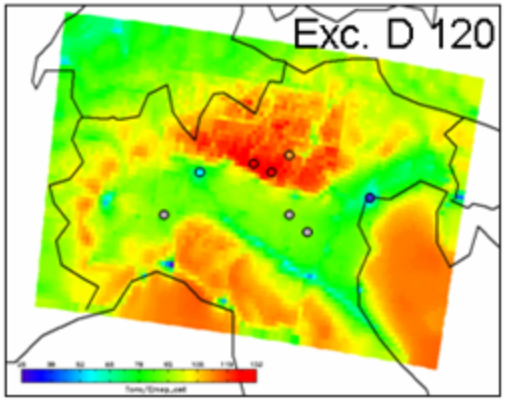
Performance summary report





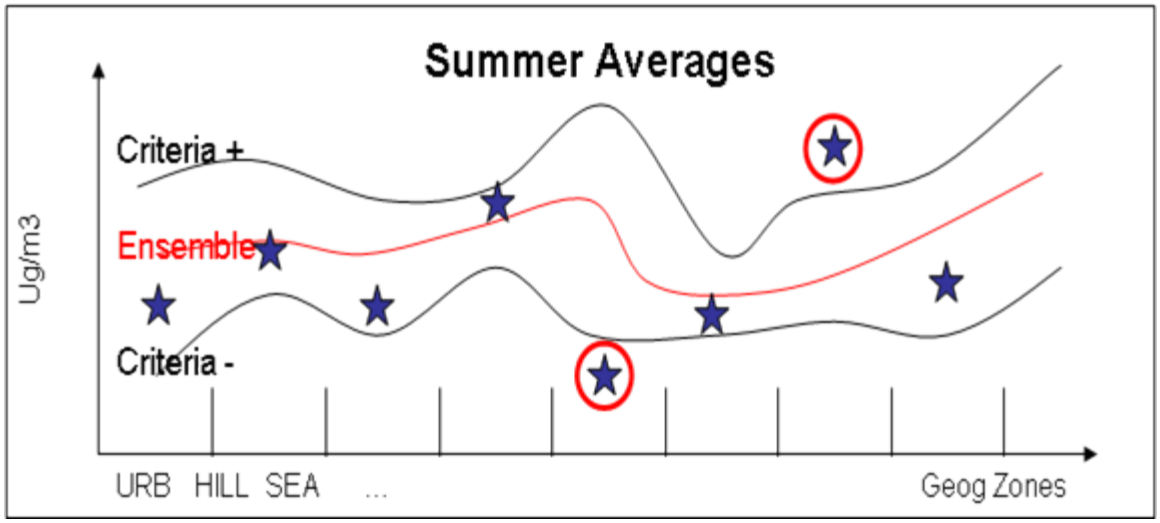


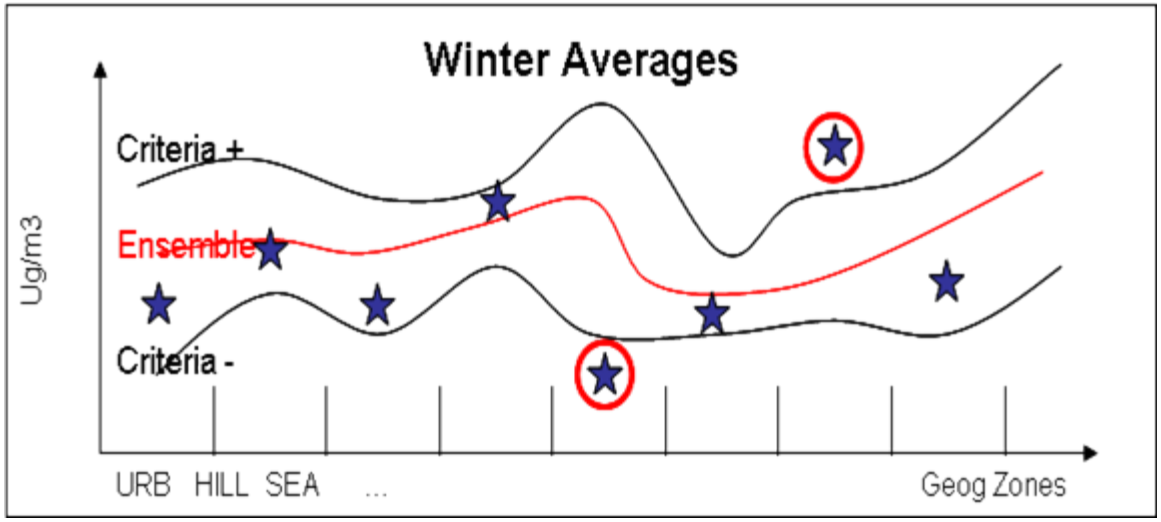




Summary Statistics		Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
Bias	Yearly mean										
Bias	Sum8h max										
RMSE	Yearly mean	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53		0.53
RMSE	Sum 8h max										
RMSEs	Yearly mean	0.53	0.53								0.53
RMSEs	Sum 8h max	0.53	0.53								0.53
RMSEu	Yearly mean	0.53	0.53								0.53
RMSEu	Sum 8h max	0.53	0.53								0.53
IOA	Yearly mean	0.53	0.53								0.53
IOA	Sum 8h max										
MRDE	Hourly										
MRDE	8h-average										

Performance summary report (multi-model)





	RMSE	Bias	IOA	…
Mod 1				
Mod 2				
Mod 3				
Mod X				

Conclusions & discussion

- **Completeness of the testing levels**
- **Composite diagrams to synthesize information**
- **Choice of relevant indicators and diagrams to define core set depending on application (model type?)**
- **Complexity, organization and size of the reports:**
  - Nb. of diagrams & indicators
  - Nb. of variables tested
  - Summary and extended report sections
- **Bounds: definition & updating process**

Work Plan

- Discussion and consensus on overall methodology (FAIRMODE meeting 09/2010)
- Development of the DELTA and ICI-MOI benchmarking service prototypes (Dec 2010)
- Testing of the prototypes on existing datasets (2011)
- Development of the JRC Web facilities (MMI-MRI benchmarking, data extraction, harmonization of output formats…)
- Set-up of a joint exercise for testing of the whole system (2012)

Contributions / Interactions

- Discussion and definition of the benchmarking service elements (species, statistics, goals and criterias…) for model performance reporting per pollutant/scale.
  - Urban/agglomerate scale: first on POMI dataset but other datasets required (even single model validation) → workshop by mid 2011
  - European scale: within the Eurodelta exercise → draft by end 2011
  - Local scale: Datasets are required → ??
- Practical organization & communication
  - Are emails sufficient?
  - Intermediate workshops?
- Links to other SGs
  - Required methodology to assess station representativness
  - Data assimilation techniques could make use of benchmark databank (in future)
- Definition of and participation to the joint activities