H14-174

PERFORMANCE CRITERIA FOR THE BENCHMARKING OF AIR QUALITY MODEL REGULATORY APPLICATIONS: THE "TARGET" APROACH

Anna Pederzoli[1], Philippe Thunis[1], Emilia Georgieva[1], Rafael Borge[2], David Carruthers[3]

[1]Joint Research Centre of the European Commission (JRC), Ispra (Italy)
[2]Universidad Politecnica de Madrid, Madrid (Spain)
[3]CERC, Cambridge (UK)

Abstract: As part of the FAIRMODE benchmarking activities (Thunis et al., 2010), performance criteria for air quality modelling in the frame of the EU Air Quality directive (AQD) 2008 are proposed and tested. The suggested approach builds on the Target indicator (Joliff et al., 2010) for quantitatively estimating model performances in air quality modelling applications. This work presents this methodology applied to NO_2, O_3 and PM_{10} concentrations on three different model-observations datasets.

Key words: *Fairmode, model evaluation, performance criteria, Target diagram.*

INTRODUCTION

Performance criteria are bounds which state whether a sufficient level of accuracy for policy applications is reached (Boylan and Russel, 2006). Up to now, only few performance criteria have been proposed in the literature and not always in a harmonized way (e.g. different statistical indicators are favoured by different Authors for the same purpose or different indicators are proposed for different variables). This prevents the inter-comparability of model performance results. One of the objectives of the Fairmode SG4 group on benchmarking is to develop a harmonised methodology to evaluate model results based on a consensus set of statistical indicators. This set needs to be reduced enough to provide an easy-to-compare synthesis but large enough to capture the main aspects of model performances. This methodology follows and builds on the recently proposed Target diagram (Joliff et al., 2010). This diagram gives the chance to visualize different aspects of model performance in one single plot. The statistical indicator directly relevant to the modelling uncertainty according to the EU AQD is the Relative Directive Error (RDE) (Denby et al., 2010). As RDE is mainly related to the exceedance of limit values, it provides only a partial view of the strengths and weaknesses of a given model application. Thus, in order to give a more comprehensive response of model performances, a series of additional statistical indicators need to be addressed. Within Fairmode the use the Target indicator (radius of target diagram) as performance criterion is also suggested. The Target diagram is very helpful because relevant information, including those coming from traditional indicators, is included in the plot, providing an almost exhaustive indication of the model response. The correspondence on the plot between the Target indicator and commonly used statistical indicators (e.g. R, Bias…) is indeed not straightforward due to normalization operations, and it needs to be addressed.

The Target diagram has been introduced inside the Delta Tool (Thunis et al., 2010), a software recently developed by the Joint Research Centre of the European Commission in Ispra (Italy) for harmonized diagnostics of air quality model performances. The DELTA tool is used here on three model-observation datasets. The first two include modelled results provided by the air quality models CMAQ and ADMS for years 2007 and 2008 respectively in the urban areas of Madrid and London while the third dataset (POMI) covers the Po valley and includes multiple model results for the year 2005.
A brief description of the target plot, the input datasets and their use for defining new performance criteria are given in the next sections. Then an analysis of the correspondence between the Target and the traditional indicators is provided for these three datasets and first guess performance criteria for these indicators are fixed.

TARGET DIAGRAM DESCRIPTION

The Target diagram, a recent evolution of the Taylor diagram (Taylor, 2001), provides information about the Mean Bias (MB), the Standard Deviation, the Root Mean Square Error (RMSE), the Centred Root Mean Square (CRMSE), the correlation coefficient R and the Model Efficiency MEF.

A schematic description of the Target plot as used in Delta is given in Figure 1. CRMSE and MB, both normalized by the standard deviation of observations (σ_o) are reported on the X-axis and Y-axis respectively. The distance between the origin and any point then represents the RMSE normalised by σ_o for a single station, and it is referred here as "Target" indicator. The outermost circle of the diagram corresponds to the performance criteria for the "Target". The target is related to the Model Efficiency score (MEF):

$$MEF = 1 - \left(\frac{RMSE}{\sigma_o} \right)^2$$

If the target value is outside this circle, MEF is negative. For points inside the circle, MEF is greater than zero. A positive MEF means that the model is a better predictor of the observations than a constant value set to the average of observations (Jolliff et al., 2009). Furthermore, for points lying within the circle of radius 1, two further conditions are automatically guaranteed: 1) the Normalized MB and normalised CRMSE are less or equal to 1; 2) model and observation data are positively correlated. Since the CRMSE is always positive by definition, the X negative axis can be used for providing some additional information, such as the sigma ratio σ_m/σ_o (Jolliff et al, 2009; Willmot et al., 1985). The X-axis is actually used to make this distinction: stations characterized by a ratio <1 are plotted on the left side (negative abscissa) whereas stations characterized by a ratio >1 are plotted on the right side (positive abscissa). Similarly to the approach followed in the AQD, we require that the performance criterion for the "Target" indicator is fulfilled for at least 90% of the available stations (which corresponds to the requirement of having 90% of the available stations inside the outermost circle). This means that the Target diagram gives an immediate overview of the model performances for the selected stations. The more stations inside the circle and the closer to the origin, the better is the model performance for a given application.

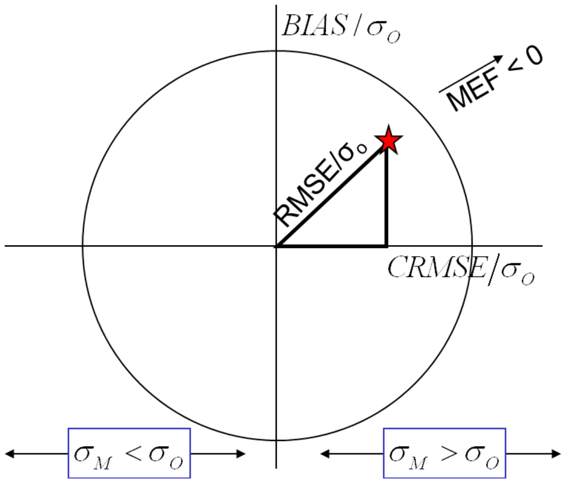

Figure 1. Schematic diagram for the Target plot.

DATASETS

Madrid test case. A modeling exercise over the city of Madrid has been carried out for 2007 by using the chemistry and transport model CMAQ 4.6 (Byun and Schere, 2006). The model domain is a 40 km x 44 km grid at 1 km^2 resolution. Emissions are based on EMEP, Spain and Portugal national official inventories as well as Madrid City Council inventory. The modeled output fields are NO_2 and O_3 concentrations (μgm^{-3}). Hourly observations of NO_2 and O_3 are from 10 urban background monitoring stations belonging to the Madrid City Council and Madrid Greater Region air quality monitoring networks.

London test case. A simulation over the urban area of London has been carried out for year 2008 by running the Atmospheric Dispersion Modelling System (ADMS). Details about the model can be found in Hanna et al. (1999) and CERC (2010). Modeled output fields are: NO_2 concentration (μgm^{-3}), O_3 concentration (μgm^{-3}) and PM_{10} concentration (μgm^{-3}). Hourly measured surface concentrations for the same pollutants are provided by 107 ground-level based stations belonging to the local network of London city (http://www.londonair.org.uk/london/asp/default.asp). They are classified as suburban, urban background, kerbside and roadside monitoring sites.

Po valley test case. This dataset contains the results from a model inter-comparison exercise performed by six air quality models for year 2005. The model domain covers the Po Valley with a 6 x6 km^2 resolution (95x65 cells) grid. NO_2, O_3 concentration and PM_{10} concentrations (μgm^{-3}) have been simulated by 8 chemical and transport models: CHIMERE, TCAM, CAMX, RCG, MINNI, AURORA. More details about the POMI exercise can be found at the POMI website http://aqm.jrc.it/POMI/index.html. Observations from 61 monitoring sites located in the Po Valley are also provided. Sites are classified as suburban, urban and rural background stations.

RESULTS

Performance criteria have been set for four traditional indicators which summarize potential errors in bias (MFB), phase (R) and amplitude (standard deviation and FAC2).The correspondence between the Target indicator and the traditional indicators is not straightforward due to normalization operations, and it is explained here through some examples.

A Target diagram for daily averaged PM_{10} concentration (POMI dataset, TCAM model) is reported in Figure 1. In order to show the correspondence between the target indicator and the traditional parameters (i.e. bias, correlation coefficient), as a first step the criterion of 60% for the Mean Fractional Bias (MFB) as suggested by Boylan and Russel (2006) is adopted. All stations for which model performances are too poor to satisfy the MFB criterion are shown as red dots. All red points remain outside a circle of radius close to 1. This value is then chosen as our reference target indicator for PM_{10}. It is also clear that some stations still remain outside this circle, even if they respect the MFB limit. These correspond to stations which behave relatively well in terms of MFB but poorly in terms of other indicators (e.g. R). We then vary the criteria on R, FAC2 and sigma progressively in order to exclude these outside points. Of course the correspondence is not univocal, but a criterion of 0.5 for σ, FAC2 and R appear to be suitable. Same responses are obtained for all models and datasets.

The same procedure can be followed for daily maximum 8-hours mean ozone concentration. For O_3, the MFB limit of 30% suggested by Chemel et al. 2010 (corresponding approximately to a factor 2 under- or over-estimation) is consistent with a target indicator (outermost circle) fixed to 0.8. This will be used as our Target indicator value. The same approach for FAC2, σ and correlation coefficient as the one for PM_{10} is then followed for progressively excluding the stations outside the circle. This lead to the following criteria: 0.5 for FAC2 and standard deviation, and 0.65 for the correlation coefficient. An example for the CMAQ model (Madrid dataset) is given in Figure 2: in this case these bounds are suitable for CMAQ and R, FAC2 and σ are not "responsible" for the sites outside the circle (the MFB only). The criteria fixed for O_3 are respected, for this specific case, by 66% of the stations.

For 1- hour maximum NO_2 concentration, no value has been suggested in the literature for MFB: in this case an arbitrary 30% is fixed, providing, like PM_{10}, an outermost target circle of radius 1. The Target analysis (not shown) suggests to adopting criteria of 0.5, 0.6 and 0.55 for FAC2, standard deviation and R respectively. An example for the ADMS model (London dataset) is given in figure 3.

All indicative values for each statistical indicator and the three pollutants are summarized in Table 1. These values must be considered as "first guess" criteria and need to be tested on more datasets covering other spatial areas and including more air quality models.

Table 5. First guess performance criteria for correlation coefficient, FAC2 and sigma as obtained by performing the Target plot analysis.

		MFB		Target		R	FAC2	σ
PM_{10}	Daily	60%	→	1.0	~	0.50	50%	± 50%
O_3	8h	30%	→	0.8	~	0.65	50%	± 50%
NO_2	1h	30%	→	1.0	~	0.55	50%	± 50%

CONCLUSIONS

In order to assess the performance of air quality models, in the frame of the FAIRMODE activities performance criteria based on the target diagram (Joliff et al., 2010) have been suggested. In this work we introduce a methodology to relate a composite indicator (Target) to a set of usual indicators synthetizing the different aspects of model performances: MFB, R, FAC2 and standard deviation ratio σ_m/σ_o. This methodology has been tested on three datasets covering different spatial areas: Madrid, London, The Po valley. Performance criteria for the Target indicator as well as indicative values for the usual performance indicators are currently proposed for daily maximum 8-hour average O_3, daily mean PM_{10} and 1-hour maximum NO_2 concentrations. The use of this set of indicators in the future will hopefully favour harmonization across EU in terms of a common evaluation standard. The proposed "first guess" performance criteria need to be verified on more datasets covering other spatial areas and including more air quality models. Similar criteria will be also developed in the future for yearly averaged values. Further developments will also include the introduction of performance goals (which indicate the optimal level of accuracy a model is expected to reach (Boylan and Russel, 2006)) inside the Target diagram.

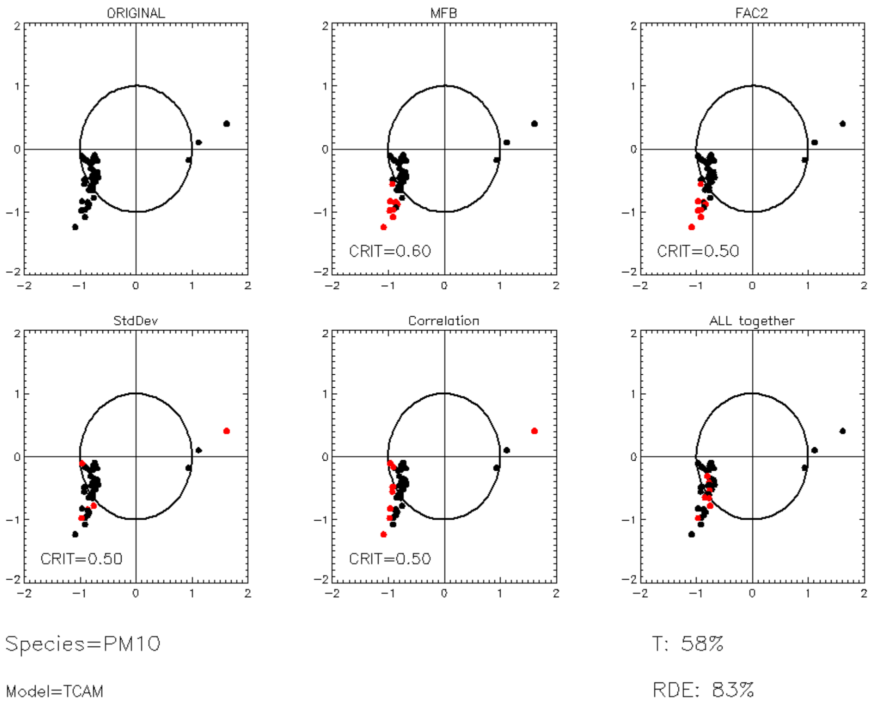

Figure 1. Target plot produced by the Delta Tool for daily averaged PM_{10} concentration, for the TCAM model (POMI dataset).

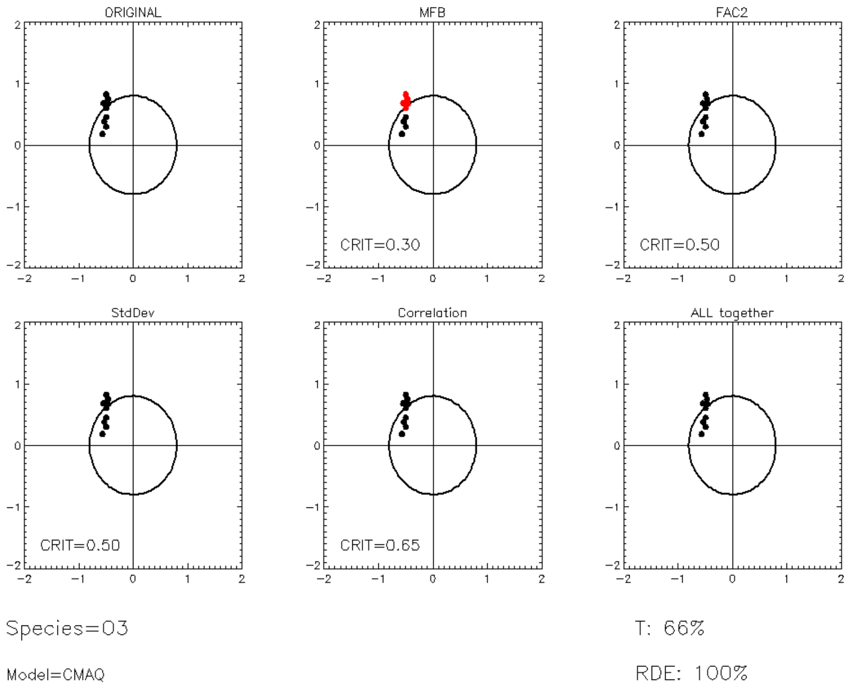

Figure 2. Target plot produced by the Delta Tool for daily maximum 8-hour mean ozone concentration, for the CMAQ model (Madrid dataset).

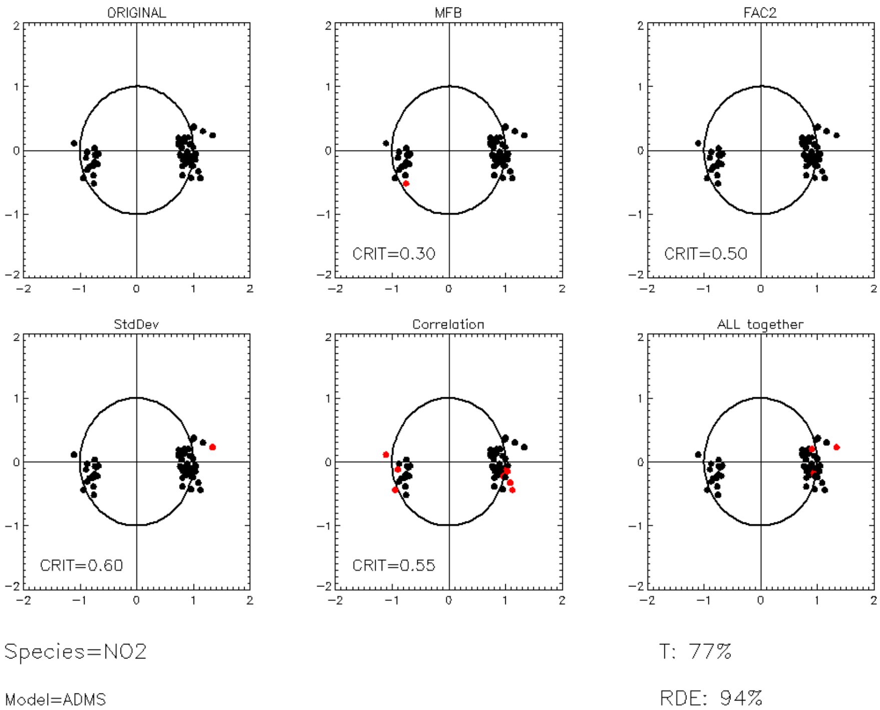

Figure 3. Target plot produced by the Delta Tool for hourly maximum NO_2 concentration, for the ADMS model (London dataset).

REFERENCES

Boylan J and Russel A. (2006). PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmospheric environment, 40, 4946-4959.

Byun, D.W., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 community Multiscale Air Quality (CMAQ) modeling system. Applied Mechanics Reviews 59, 51–77.

CERC (2010), ADMS 4 Technical Specification, www.cerc.co.uk

Chemel C., R.S. Sokhi, Y. Yu, G.D. Hayman, K.J. Vincent, A.J. Dore, Y.S. Tang, H.D. Prain, B.E.A. Fisher (2010) Atmospheric Environment, Volume 44, Issue 24, Pages 2927-2939

Denby B. (2010). Guidance on the use of models for the European Air Quality Directive (A FAIRMODE working Document), ETC/ACC report, ver 6.2. http://fairmode.ew.eea.europa.eu/

Hanna SR, Egan BA, Purdum J and Wagler J. (1999). Evaluation of the ADMS, AERMOD and ISC3 Models with the Optex, Duke Forest, Kincaid, Indianapolis and Lovett Field Data Sets. International J. of Environment and Pollution (Volume 16, Nos. 1-6, 2001)

Jolliff J.K., Stow C.A., D.J. McGillicuddy Jr., S.C. Doney, J.I. Allen, K.A. Rose and P. Wallhead (2009) Skill assessment for coupled biological/physical models of marine systems, Journal of Marine Systems 76, pp. 4–15.

Taylor K.E. (2001) Summarizing multiple aspects of model performance in a single diagram Journal of geophysical 509 research. **106**, no D7: 7183-7192

Thunis P., E. Georgieva, A. Pederzoli (2010) The DELTA tool and Benchmarking Report template. Concepts and User guide, version 2. Available at http://aqm.jrc.it/DELTA/data/DELTA_SG4_V2.pdf