H14-235
BENCHMARKING OF A REGULATORY URBAN AIR QUALITY MODELLING SYSTEM UTILIZING PROBENCH

Karppinen A., Kauhaniemi M., Kangas L, Härkönen, J. , Kukkonen, J.,

Finnish meteorological Institute, Air Quality Research , Finland

Abstract: We present a first practical application of "Procedure for air quality models benchmarking" PROBENCH for FMI-regulatory models benchmarking , provide example evaluation results based on archived model and measurement data and finally discuss the added value this benchmarking procedure provides compared to regulatory modelling and model development.

Key words: *model evaluation, air quality, PROBENCH, DELTA, modelling*

INTRODUCTION

This study is based on existing database of predicted and measured hourly air quality data during 1993-2009 , consisting of air quality model results for 10 full years during the period and measurement data for 7 monitoring stations in Helsinki metropolitan Area for the full 17 years . The air quality monitoring stations operated by Helsinki Region Environmental Services Authority (HSY) utilized in this study are located in traffic or urban background locations. The full database of monitored pollutants consists of nitrogen monoxide NO, nitrogen dioxide NO2, ozone O3, sulphur dioxide SO2, thoracic particles PM10 and fine particles PM2.5. For this study the selected pollutants were NO2 and PM2.5 and study years 1993 and 2002 correspondingly. This re-evaluation is useful in assessing the completeness and reliability of the earlier model performance assessments done in the same area. As all the earlier evaluation studies for some specific years and some specific pollutants can be considered as a subset of this new comprehensive evaluation, we can also assess the consistency of the old evaluations with the new benchmarking results.

MATERIALS AND METHODS

Dispersion modelling is based on the combined application of the Urban Dispersion Modelling system (UDM-FMI) and the road network dispersion model (CAR-FMI), developed at the FMI (Karppinen et al., 2000a). Emissions from approximately 6000 road and street links in the Helsinki Metropolitan Area are included in the computations. The model uses meteorological parameters from the FMI database and calculates ambient air pollution concentrations for each hour over the whole year. The concentrations are calculated in an adjustable grid, spatial resolution of the grid being dependent on the distance of calculation points to nearest traffic sources and varying from 5 meters to 1 km. For this re-evaluation study, we did not perform any new model calculations; instead we utilized the existing data-base of model results for Helsinki area. This made it also possible to partially verify the calculated statistical measures against earlier

We analyse the agreement between measured and predicted concentrations utilizing the new "Procedure for air quality models benchmarking" PROBENCH (Thunis, 2010) and specifically on of the main elements of PROBENCH, the DELTA-tool. The DELTA software is IDL-based evaluation software which includes the main assets of the EuroDelta, CityDelta, and POMI tools. It allows the user to perform a rapid diagnostic of the model performances. DELTA focuses on the pollutants mentioned in the AQ directive and addresses all relevant spatial scales and it works temporal series of modelled and monitored data at selected ground level locations. The tool provides complete evaluation measures independently of model gridding and spatial scale but the user must be careful to ensure comparability between grid-cell averaged model results and point measurements. In our case scales were not on issue, as CAR-FMI results are natively calculated to a set of given calculation points, not averaged to grid cells.

RESULTS

The model-measurement database available for Helsinki area covers 10 full years during the period 1992-2010. We present here results for 2 specific case-studies for this period (1993 and 2002). In the first case we evaluated the NO2 model performance for 3 different stations in Helsinki Metropolitan Area: Töölö, Vallila, and Tikkurila and in the second case the PM2.5 model performance against observations at 2 station, Vallila and Kallio is evaluated. Only some extracts of the complete evaluation results are presented here.

Table 1. Evaluation statistics for the NO2 modelling 1993 based on the original study (*Karppinen et al, 2000b*)

Statistical Parameter	Töölö		Vallila		Tikkurila	
	Predicted	Observed	Predicted	Observed	Predicted	Observed
Mean (µg/m3)	43	45	44	40	27	24
Maximum (µg/m3)	191	167	211	176	144	171
Standard deviation (µg/m3)	22	26	23	21	18	18
Index of agreement (IA)	0.75		0.69		0.79	
Pearson's correlation coefficient (COR)	0.57		0.50		0.65	
Normalised mean square error (NMSE)	0.28		0.26		0.33	
Fractional bias (FB)	- 0.045		0.095		0.118	

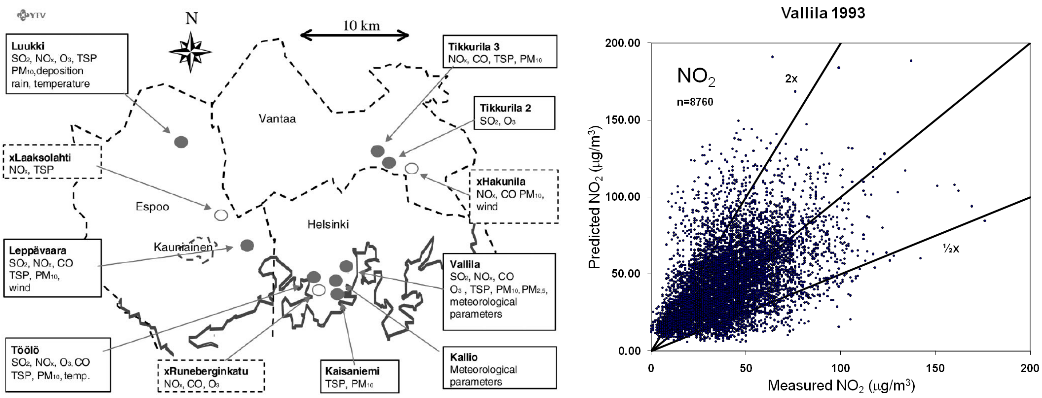

Figure 1. Measurement stations in the Helsinki metropolitan area and a scatterplot of predicted vs. measured NO2 concentrations at Vallila monitoring station

Second case study evaluated the model performance for PM2.5 concentrations in Helsinki area. (*Kauhaniemi et al, 2008*).

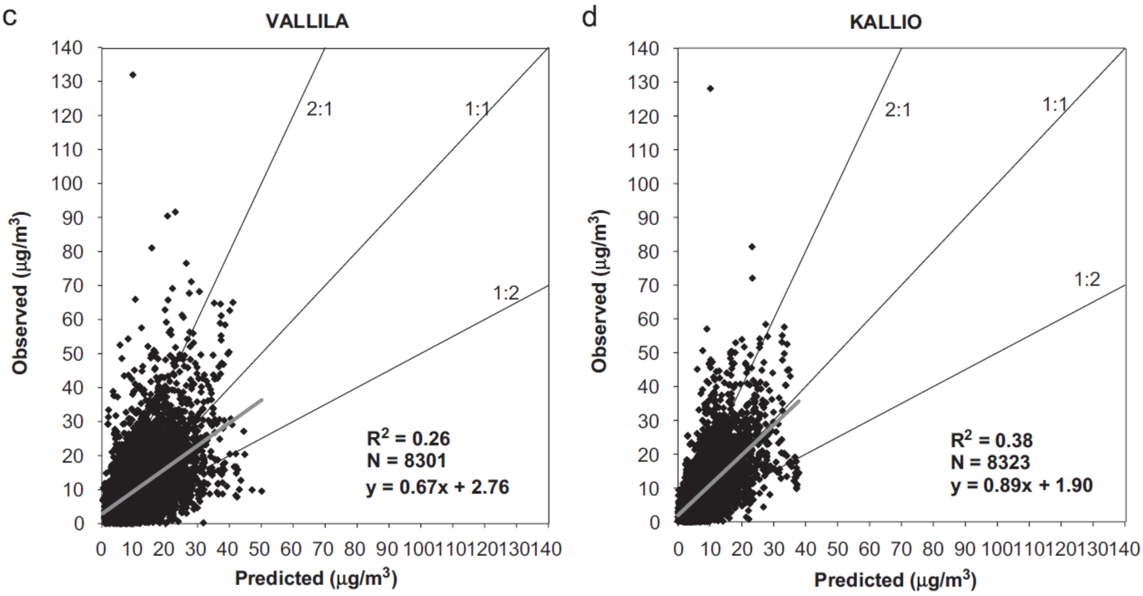

Figure 2. Scatterplots of measured vs. predicted PM2.5 concentrations at Vallila and Kallio monitoring station in 2002.

Table 2. The statistical analysis of the predicted and observed hourly average time series of PM2.5 concentrations at the Vallila and Kallio monitoring stations for 2002.

Statistical Parameter	Vallila		Kallio	
	Predicted	Observed	Predicted	Observed
Mean (µg/m3)	11.3	10.4	8.06	9.07
Maximum (µg/m3)	50.2	132	37.9	128
Standard deviation (µg/m3)	6.47	8.48	5.13	7.36
Index of agreement (IA)	0.69		0.74	
Pearson's correlation coefficient (COR)	0.51		0.62	
Normalised mean square error (NMSE)	0.40		0.36	
Fractional bias (FB)	0.09		-0.12	
Number of data	8301	8301	8323	8323

Tables 1-2 present the complete set of statistical measures used for the model evaluation and *Figs. 1-2* present the typical graphical scatter-plot analyses performed with every model evaluation exercise. In addition to the fixed set of statistical parameters also some additional, more detailed analyses on e.g. the seasonal variation of the model skill is typically performed (e.g. *Karppinen et al, 2000b, Kauhaniemi et al.,2008)* , but the need and content of the additional analyses is very

much decided case-by-case. The model performance was assessed generally "satisfactory" in the earlier studies. The re-evaluation gives us a detailed view on the performance of models, states clearly which statistical indicators show acceptable model skill, and also points clearly out if the model skill is consistent at all different monitoring locations.

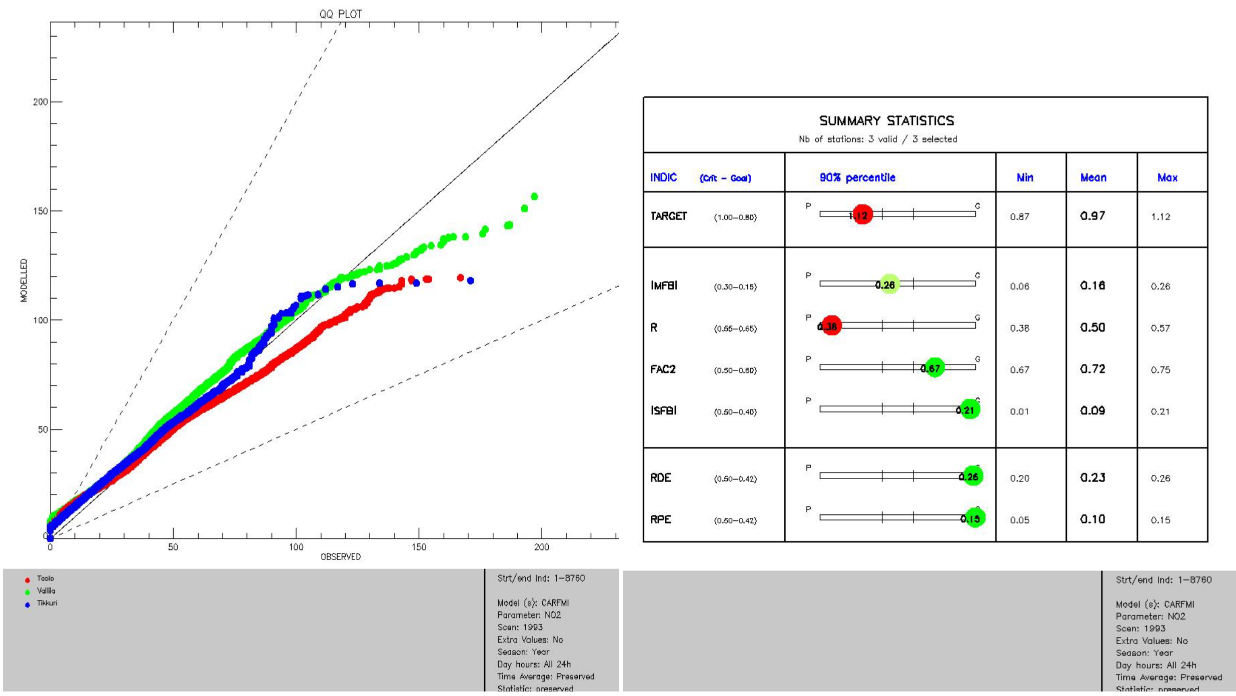

Figure 3. DELTA QQ-plot and summary statistics for CAR-FMI's NO2 prediction skills for 3 monitoring stations.

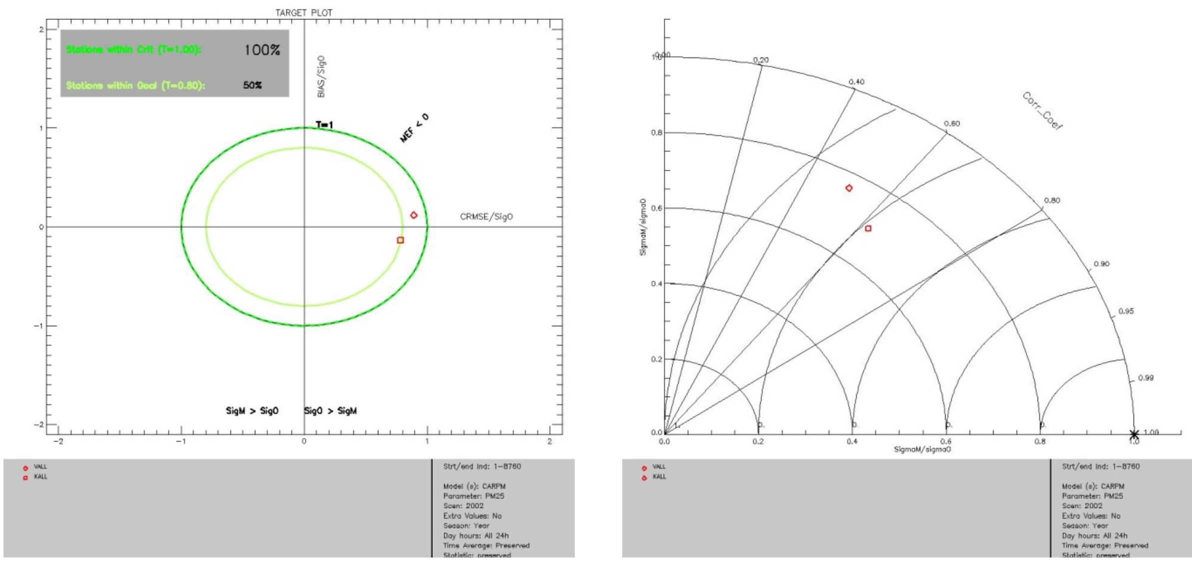

Figure 4. Target diagram and Taylor diagram for the CAR-FMI/PM2.5 –model prediction skills for 2 monitoring stations.

Corresponding evaluation results calculated by DELTA-tool are presented in *Figs. 3-4*. Especially the Taylor and Target diagrams are very interesting additions to the visual model-skill assessment, as they are showing several crucial statistical parameters : correlation coefficient, cantered root-mean-square-error, standard deviation (Target: BIAS and model efficiency score - MEF) in one, easy-to-read diagram.

The re-evaluation results agreed nearly completely with the original evaluations on all the common statistical measures. However, for one specific case (Vallila/NO2/1993) the re-evaluation showed completely different values for all the statistical parameters compared to the original study. This turned out to be an issue with the original measurement data. The archived data used in the DELTA-analyses was based on the original, non-corrected data extracted directly from the non-quality confirmed measurements. A re-run with measurement data where all erroneous values were removed from the original data corrected the discrepancy completely. The visualisation options of DELTA-tool proved to be very useful in finding out the real reason behind the erroneous evaluation statistics. The evaluation work with the complete database will continue, but already these first results were enough to proof the value of the PROBENCH concept and DELTA tool.

DISCUSSION

The basic statistical parameter set utilized in the FMI air quality model evaluations has been quite consistent during the last 15 years as illustrated eg. by the tables 1-2) . Obviously it is only a subset of the parameters recommended in the state-of-art PROBENCH methodology. Even more important difference in the methodologies is that the earlier evaluations did not clearly specify any clear acceptance criteria for the model behaviour, while PROBENCH /DELTA has added a multitude of statistical indicators with clear acceptance levels to help to identify unacceptable model performance. For each statistical indicator, two quality bounds are proposed: a performance criterion which states whether sufficient quality for policy application is reached and a performance goal which points to the optimum quality level that a model is expected to reach. These two quality bounds will greatly assist the user in assessing the quality of the model performances for a given AQD application.

One of the most important differences between earlier evaluation methods and the PROBENCH- methodology is not so much a scientific one but more a very practical one. Use of the new tool and reporting schemes will obviously make it much easier and practical to assess and evaluate regulatory models with the exactly same set of measures. PROBENCH will most probably bring also some added value to research and development as the new statistical toolbox for evaluation will make it very easy to study in more detail all relevant aspects of the data available. In most cases the direction for model development is guided by results from evaluation studies. As PROBENCH/DELTA will make generating state-of-art evaluation statistics easy for even an inexperienced model developer, this can be foreseen to help and enhance the practical model development work in the future.

CONCLUSIONS

DELTA tool has been utilized to re-assess model calculations for Helsinki Area based on archived model and measurement data for the area. Although the new evaluation did not reveal any major discrepancies in the calculated statistical parameters, the new concept and tool already proved to be suitable tool for processing data created long time before any format or content requirements for DELTA tool were created.

As with any new tool, some time and resources are needed to learn to use the tool effectively, but the first experiences with the tool are promising enough to prove, that the resources invested are going to be returned in the future. Assuming that a continuous support and resources for further development for the tool can be guaranteed, it will be an important step towards harmonizing the European model evaluation practices in the future. Although the main aim of the tool is to help in evaluating regulatory model calculations , it can be foreseen to be a useful tool for research model development too, although neither DELTA nor any other tool can ever completely automate the complicated process of full model evaluation.

REFERENCES

Karppinen, A, Kukkonen, J., Elolähde, T., Konttinen, M., Koskentalo, T. and Rantakrans, E., 2000a. A modelling system for predicting urban air pollution, Model description and applications in the Helsinki metropolitan area. *Atmospheric Environment* 34, pp. 3723-3733.

Karppinen, A., Kukkonen, J., Elolähde, T., Konttinen, M. and Koskentalo, T., 2000b. A modelling system for predicting urban air pollution, Comparison of model predictions with the data of an urban measurement network. *Atmospheric Environment* 34, pp. 3735-3743.

Kauhaniemi M , Karppinen A , Kukkonen J, Härkönen J , Nikmo J , Kousa A, Alaviippola B., Koskentalo T and Aarnio P, 2008. Evaluation of a modelling system for predicting the concentrations of PM2.5 in an urban area, *Atmospheric Environment* 42 (19), pp. 4517-4529.

Thunis P., E. Georgieva, S. Galmarini, 2011: A procedure for air quality models benchmarking. (http://fairmode.ew.eea.europa.eu/models-benchmarking-sg4/wg2_sg4_benchmarking_v2.pdf/download)