A MODEL EVALUATION PROTOCOL FOR URBAN SCALE FLOW AND DISPERSION MODELS

Silvana Di Sabatino[1], Helge Olesen[2], Ruwim Berkowicz[2], Jörg Franke[3], Michael Schatzmann[4], Rex Britter[5], Heinke Schlünzen[4], Alberto Martilli[6] and Bertrand Carissimo[7]

[1]Dipartimento di Scienza dei Materiali, University of Salento, Lecce, Italy
[2]Department of Atmospheric Environment, National Environmental Research Institute, Roskilde, Denmark
[3]Department of Fluid and Thermodynamics, University of Siegen, Germany
[4]Meteorological Institute (ZMK/ZMAW), University of Hamburg, Germany
[5]Department of Engineering, University of Cambridge, United Kingdom
[6]CIEMAT, Madrid, Spain
[7]CEREA, Teaching and Research Centre in Atmospheric Environment, CHATOU Cedex, France

Abstract: This paper reports on a comprehensive model evaluation protocol for urban scale flow and dispersion models that has been developed within the framework of the COST action 732 on Quality Assurance and Improvement of Micro-Scale Meteorological Models.
It briefly discusses the different components forming model evaluation with particular emphasis on model validation and on the implementation of the protocol for a specific test case: the MUST (Mock Urban Setting Test) experiment.
The protocol was first developed with building-resolving models in mind, but more traditional integral models have also been included. Currently the Action is finalising the MUST exercise results and will suggest the best approach for further model evaluation and for the standardization of CFD modelling practise for micro-scale meteorological applications.

Key words: *COST Action 732, model evaluation protocol; MUST experiment, CFD and non-CFD models*

1. INTRODUCTION

Very recently there has been an increase in the development and use of models within Europe for urban air quality applications. These models have begun to play a crucial role in environmental assessment and urban climate studies; studies that are undertaken to investigate and to quantify the effects of human activity on air quality and on the local climate. Until now only limited work has been done to check the performance of these kinds of models and currently there is no standardized modelling practise for atmospheric applications.

The role of the COST Action 732 (2005-2009) has been to overcome this lack of information. Particular attention has been given to producing a methodology for assuring the quality (fitness-for-purpose) of micro-scale meteorological models and providing the basis for a standardization of modelling practise for flow and dispersion applications within urban areas. In this context the Action has interpreted "micro-scale meteorological models" as models for the prediction of the flow and/or the dispersion of pollutants within and near the urban canopy or industrial landscape.

The impact of the COST Action 732 depends upon whether the evaluation procedures suggested by the Action become widely accepted by the community of model developers and model users. In May 2007 the first version of the evaluation procedure was released in three official documents that are publicly available on line at http://www.mi.uni-hamburg.de/Official-Documents.5849.0.html. These documents are:

- Background and Justification Document to support the Model Evaluation Guidance and Protocol, Version 1, May 2007 (Britter, R., and M. Schatzmann, 2007a).
- Model Evaluation Guidance and Protocol Document, Version 1, May 2007 - a stand-alone document to assist the setting up and executing of a model evaluation exercise (Britter and Schatzmann, 2007b).
- Best Practice Guideline for the CFD simulation of flows in the urban environment, Version 1, May 2007 - based on published guidelines and recommendations, which mainly deal with prediction of the statistically steady mean flow and turbulence in urban areas under conditions of neutral density stratification (Franke et al., 2007).

The Model Evaluation Guidance and Protocol Document is a condensed version of the background document. It gives specific guidance to model developers and users on how to evaluate and assure the quality of the model in an objective and defendable manner. The evaluation process includes various components and, particularly, a scientific assessment, model verification and a model validation. To conduct a validation (comparison of model predictions with experimental observations) one will have to decide for which purpose the model results should later be used and thus to decide the variable(s) whose prediction is the most important. In other words the validation objectives have to be explicitly specified. Briefly, a possible (very simple) baseline approach to a model validation (Britter and Schatzmann, 2007b) is to:

- decide to allow those running the models to either have access to the experimental results or not (that is non-blind or blind simulations)
- select the mean velocity and mean concentration (based on certain temporal and spatial scales) for comparison
- look at the streamline pattern and the concentration pattern to provide a qualitative view of the model quality

- produce a quantitative validation of the model quality by comparing the experimental and model data "paired in space and time" and as "arc maxima". The complexity of the flow may make the latter choice less feasible where local maxima could be distributed widely; to use the metrics of FAC2 (for its transparency) and FB/NMSE (for information on bias and variance). The FB/NMSE weights the higher magnitudes at the expense of the smaller

The documents will be updated with recommendations for datasets that are currently being used in the validation work. The recommendations given in the documents listed above are presently being tested by the COST 732 Action itself. Many European academic and research groups have participated in this action with the objective of building a consensus within the scientific community on evaluation of micro-scale meteorological models. Currently the Action is finalising the results for the Mock Urban Setting Test (MUST) (Yee and Biltoft, 2004), which comprises both field and wind tunnel experiments from flow and dispersion experiments carried out within and above a simulated urban setting made up from 120 standard size shipping containers.

The wind tunnel measurements with a scaled model (1:75) of that configuration were carried out at the University of Hamburg (Bezpalcova, 2007). So far, several European groups of numerical modellers (using both CFD and non-CFD models) have simulated the wind tunnel MUST experiments following the model evaluation guideline. The experiments used were those with two main wind directions, 0° and -45° (and these correspond to 270° and 315° respectively, in meteorological terminology) of the approaching flow. This study was launched in Athens in October 2006. Some of the results achieved are discussed in companion papers (Franke et al., 2008; Olesen et al., 2008).

The next phase will be the launch of an exercise based on field experiments in a real city. This exercise will be devoted to the simulation of dispersion experiments over Oklahoma City, USA. Several groups are currently involved in preparing simulation grids; simulation results are expected to come from about 5 to 10 teams. Initially, the simulations will be done in form of a blind test, i.e. only the geometry and the input data will be given to the participants.

2. PROTOCOL IMPLEMENTATION AND METHODOLOGY

Computational Fluid Dynamics (CFD) models are an appropriate tool for this application and their evaluation is central to COST 732. However, the application of simpler models to these problems is of direct interest to many participants in COST 732. Our approach has been to develop a methodology that can be used for CFD models and can also be modified to accept simpler models.

Several CFD models have been used by different groups from many European countries to simulate the MUST experiment. They are: Miskam, Fluent, ADREA , Star-CD, Finflo, Cfx, Mitras, Tsu/M2UE, VADIS, Code_Saturne. Only recently, non-CFD models, such as Lasat, ADMS-Urban, RAMS, OML, ESCAPE, CALPUFF, have received increasing attention within COST 732 and results for them are in progress.

For the comparison of numerical results with experimental data, both qualitative and quantitative approaches have been chosen. There is a common understanding that exploratory qualitative data analysis using graphical comparison between model and data and an inter-comparison among models gives a simple, useful and transparent way of showing the strengths and weaknesses of models. For the evaluation of a model both qualitative (through profiles and contours) and quantitative (through statistical analysis) approaches are needed, otherwise statistical parameters alone could obscure deficiencies of the model. This particular aspect is being investigated in Olesen et al. (2008).

In particular, in our proposed methodology model results needs to be analysed in a combined way by means of:

- contours of velocity components, turbulent kinetic energy (TKE) and Reynolds stress components;
- vertical and horizontal profiles of velocity components and TKE;
- profiles of dimensionless concentration. In the example provided we only use the -45° approach flow case as concentration measurements were not available for the 0° case;
- statistical analysis.

The first three are essentially a qualitative analysis, in which model results are evaluated using direct point-by-point comparisons with wind tunnel data. This approach was preferred over the alternative of using manipulated data such as estimating a maximum concentration on an arc and using it for model comparison purposes. The fourth point is indeed quantitative and is used by employing several statistical measures, such as the fractional bias (FB), the normalized mean square error (NMSE), the fraction of predictions within a factor of two of observations (FAC2), the geometric mean bias (MG) and the geometric variance (VG). Typical magnitudes of the above performance measures and estimates of model acceptance criteria have been summarized by Chang, J. and S. Hanna, (2004) based on extensive experience with evaluating many models with many field data sets. The commonly accepted values for "state of the art" model performance are: $-0.3<FB<0.3$; $NMSE<4$; FAC2 0.5; $MG<1$; $VG<1.5$. Also the hit rate evaluation test (VDI guideline 3783 Part 9, 2005-11) should be performed using a fractional deviation D=0.25 and specific absolute deviations W for the different variables analysed (the hit rate must not fall below 66% for the comparison with wind tunnel data).

The set of models involved in COST 732 is a representative sample of the micro-scale meteorological models currently available and widely used in Europe and with the MUST exercise the Action's intention is to suggest criteria for the “state of the art” of the modelling process. The state of the art is a dynamical concept; models constantly improve and the state of the art consequently changes. So the methodology which the Action is following will contain a procedure to update the criteria, so that if, in the future, new models are run using the COST 732-MUST case or other data, the value of the metrics reflecting the state of the art will can be modified.

A somewhat different question concerns the “fitness for purpose” criteria as this changes with the intended purpose of the model. An important point to be addressed by a model user is whether the “state of the art” will satisfy the “fitness for purpose” criteria for the particular purpose of the modeller. When determining model quality, it is obviously essential to consider and specify what the purpose of the model is. For example models with a similar scientific basis may be required for quite different purposes such as:

- a model for planning or regulatory purposes may need to be run several thousands of times;
- a model for emergency response may need real-time predictions or access to pre-calculated real-time output;
- a model for post-accident investigation or air quality hot spot analysis could be very complex with less concern for computational cost or resource requirement.

And an assessment of the “fitness for purpose” of the same model could be very different for each of the above purposes.

3. EVALUATION EXERCISE

The MUST experiment

Vertical and horizontal profiles of wind tunnel data and the results from the various model simulations have been collected in Excel spreadsheets that include a macro tool, which allows easy graphical comparisons. The tool was developed within this Action by Berkowicz et al.: (http://www.dmu.dk/International/Air/Models/Background/MUST.htm). This tool was found to be extremely useful for exploratory data analysis to highlight both large errors and subtle differences among the models.

In particular, there is a group of Excel files that allows easy graphical inspection of the details of every case for all of the models. Another group of Excel workbooks contain essential information extracted from the above Excel files which summary plots and metrics. As an example, the graphs are in the two sheets PlotsX (Fig. 1) and PlotsZ

For the 0° approaching flow case (not shown here), the developed macro allows us to note that the qualitative behaviour of the models is different. Some of them seem to underestimate the wind velocity in the layer occupied by the buildings, while others overestimate the velocity. No model represents the z-velocity (vertical velocity) well at positions behind buildings.

As an example of the qualitative evaluation, some profiles of the dimensionless concentrations for the -45° case are shown in Figure 2, where concentration results at the beginning, middle and end of the array are plotted at z=0.5H (where *H* is the height of the building). The qualitative behaviour of the pollutant plume seems to be acceptable.

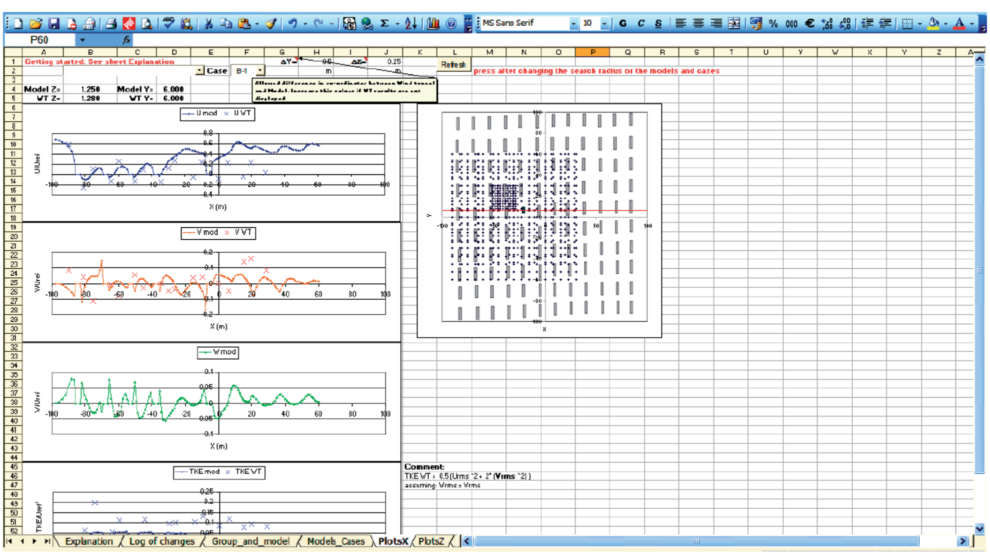

Figure 1. Example of the Excel spreadsheet including a macro tool which allows easy graphical comparisons.

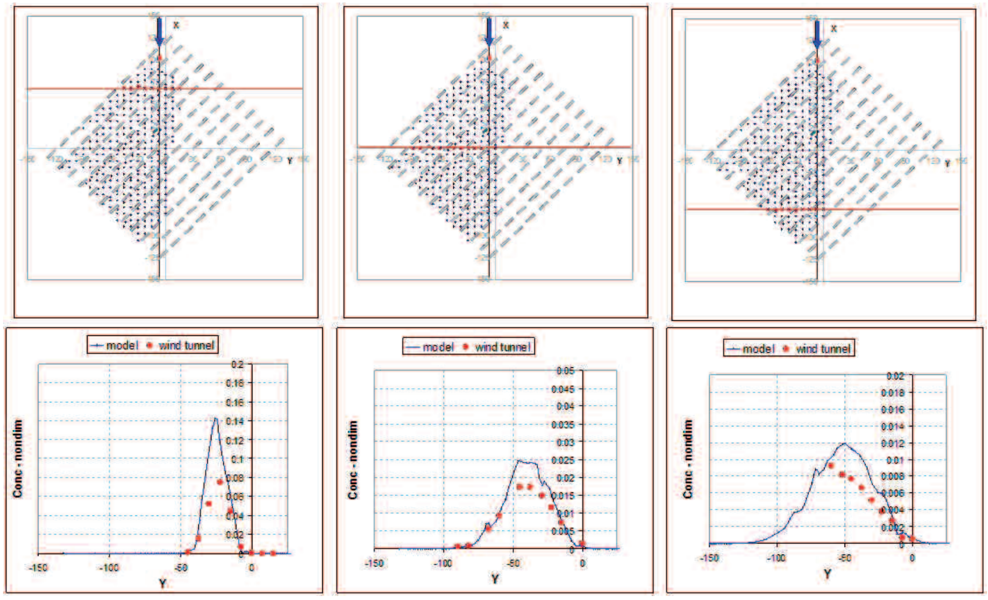

Figure 2. -45° case, samples of horizontal profiles (bottom) of dimensionless concentrations from one CFD model at z=0.5H at the beginning (left), middle (middle) and end (right) of the array. In the figure the array is rotated and the source is in the upper part (top).

Companion papers presented at this Conference deal with detailed CFD and non-CFD model results and specific statistical analyses. Those papers will discuss the differences in the metrics for different models for the MUST data and for other datasets and the criteria for the "state of art" and for "fitness for purpose" for typical purposes. This will help to formulate a Best Practice Guideline specifically for using the MUST data and to revise the Model Evaluation Guidance and Protocol Documents. Further detailed qualitative and quantitative results will be presented during the Conference and publications on specific aspects of the exercise have been planned.

4. CONCLUSIONS AND WORK IN PROGRESS

The comparison carried out in the COST 732 shows that flow and concentration model results compare relatively well with the measurements. The prediction for the x-velocity component is better than for the z-velocity one. The Excel tool developed within the COST 732 has allowed us to make detailed studies of the differences in model results and has helped us to emphasized strength and weakness of synthetic statistical parameters. In some specific cases the models show some weaknesses in predicting the complex flow especially the turbulent structure of flow. Correct specification of the inlet profiles and specific aspects of two-equation turbulence models were thought to require further investigation.

For the evaluation of the accuracy of a model both qualitative and quantitative approaches are needed. Statistical measures alone could lead to wrong conclusions. This could be true especially when measurements are limited to few points/profiles or when raw data are used without special treatment of the values that are smaller than the allowed deviation considered in the statistical analysis.

Currently the Action is finalising the MUST exercise and will suggest the best approach for further model evaluation for the standardization of CFD modelling practise for micro-scale meteorological applications. This will include a critical review and refinement of the numerical results before proceeding to the next model evaluation exercise. To assist with this several small working groups have been formed to investigate specific aspects including boundary conditions, statistical measures and non-CFD model evaluations.

The result of the protocol implementation through the large MUST experiment exercise have allowed us to:

- develop a coherent and structured quality assurance procedure for these types of models that gives clear guidance to developers and users of such models as to how to properly assure their quality and their proper application;
- provide a systematically compiled set of appropriate and sufficiently detailed data for model validation work in a convenient and generally accessible form.

The next phase, based on the Oklahoma City experiments will help to further strengthen our previous results and to finalize our original objectives that are:

- to build a consensus within the community of micro-scale meteorological model developers and users regarding the usefulness of the procedure;

- to stimulate a widespread application of the procedure and the preparation of quality assurance protocols which prove the 'fitness for purpose' of all micro-scale meteorological models participating in this activity;
- to determine the current "state of the art" of the modelling process and to give recommendations for the improvement of present models and, if necessary, for new model parameterisations or even new model developments.

The discussion of the quality assurance procedure, the use of specific datasets and the recommendations specified in the Best Practice Guideline should lead to a harmonized approach accepted at least at the European level. It is to be expected that the very existence of a widely accepted European standard for quality assurance in the field of micro-scale meteorological models in combination with the provision of suitable validation data will significantly improve "the culture" within which such models are developed and applied.

Acknowledgements: *The authors would like to thank DTRA/DPG for providing MUST and Oklahoma data for use in COST732.*

REFERENCES

Bezpalcova, K., 2007: Physical modelling of flow and diffusion in urban canopy. Ph.D. thesis, Faculty of Mathematics and Physics, Charles University in Prague.

Britter, R., and M. Schatzmann, (Eds.), 2007a: Background and Justification Document to support the Model Evaluation Guidance and Protocol Document. COST Office Brussels.

Britter, R., and M. Schatzmann (Eds.), 2007b: Model Evaluation Guidance and Protocol Document. COST Office Brussels.

Carissimo, B., S.F. Jagger, N.C. Daish, A. Halford, S. Selmer-Olsen, K. Riikonen, J.M. Perroux, J. Wuertz, J.G. Bartzis, N.J. Duijm, K. Ham, M. Schatzmann and R. Hall, 2001: The SMEDIS Database and Validation Exercise. *Int. Journal of Environment and Pollution,* **16**, 614-629.

Chang, J. and S. Hanna, 2004: Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, **87**, 167-196.

COST Action 732, 2005-2009: Quality assurance and improvement of micro-scale meteorological models. http://www.mi.uni-hamburg.de/Home.484.0.html.

Franke, J., A. Hellsten, H. Schlünzen, and B. Carissimo (Eds.), 2007: Best Practice Guideline for the CFD simulation of flows in the urban environment. COST Office Brussels.

Franke, J., J. Bartzis, F. Barmpas, R. Berkowicz, K. Brzozowski, R. Buccolieri, B. Carissimo, A. Costa, S. Di Sabatino, G. Efthimiou, I. Goricsan, A. Hellsten, M. Ketzel, B. Leitl, R. Nuterman, H. Olesen, E. Polreich, J. Santiago and R. Tavares, 2008: The MUST model evaluation exercise: Statistical analysis of modelling results. *Proc. 12th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes,* Cavtat (Croatia), 6-9 October.

Olesen, H.R., A. Baklanov, J. Bartzis, F. Barmpas, R. Berkowicz, K. Brzozowski, R. Buccolieri, B. Carissimo, A. Costa, S. Di Sabatino, G. Efthimiou, J. Franke, I. Goricsan, A. Hellsten, M. Ketzel, B. Leitl, R. Nuterman, E. Polreich, J. Santiago and R. Tavares, 2008: The MUST model evaluation exercise: Patterns in model performance. *Proc. 12th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes,* Cavtat (Croatia), 6-9 October.

VDI guideline 3783 Part 9, 2005-11: Environmental meteorology – Prognostic micro-scale wind field models – Evaluation for flow around buildings and obstacles. Beuth Verlag, Berlin.

Yee, E., and C. Biltoft, 2004: Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles. *Boundary Layer Meteorology*, **111**, 363–415.