Towards a model evaluation protocol for urban scale flow and dispersion models

S. Di Sabatino*

Dipartimento di Scienza dei Materiali,
University of Salento,
Via Monteroni, 73100 Lecce, Italy
E-mail: silvana.disabatino@unisalento.it
*Corresponding author

H.R. Olesen and R. Berkowicz

National Environmental Research Institute,
Aarhus University, P.O. Box 358,
Frederiksborgvej 399, DK-4000 Roskilde, Denmark
E-mail: hro@dmu.dk
E-mail: rb@dmu.dk

J. Franke

Faculty of Mechanical Engineering,
Department of Fluid- and Thermodynamics,
University of Siegen,
Paul-Bonatz-Str. 9-11, 57076 Siegen, Germany
E-mail: franke@ift.mb.uni-siegen.de

M. Schatzmann, B. Leitl
and K. Heinke Schlünzen

Meteorological Institute (KlimaCampus),
University of Hamburg,
Bundesstrasse 55, D-20146 Hamburg, Germany
E-mail: schatzmann@zmaw.de
E-mail: bernd.leitl@zmaw.de
E-mail: heinke.schluenzen@zmaw.de

R. Britter

Visiting Scientist, DUSP,
MIT, Boston, MA 02139, USA
E-mail: rb11@eng.cam.ac.uk

A. Martilli

Environment Department,
Research Center for Energy,
Environment and Technology (CIEMAT),
Avenida Complutense 22, 28040 Madrid, Spain
E-mail: alberto.martilli@ciemat.es

B. Carissimo

CEREA, Teaching and Research Centre in Atmospheric Environment
(ENPC/EDF R&D), 6 Quai Watier,
78400 CHATOU Cedex, France
E-mail: bertrand.carissimo@edf.fr

Abstract: This paper reports on the stages forming a model evaluation protocol for urban flow and dispersion models proposed within the COST Action 732 on "Quality Assurance and Improvement of Micro-Scale Meteorological Models". It discusses the different components forming model evaluation with emphasis on validation and implementation of the protocol for the test case Mock Urban Setting Test (MUST). The protocol was proposed with building-resolving models in mind, but integral models have also been included. The suggested approach can be used for further micro-scale model evaluation and for the standardisation of their applications.

Keywords: COST action 732; model evaluation protocol; MUST experiment; CFD; non-CFD models.

Reference to this paper should be made as follows: Di Sabatino, S., Olesen, H.R., Berkowicz, R., Franke, J., Schatzmann, M., Leitl, B., Schlünzen, K.H., Britter, R., Martilli, A. and Carissimo, B. (2011) 'Towards a model evaluation protocol for urban scale flow and dispersion models', *Int. J. Environment and Pollution*, Vol. 47, Nos. 1/2/3/4, pp.326–336.

Biographical notes: Silvana Di Sabatino received her PhD (Engineering) from the University of Cambridge (UK). She holds an MPhil from the same University and an MSc in Physics from Bologna University. Currently, she is a Lecturer in Atmospheric Physics at the Dipartimento di Scienza dei Materiali, University of Salento. Her research interests are in the field of atmospheric dispersion over complex geometry, urban boundary layer modelling and data analysis. She has published several research papers at international journals, conference proceedings as well as book chapters.

Helge Rørdam Olesen is a Senior Adviser at the Department of Atmospheric Environment of the Danish National Environmental Research Institute, Aarhus University. He has been working with development and implementation of atmospheric short-range models for more than two decades. His main research interest is model evaluation. An important aspect of his work has been efforts on an international level to establish common tools within dispersion modelling. He is chairman of the European initiative on *Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, which is responsible for the series of international harmonisation conferences.

Ruwim Berkowicz, PhD, is a Retired Senior Scientist from the National Environmental Research Institute (NERI), Denmark. His main research interest was development of atmospheric dispersion models with special emphasis on models for practical applications.

Jörg Franke received his PhD in Mechanical Engineering from the University of Karlsruhe. He is a Senior Lecturer at the Department of Fluid- and Thermodynamics, University of Siegen. His research interests include the application of computational fluid dynamics in the fields of building aerodynamics and pollution dispersion with special emphasis on quality assurance of simulation results. He is the author of several research studies published in national and international journals, conference proceedings as well as a book chapter.

Michael Schatzmann is a retired Professor from the Meteorological Institute of the University of Hamburg. He has been in charge of the Technical Meteorology Division, which develops meso- and micro-scale meteorological models and operates a boundary layer wind tunnel laboratory. He chaired action COST 732.

Bernd Leitl is a Professor for Technical Meteorology and the head of the Environmental Wind Tunnel Laboratory at Hamburg University. He received his PhD in Mechanical Engineering from the University of Technology Dresden and gained expertise in experimental fluid mechanics and physical modelling of environmental flow and dispersion phenomena. His research is primarily focused on turbulent flow and dispersion within the lower atmospheric boundary layer.

K. Heinke Schlünzen is a Professor at the University of Hamburg and head of the meso-scale and micro-scale modelling group. Her research interests are in the field of meso- and micro-scale meteorological processes and phenomena with particular emphasis on model evaluation and model uncertainty. She investigates scale interaction, meso- and micro-scale dynamics and chemistry, and the interaction of atmosphere, ocean, biosphere and chemisphere. She has published more than 50 peer-reviewed papers and numerous technical reports and conference papers.

Rex Britter is a Visiting Scientist at the Department of Urban Studies and Planning, MIT, Boston. He has been Professor of Environmental Fluid Dynamics in the Department of Engineering at the University of Cambridge, UK. His research interests include fundamental studies into turbulent fluid dynamics, particularly those involving buoyancy. This is paralleled with operational interests in the flow and dispersion of hazardous materials, conventional pollutant dispersion problems in complex geometries such as cities, formalised model evaluation procedures, urban air quality, sustainable energy use in cities and security issues.

Alberto Martilli graduated in Physics in 1995 (Università Statale of Milano, Italy) with a dissertation on numerical modelling of meso-scale circulations over complex terrains. In 1996, he moved to the Swiss Federal Institute of Technology (EPFL, Lausanne), where, in 2001, obtained his PhD with a dissertation on the development of a new urban surface parameterisation for meso-scale models. From 2001 to 2004, he worked as a Post-doctoral fellow at the University of British Columbia (Vancouver, Canada), on meso-scale air pollution modelling. Since 2005, he is a Researcher at CIEMAT (Spain), where he works on micro- and meso-scale atmospheric modelling over urban areas.

Bertrand Carissimo received his PhD in Atmospheric Science from Princeton University. He is a Senior Researcher at the Teaching and Research Centre in Atmospheric Environment (CEREA) a joint laboratory of the Ecole Nationale des Ponts et Chaussées (ENPC) and Electricité de France (EDF). His research interest is applied meteorology at local scale, including air pollution in the urban and industrial environment. He is the author of many publications in international journals and conference proceedings.

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 and 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 standardised modelling practice for atmospheric applications.

The role of the COST Action 732 (2005–2009) has been to overcome this lack of information by also considering outcome from the previous European project on heavy gas model evaluation called SMEDIS (Carissimo et al., 2001). 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 standardisation of modelling practice 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 or the dispersion of pollutants within and near the urban canopy or industrial landscape.

The impact of the COST Action 732 depends on 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 online 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 and 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 (i.e., 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 have been updated with recommendations for data sets used in the validation work. The recommendations given in the documents listed earlier have been 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, participants of the Action are continuing to analyse the results for the MUST case (Yee and Biltoft, 2004), which comprises both field and wind tunnel data 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 Franke et al. (2008) and Olesen et al. (2008).

The next phase was the launch of an exercise based on field experiments in a real city. This exercise was devoted to the simulation of dispersion experiments over Oklahoma City, USA. Several groups were involved in preparing simulation grids; simulation results were provided by about five teams. Initially, the simulations were done in the form of a blind test, i.e., only the geometry and the input data was given to the participants. Results from this are not included in this paper.

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 and CALPUFF, have received increasing attention within COST 732 and results for them are available at the Cost Action website.

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 need 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. The fourth point is indeed quantitative and is used by employing several statistical measures, such as the Fractional Bias (FB), the Normalised 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-mentioned performance measures and estimates of model acceptance criteria have been summarised by Chang and Hanna (2004) based on extensive experience with evaluating many models with many field data sets. In terms of metrics, their criteria can be formulated as follows for good performing models: $-0.3 < FB < 0.3$; $NMSE < 4$; $FAC2 > 0.5$; $MG < 1$; $VG < 1.6$. These criteria can be used as guidance, although it should be noted that the values of Chang and Hanna refer to comparisons of concentrations from research-grade experiments in terms of maximum concentrations on arcs, i.e., unpaired in space. If used in other contexts (e.g., data paired in space and time, or for other parameters than concentration) less strict criteria might be applied. Performance can also be evaluated in terms of 'hit rate', as defined in VDI guideline 3783 Part 9, (2005–2011). In this guideline, a fractional deviation $D = 0.25$ is allowed in conjunction with 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 ha been 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 has followed, contains 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 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-mentioned purposes.

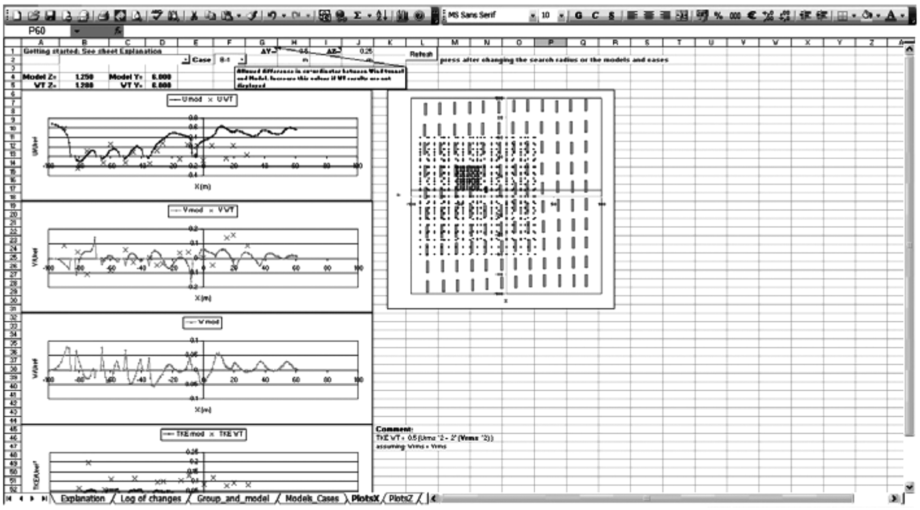
3 Evaluation exercise

3.1 *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/en/air/models/background/exceltools/). 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 contains essential information extracted from the above-mentioned Excel files with summary plots and metrics. As an example, one of the sheets with detailed information (PlotsX) is shown in Figure 1.

Figure 1 Example of the Excel spreadsheet with a macro tool for easy graphical comparisons



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 along cross-sections at, respectively, the beginning, middle and end of the building array are plotted for the height $z = 0.5H$ (where H is the height of the building). The qualitative behaviour of the pollutant plume seems to be acceptable.

Results from this exercise has helped to formulate a Best Practice Guideline specifically for using the MUST data and to revise the Model Evaluation Guidance and Protocol Documents.

Figure 2 –45° case, samples of horizontal profiles of dimensionless concentrations from one CFD model at $z = 0.5H$ at the beginning, middle and end of the array (first/fourth, second/fifth and third/sixth picture, respectively). Please note that the building array is rotated so the *x*-axis points downwards and the source is on top

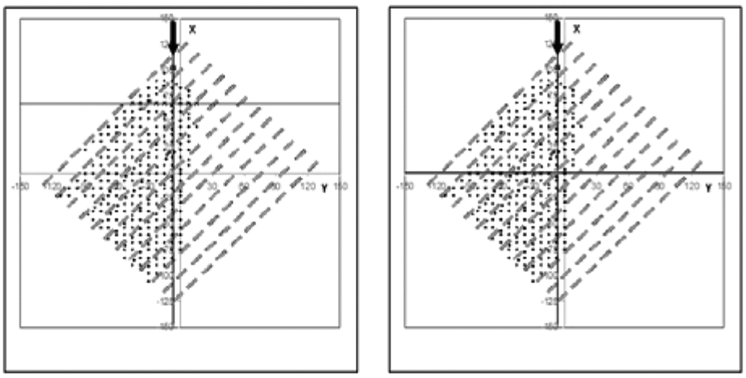
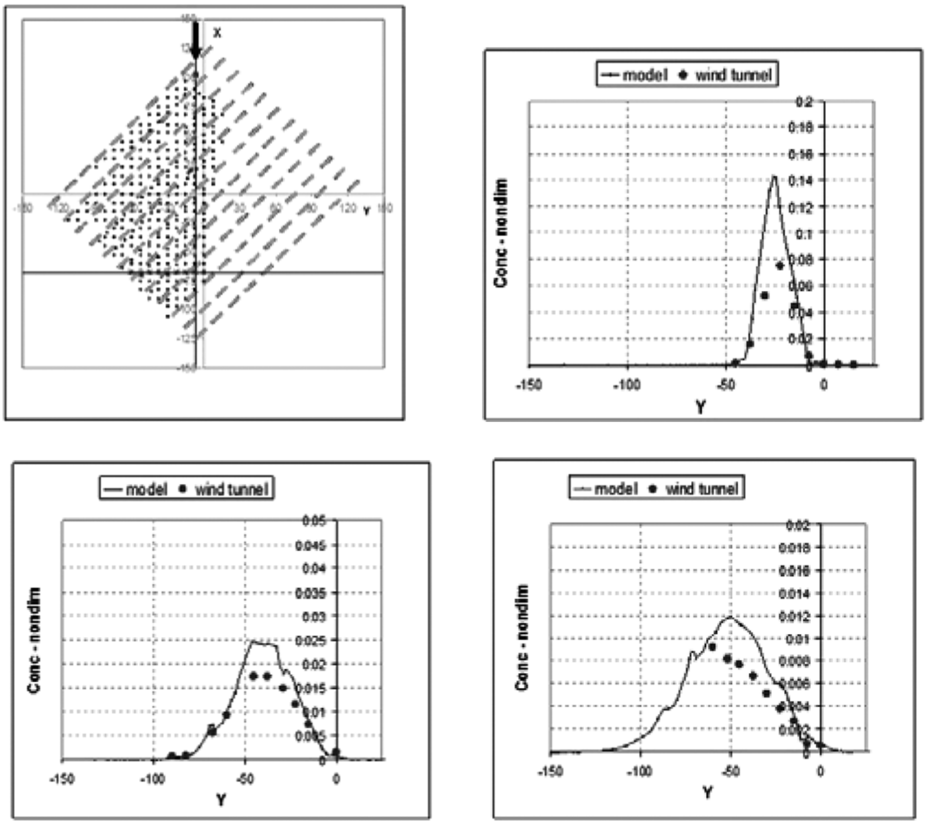


Figure 2 –45° case, samples of horizontal profiles of dimensionless concentrations from one CFD model at $z = 0.5H$ at the beginning, middle and end of the array (first/fourth, second/fifth and third/sixth picture, respectively). Please note that the building array is rotated so the x-axis points downwards and the source is on top (continued)



4 Conclusions and work in progress

The comparison carried out in the COST 732 action shows that flow and concentration model results compare relatively well with the measurements, although they share some common problems in predicting certain features of the flow. For example, all models have difficulty in predicting the vertical component of wind velocity, whereas prediction for the x-velocity component is much more successful. 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 emphasise 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.

The Action has finalised the MUST exercise and has suggested the best approach for further model evaluation for the standardisation of CFD modelling practice for micro-scale meteorological applications. This includes a critical review and refinement of the numerical results. 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 Oklahoma City exercise has finally helped us to further strengthen our previous results and to finalise 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 data sets and the recommendations specified in the Best Practice Guideline should lead to a harmonised 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.

References

Bezpalcova, K. (2007) *Physical Modelling of Flow and Diffusion in Urban Canopy*, PhD thesis, Faculty of Mathematics and Physics, Charles University, Prague.

Britter, R. and Schatzmann, M. (Eds.) (2007a) Background and Justification Document to Support the Model Evaluation Guidance and Protocol Document, COST Office Brussels, Belgium, pp.1–88.

Britter, R. and Schatzmann, M. (Eds.) (2007b) *Model Evaluation Guidance and Protocol Document*, COST Office Brussels, Belgium, pp.1–28.

Carissimo, B., Jagger, S.F., Daish, N.C., Halford, A., Selmer-Olsen, S., Riikonen, K., Perroux, J.M., Wuertz, J., Bartzis, J.G., Duijm, N.J., Ham, K., Schatzmann, M. and Hall, R.

(2001) 'The SMEDIS Database and Validation Exercise', *Int. Journal of Environment and Pollution*, Vol. 16, pp.614–629.

Chang, J. and Hanna, S. (2004) 'Air quality model performance evaluation', *Meteorology and Atmospheric Physics*, Vol. 87, pp.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., Bartzis, J., Barmpas, F., Berkowicz, R., Brzozowski, K., Buccolieri, R., Carissimo, B., Costa, A., Di Sabatino, S., Efthimiou, G., Goricsan, I., Hellsten, A., Ketzel, M., Leitl, B., Nuterman, R., Olesen, H., Polreich, E., Santiago J. and Tavares, R. (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, www.harmo.org/Alias/Cavtat_proceedings

Franke, J., Hellsten, A., Schlünzen, K.H. and Carissimo, B. (Eds.) (2007) *Best Practice Guideline for the CFD Simulation of Flows in the Urban Environment*, COST Office Brussels, Belgium, pp.1–52.

Olesen, H.R., Baklanov, A., Bartzis, J., Barmpas, F., Berkowicz, R., Brzozowski, K., Buccolieri, R., Carissimo, B., Costa, A., Di Sabatino, S., Efthimiou, G., Franke, J., Goricsan, I., Hellsten, A., Ketzel, M., Leitl, B., Nuterman, R., Polreich, E., Santiago, J. and Tavares, R. (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, www.harmo.org/Alias/Cavtat_proceedings

VDI guideline 3783 Part 9 (2005–2011) *Environmental Meteorology – Prognostic Micro-scale Wind Field Models – Evaluation for Flow around Buildings and Obstacles*, Beuth Verlag, Berlin.

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