BACKGROUND AND JUSTIFICATION DOCUMENT TO SUPPORT THE MODEL EVALUATION GUIDANCE AND PROTOCOL

Edited by:

Rex Britter and Michael Schatzmann

COST Action 732

QUALITY ASSURANCE AND IMPROVEMENT OF MICROSCALE METEOROLOGICAL MODELS

1 May 2007

Contact:
Carine Petit
Science Officer
COST Office
Avenue Louise 149
1050 Brussels
Belgium
Tel: + 32 2 533 38 31
Fax: + 32 2 533 38 90
E-mail: cpetit@cost.esf.org

Edited by Rex Britter and Michael Schatzmann

Contributing authors: Alexander Baklanov, Photios Barmpas, John Bartzis, Ekaterina Batchvarova, Kathrin Baumann-Stanzer, Ruwim Berkowicz, Carlos Borrego, Rex Britter, Krzysztof Brzozowski, Jerzy Burzynski, Ana Margarida Costa, Bertrand Carissimo, Reneta Dimitrova, Jörg Franke, David Grawe, Istvan Goricsan, Antti Hellsten, Zbynek Janour, Ari Karppinen, Matthias Ketzel, Jana Krajcovicova, Bernd Leitl, Alberto Martilli, Nicolas Moussiopoulos, Marina Neophytou, Helge Olesen, Chrystalla Papachristodoulou, Matheos Papadakis, Martin Piringer, Silvana di Sabatino, Mats Sandberg, Michael Schatzmann, Heinke Schlünzen, Silvia Trini-Castelli.

ISBN: 3-00-018312-4

Distributed by

University of Hamburg
Meteorological Institute
Centre for Marine and Atmospheric Sciences
Bundesstraße 55
D – 20146 Hamburg, Germany

Contents

1 Background

The Framework Directive on Air Quality Assessment and Management together with its ‘daughter’ directives is a key element of present-day European environmental legislation. It addresses air quality within conurbations and near to major sources. It is the first time that a European directive requires the use of models as tools for the execution of air quality policy. The directive only considers air quality modelling as a supplement to measurements for cases in which concentrations are likely to fall below air quality limit values. However there is now the possibility and likelihood for the extensive use of models also in related fields.

Mathematical models will increasingly be used to investigate the reasons for limit value exceedances, to decide on the proper siting of monitoring stations in urban areas and to support decisions within licensing procedures. These additional purposes will commonly require models that are more advanced than those usually simpler tools used for checking compliance with air quality limit values.

Though the directive does not require the harmonisation of models across Europe, the performance, representativeness and accuracy of results should be based on quality assured models that can be inter-compared across national borders in order to ensure sound, equitable and effective protection and/or mitigation measures.

For the short-range local problems (0 km to 5 km) of concern here, simple Gaussian type models have generally been used. These models are applicable for pollutant emissions into uniform atmospheric flows (for example tall stack releases in flat, unobstructed terrain and averaged over a large number of atmospheric conditions). It is accepted that these models are not appropriate for predicting flow, dispersion and the resulting concentrations in complex urban or industrial areas and this, unfortunately, is where pollutants that are of major concern at present (NO_2, $PM_{2,5}$, PM_{10}, O_3, VOC) are emitted and where the population are present. The pollutants are emitted from traffic and low elevation domestic or industrial sources. Additionally, hazardous and critical situations occur under specific flow conditions where the meteorological conditions and the effects of buildings make the flow far from uniform or stationary; the conditions for which these models have been developed.

Other types of emissions that deserve attention are toxic or flammable gases that escape during accidents or that are deliberately released in the context of terrorist attacks. It is common in all these cases that the dispersion takes place near to the ground within or slightly above the urban or industrial canopy layer where, again, the flow is non-uniform and non-stationary.

The emergence of increasingly powerful computers enabled the development of more powerful computational tools that have the potential to meet new demands for predictions from models. These new tools are termed microscale meteorological models and they are of a prognostic or a diagnostic type. Prognostic models are based on the Reynolds-averaged Navier-Stokes (RANS) equations, whereas diagnostic models are less sophisticated and only ensure the conservation of mass. These two model types are presently supplemented by simpler tools that are based on a statistical description of the urban area as distinct from the resolution of specific buildings and obstacles. It is to be expected however, that the latter will sooner or later be replaced by RANS based models or the even more complex unsteady RANS (URANS) and Large Eddy Simulation (LES) models. The timescale of this replacement will depend upon the requirements of model applications and constraints. RANS, URANS and

LES models belong to the family of Computational Fluid Dynamics (CFD) tools; tools commonly used in various engineering contexts.

Microscale meteorological models in the context of COST 732 are those used specifically for urban areas; distinct from other uses of the term "microscale meteorological models" such as in the area of cloud modelling. For urban applications the models are adjusted to domain sizes of the order of several tens of metres to a few kilometres (street canyons, city quarters). In this sense they are at the intersection of the engineers and the meteorologists perspectives of cities. The models use boundary and initial conditions that are based on descriptions of surface characteristics like land use and parameterized in terms of a surface roughness lengths and displacement thicknesses. They may contain modules that have the potential to simulate chemical transformations, aerosol formation or other important atmospheric physico-chemical processes. Typically these models contain a substantial amount of empirical knowledge, not only in the turbulent closure schemes but also in the use of wall functions and in other parameterizations.

Models have begun to play an important and often dominant role in environmental assessment and urban climate studies that are undertaken to investigate and to quantify the effects of human activity on air quality and the local climate.

The increasing use of microscale meteorological models is paralleled by a growing awareness that the majority of these models have never been the subject of rigorous evaluation. Consequently, to a certain degree, there is a lack of confidence in the modelled results.

To cast doubt on the results is perfectly justified, as was shown by systematic studies in which applications of the same model by different modellers to a given problem (Hall et al., 1997) and applications of different models by either the same or different modellers to the same problem (Ketzel et al., 2001) revealed significant differences. Nevertheless, these models are used in the preparation of decisions with profound economic and political consequences.

It is timely that this Cost Action 732 was initiated in 2005 because of the wide range of associated collaborative activities which recently were or still are under way within the European research community. Some of these associated activities include:

- The Clean Air for Europe project (CAFÉ) under the 6th Environment Action Programme that strives to develop a thematic strategy on air pollution the main elements of which are: (i) identify gaps and priorities for further action (e.g. particulate matter, smog, NO_x) taking account of risks to vulnerable groups; (ii) review and, if necessary, update existing air quality standards and national emission ceilings (with attention to vulnerable groups); and (iii) better systems of gathering information, modelling and forecasting. Additionally, it looks at the implications of international policy such as the national emissions ceiling directives and the Gothenburg Protocol on ambient air quality.

- The City-Delta project organised by the Joint Research Centre of the EC which focuses on urban background concentrations in several European cities.

- The FP6 Network of Excellence ACCENT on atmospheric composition change that includes a group on transport and transformation of air pollutants on a range of scales including the local scale.

- The ENV-e-City project which aims to improve access to environmental data; meteorology for air pollution assessments is a pilot application area.
- The EU-FUMAPEX project (Integrated systems for Forecasting Urban Meteorology, Air Pollution and Population Exposure) that was about to improve meteorological forecasts for urban pollution, coupling weather prediction models to urban air pollution and exposure models in cities subject to various European climates.
- National funded projects as, e.g., DAPPLE (UK), BUBBLE (Switzerland) and VALIUM (Germany) that have carried out intensive measurements within or closely above city quarters in London, Basel and Hanover, respectively.
- The COST Action 728 (Enhancing Mesoscale Meteorological Modelling Capabilities for Air Pollution and Dispersion Applications) which will consider the scale range immediately above the microscale, but can provide boundary conditions for the models treated under this COST Action.

In addition to COST Action 728 there are three recently completed COST Actions, namely COST 615: Science and research for better air in European cities (CITAIR), COST 710: Harmonization in the pre-processing of meteorological data for dispersion models and COST 715: Meteorology applied to Urban Air Pollution Problems, which are relevant.

However, none of these activities addressed the focus of COST Action 732: The Quality Assurance of Microscale Meteorological Models in a standardised European-wide accepted form and the subsequent improvement of such models. The aim of COST Action 732 is to involve and to support, but not to duplicate these other activities. Additionally, COST provides a suitable forum for the proposed activities as these have been performed so far at national level in a non-concerted and scattered way.

There is substantial experience and background in the formal evaluation of model quality for non-CFD codes. These include application to general atmospheric dispersion problems using the Model Validation Toolkit (Olesen, 1995; Olesen, 2005), mesoscale transport and dispersion models (Chang et. al., 2003), dense gas dispersion models (Hanna et al. 1993), the EU SMEDIS project (Carissimo et al., 2001) and more. This is particularly so in the dispersion and air quality areas and this has been driven by the regulatory context in Europe, the US and elsewhere. There is less evidence in Europe of formal evaluation of meteorological (non-CFD) models though this may be due to the small number of such models (linearised models, porosity models). However the surface parameterizations incorporated into CFD models can be interpreted as microscale non-CFD models. It would be useful to evaluate these parameterizations as well.

There is only limited evidence of formal model evaluation for “microscale” models used near and within the urban canopy. COST 615, COST 715 and the SATURN project (Moussiopoulos, 2003, Borrego et. al. 2003) initiated work in this area. The DAPPLE project in the UK (Arnold et al., 2004) and VALIUM project in Germany (Schatzmann et al., 2006) are recent studies directed towards model evaluation. This said we are aware of other related initiatives in various European countries and a major goal of COST 732 is to raise a wider awareness of these and to build a coherent, comprehensive and accessible body of knowledge.

The principal context for the use of "microscale" meteorological models of interest to COST 732 is dispersion modelling. As a consequence, COST 732 also addresses the evaluation and improvement of dispersion models for use at the microscale but will not directly consider any chemistry components of the dispersion models.

A COST 732 Workshop, held in Hamburg on July 28/29 2005 (Schatzmann and Britter, 2005), was structured into six parts:

- Model Evaluation Strategies in General
- Initiatives on Model Evaluation
- Evaluation Requirements for Specific Model Types
- Relevant Specific Case Studies
- General Data Requirements, and
- Availability and Experiences with Data Sets/Data Bases

After two days of discussions the Workshop participants agreed on the following points:

1) The existing flow and dispersion models for urban applications have indeed not yet been subject of a rigorous and structured evaluation.
2) The reason that most of the models lack quality assurance is not due to insufficient efforts spent by the model developers, it is mainly caused by
    - the lack of a generally accepted quality assurance procedure for such models, and
    - the lack of data sets that are quality checked and generally accepted as a standard for model validation purposes.
3) The participants recommended
    - to develop coherent and structured quality assurance procedures for all types of microscale meteorological models which give clear guidance to developers and users of such models as to how to properly assure their quality and their proper application,
    - to provide a systematically compiled set of appropriate and sufficiently detailed data for model validation work in a convenient and generally accessible form (www data bank),
    - to invite scientists and users from all participating states to apply the procedure and to prove its serviceability,
    - to build a consensus within the community of microscale 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 microscale meteorological models participating in this activity,
    - to contribute to the proper use of models by disseminating information on the range of applicability, the potential and the limitations of such models,
    - to identify the current weaknesses of the models and data bases,

- to give recommendations for focussed experimental programmes in order to improve the data base and
- to give recommendations for improvements of present models and, if necessary, for new model parameterisations or even new model developments.

It was felt that the evaluation procedures need to be tailored to the specific needs of individual groups of models (non-CFD, CFD including LES-codes). The more sophisticated a type of model is the more detailed and elaborate the procedure needs to be.

4) Particular importance should be given to the proper use of models. The participants recommended establishing in addition to the guidance for model developers a standard of ‘good practices’ for model users (model type specific as well). It would be ideal to establish a European trainee project within which young researchers and experienced model users could work together and transfer knowledge.

5) The participants recommended defining particular scenarios in combination with Design or Assessment Parameters (DOAPs) to evaluate the ‘fitness for purpose’ of individual models. The complexity of data must correspond to the complexity of the scenarios.

6) It was recommended that notice be taken of existing guidelines e.g.

- Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (American Institute of Aerodynamics and Astronautics),
- VDI-Guideline on Prognostic Meteorological Models (Verein Deutscher Ingenieure),
- Q-Net Best Practice Guidelines in the Industrial Application of CFD.

Furthermore, it was regretted that the COST 732 initiative is presently restricted to COST member states only. Other parts of the international community should be invited to join the initiative.

7) Some participants strongly recommended organizing a model contest which includes test cases for which the calculations are made prior to the experiments (‘blind testing’ of models).

8) With respect to the data it was made clear that validation data sets are not just data but must fulfil additional requirements (completeness, representativeness, known uncertainty ranges, etc.). Due to the unsteady nature of the weather, field data collected inside the urban canopy layer have the character of random samples only. Wind tunnel data, on the other hand, represent idealized situations. It was recommended to use combined data sets that are based on both field and laboratory experiments.

9) It was recognized that the complexity of models and data sets must correspond to each other. Whereas the validation of non-CFD-codes can be based on mean values alone, LES-models will need time series of flow and transport properties.

10) It was felt that a glossary of terms needs to be formulated since words like ‘validation’, ‘verification’, ‘evaluation’, ‘quality assurance’ etc. are not unambiguously defined and used.

These agreed points were used to guide the development of this Model Evaluation Report with the following Report Structure. In the second chapter four “drivers” for quality assurance are identified. In chapters 3 and 4 an outline of the general evaluation philosophy is given together with some recent examples. Chapter 5 discusses the elements of the model evaluation procedure, while Chapter 6 considers the sources of experimental data for model evaluation purposes, the assessment of the quality of data and its archiving and accessibility. The report closes with conclusions and recommendations (Chapter 7), a list of references (Chapter 8) and a glossary of terms (Appendix).

2 Model Evaluation Drivers

2.1 Model Evaluation Drivers: National Participants

The COST 732 Action is targeted towards developing a coherent, structured and widely-accepted quality assurance procedure throughout Europe for microscale meteorological models and their application to pollutant dispersion concerns in cities. At the inaugural COST 732 Workshop, a questionnaire was circulated to the national consortium members, prior to the meeting, addressing three questions:

- What are the typical areas of application of microscale meteorological models?
- Are you aware of any particular activities, past, present or planned to address the evaluation of the quality of microscale meteorological models?
- What actions arising from COST 732 would be most helpful to you in your national activities?

Responses were received from the national representatives of the 14 countries that were participating in COST 732 at the time. The detailed responses are contained within the Proceedings of the COST 732 Workshop (Schatzmann and Britter, 2005).

Some overall remarks and summary are provided below addressing each question.

- As to typical application areas, all consortium members include air quality simulations ranging from street scale to small city areas; many (7 out of 14) include wind-related applications (pedestrian comfort, wind loading problems/building engineering, wind-driven rain), some (2 out of 14) include indoor/outdoor air pollution, and 1 out of 14 include specific cases such as car parks with emissions from cold engine start.
- Regarding model evaluation activities, some consortium members (3 out of 14) report that there has been no model validation performed at all; many (6 out of 14) report that they do not have validated models against any field datasets, some (3 out of 14) report that they are using alternative validation procedures (e.g. from the CFD community within the ERCOFTAC framework), and some others (6 out of 14) are going through or have just started a validation procedure using field data sets.
- Finally, COST 732 is expected to develop a rigorous, detailed, widely-accepted model evaluation methodology, and the provision of field datasets, appropriate for the validation of models (all consortium members), while some consortium members (4 out of 14) would also expect the generation of field

datasets through funded joint field campaigns and improvement of currently available models.

2.2 Model Evaluation Drivers: Industry and Local Authorities

The application of numerical models to air pollution problems, such as the environmental impact of planned developments or the effectiveness of possible mitigation strategies on both long and short time scales is widespread in industry and local authorities as part of the regulatory process. The European Directives require accurate predictions of concentrations for various pollutants to allow for comparison with air quality objectives. In addition, air quality strategies necessitate the development of plans to reduce pollution in specific areas. This requires models that can calculate pollutant concentrations for different mitigation scenarios. Thus models must be able to be run with various, typically very large emission inventories i.e. large databases containing emission information for the many sources in an urban area. Different scenarios and corresponding emission inventories could be those resulting from a traffic management policy which includes low emissions zones, clear zones or multiple occupancy vehicles and emission reduction strategies for other sources.

On a shorter time scale models are also used to predict the impact of a change in atmospheric conditions on pollutant concentration. For example, when meteorological conditions are expected that might lead to high pollutant concentrations, local authorities need to predict the expected maximum concentrations and take timely action to prevent them from exceeding the regulatory limits.

In a similar manner, but in response to an accidental release of a hazardous material, fast response models are required by emergency responders to predict the extent of the hazardous area and the duration of the hazard

The use of “operational models” is often preferred; a class of models that can be applied routinely to assess air quality and/or to enable emergency services to make decisions. Operational models can be of different types but they are commonly required to run quickly in order to be able to investigate very many different possible scenarios in a reasonable time or a smaller number of scenarios but in a relatively short time.

Many operational models for urban air quality are currently available ranging from simple empirical models to advanced computational fluid dynamics (CFD) models. Non-CFD models are based on the description of the flow and dispersion processes at various spatial scales (such as the street scale and the city scale). CFD models could allow the study of dispersion at the neighbourhood scale, simply by representing the geometry of every individual building in the neighbourhood. This approach has not been exploited in a systematic way, partly because in the past CFD models have been regarded as too computationally intensive to be able to run with such complicated geometry. However, with improved computer technology this is no longer the case. Within Europe the use of CFD codes for decision making is increasing both by industry and by national and local government agencies.

It is important to observe that the models are used to guide various societal, business and political decisions. If the models are not of the appropriate quality (fitness-for-purpose) then it is likely that the consequent decisions will be inappropriate. If the models produce results that are too optimistic then bad decisions will likely be reached. If the models are too pessimistic then bad decisions will again be reached, possibly to the missing of opportunities, but definitely reducing business profitability.

Developers of numerical models, analysts who use numerical models and decision makers who rely on the results of the analyses face the critical question: how should confidence in modelling be objectively assessed?

Evaluation (typically encompassing scientific assessment, verification and validation) of computational models are the main methods for quantifying this confidence. This activity is best performed in a structured manner to a clear protocol. Thus models need to be formally evaluated prior to their use but often this is not undertaken in a structured way by industry or by local authorities because of the need to provide solutions quickly, with constrained resources. And it is exceedingly inefficient for every user of a model to undertake an extensive and expensive evaluation study each time they wish to use a new or modified model.

Thus industry and government agencies must either perform model evaluations in a communal way or require the model developers to perform this activity under guidance. In either case the evaluation process will have to satisfy some broad consensus of the informed community as to its correctness and suitability.

Another concern, somewhat, different from other model applications, is that of the legal aspect associated with the use of any model for critical decision making. Formalized model evaluation procedures may be of great importance and value here for both industry and Government.

Finally, in addition to ensuring that the quality of the model is adequate, it is equally important that the user and the user-model combination are of adequate quality. It is more difficult to set up an evaluation protocol to address this concern but a useful approach has been to develop guidelines for best-practice in using models.

In summary then; models should always be accompanied by a set of evaluation documents that allows the user to assess the fitness-for-purpose of the model for common applications and also documentation reporting on the best practice in the use of the models. In the absence of such documents the user, particularly those making decisions of consequence, should be cautious.

2.3 Model Evaluation Drivers: the European Commission Perspective

2.3.1 Introduction

Air quality models are powerful tools to describe the dispersion and fate of pollutant gases or aerosols upon their release into the atmosphere. Dispersion is primarily controlled by turbulence in the atmospheric boundary layer. Turbulence is random by nature and thus cannot be precisely described or predicted. As a result, there is spatial and temporal variability that naturally occurs in the observed meteorology and in the observed concentration field. Additionally, uncertainty in the model prediction can also be due to factors such as errors in the input data and in the model formulation. Because of the effects of uncertainty, it is not possible for an air quality model to provide “perfect” results, and there is always a base amount of scatter that cannot be removed (Chang and Hanna, 2004). Nevertheless, air quality models need to be properly evaluated before their predictions can be used with confidence, since model results often influence decisions that have large public health and economic consequences. This is clearly a key issue for national regulatory bodies and for those determining or selecting the intent and structure of the regulations. In a European context this body is the European Commission.

2.3.2 Definition of Uncertainty and Data Quality Objectives according to the Air Quality Framework Directive

The Air Quality Framework Directive, FWD, (official title: Council Directive 96/62/EC of 27 September 1996 on ambient air quality assessment and management) sets a general policy framework for dealing with ambient air quality. The prime goal defined by this Directive is protection of human health and the environment as a whole. For this purpose, a set of long-term objectives for air quality is established by the legislation.

Monitoring and modelling are identified as crucial components of air quality management. While monitoring of the atmospheric pollutants provides essential information about air quality in certain points, modelling tools provide an estimation of a spatial distribution of the pollutants over large areas. The accuracy of the monitoring data and modelling results is one of the essential issues of the FWD.

Uncertainty of modelling results is of critical concern. For this reason the Framework and Daughter Directives establish new requirements for air quality modelling, including the definition of the Modelling Quality Objectives as a measure of the acceptability of modelling results. In this context, the uncertainty for modelling and objective estimation is defined as the *maximum deviation of the measured and calculated concentration levels over the period considered, by the limit value, without taking into account the timing of the events*. It should be noted here that the measured value is assumed to represent the ‘truth’ (which is a courageous assumption, see chapter 6). The quality objectives defined for each quality indicator are listed in Table 2.1.

Model quality measures described in the EU Directives can be interpreted as the relative maximum error (RME) without timing, which is the largest concentration difference of all percentile (p) differences normalized by the respective measured value. Co_p and Cp_p are the concentration observed and predicted values at the percentile (p)

$$RME = \frac{\max(|Co_p - Cp_p|)}{Co_p} \tag{2.1}$$

The question of timing is relevant for those hourly and daily limits, or target values, which are defined as a number of allowed exceedances of a given threshold concentration. Besides that, the model quality objectives for the allowed uncertainty are given as a relative uncertainty, without clear guidance on how to calculate this relative uncertainty. It could be assumed that the respective measured value shall be used to normalize the absolute difference between the maximum deviation of the measured and calculated concentration levels. Another possibility would be to take the maximum relative deviation, but this approach could shift the emphasis to the very low measured concentration ranges, where usually the largest relative deviations between observations and calculations occur, which could be the main reason for non-compliance of annual mean values accuracy requirements. Besides that, other problems of the interpretation of the model accuracy requirements, according to the EU Directives could occur since there are no differences between a short-term and long-term model application accuracy analysis, being the first one in advantage due to the number of paired-in-time results.

In case of PM_{10}, Pb and Benzene pollutants, the averaging period for modelling uncertainty estimation is defined for 1 year. This fact can be seen as an important gap of the Framework Directive since it restricts the uncertainty estimation process of models with feasible temporal applications of several days, such as CFD models (Borrego *et al.*, 2005).

Pollutant	Quality Indicator	Quality Objective	Directive
SO_2, NO_2, NO_x	Hourly mean	50-60%	1999/30/EC
	Daily mean	50%	
	Annual mean	30%	
PM_{10}, Pb	Annual mean	50%	
CO	8-hour mean	50%	2000/69/EC
Benzene	Annual mean	50%	
Ozone	8-hour daily maximum	50%	2002/3/EC
	1-hour average	50%	

Table 2.1 Modelling Quality Objectives established by European Directives

Also, it is possible to question the established Quality Objective values of the FWD. Does a modelling uncertainty value of 50% indicate a good model performance or can it guarantee reliable modelling results for decision-makers?

An alternative model error measure was already proposed by Stern and Flemming (2004), defining the quality indicator as the concentration difference at the percentile corresponding to the allowed number of exceedances of the limit value normalized by the observation (= Relative Percentile Error RPE)

$$\mathrm{RPE} = \frac{\left|\mathrm{Co_p} - \mathrm{Cp_p}\right|}{\mathrm{Co_p}} \tag{2.2}$$

This measure can be more robust than the error defined in the EU Directive and also evaluates the model performance in the high concentration ranges, but without the sensitivity to outliers. Since the model accuracy is examined in the concentration range of the limit values, there is also a direct link to the EU Directives.

Another problem related to model evaluation in the EU Directives is the heterogeneity of the observed concentration fields and the importance of selecting the adequate and representative monitoring sites for model resolution, since it is presently hardly possible for a grid model to simulate all stations with the required accuracy. In spite of this the European Air Quality monitoring network (EUROAIRNET) considers that both spatial and temporal representativeness of monitoring stations should be addressed in uncertainty estimation procedures, in order to guarantee a more accurate comparison with air quality standards, the Daughter Directives say nothing about the monitoring stations representativeness and the selection of criteria for the number and type of stations to be used on model accuracy evaluation. Nevertheless, there is a need for pre-selecting the stations to be used for model evaluation and that should be relied on the sites classification or on the prior knowledge of the air quality regime of the measurement sites (based on daily mean and the daily variation of each pollutant). Besides the monitoring stations representativeness, there is absence of any guidance in the EU Directives about measurement inaccuracy and incomplete data coverage that should all be taken into account in the context of a model evaluation. Regarding the data coverage, the EU Directives require a minimum of 90 % data coverage of the hourly or daily values.

2.4 Model Evaluation Drivers: Scientific Driver - Uncertainty and its Quantification

2.4.1 Introduction

2.4.1.1 The uncertainty concept

In order to have confidence in the usage of model results it is necessary to estimate their uncertainty. The aim of uncertainty analysis is a quantification of the confidence that can be expected from the model results. Uncertainty analysis can be defined, following Morgan and Henrion (1990) as "the computation of the total uncertainty induced in the output by quantified uncertainty in the inputs and model, and the attributes of the relative importance of the input uncertainties in terms of their contributions." Total uncertainty is, then, the result of the model uncertainty (theory, numerics, etc.), variability (in the current context this refers to the role of turbulence or short scale spatial and temporal variations) and data uncertainty (measurement error, representativeness, etc.).

Thus, total model uncertainty can be defined by the sum of the model uncertainty, variability and uncertainty on input data, as showed in Figure 2.1.

Figure 2.1 Definition of Total Uncertainty

In the following sections we will further classify the uncertainties, define model sensitivities and list some methods on how to quantify uncertainties.

Saltelli et al. (2000) define "sensitivity" analysis as a step beyond "uncertainty" analysis as the apportioning of uncertainty according to sources (i.e. the factors). For simplicity, we will keep this uncertainty terminology in the following.

Uncertainties associated with model formulation may be due to erroneous or incomplete representation of the dynamics and chemistry of the atmosphere, incommensurability, numerical solution techniques, and the choice of modelling domain and grid structure.

Variability (sometime mentioned as the stochastic uncertainty) refers to stochastic atmospheric and anthropogenic processes. It contributes to uncertainties discussed previously, notable those associated with emissions estimation and representations of chemistry and meteorology. These uncertainties can be discussed in two aspects: the implications of using means to represent values that vary in space and time and the inability to treat inherent variability. The deterministic treatment of stochastic processes using nominal mean values is a source of uncertainty. Although it may be possible to represent stochastic processes using probabilistic methods, doing so does not eliminate the uncertainty inherent to variability, since, for example, extreme events are not represented in such cases (Borrego et *al.*, 2005).

Uncertainties in data are described in terms of model input (emissions, meteorology, chemistry and model resolution) and observational data used for model output comparison (measured concentrations, dosages etc.). If a data assimilation procedure is included in the modelling, this additional uncertainty needs also be taken into account.

The total model uncertainty, resulting as the sum of three distinct components: uncertainty due to errors or assumptions/approximations in the model physics, uncertainty due to stochastic processes in the atmosphere and uncertainty due to input data errors can be determined by comparison between observations and model predictions through the application of Quality Indicators. Quality Indicators reflect the ability of a model to simulate real world phenomena. It could be impossible to define a unique and universal quantitative Quality Indicator for model evaluation. Nevertheless, application of different indicators helps to understand model limitations and provides a support for model evaluation and the intercomparison of models (Borrego *et al.*, 2005).

There can be three components to the evaluation of air quality models: scientific, statistical and operational (Chang and Hanna, 2004). In a scientific evaluation, the model algorithms, physics, assumptions and codes are examined in detail for their appropriateness, accuracy, efficiency and sensitivity. This exercise usually requires in-depth knowledge of the model. For statistical evaluation, model predictions are examined to see how well they match observations. The operational evaluation component mainly considers issues related to the user-friendliness of the model (user's guide, user interface, etc).

2.4.2 Classification of model uncertainties

Following Oberkampf et al. 2002, uncertainties can be grouped in two classes:

Epistemic: Arising due to lack of knowledge of the model physics and chemistry. An increase of knowledge can lead to a reduction of the uncertainty of the model's predictions. Sources of this uncertainty include: (a) Limited understanding of a complex physical process. This is what Irwin (2000) defined as *theory uncertainty* (there can be more than one theory that adequately describes the data). In this class there are also the uncertainties induced by the lack of accuracy of the parameterizations used or by the set of assumptions made in order to simplify the equations. (b) Model input e.g. the limited or non-existent experimental data for a fixed (but unknown) physical parameter or insufficient knowledge of initial or boundary conditions. (c) Lack of accuracy of numerical algorithms (called *numerical uncertainty* by Irwin 2000).

Aleatory: inherent variation associated with the physical system or environment considered (e. g. variability induced by turbulence). Many models are considered as ensemble averaged models, while very often measurements, in particular from real scale experiments, are representative only of one 'realization'. In other words, there is a problem because every model is based on assumptions of spatial and temporal homogeneity which the model inputs as well as data for model evaluation are supposed to satisfy (called *representativeness uncertainty,* Irwin, 2000, Schatzmann et al., 2003).

Epistemic uncertainty can be reduced, in principle, by an increase of knowledge and modelling skill, while aleatory uncertainty cannot.

2.4.3 Quantification of model uncertainties

In order to quantify uncertainties three steps are proposed:

2.4.3.1 Characterization of the sources of uncertainty

For the specific case of the micrometeorological models, the following sources can be listed:

- *Theory uncertainty*. Uncertainty due to the lack of accuracy of some physical parameterizations. An example can be different types of k-ε closures for RANS codes. Information on this aspect of model evaluation will arise in a Scientific Evaluation
- *Numerical uncertainty*. Uncertainty due to numerical algorithms, grid structure and resolution, and the numerical convergence criteria adopted. Information on this aspect of model evaluation will arise in the Verification stage of a model evaluation.
- *Uncertainty due to input and output data*. This includes
    - Uncertainty in the initial values and boundary conditions for meteorological variables. To quantify these it is important to have the statistical measurement error and information about the variability in time and space of the meteorological values used as inputs.
    - Uncertainty in the urban morphology, terrain and local features data of the location investigated - this may be due to limited data or a question of representativeness.
    - Uncertainty in source or emission characteristics.
    - Uncertainty in the measurements used for validation. It is important to have a statistical estimate of the measurement error.
- *Uncertainty linked with the possible different 'meaning' of the measurements and model results*
    - For example, the use of time averages vs. ensemble averages vs. space averages. The unresolved natural variability may result in large deviations between individual observations and model results which are usually characterizing an ensemble average result. Therefore, a definition is required, under which conditions model results and measurements can be considered equal (Schatzmann and Leitl, 2002). What is the spatial or temporal scale of interest? For dispersion models, is there a background concentration that needs to be considered?
    - These uncertainties may be interconnected. Irwin (2000) points out that as the model formulation increases in complexity to explicitly treat more and more physical processes, the number of input variables is increased and thus the likelihood of degrading the model's performance due to the increase in uncertainty in the data representativeness.

2.4.3.2 Quantification of model uncertainty

To quantify uncertainty it is necessary to define one or more metrics (a ‘measure’ of the differences between modelled and observed values, or among the predictions from several model results, in the case of a model intercomparison study). The metric should follow from a clear definition of the objectives or application of the simulations and identification of the phenomenon investigated. This is a crucial point of the process of Validation and must be carried out in close connection with that process.

The properties needed for such metrics for meteorological variables are:

(a) The metric must account for the model and experimental uncertainties.

(b) The metric must be devised such that it addresses the relevant model outputs for the intended application of the model. It must address *fitness-for-purpose* and cannot be *universal*. For example, the metric needed for an investigation of peak concentrations may be different from the metric needed for an investigation of averaged values. Ideally the metric should involve the most important variables required for the application.

In the previous chapter the Data Quality Objectives of the Air Quality Framework Directive of the European Community for dispersion models were considered. There the objective estimation is defined as the *maximum deviation of the measured and calculated concentration levels (over the period considered for the limit value), without taking into account the timing of the events*. This definition as well as the statistical formulae mentioned in the previous chapter is considered the best for health impact studies. They are *fitness-for-purpose*, not *universals*. However, in the EU directives nothing is specified for meteorological variables.

2.4.3.3 Sensitivity studies

As a refinement, the Sensitivity Study provides additional information relevant to uncertainty estimation. A set of model simulations can be performed in order to asses the sensitivity of the model output to the uncertainty of each model input. For any set of simulations different aspects of the model (physics, numerics, inputs etc.) can be tested for their influence on model outputs. The “set” of simulations can be undertaken in several ways e.g.

(a) Using Monte Carlo or Latin Hypercube Sampling to individuate the variability.

(b) Using Direct Decoupled Methods (see Isukapalli,1999). This type of methodology has lower computational cost than the techniques in a).

(c) By developing mathematical formulations to ‘transport’ the input uncertainty through the simulation (this is equivalent to the propagation of errors, e.g. Beychock, 2005, and also Peltier et al., 2006).

Sensitivity studies can assist in the definition of the criteria for the validation. If the variability among the model results is greater than the validation criteria chosen, work must be done in order to reduce the uncertainties.

Sensitivity studies can also help in the definition of the metric for model output. For example, let us imagine that the final aim of the model simulation is the estimation of pollutant dispersion. The following procedure can then be used:

(a) First, a set of simulations (of meteorology plus dispersion) can be performed in the framework of the sensitivity study.

(b) Secondly, the different simulations results can be clustered based on the pollutant concentration using the metric defined, for example in 2.4.1

(c) Then, other metrics involving measurable meteorological variables can be used to form different clusters of simulation results.

(d) The metric, based on meteorological variables, that gives the cluster that is closest to the cluster obtained with the metric based on pollutant concentration, is the most appropriate metric for the application (at least the one in which we are most confident)

3 Model Evaluation Philosophy

The rapidly growing need for mathematical models of the appropriate quality (fitness-for-purpose) has been paralleled by the considerable experience that has been gained over recent decades in model development and model use. This has led to the realisation from the scientific and the wider stakeholder communities that models of any type will have to undergo thorough evaluation in order to ensure that they are of appropriate quality for a specific purpose. At the same time the range of model types being used has expanded considerably. To the more conventional and widely used empirical and semi-empirical models, CFD codes based on the full set of conservation equations have been added. These include Reynolds-Averaged Navier-Stokes (RANS) models, unsteady RANS models and Large Eddy Simulation (LES) models. Direct Numerical Simulation (DNS) models are sometimes added as well. Each of those comes in a seemingly infinite variety of sub-species. Furthermore there are other dimensions of type to be considered including various mesh gridding structures, numerical solution schemes and others. As a result, from the perspective of a potential user, it is significant that all these models undergo the same, or similar, evaluation procedures, even if only in a generic sense.

Models of whatever type are only of use if their quality (fitness-for-purpose) has been quantified, documented, and communicated to potential users. Any model should be evaluated on the basis of the range of problems that it has been designed to address. This is because different problems may lead to different technical approaches and may therefore require the use of models which have been based on different numerical techniques employing different mathematical modelling of the physical processes involved. At present the choice of the type of model required is often still problem specific.

When evaluating model quality (fitness-for-purpose), it is essential to consider and specify what the purpose of the model is. Experience has shown that a clear and formal protocol is required for calculating model predictions and for comparing these with experiments. The protocol should reflect a consensus among the various interested parties.

Model evaluation is the sum of processes that need to be followed for mathematical models of any type, in order to determine and quantify their performance capabilities, weaknesses and advantages in relation to the range of applications that they have been designed for; that is, their purpose. This is consistent with the understanding that "quality" refers to "fitness-for-purpose", a concept quite different to measures such as "accuracy".

Designing a model evaluation procedure which is more or less common across, and widely accepted by, the scientific, industrial and government communities will provide the potential user with a powerful tool to assist in choosing the most appropriate model consistent with the problem requirements. This would include ensuring the presence of relevant processes in the model, the calculation of specific physical quantities (predicted flow and pollutant concentration fields, etc) and the spatial and temporal domains required that need to be simulated. In essence, model evaluation must focus on the adequacy of the numerical results generated by the model for the intended uses of the model (Oberkampf et. al., 2002).

Depending on the nature of the problem the requirements can differ significantly. For example:

- For regulatory purposes, such as the assessment of the impact of a new mitigation strategy on urban air quality or on human exposure, the principal requirements are that the models are able to simulate extended periods of time (in the order of years) and can be run very many times (possibly several thousands for the many possible scenarios) and over a short time period (of the order of hours to days). The expected results (usually pollutant concentrations) would need to be "accurate" in the sense that the predictions would (statistically) lie within a range defined by the appropriate regulatory authorities (see Table 2.1). Similar use will be found in the "safety cases" required by regulatory authorities. Safety cases involve the investigation of many and various hypothetical scenarios for safety planning issues, such as, the accidental release of toxic materials, fire safety and the dispersion of toxic combustion products, and many others. These scenarios may have severe impacts on humans and on the environment generally.
- For post accident investigations, possibly involving litigation or Government inquiries, accurate, robust and easily interpretable predictions would be required. In such cases, complex models which can generate time dependent solutions and have the ability to take into account the combined complex effects of all the physical and chemical processes involved would need to be employed, with little concern for the required computing power and resources.
- For emergency response such as the accidental release of toxic gases near a populated area, real-time predictions or access to pre-calculated real-time output may be required. In such cases, models that can generate quality assured results (or results with known uncertainty) of the development of the concentration field over a specified period of time would be required. In cases like that it is essential to produce timely results of known accuracy that are presentable in a form that aids the decision maker working in real time. There may be little concern about the computational cost, or resource requirements.
- A further, often not considered, purpose of models is where more sophisticated models are used to provide results that assist in the parameterisation of various processes in simpler models. An extension of this activity is to locate scenarios where specific effects will be negligible allowing very simple (and much faster running) models to be developed. An example of this in the context of this

document would be to determine wind speeds above which thermal effects within a city might be ignored in modelling.

On the basis of the statement that a model should be evaluated in the context of its purpose and in order to successfully correlate the technical capabilities of each model with the actual requirements of the model, it is important that a problem categorisation is made by the user.

In microscale modelling problems, the natural variability of environmental flows leads to complexities that are not met within more conventional engineering flows and this may lead to misunderstandings. For example a meteorologist considering measured or predicted velocities or concentrations would immediately ask as to the averaging time being used. The engineer most likely would be unaware of this consequential issue since most engineering sourced models (CFD or otherwise) do not address this aspect of environmental flows. The natural variability of environmental flows generally creates a level of inaccuracies and uncertainties much larger than the accuracy of the instrument used, and this makes the sound evaluation of a model a really difficult task. In order to address the unique features of the environmental flows, models have become ever more complex and costly in terms of computing power and resource requirement. The latter generally manifests itself in the cost of skilled people to set up and run the model. However, as the complexity of a model increases, its accuracy and range of applicability becomes a lot more difficult to assess. A good example of this is the increase in use of sophisticated Large Eddy Simulation models. These models have the potential to produce accurate time dependent solutions and have helped to address issues such as the aforementioned one of averaging times. But the effort required determining acceptable, necessarily non-stationary initial and boundary conditions and validation data (Adrian et al., 1999) can be a major inhibitor for the use of LES models.

It may not be appropriate to talk of a valid model in the true meaning of the word 'valid', but only of a model that has agreed upon regions of applicability and quantified levels of performance (accuracy) when tested upon certain specific and appropriate data sets. It is possible to build up evidence about the behaviour of the model using comparisons with field and laboratory data and to determine if the model performance is within agreed upon acceptability criteria.

Various attempts to provide sound guidelines for the various processes that a Model Evaluation procedure should comprise have been made within the EU. One of the most notable was under the Model Evaluation Group. Under this specific action it was argued that in order to evaluate a model properly, the following steps need to be followed:

- Model description: this should be a brief description of the characteristics of the model, the intended range of applicability, the theoretical background on which the model development was based, the software and hardware requirements, etc. Any experimental data that has been used in the development of the model should be noted in order that it may be separated from data that has not been used for model development. Both data sets may be used in the model evaluation though data previously used in model development should be flagged.
- Database description: a complete description of the database which is to be employed for the evaluation of the model, including the reasons why this

specific database was chosen. It is necessary for users to have access to quality assured experimental measurements relevant to the intended range of applicability of their models. Datasets of field and laboratory experimental measurements need to be submitted to an evaluation procedure themselves. An estimation of the data uncertainty is required.

- Scientific Evaluation: This is a description of the equations employed to describe the physical and chemical processes that the model has been designed to include. If appropriate it should justify the choice of the numerical modelling procedures and it should clearly state the limits with respect to the intended applications.
- (Code) verification: this process is to verify that the model produces results which are in accordance with the actual physics and mathematics that have been employed. This is to identify, quantify and reduce errors in the transcription of the mathematical model into a computational model and the solution (analytical or numerical) of the model.
- Model validation: this is a structured comparison of model predictions with experimental data and is based on statistical analyses of selected variables. It seeks to identify and quantify the difference between the model predictions and the evaluation datasets; it provides evidence as to how well the model approximates to reality. A quantification of the uncertainty of the model predictions should be produced.
- User-oriented assessment: is there a readable, comprehensive documentation of the code including technical description, user manual and evaluation documentation? The range of applicability of the model, the computing requirements, installation procedures, and troubleshooting advice should be available. This could also include best-practice guidelines to assist the user in working with the model.

Five of the steps of the evaluation procedure described above are relatively straightforward but the model validation is complex and requires more attention. Unfortunately this has led to the often-seen model evaluation study that is no more than the validation step. At the heart of the complexity of the model validation process is the stochastic nature of atmospheric flows, whether real or physically modelled. For example, and prior to any comparison between mathematical model and experimental results, the user or model evaluator needs to address issues such as:

- Which quantities should be compared?
- At which point within the area of interest should the comparison take place?
- Should the comparison take place on a point to point basis or on an area averaged basis?
- Should the compared quantities be averaged over a specific period of time and if so what is the time over which the averaging should take place?
- Should the quantities be compared at the same time or at different times?

The answers to these questions become clearer when the purpose of the model is precisely stated. The answer may also become clearer within the evaluation process if we are working in an "application challenge" mode. This is in a mode for which the specific problem or sets of problems that are to be used for the evaluation are clearly defined. Furthermore, the various metrics that will be used need to be carefully selected and agreed upon. Experience has shown that there may be some generally expected values for these metrics for “state of the art/science” models when applied to particular data sets subject to a specified protocol. If the data set, the protocol and the expected values of the metrics are available it is possible for anyone to evaluate the quality of a model by himself. After the evaluation process it should be possible to provide guidance as to what an evaluated model could be used for, and, as importantly, not used for.

Overall, the rapidly expanding need for the use of mathematical models of all types in microscale meteorological problems dictates the need for the proper evaluation of these models through a single procedure accepted within the scientific and wider communities. In addition, a model should be evaluated in accordance to its intended purpose(s). Although the natural variability of environmental flows lead to complexities which make the evaluation of a model a difficult task, it is possible to design such a procedure which comprises various steps through which all of the critical issues can be addressed. The goal of COST 732 is to design such a procedure, implement it and have it accepted for microscale meteorological modelling throughout Europe.

4 Recent Initiatives for the Evaluation of Microscale Meteorological Models

There has been considerable recent experience, particularly within the EU and supported by the EC, on the evaluation of microscale meteorological models and particularly those using CFD approaches. COST Action 615 was a seminal project in the study of the evaluation of models for use in urban environments, although it did not produce a formal protocol for implementation. It did, however, lead to an inventory and categorization of models used within Europe and databases available for comparison with model results.

The **EMU** (Evaluation of Modelling Uncertainty) model comparison study treated the near-field dispersion of toxic and hazardous gases near buildings. The overall objectives were to evaluate the variability in results due to the way that Computational Fluid Dynamics (CFD) models are used and to assess the accuracy of such codes for complex situations. The study involved four partners and 14 test cases were chosen. These ranged from single building on flat terrain scenarios right through to cases associated with a specific, complex topography industrial site. One of the most important parameters from a practical perspective, arising from the release of any contaminant, is the “hazard range”. This is usually defined as the downwind distance from the source to the span-wise vertical plane through the downwind tip of the particular iso-concentration surface. In this study separate teams of experienced CFD modellers were all working with the same CFD model and working under realistic time and resource constraints. Of particular interest was the large variation among the results and that a major part of this variation arose from the different selections of model inputs made by the teams (Hall, 1997).

SMEDIS (Scientific Model Evaluation of Dense Gas Dispersion Models) was a project to develop a protocol for model evaluation with particular emphasis on the complex effects of obstacles, terrain and aerosols and then apply the protocol to some 30 models then used in Europe. Daish et al (1999) describe the project and the implementation of SMEDIS. An important observation was that the validation should be a staged procedure to allow identification of any potential problems with the protocol. Within SMEDIS, a database suitable for the validation of dense gas dispersion models was developed (Carissimo et al., 2001)

SATURN was one of the first large pan-European projects that focussed on flow and atmospheric pollution in urban areas utilising Computational Fluid Dynamics (CFD) modelling together with more conventional mathematical modelling. The project was structured into three clusters and one of them addressed microscale modelling on the street and neighbourhood scales. These scales were also investigated using wind tunnel and field experiments. Applications included simulations of the flow, turbulence, heat fluxes close to building walls, vehicles motion, pollution dispersion and the assessment of road-user exposure to fine particulate air pollution using full-scale measurements and physical modelling in wind tunnels. Numerical modelling of airflow characteristics and particulate matter was evaluated against collected data. Microscale models were applied to various street-canyon geometries, in some cases taking into account windward-facing wall heating. Corresponding results were further assessed and analysed in comparison with high-quality wind tunnel data. Two models that allowed for buildings were compared for assessing the uncertainty associated with their application. A Lagrangian particle model and an Eulerian-Lagrangian method accounting for moving vehicles were further developed and evaluated. Moreover, urban aerosol and pollution exposure models were refined and evaluated. Finally, episode forecasting models were improved and tested (Moussiopoulos, 2003).

Within the **TRAPOS** network, several CFD codes were used for flow and dispersion calculation inside street canyons. The project provided an excellent opportunity to launch an inter-comparison exercise for different CFD codes and validate the codes with available data sets. The exercise comprised six CFD models used for the numerical simulation of the three-dimensional flow field and the dispersion of pollutants at the microscale. Five test cases were defined, ranging from simple 2-dimensional cases such as a flat plate approached by a homogeneous wind field, a single cavity case, and a simple 3-dimensional case. Additionally a real scenario (Podbielski Strasse in Hanover, Lohmeyer et al., 2002) was modelled with CFD models and more conventional operational models. The aims of the model inter-comparison were:

- to assess and allocate the source of differences that appear when different CFD codes using the same turbulence model are applied to well defined test cases,
- to improve the knowledge base for model development and application,
- to demonstrate the level of agreement that can be expected from CFD modelling in the urban environment,
- to give guidance for the procedure of the case set-up e.g. grid and inflow definitions, boundary conditions etc.,
- to prove the CFD codes to be a powerful and reliable tool for application in practical situations and for the improvement of practical street pollution models.

As a result of this exercise, recommendations for practical CFD models and their use have been made. It was suggested that CFD models should not be used for a very local purpose (single observation points). The accuracy of the CFD modelling results in complex building configurations based on only one location that can be affected by local gradients should be treated with special care. For practical purpose, it is possible that an estimation of averages in time (over different inflow situations) or averages in space (to avoid local gradients) are more appropriate. In addition recommendations as to the gridding of a building configuration were made.

The on-going **QNET-CFD** network comprises 43 members drawn from industry, government research establishments, academia and the principal code vendors. Activities were organized across six thematic areas each of which was broadly aligned with one or more industrial sector including environment. The complete output comprises 44 application challenges and 42 underlying flow regimes each of which has been fully documented and quality reviewed. Some of these are relevant to microscale studies in the urban environment. Each partner calculated the case using the same commercial code and the same turbulence model. The aim was to ensure fitness for purpose without being over-restrictive. In the event, for many of the studies examined it was not possible to achieve compliance with all the quality criteria (Bartzis et al., 2004).

The Basel UrBan Boundary Layer Experiment (**BUBBLE,** Rotach et al., 2005) focussed on the investigation of the boundary layer flow characteristics above a typical European urban quarter. It was carried out in the city of Basel in Switzerland, within the Rhein valley. The surrounding topography is gentle but affects the meteorological and climatological conditions of the valley. The general philosophy of BUBBLE was to establish a long-term (one year) observational network to supplement the already quite dense permanent observations from routine stations. Additional measurements were made at two urban towers, with a Lidar and with a Wind Profiler. All operated between summer 2001 and summer 2002. Turbulence profiles were measured at the towers from street level up to 2 to 2.5 times the mean building height. For an intensive observation period (IOP) between June 10 and July 12 2002, the number of measurement positions was increased through a RASS system, several SODAR instruments, and a tethered balloon at a down town site. Sensors were operated which were specifically designed to provide information for the testing of algorithms to retrieve surface information from satellite data. All data are documented and stored in the BUBBLE-database, which includes a web-interface for data selection and download. Emphasis was given to meteorological and air quality variables measured above roof level. The field experiments were complimented by a physical model study carried out in the Large Boundary Layer Wind Tunnel at Hamburg University (Feddersen et al., 2003, 2004).

The **VALIUM** project (Schatzmann et al., 2006) was devoted to the generation of a set of high quality data for the validation of numerical models. The validation data are based on a combination of field experiments, tracer dispersion studies and corresponding wind tunnel experiments. The field experiments were carried out inside and around a street canyon in a city district of Hanover/Germany. In parallel to the measurements a system of consistent coupled numerical models has been developed. The purpose of the model system was to serve as a tool for the execution of European urban air quality regulations.

The **DAPPLE** project in London, UK included a model intercomparison and a comparison with results from tracer release experiments with predictions from four non-CFD and one CFD urban dispersion model, each of them having different complexity. The exercise examined how tracer concentrations varied at different receptor points and thereby deduced how pollutant concentrations varied with distance from the emission source. The sensitivity of operational modelling to the uncertainty of input variables (e.g. surface roughness) was also observed (Arnold et al 2004).

The **Joint Urban 2003** project in Oklahoma City (Allwine, 2004, Leitl et al., 2004 and 2005, Kastner-Klein et al., 2004, Klein et al. 2007) was probably the up to now largest urban flow and dispersion experiment in a typical American business district with high rise buildings. Instantaneous and short term releases from a variety of ground level and elevated sources were investigated. In addition to the field measurements intensive wind tunnel experiments were carried out prior and subsequent to the field experiments.

These projects addressed different objectives and used different types of model inter-comparison. The CFD models used are complex and many-facetted and it can be difficult to develop optimum evaluation procedures that also allow the apportionment of the error of model results to specific aspects of the model. But in general, it was found that model inter-comparison and model evaluation against the same databases assisted in model development partly by determining gaps in knowledge in microscale modelling within urban areas. It was generally observed that although metrics such as the positions of stagnation, separation or reattachment points were easily definable for single obstacles it was far more difficult to develop metrics for arrays of several or many obstacles or buildings. Sometimes only a qualitative assessment of the correctness of the flow pattern can be made. Quantitative comparison at fixed location was a severe test of the models. It was also found that choosing the inflow and boundary conditions for numerical simulations including group of buildings, is a big challenge, especially for comparison with field measurements. The inflow and boundary conditions are far more controlled in laboratory experiments but these cannot easily cover the range of atmospheric stabilities occurring naturally. It is tempting here but dangerous to use the evaluation of dispersion modelling against dispersion experiments as a surrogate for the evaluation of the underlying meteorological model. It is unrealistic to have a universal model for all applications and a more reasonable expectation is the identification or provision of various fit-for-purpose models that have been comprehensively evaluated within their regions of applicability. Development of the comprehensive evaluation procedure and protocol are the goals of COST 732.

Further recent, current and planned activities in the countries of the participants of COST 732 are available in the proceedings of the 2005 COST 732 Workshop (Schatzmann and Britter, 2005) and will not be repeated here. In the US a recent paper (Hanna et al, 2006) summarised the efforts of five major groups using CFD codes to simulate the flow and dispersion of tracer gases in a large region of New York City. Only qualitative comparisons are made but this paper reflects much of the current status of CFD-based studies of microscale meteorological models in urban environments and the problems related to their evaluation.

5 Model Evaluation

A model evaluation has several distinct components; scientific evaluation, code verification, validation with data and operational issues. The six specific steps that are required for a model evaluation were presented in Chapter 3. They are now developed in more detail. The need for a clear model description is considered in section 5.1 while the scientific evaluation and code verification are dealt with in sections 5.2 and 5.3. Section 5.4 addresses model validation using experimental data (the provision of the data is discussed separately in Chapter 6), the subject of model sensitivity analysis and model intercomparison is addressed in section 5.5. The sections 5.6 and 5.7 extend the discussion of operational issues initially considered in Section 5.1 that involve the model, the model developer and the model user. Finally, some aspects of model evaluation specific to CFD and non-CFD models are discussed (sections 5.8 and 5.9).

This report refers to both the flow and the dispersion of material at the microscale, interpreted here as the street and neighbourhood scales within urban environments. The methodologies developed here are to be used with non-CFD models and with CFD models. Our approach has been to write this with both model types in view though recognizing that there is more experience and literature available for the non-CFD models applied to urban environments. Consequently the report is comprehensive as far as CFD models are concerned but not all issues addressed (e.g. arguments about developing a computational grid for CFD models) will be applicable to the evaluation of non-CFD models.

5.1 Introduction and User Requirements

This section considers the interests and expectations of the user of a model; a model provided either by a commercial company ("black box" models with limited or no user access to the source codes) or publicly (e.g., US EPA developed and supported models that are provided at no or modest cost and for which the source code is available for inspection).

The section attempts to define how far the model developer is responsible for the evaluation of the model and how far the user can or has to take part in this quality assurance and improvement process.

5.1.1 Model developer's responsibilities

5.1.1.1 Documentation and Comprehensibility

The model developer should provide an extensive documentation. The documentation should be in the form of a User Manual (including a short description of the model) and a Technical Reference Manual (including a more detailed description of the model). It should include:

- Detailed documentation of the physical processes and numerical methods as well as approximations and parameterizations used in the model. In general, it should also enable potential model users to quickly review models for applicability to their planned project. For this choice, information about the following topics are essential:
    - model use e.g. temporal and spatial scales, expertise required , typical applications
    - processes modelled, model approach, solution procedures
    - model documentation
    - model performance and limitations
- Comprehensive and well documented evaluation activities. A potential user needs some indication of the expected applicability of the model. Examination

of previous applications of the model and documentation of the evaluation process provide the user with an understanding of the model's predictive ability. The documentation of the evaluation should include:

    - The range of the model tests - a summary of typical (previous) applications of the model
    - Enough information about definition, design and analysis should be included in order to allow reproduction of the validation experiments.
    - A sensitivity analysis of the influence of the input data on model results
    - The open publication of evaluation studies in peer-reviewed literature is expected

- The model documentation has to identify shortcomings and restrictions of the model. There may be known situations or scenarios where the model is not applicable. A summary of the range of testing gives some indication of the situations that the potential user should be most confident with the results.
- Many models provide tables to assist the user in parameter selection. There should also be some guidance indicating the range of the parameters which should be measured in the field. In cases where data is not available, it should provide some guidance to the selection of the most appropriate sources of input data along with suggested procedures for obtaining this data.
- Guidance on mesh generation. This is an important issue for CFD models, and many models are distributed together with mesh generation software. The user manual has to describe not only the methodology as to how to build the mesh, but also explain the restrictions due to mesh number and corresponding CPU time, and other possible "pitfalls" of the model setup.

5.1.1.2 User training and support

Sufficient training and education of the customer by the software company is also desirable. The possibilities of user errors (e.g. in input and definition of the model run) increase with increasing model complexity. User training and support are essential to minimize this effect. It may include:

- User friendly well documented model installation procedure including test case files
- Web pages providing on-line manuals, updates, courses for users, FAQs, on-line forums
- Telephone/email support

Many commercial code developers have a good policy of differentiated user approach, offering packages of different services such as training, technical support, consulting and university programs. They also organize meetings, workshops and conferences where users can present results of their numerical simulations and discuss different model applications. This approach produces useful feed-back that creates opportunities for further model development.

5.1.1.3 Graphical user interfaces (GUIs) and input data formats

The user interface to the model is also an important consideration. A user-friendly GUI with plausibility checks of the model input helps to decrease inconsistencies in the model set-up.

A model developer should keep in mind that the user may use different databases of input data and may need to create their own special data pre-processors. Therefore, a comprehensive detailed description of all input data files needs to be provided to the user, together with example files.

5.1.2 User responsibilities

As Oberkampf et al. (2002) points out, it is impossible for the model developer to cover all possible combinations of physical, boundary and geometrical conditions during the evaluation of the model because of the extremely wide range of possible applications. Furthermore, the software is ported to a wide variety of computer platforms under different operating systems and must be verified for a range of compiler options. This requires a large portion of the responsibility for evaluation to be placed on the model developer but also the user takes an active part in order to achieve sound application of the model in a particular situation.

The extensive documentation is essential to enable the user to decide whether the model is the proper tool for a specific application and to assess its strength and weaknesses. Commercial models sold as “black boxes” should not be used for regulatory purposes at all, if an extensive documentation is not provided. On the other hand, if a user decides to modify the codes provided in case of freely accessible models, they must be aware that it may lead to invalidation of any prior model evaluation results.

Model users have to be aware that uncertainties still remain. The user has to take notice of the documentation provided with a model and has to be informed about consequences of model updates. Updates may lead to significant differences between new and old model results for the same setting. If unexplainable differences arise or model results seem questionable to the user, interaction between developer and user is required for clarification and may lead to the detection of errors and bug-fixing.

It is to the benefit of the user if they take an active part in obtaining of all necessary information. A participation in a forum (on-line or by e-mail list) is a very useful tool for discussing different model applications not only with the model developer but also between different users. It can help to extend the area of model applications, to fix the optimal input parameters such as, e.g., the best choice of domain sizes, different mesh architectures, extension of the embedded refined regions (if used at all), mesh resolution near buildings, etc. Sharing user experience within this area can reduce uncertainty in model results due to the human factor.

5.2 Scientific Evaluation

5.2.1 Introduction

When simulating an air flow or air quality problem with a model, the model predictions rarely coincide with reality (or our measurement of reality). There are several sources for the differences (see Figure 2.1). One of them is Model Uncertainty and this arises from:

- The inability to include all the physical and chemical processes present within the model; pragmatically the model developer will include processes with a priority based on their importance for the intended broad area of application
- Some of the processes not being well understood or not easily parameterised
- The fact that those processes that are included will typically be simplified or approximated to a level commensurate with other model requirements, such as run time etc.
- The difficulty of representing a stochastic problem in a deterministic way
- The numerical methods are always of an approximate nature.

The model developer or the model user will apply their judgment in addressing the above issues and may be wise or unwise when applying the model to a particular problem.

When providing a solution to an air flow or air quality problem there are two important considerations:

- The provision of a timely solution within the constraints of the resources (computing, financial and human) available and with the minimum error possible (or practicable or acceptable).
- The provision of an error estimate associated with this solution.

5.2.2 Background

Based on the above remarks it is commonly unrealistic to have a universal model for all applications. A more reasonable expectation is the provision of various fit-for-purpose models that have been comprehensively evaluated within their regions of applicability. One of the essential elements in any model evaluation is the Scientific Evaluation. Any model evaluation has to begin with a Scientific Evaluation and the primary objectives of the Scientific Evaluation of a model are that:

- all important phenomena within the model's range of application are included
- the mathematical modelling of these phenomena and the associated simplifications are well justified in terms of science and model practicality
- the limits of model applicability are clear and explicit
- the numerical techniques are appropriate to the intended applications

Scientific Evaluation issues are addressed in the various Best Practice Guidelines for CFD codes (e.g. ERCOFTAC, 2000, Franke et al., 2007). Specifically in the area of air pollution an early attempt in Europe to give structured guidance on Scientific Evaluation was realized by the Model Evaluation Group (1994). Concerning the Scientific Evaluation their suggestion was that it required the detailed description of the model physics and chemistry including special features, the regions of applicability and inapplicability of the model and an assessment of the appropriateness of the scientific content.

In the SMEDIS Project funded by the European Union a methodology for evaluation of the dense gas dispersion models was developed. It was mainly based on a questionnaire to be filled by each modeller, covering all the evaluation aspects. The scientific evaluation questions refer to issues such as the problems addressed by the model, the physical and chemical processes modelled and the mathematical formulation of the problem, the solution method, the output variables and any planned scientific developments.

Additionally the recent guideline of the German Association of Engineers (VDI, 2005) addresses the issue of Scientific Evaluation mainly in terms of the input data, the domain and grid description, the general equation system and its simplifications, the various parameterisations that are important in microscale modelling, the turbulence closure, the boundary and initial conditions and the output data.

For atmospheric flows at the street or neighbourhood scale it is worth mentioning also the recommendations as to the use of CFD (RANS type) in the area of wind engineering performed under COST C14 (Franke et al., 2004). This paper addresses the problems of the basic equations, the turbulence models, the computational domain and the numerical approaches.

5.2.3 Details of the Scientific Evaluation:

The Scientific Evaluation (SE) has to be performed in such a way that:

- The user is able to judge the suitability of the model for his purpose.
- It is clear how the model, with its type and range of application, compares with the state of the art in all the modelling aspects.
- It is clear that any gaps between the model and the state of the art are being minimized.

For the SE to most effectively meet its purpose it is preferred that both CFD and non-CFD models be addressed in a uniform way. The problem in general terms is the same and in many cases the parameterization schemes for many processes are the same. Additionally such a unified approach can contribute towards model improvement and harmonization through the exchange of ideas and experiences.

The SE must address all aspects of the modelling:

- The inclusion of all necessary physical and chemical processes
- The acceptable parameterization of the important physical, thermodynamic, chemical and radiological processes (where applicable)
- The appropriateness of the model deliverables (concentrations, concentration moments, doses, concentration statistics, probability density function, etc)
- The appropriate conservation equations
- The boundary and initial conditions including source(s) descriptions
- The adequacy of the spatial domain
- The adequate representation of the geometrical details (topography and obstacles)

The parameterisation of the processes refers to the set of empirical relationships that commonly describe the effects of important physical, thermodynamic, chemical and radiological processes in simple terms.

The conservation equations refer mainly to those of mass, momentum and energy (both mechanical and thermal as appropriate), and possibly entropy. In the case of RANS CFD one could add the turbulent kinetic energy, the turbulent energy dissipation, the Reynolds stresses etc. depending on the model complexity. Simpler models, especially non-CFD ones, replace the conservation equation for a certain parameter (based on differential equations in an Eulerian framework) with a parameterization of the integral of the variable (typically in a control volume context) using either an empirical relation or the introduction of similarity arguments or similar. In the simple dispersion models parameterization schemes are used also for the wind variance, the temperature variance and the concentration variance

Integral to the scientific evaluation is the provision of a brief model description. This description should reflect not only the client information needs but also the needs of an evaluator (or user) to provide a quick overview of the model.

An issue of some operational importance is the inclusion or not of chemical processes in the model and its evaluation. The microscale meteorological models themselves are unlikely to require or include any treatment of chemistry. However many dispersion models at the microscale, particularly those being used for the prediction of NO_2, will require the inclusion of chemistry or, as a grosser approximation, a chemistry post-processing. Both of these approaches are widely used. A major difficulty however does arise with the latter approach and its evaluation because the chemistry relies on the absolute concentration of the reacting species, including any background concentrations, and the datasets for the evaluation of dispersion models are those relative to any background concentration. It is unclear how best to evaluate dispersion models that involve chemistry. The pragmatic approach may be to evaluate the microscale meteorological models and the non-chemistry microscale dispersion models along the lines described by the COST 732 documentation leaving the evaluation of any chemistry elements to field data taken in an operational context, that is in terms of field data taken for regulatory air quality monitoring.

5.2.4 The Proposed methodology

A mathematical model is in fact a set of mathematical expressions each one describing a particular process that is considered important. These equations attempt to approximate reality. In other words they represent, to a degree, simplified concepts of the reality. The underlying assumptions leading to a particular simplification have to be evaluated with respect to the range of applications, for each one separately and in connection with the whole approach. Additionally every mathematical expression has to be defended against the state-of-the-art taking into consideration the model type and the range of application.

Based on the above discussion the Scientific Evaluation should be based on the SMEDIS methodology together with the documentation format used for the COST 728 Model Inventory with some modifications to meet more precisely the present objectives.

In this case, the Scientific Evaluation has to be based on a questionnaire. The idea is the items in the questionnaire to be a complete set of the phenomena or/and processes that a modeller can ever face. The items need to be as precise and detailed as possible but at the same time not too extensive so to discourage the people from completing the questionnaire. Special care should be taken in the questionnaire design so that the answers are as short as possible without loosing completeness and utilizing at the same time internet interlinks and peer review references. The approaches/processes that are irrelevant for a particular model can be characterized in the questionnaire as N/A.

An additional important consideration is that the model Scientific Evaluation is done objectively. Relevant best practice Model Evaluation guidelines are required that deal specifically with Scientific Evaluation. It is better that our guidelines are not too general but focus on the local scale atmospheric flows and pollutant dispersion and take into consideration the type of the models that are to be evaluated. The evaluation questionnaires have to be an integral part of the guidelines.

As it has been pointed out above, several issues concerning model Scientific Evaluation are addressed in the various guidelines and recommendations focused mainly on CFD-RANS models and on model physics. There is a need the work to be extended to include non-CFD, CFD-LES and processes regarding chemistry and thermodynamics as well.

5.3 Verification of CFD Codes and Numerical Simulation Results

5.3.1 Introduction

In the context of quality assurance of CFD codes verification deals with the relationship between the conceptual and the computerised model (Oberkampf et al., 2004). The conceptual model comprises all the equations that are necessary to describe the physical system, including initial and boundary conditions. The implementation of these equations into an operational computer program is called the computerised model or CFD code. Verification therefore is purely mathematic. Contrary to that validation deals with physics and is based on the comparison of the results of a numerical simulation with experimental measurements. Validation is therefore concerned with the question whether the conceptual models together with the computerised model are an appropriate representation of reality while verification is concerned solely with the question whether the CFD code is an appropriate representation of the conceptual model. Or as Roache (1997) has formulated it succinctly, verification is used to check whether the equations are solved right and validation is used to check whether the right equations are solved. Section 5.3 deals with verification. Validation issues are discussed in section 5.4.

There exist two distinct types of verification. One is code verification which is used to demonstrate that the computerised model is consistent with the CFD code as stated above, i.e. that there are no programming errors or inconsistencies in the solution algorithm (Roy, 2005). This is normally done by the code developers. The other kind of verification is solution verification which is the estimation of the numerical error (Roache, 1997; Oberkampf et al., 2004; Roy, 2005) or uncertainty (Stern et al., 2001) of a specific simulation result and is to be done by the code user. Solution verification is also known as numerical error estimation (Oberkampf et al., 2004).

Both kinds of verification need to quantify the discretisation error which results from the fact that a system of partial differential equations is solved with finite discretisation in space and time. The most general method for estimating the discretisation error is Richardson extrapolation (Richardson, 1910; Richardson, 1927) which is used in code verification and solution verification. Therefore generalised

Richardson extrapolation is briefly introduced first and afterwards code and solution verification are discussed in general.

It is widely accepted that verification and validation of computational simulations are the primary method of building and quantifying confidence in modelling and simulation. Therefore they are core activities in the evaluation procedure of general purpose CFD codes and microscale obstacle accommodating meteorological models, which is one of the main targets of this COST Action 732. While the procedures described in the following are well established for CFD codes based on the steady RANS approach, different methods have to be used for LES. The reason for this is that in LES both the numerical error from spatial discretisation and the subgrid-scale parameterisations for modelling the unresolved flow features are proportional to the grid size. So a grid independent implicit LES is a DNS (Direct Numerical Simulation) and cannot be used to assess the discretisation error including the subgrid terms. This problem is also present in microscale meteorological models with their corresponding subgrid parameterisations (Franke et al., 2006). However, recent contributions to the quality assessment of LES from Klein (2005) and Celik et al. (2005) also employ the concept of Richardson extrapolation, which is introduced next.

5.3.2 Generalised Richardson extrapolation

Richardson extrapolation is an a posteriori error estimator that is independent of the numerical method used to obtain the numerical solutions. It can be applied to the local flow variables as well as to derived integral quantities. The method can be used for the spatial discretisation as well as for the temporal discretisation. Here it will be introduced for the spatial discretisation. If f_{ex} is the smooth exact solution and f_k the result of a numerical solution on the mesh indexed by k then these two can be related by a series expansion,

$$f_k = f_{ex} + g_p h_k^p + g_{p+1} h_k^{p+1} + g_{p+2} h_k^{p+2} + \dots \, . \qquad (5.1)$$

h_k is a (linear) measure of the grid width of mesh k, p is the order of accuracy and g are coefficients. When the solution on mesh k is in the asymptotic range then all terms of higher order than p can be neglected and p and g do not depend on h_k (Stern et al., 2001). The only unknowns that remain on the right hand side of (5.1) are then f_{ex}, g_p and p. In the most general case (which is the one encountered in solution verification) none of these is known and three equations corresponding to solutions on three different meshes are necessary to estimate f_{ex}. If k=1 denotes the fine, k=2 the medium and k=3 the coarse grid, two grid refinement ratios can be introduced, $r_{21} = h_2/h_1$ and $r_{32} = h_3/h_2$. With these ratios the series expansion (5.1) can be written for the solutions on the three meshes. From the three equations f_{ex}, g_p and p can be computed. The neglect of the higher-order terms in the series for the medium and coarse grid requires that these solutions are also in the asymptotic range. Another criterion for the applicability of the generalised Richardson extrapolation with solutions from three meshes is that the solution displays monotonic convergence (Stern et al., 2001).

Determination of f_{ex}, g_p and p is simplified with a constant refinement ratio $r = r_{21} = r_{32}$. The order of the numerical solution can then be calculated explicitly from

$$p = \frac{ln\left[\left(f_3 - f_2\right)/\left(f_2 - f_1\right)\right]}{ln(r)} \qquad (5.2)$$

The estimate for the exact solution and the discretisation error DE_1 on the fine mesh are then:

$$f_{ex} \approx f_1 + \frac{f_1 - f_2}{r^p - 1} \tag{5.3}$$

$$DE_1 = f_1 - f_{ex} \approx \frac{f_2 - f_1}{r^p - 1} \tag{5.4}$$

How the described Richardson extrapolation is used also in code and in solution verification will be shown next. Here the main prerequisites which can also be viewed as disadvantages of the method are briefly restated.

- The applicability of the method requires smooth solutions. For solutions with discontinuities or singularities the effectiveness of the method is reduced (Roy, 2005).
- The method relies on having multiple solutions in the asymptotic range which can be very expensive.
- The method does not work with divergent changes in the solution. Oscillatory changes in the solution might not be detected.
- The method tends to amplify other sources of numerical errors like round-off and incomplete iterative convergence errors. Roy (2005) states that these two errors should be at least 100 times smaller than the discretisation error.

The advantages of the method are the following:

- As a post-processing tool it can be applied with every discretisation method (Finite Difference, Finite Volume and Finite Element).
- No intrusion into the code is necessary.
- The global error or estimates of this error can be calculated for every quantity.

5.3.3 Code verification

As stated in the beginning code verification is used to analyse whether the conceptual model is correctly implemented in the computerised model or CFD code. The correct implementation has to be demonstrated (Oberkampf et al., 2004).

If the numerical method is consistent then the basic partial differential equations are recovered from the discrete equations in case of vanishing grid and time step size. The rate at which the basic partial differential equations are approached is determined by the truncation error. For example, if the smallest exponent of the grid width in the truncation error is 2 then the method is said to be of second order (accuracy) in space. Halving the grid size will therefore reduce the truncation error by a factor of 4 if the solution is already in the asymptotic range as defined above. The formal truncation error and thus the formal order of the computerised model can be found by using Taylor series expansion and subtracting the basic partial differential equations from the expanded discrete equations. Whether the formal order is observed in actual applications of the code is analysed with the aid of code verification by determining the observed order of accuracy. This is the most rigorous and therefore recommended acceptance test for code verification (Knupp & Salari, 2003).

The observed order of accuracy is determined with the aid of Richardson extrapolation as described above. Assuming that the exact solution to the partial differential equations is known only solutions on two meshes are required. From these the observed order of accuracy p can be calculated.

The observed order of accuracy is defined at every node where both solutions are available. Assuming $r_{21} = 2$ which is the general but not necessary choice for code verification this requirement is fulfilled for the coarser mesh 2 without interpolation. For the verification of the code the computation of a global discretisation error suffices to calculate the observed order of accuracy, Roy (2005).

If the observed order and the formal order coincide code verification is achieved. There are several possible reasons for the case that the observed and the formal order do not agree. The most important one is that there are programming errors. Indeed order of accuracy testing is an efficient tool to detect these mistakes. To that end the other possible sources of disagreement between the observed and formal order of accuracy should be eliminated. These sources mainly relate to the Richardson extrapolation and are solutions which are not smooth enough and round-off or incomplete iterative convergence errors. By assuring smooth solutions as well as negligible round-off and iterative convergence errors (at least 100 times smaller than the discretisation errors, see Roy (2005)), failure of the order of accuracy test can be safely attributed to programming errors.

The method described above relies on the availability of exact solutions for the basic partial differential equations. Analytical solutions of the Navier-Stokes equations do only exist for simple problems or are obtained after substantial simplification of the basic equations.

As an alternative for the use of analytical solutions to the Navier-Stokes equations the Method of Manufactured Solutions (MMS) is advocated as the best choice in code verification (Roache, 2002, Oberkampf et al., 2004, Roy, 2005). This method is based on the prescription of an analytical solution for all variables that are computed. These solutions do not of course fulfil the basic conservation equations but lead to additional source terms when inserted in the basic equations. Thus with MMS not the original system of equations is solved but a modified system of equations. However, the additional terms are known and can be implemented into the code in the exact analytical form. The corresponding initial and boundary conditions are also obtained from the prescribed analytical solutions. When the original code is run with these extensions, then results of the simulation must approach the prescribed analytical solutions at a rate with the formal order of accuracy when the grid or the time step is refined. The observed order test described above must therefore be applied to the solutions obtained with the modified equations. As the modification (hopefully) only introduces analytical, i.e. exact terms in the code, the untouched original part of the code is tested for programming errors. Roy (2005) summarises code verification with MMS in the following six steps:

- Choice of the form of the governing equations.
- Choice of the form of the manufactured solutions.
- Derivation of the modified governing equations.
- Solution of the discrete form of the modified equations on multiple meshes.
- Evaluation of the global discretisation error in the numerical solutions.

- Application of order of accuracy test to determine whether the observed order matches the formal order.

He also formulates the following requirements of the manufactured solutions:

- The analytical functions and all their derivatives should be smooth (trigonometric and exponential functions recommended). Thus the observed order can be determined on relatively coarse meshes.
- The analytical functions are not allowed to lead to vanishing derivatives (also cross derivatives) in the governing equations.
- After insertion of the analytical functions all terms in the original equations should be of similar order.
- It must be certified that the analytical functions lead to realizable variable values only, e.g. the turbulent kinetic energy must be non negative.

The MMS for code verification is a powerful set of procedures to determine the correct implementation of the conceptual model in the code. It is independent of the basic discretisation method (Finite Difference, Finite Volume or Finite Element) and can deal with coupled sets of nonlinear partial differential equations. It can also be applied to other software than CFD codes. However, MMS depends on the possibility to implement arbitrary source terms as well as initial and boundary conditions into the code and is therefore code intrusive. While this is certainly no problem for code developers, mere code users may not be able to perform code verification. Another weakness of the method is its restriction to smooth solutions.

With regards to the COST Action 732 it is recommended that code verification activities should be included in the evaluation protocol. The code developers should provide information about their verification strategies. Ideally the above described observed order of accuracy test should have been performed. Another outcome of the COST Action could be to provide/recommend some manufactured solutions which can then be used by the developers to verify their code, in the same manner as the validation dataset that are provided.

5.3.4 Solution verification (numerical error estimation)

As stated in the beginning, solution verification deals with the estimation of the numerical error or uncertainty of a given simulation result. It has been indicated previously that there exist several sources of the numerical error or uncertainty. This section deals only with the discretisation error. Numerical errors due to computer programming round-off or incomplete iterative convergence are not addressed. Rather it is implied that these errors have been reduced to a negligible amount. The remaining numerical error can then be attributed to the finite resolution in space and time. The following methods for the estimation of this error can be applied to the space discretisation and to the time discretisation. The presentation will however only describe the estimation of the spatial discretisation error.

Solution verification is also performed with the aid of generalised Richardson extrapolation. As the exact solution to the partial differential equations is not known solution verification requires at least solutions on three systematically refined or coarsened meshes, i.e. the refinement or coarsening must be constant in the entire computational domain. Then the observed order of accuracy can be computed in a manner similar to the code verification.

The necessity to use solutions on three meshes makes the method expensive because all three solutions must be obtained on meshes which are fine enough for the solutions to be in the asymptotic range, which has to be analysed. This requirement raises the question about the minimum refinement ratio r that should be used in the grid convergence study as it determines the required number of nodes or cells. For code verification it was stated that the ideal case is $r = 2$ corresponding to a doubling of cells in each coordinate direction. This increases the number of cells from the coarse to the fine grid by a factor of 64 and is therefore very demanding concerning the computational resources for practical applications. Ferziger & Peric (2002) recommend at least an increase of 50% of the cells in each coordinate direction, corresponding to $r \approx 3.4$. Stern et al. (2001) state that for industrial applications $r = 2^{1/2}$ is an appropriate choice and Roache (1998) shows that even $r = 1.1$ is enough for simple meshes.

With regards to the COST Action 732 it should be required that every simulation performed for validation has to provide an estimate of the discretisation error of the solution. The error estimate must be computed for all the target variables, i.e. those variables that shall be compared with experimental data. Although this is very demanding due to the need to perform multiple simulations and to aggregate the results on the coarse mesh, with or without interpolation, it is the only way to quantify the discretisation error in the results. If the grid convergence study should fail because of the already listed reasons or because the solution on the fine mesh displays previously unresolved flow features like e.g. vortex structures or even unsteadiness, then at least a qualitative estimate of the solution changes with grid refinement is possible. Single grid solutions should not be allowed as validation simulations.

5.4 Validation

5.4.1 Introduction

This chapter deals with the validation of models for flow and concentration predictions, as well as with the validation of interim variables relevant for dispersion, such as the velocity field. In order to validate these models, the model predictions should be compared to experimental data. This can be done for idealised test cases as well as for realistic test cases. Both aspects are discussed in subsequent sections. It is important that there is consensus amongst the scientific and wider community as to the appropriateness of the validation process. A useful reference for further reading on the matters raised in this chapter is Chang and Hanna, 2004. This report is considering only the microscale (street and neighbourhood scales) in urban areas. There are currently parallel activities in Europe dealing with the mesoscale and larger scales (Schlünzen, 1997, COST 728, ACCENT network). Concepts existing for other scales can only be transferred in a generic structure but not in detail.

5.4.2 Validation objectives

In order to conduct a validation, 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 e.g. is it the maximum concentration at various points in space or is the time for which the concentration is above some limit value or is it both of these. In other words the *validation objectives* have to be determined. The validation objectives will depend on the application that is considered, so it is not possible to devise a simple and universally applicable procedure for model validation. Many of the questions raised here and below become much easier to answer when the purpose of the model and the validation objectives are clear and documented.

However, if a model is able to predict not only concentrations but the flow field as well (CFD-codes), it is a must to include into the validation the mean and turbulent velocities as well. A model that properly predicts the concentrations but fails with respect to the flow field would be ‘right for the wrong reasons’ and must not be regarded as an applicable tool.

For a validation based on field experiments there will be practical constraints due to a lack of data. Often it is the case that not all the desirable validation objectives can be examined. For such reasons, the following does not stipulate a simple cookbook procedure, but outlines *a variety of useful considerations and procedures*. The consensus of those involved in the validation process will determine the particular process to be adopted.

5.4.3 What variables to compare?

These might be put into a hierarchy of:

- Those directly relevant to the model purpose such as the concentration or the velocity (taken over a specified averaging time).
- Those variables that are intermediaries in determining the variables which are directly relevant to the model purpose.
- Those that are neither of the above but are a useful diagnostic of the performance of the model such as the turbulence intensity.

There may also arise the question as to what variables might be compared that will provide information on the uncertainty of the principal variable, that is compare velocity or concentration fluctuations as well as the mean velocities or mean concentrations.

The question of what variables to compare is then answered by the available resources and the detail that is thought to be required.

5.4.4 How should the variables be compared?

Several questions arise here that, again, are difficult to consider in the abstract without a specific model purpose in mind. Consider an example where concentrations from model predictions are to be compared with experimental data. Should they be compared:

- when paired in space or time, or only in space, or only in time. All three are commonly used. Paired in space and time is very important for the assessment of the turbulence model within CFD models. However the stochastic nature of atmospheric flows makes a comparison based on pairing in space and time a very severe test,
- on a point-wise basis or a footprint basis; the former providing a more detailed interrogation of the model while the latter may be satisfactory in an overall operational sense,
- with greater weight be given to agreement of, for example, the larger concentrations rather than the smaller, near zero ones,

- in a way that weights false positives and false negatives differently. This is a commonly met case where the consequences of over and underestimation are asymmetric,
- as a difference between the variables or as the ratio of the variables.

5.4.5 Processing of experimental data to be used in the validation process

5.4.5.1 Familiarity with data

The field and/or laboratory data to be used for validation should undergo a preliminary analysis, such as plotting space and time patterns or testing similarity relations, so the user becomes familiar with the data and their attributes and problems.

5.4.5.2 Data Quality

Field data are often incomplete in some respects, and there may be good reason to discard some data. It is recommended that quality control and selection of data are not left to individual users of data, but is primarily done by the responsible data provider.

It can be useful to assign quality flags to data, so that it is possible to work with well-defined subsets of data - as opposed to the situation, where every data user establishes their own selection criteria. This issue is dealt with thoroughly in Chapter 6.

5.4.5.3 Processing of Laboratory data

Laboratory experiments (e.g. from a small scale simulation in a wind tunnel) produce far more extensive and controlled data but may lack specific phenomena of interest such as atmospheric stability. When using laboratory data for a model validation it must be decided whether,

- the comparison is to be made between the mathematical model and the wind-tunnel model directly, or
- the comparison is to be made between the mathematical model and the results from the wind tunnel model that have been scaled up to a real scenario.

Whatever approach is adopted there will then need to be an accompanying document describing what can be determined by this validation approach when the model is to be applied to real scenarios

5.4.5.4 Treatment of zeros or relatively small data

In many atmospheric boundary layer experiments there are many very small observed concentrations or wind speeds. The way such data are handled may have a decisive influence on quantitative performance metrics. Experience has shown that this is often a major concern in a model validation study and can be subject to somewhat arbitrary decisions. If the ratio of a predicted to measured concentration is used as a comparator then difficulties will arise and the overall validation metric can be dominated by the smallest concentrations that are being compared. Caution should be paid to this problem. A possible workaround is to impose a filter so that the ratio is considered unity when both predicted and measured concentrations are below a certain threshold. Knowledge of the uncertainty of the experimental data can be used as a guide in determining what data should be discarded or assumed to be of lower quality.

5.4.5.5 The use of raw or manipulated experimental data

In addition to using the raw point measured data in a validation exercise it may prove useful to manipulate the data into a form that overcomes the scarcity or inaccuracy of data, particularly from the field. For example the use of a "footprint" approach, in which areas within a concentration contour are compared, is common and this enables uncertainties in the advection process and the diffusion process to be disentangled. Another common approach is to use data from arcs around the source, fit this with a Gaussian curve and then extract the maximum concentration and standard deviation and use these as the data for comparison in the model validation. This is particularly favoured for the validation of non-CFD models partly for its ease and transparency. Manipulated data is often used when it is found that a comparison based on raw data "paired in space and time" produces such extremely poor performance statistics that it is not possible to discriminate among models.

5.4.5.6 Stochastic variability of velocity and concentration measurements

It is a prominent feature of data from the real atmosphere that there are large stochastic variations. For model validation, this has the consequence that it is recommended to group scenarios into *regimes* with certain physical characteristics (e.g. high wind speed, cloudy day time). With this approach the amount of experimental data available within a particular scenario type can be increased to aid in providing stable statistical results. This process is sometimes called "stratifying" the data. One should be very cautious when comparing parameters from *individual scenarios* with model results (Schatzmann et al, 2003).

5.4.6 Provision of data from modelling to be used in the validation process

The model predictions will be based on the inputs to the modelling, the setting up and running of the model and any manipulation of the model outputs thought necessary. The points made below may seem relatively straightforward if a particular model is being evaluated against a particular data set. However, preparation of input data is often far from trivial. There may be many options for choice and processing of measured data when model input is being prepared. Further difficulties arise and need to be overcome when considering the intercomparison of several models when evaluated against a particular experimental dataset. The models may have different input requirements, different procedures for setting up and running the model and require different treatments of the model outputs. These will need to be reconciled in an equitable way for all models. This issue is of less concern when working with well controlled and extensively measured laboratory data than when using field experimental data. In the latter case there will be little control over the experiment and restricted availability of meteorological and other data. The most appropriate way to work with field data is a subject of much debate within and outside Europe.

5.4.6.1 Modelling inputs

The inputs required for the model will require specification based on the experiment to be modelled. These would include:

- Terrain, buildings, obstacles either precisely or in some surrogate form such as land use or surface roughness lengths etc.
- Meteorology at whatever level of detail is required for the model. These may be specified based on a particular averaging time.

- Boundary conditions or parameterizations of thermal or mass transfer processes as appropriate.

5.4.6.2 Setting-up and running of the model

This will depend on the type of model under evaluation but for CFD models would include:

- Deciding on level of detail to be used for the modelling
- Computational domain
- Gridding/meshing techniques
- Choice of turbulence model if a choice is available
- Numerical schemes
- Use of wall functions
- Allowed run time on allowed computer type
- The handling of the "averaging time"

Specific Guidance on these matters is provided in the Best Practice Guideline for the CFD Simulation of Flows in the Urban Environment (Franke et al, 2007).

5.4.6.3 Manipulation of model outputs

The outputs may need to be manipulated into the same format as that for the raw or manipulated experimental data prior to the comparison of model prediction and experimental data.

5.4.7 Communicating experience

A mechanism for reporting experiences with data has been established. It makes use of the so-called *Wiki* concept, which allows everybody to contribute with experiences to a public pool of information on the web. This Wiki is also used to compile a list of "pitfalls" related to quality assurance of dispersion models (see http://atmosphericdispersion.wikia.com/wiki/ Experimental_data_sets and http://atmosphericdispersion.wikia.com/wiki/Pitfalls_related_to_quality_assurance_of_dispersion_models).

5.4.8 Exploratory data analysis

A convenient outcome of a model validation exercise is statistical performance measures However; such statistical measures should not stand alone as the outcome of a validation process. In parallel to the calculation of validation metrics, it is recommended to perform *exploratory data analysis*, where modelled and observed data are plotted in various ways. Such exploratory data analyses are well suited to highlight notable features in data and reveal shortcomings of models.

Commonly used types of plots are:

- Scatter plots
- Quantile-quantile plots
- Residual analysis through residual plots (box plots and scatter plots).

Scatter plots are just the plotting of predicted concentration against observed concentration and will show the degree of correlation, any offsets or any trends. Below we see the scatter plots for the concentrations from a dispersion experiment using four different sets of input meteorology. The observed data and predictions have been paired in space and time. Note that the level of agreement might be deemed poor, but this is not atypical.

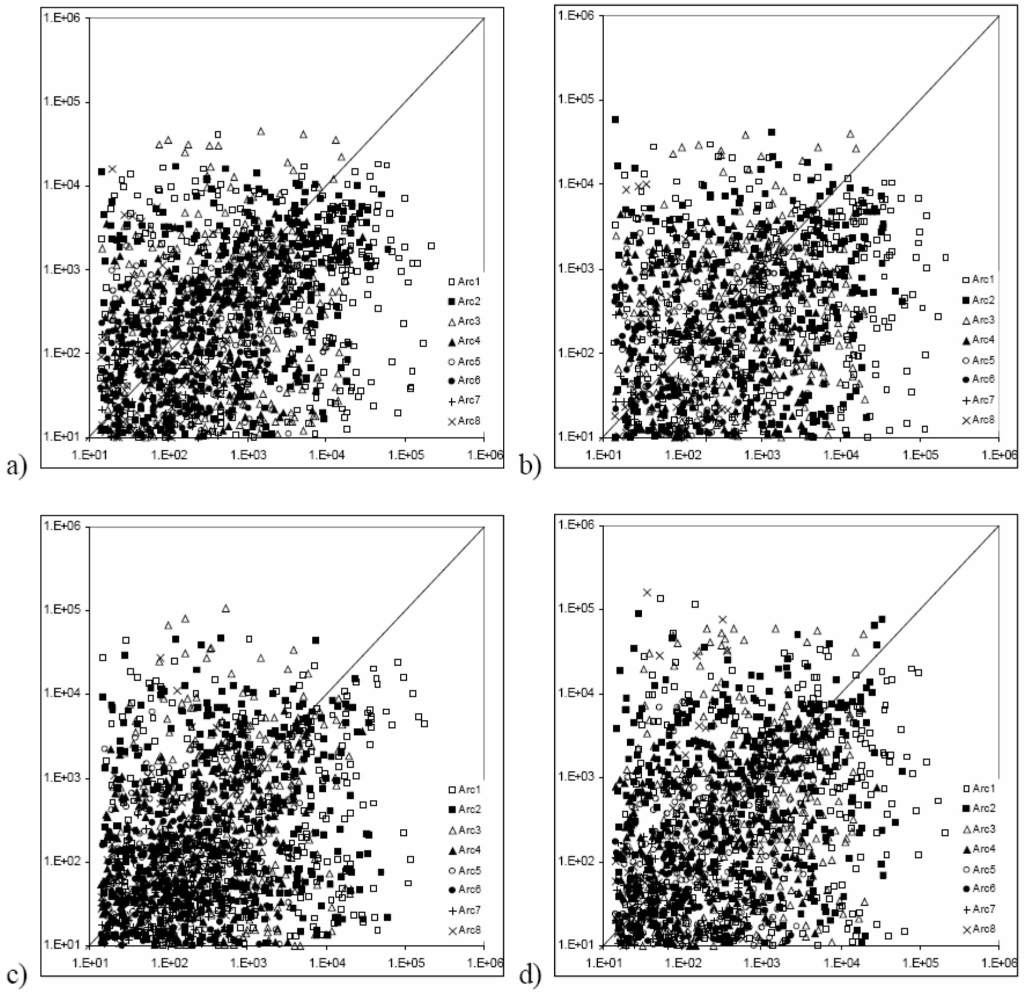


Figure 5.1a-d Scatter Plots for one particular model with data paired in space and time and using four different input meteorology data sets. Note that the axes are logarithmic.

When such data are manipulated (in this case by objectively estimating the maximum concentration on various arcs at different distances from the source and using these data for the comparison with the model results) then more can be deduced from the comparison. As shown in Figure 5.2 the spread or variance of the data is clear and it is clear that some models overpredict, some underpredict and some have little bias. These effects are evident at all concentrations.

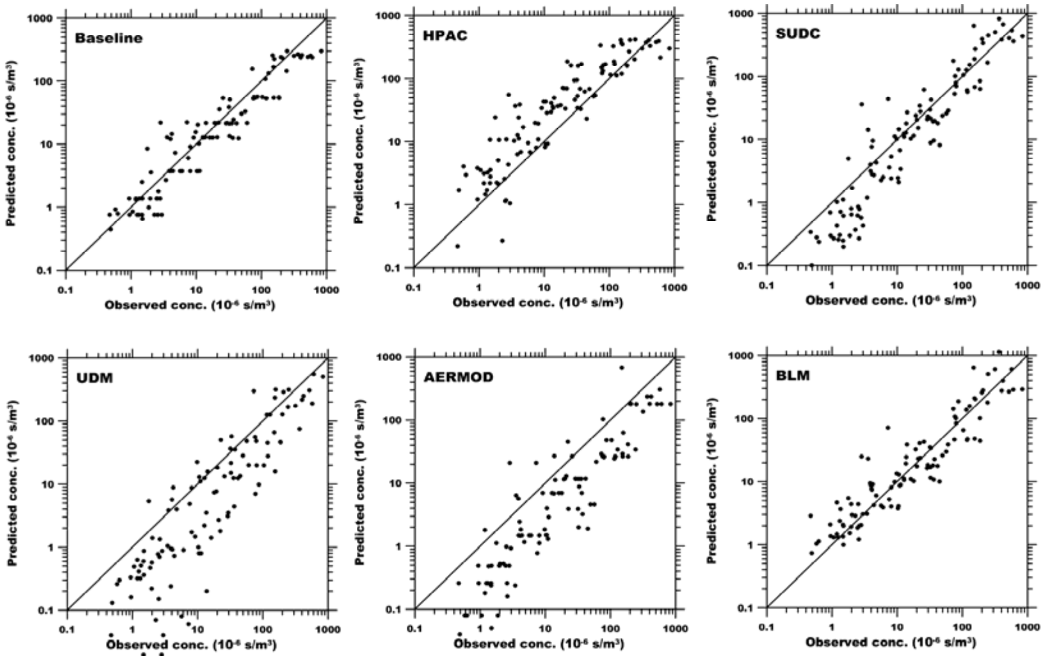


Figure 5.2 Scatter plots from six different models of predicted against observed concentration from a field experiment after manipulation of the raw data into “estimated peak concentration on an arc” (Hanna et al, 2004)

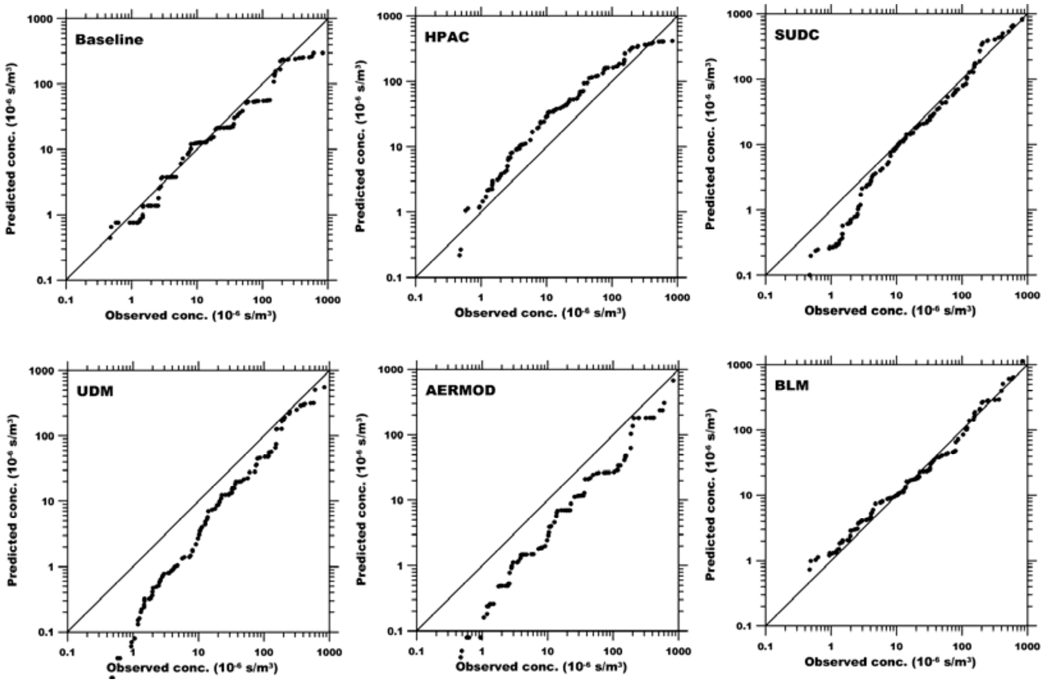


Figure 5.3 Typical Quantile-Quantile plots from six different models compared with the same field experiment results (Hanna et al, 2004).

Quantile-quantile plots (Figure 5.3) take the sets of predictions and observations and order each set from highest to lowest. The two ordered sets of data are then plotted as predicted against observed. This plot removes the "pairing in space" aspect of the comparison. This will be of use in a regulatory context that is based on the highest or, say, the 10 highest concentrations recorded. That is, where the validation objective is prediction of the higher concentrations only without concern as to when or where they occur.

Residual plots (Figure 5.4) are particularly instructive. In residual plots, the ratio of prediction to observation (P/O) is studied as a function of various physical parameters. Residual plots can reveal model weaknesses such as undesirable trends in model performance, for example such as over-prediction at low wind speeds. They can be used with raw or manipulated data.

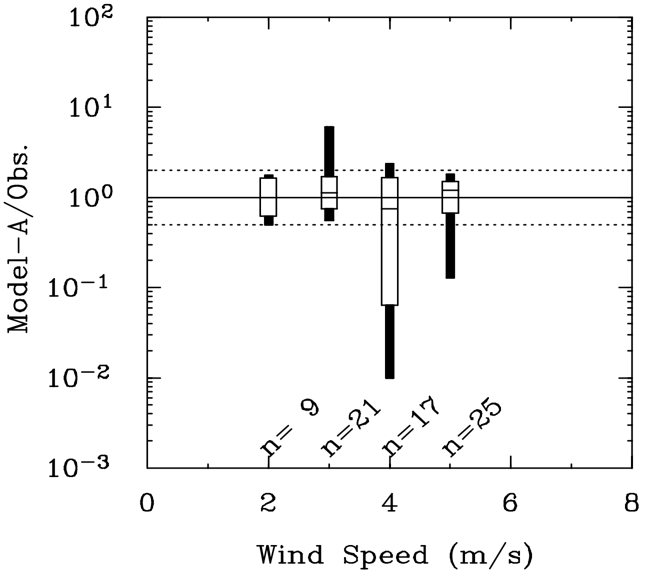


Figure 5.4 Sample residual (ratios of predicted to observed concentrations) plots for a model as a function wind speed (m s^{-1}). The significant points for each box indicate the 2nd, 16th, 50th, 84th, and 98th percentiles of the cumulative distribution of the n points considered in the box. Dashed lines indicate factor-of-two scatter. (From Hanna, Chang, Britter and Neophytou, 2004)

5.4.9 Metrics for a Model Validation

After a paired set of experimental data and model predictions have been obtained, one or more methods of comparison must be selected and quantified.

There are several standard metrics of which the correlation coefficient is the most obvious. Some other frequently used examples are:

FAC2 = Fraction of predictions within a factor of two of the observations.

FB (Fractional Bias) = $(\overline{C_o} - \overline{C_p}) / \left[0.5\,(\overline{C_o} + \overline{C_p})\right]$

NMSE (Normalized Mean Square Error) = $\overline{(C_o - C_p)^2} / \overline{C_o}\;\overline{C_p}$

MG (Geometric Mean) = $\exp\,(\,\overline{\ln C_o} - \overline{\ln C_p}\,) = \exp\,(\overline{\ln(C_o / C_p)})$

VG (Geometric Variance) = $\exp\left[\overline{(\ln C_o - \ln C_p)^2}\right] = \exp\left[\overline{(\ln(C_o / C_p))^2}\right]$

Figure of Merit (FoM) is defined as the ratio of the area marked by the intersection of the predicted and observed concentrations or dosages within a prescribed contour, divided by the union of the two areas

Measure of Effectiveness (MoE) proposed by Warner et al. (2001) assigns different weights to false positives and false negatives.

Hit Rate is another metric that has recently been found useful. The German VDI Guideline on prognostic mesoscale wind field models sets requirements in terms of this metric (VDI 2005). To evaluate the model performance, normalized values are compared, with the wind speed used for normalisation. From the normalised model results P_i and normalised comparison data O_i a hit rate q is calculated from the equation below, which specifies the fraction of model results that differ within an allowed range D from the comparison data. D accounts for the relative uncertainty of the comparison data. Only those differences are counted that are above a threshold value W, which describes the repeatability of the comparison data.

$$q = \frac{N}{n} = \frac{1}{n}\sum_{i=1}^{n} N_i \quad \text{with} \quad N_i = \begin{cases} 1 \ \text{for} \left|\dfrac{P_i - O_i}{O_i}\right| \le D \ \text{or} \ |P_i - O_i| \le W \\ 0 \ \text{else} \end{cases} \tag{5.5}$$

Another “set” of scalar metrics (for details see *Wilks, 2006*) reflects different aspects of model quality:

- *accuracy:* the average correspondence of individual model results and the observed values
- *bias: (or unconditional/systematic bias)*: the correspondence of the average model prediction and average observed value
- *reliability: (or calibration/conditional bias):* relationship of the model prediction to the average observation for specific (conditional) values of the model predictions
- *resolution*: for the model the degree of sorting the observed events in to different groups
- *discrimination*: related to the differences in conditional averages of model predictions for different values of observation
- *sharpness*: only model related (no reference to observations) ; measures the sharpness of the unconditional distribution of model predictions

Utilizing these scalar metrics for 2-D or 3-D meteorological model evaluation is simple and straightforward, however it should be remembered that these measures always mask very much relevant detail. Especially in the field of numerical weather forecasting huge amounts of different scalar and non-scalar measures for non-probabilistic and probabilistic models have been developed and tested for typically regional model evaluation studies starting as early as (Murphy, 1995) in the late eighteenth century. Many of these methods, like the new method for decomposition of the fields/contours into displacement, amplitude and residual components (Hoffman et al., 1995) can be a valuable tool also for microscale meteorological model evaluation in the future.

5.4.10 Presentation of validation results

The results from a model validation process are often presented in the form of a table as shown in Table. 5.1:

	Obs.	Baseline	HPAC	UDM	AERMOD	SUDC	BLM
Highest (10^{-6} s/m^3)	836	299	418	557	684	822	1134
2nd Highest (10^{-6} s/m^3)	606	299	410	508	305	674	646
FB	n/a	0.371	-0.221	0.404	0.678	-0.100	-0.054
NMSE	n/a	1.26	0.88	1.16	3.21	0.97	1.66
MG	n/a	1.12	0.49	2.17	2.62	1.00	0.99
VG	n/a	1.61	3.15	6.60	6.17	1.71	1.83
FAC2	n/a	0.696	0.478	0.478	0.304	0.696	0.710

Table 5.1 Summary of metric values for a validation study using six non-CFD models and manipulated data from a field experiment. The variable being used is C/q, the mean concentration divided by the source emission rate (Hanna et al, 2004).

A useful way of presenting information is to plot the results for a comparison of models as a graph of “bias” and “variance” (FB, NMSE) or (MG, VG) as shown in Fig. 5.5.

The results from Table 5.1 are plotted in this form in Figure 5.6. Here it is apparent that three models clearly show better predictive skill than the other three models. But the validation activity cannot stop here. At least two further actions are required. The first is that a decision has to be made whether none, any or all of the models under validation are “fit-for-purpose” and this is dealt with in section 5.4.12. The second action is to determine whether the results shown in Figure 5.6 reflect some oddity of that particular experimental programme. This requires the same process to be performed on other independent sets of experimental data. The outcome of this for one further data set might be that, for example, the model HPAC underpredicted by a factor of two but had the same variance. If we gave equal weight to these two validation activities then the overall result for HPAC would be a MG of 1.0 (no bias) but a large increase in VG (the variance).

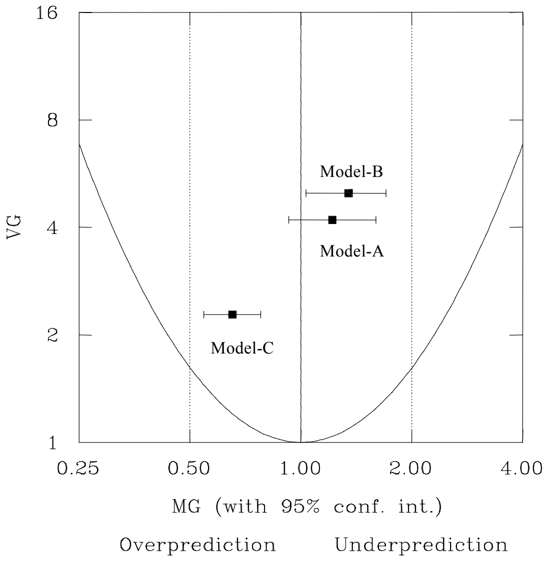


Figure 5.5 Sample residual (ratios of predicted to observed concentrations) plots for a model as a function wind speed (m s^{-1}). The significant points for each box indicate the 2nd, 16th, 50th, 84th, and 98th percentiles of the cumulative distribution of the n points considered in the box. Dashed lines indicate factor-of-two scatter (From Hanna, Chang, Britter and Neophytou, 2004).

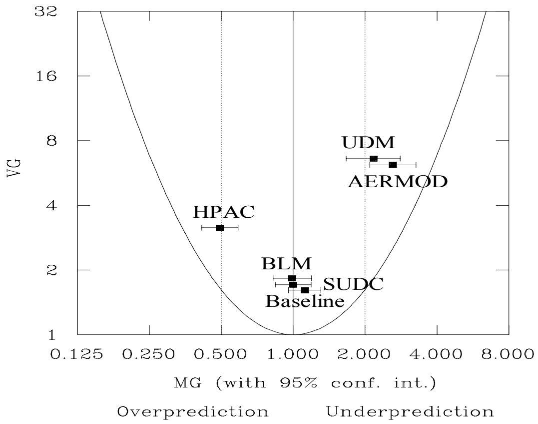
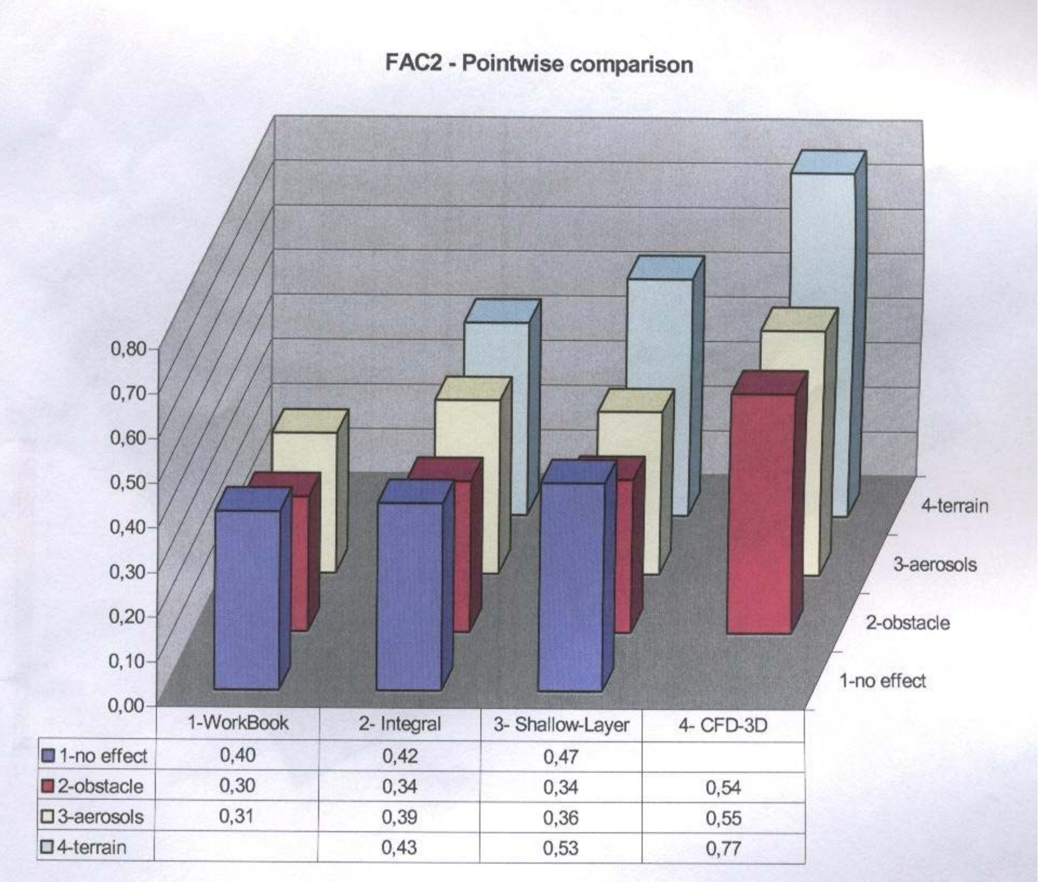


Figure 5.6 Tabulated results from Table 5.1 plotted in the manner of Figure 5.5

Another presentation method was used in the SMEDIS project that was based on the "factor of two" metric. Within the project 30 models were being evaluated and these models fell into four broad classes of model complexity. The experiments available also could be classified within four different groups of complexity. Figure 5.7 shows an interesting outcome of the study.



	1-WorkBook	2- Integral	3- Shallow-Layer	4- CFD-3D
🟦 1-no effect	0,40	0,42	0,47	
🟥 2-obstacle	0,30	0,34	0,34	0,54
☐ 3-aerosols	0,31	0,39	0,36	0,55
☐ 4-terrain		0,43	0,53	0,77

Figure 5.7 The value of the "factor of two" metric for four different categories of model complexity and four different categories of application complexity (Carissimo et al., 2001).

For the simpler problems the level of model complexity has little impact on the model performance, but for more complex problems the more complex models do provide better model performance. It is up to the model user whether the gain in model performance warrants the use of the more complex model. Additionally note that the overall average of all the performance figures was 0.48, suggesting that the "state-of-the-art" is that skilled model users in a research context are able to predict around 50% of the concentrations within a factor of two (that is a range of predictions of a factor of four from smallest to largest). Averaging each column above shows a definite improvement in model performance as the value of the performance metric changes from 34% to 62% from the workbook/rule-of-thumb model to the CFD model.

5.4.11 The BOOT software

A tool that can be useful is the free BOOT software. BOOT is a tool designed for statistical performance evaluation, and developed by Chang and Hanna. It is accompanied by a comprehensive, User's Guide (Chang and Hanna, 2005). Besides detailed technical description of performance measures and the use of the software, the User's Guide also provides a discussion of model evaluation objectives and exploratory data analysis. The BOOT package is flexible and general in nature. It seems wise to use free, verified and common software for the data analysis.

The BOOT package is capable of computing performance measures such as the Fractional Bias (FB), the Normalised Mean Square Error (NMSE), the Geometric Mean Bias (MG), the Geometric Variance (VG), the fraction within a factor of 2 (FAC2), the Measure of Effectiveness (MOE), as well as several others. (FB and MOE are in fact closely related.) BOOT allows FB and MG to be separated into over-predicting and under-predicting components. Bootstrap resampling is used to estimate the confidence limits of a performance measure - hence the name BOOT of the package. BOOT is distributed as part of the Model Validation Kit for point source dispersion, and is available at www.harmo.org/kit.

5.4.12 Quality acceptance criteria

The results in Figures 5.5 to 5.7 require interpreting in terms of quality. This is not possible unless a clear statement of the purpose of the model is stated and that purpose is converted into limit values for one, more, or all the metrics that are chosen to be relevant. This step must be performed early in any model evaluation process to "ground" the evaluation appropriately. One useful step that can be made is to establish, with experience, values of the metrics that are typical of particular model types when applied to specific problems. In a sense, these metrics and their limit values are a quantitative statement concerning the much-used, but rarely defined concept of the "state-of-the-art". As examples some opinions are expressed here:

- For the maximum concentration on an arc, the acceptable relative mean bias is + or – 50 % and the acceptable relative scatter is factor of two for research grade experiments with source well-known and on-site meteorology This is a very general statement. More precise statements can be formulated when specific test cases are considered.
- For comparisons with wind tunnel data the German VDI Guideline on windfield models (VDI, 2005) requires a certain hit rate for a number of specific test cases. For example, for wind components at specific points behind a building, the Guideline requires a hit rate of $q > 66\,\%$ with an allowed deviation $D=0.25$.

5.4.13 Validation test cases

A set of test cases with complete documentation and accessibility is required for the validation process. These should range from the simple to the complex. Three categories can be identified; simple laboratory, complex laboratory and field experiments. Each has their place. The laboratory experiments are ideal in that there is nearly complete control over the experiment. However some processes are difficult to model in the laboratory. The probably best solution is a combined data set comprising laboratory and field experiments.

5.4.13.1 Processing of input data for test cases

The initial and boundary data are very relevant drivers for the model output (initial and boundary value problem to be solved). Thus, even a perfect model can easily produce unreliable data, if input and boundary values are not correctly used. Thus, when repeating pre-defined test cases and comparing with measurements it is advisable to carefully check the initial and boundary data, and to ensure that both follow the recommendations given in the test cases. The initial data are the more relevant the shorter the forecast time, while the boundary values are the more relevant the smaller the model domains are. Thus, both are very relevant in microscale model validation. Chapter 6 describes the way which will be chosen in COST 732.

5.4.13.2 Physically modelled test cases (wind tunnel, water flume)

The German VDI Guideline on *Evaluation of prognostic microscale wind field models for flow around buildings and obstacles* (VDI, 2005) specifies several test cases, including comparisons to wind tunnel data. These are summarized in Table 5.2. All of these are considered to be simple laboratory test cases except for the last in which several buildings were used and this is considered to be a complex laboratory test case. This VDI Guideline is available at modest cost.

For these test cases the metric considered in the VDI Guideline is the Hit Rate for the three components of the wind. For comparisons with wind tunnel data a hit rate of $q > 66$ % is demanded, while for comparisons with model results or analytic solutions a hit rate of $q > 95$ % is demanded. For the cases considered, the allowed deviation D is 0.25 and 0.05 respectively.

Test case	Kind of building	Tested quality	Comparison data set
a1-1	quasi 2d building	2-dimensionality	M a1-1
a1-2	quasi 2d building	Scaling	M a1-1
a2	quasi 2d building	Stationary	M a1-2
a3-1	1 building	Symmetry	M a3-1
a3-2	1 building	Grid size dependence	M a3-1
a4-1/2	1 building	Building orientation in coordinate system	M a4
b-1	no building	Development of boundary layer	A b-1
b-2…6	no building	Direction of incoming flow (5 different directions)	M b-1
b-7	no building	Coriolis force	A b-7, M b-1
b-8	no building	Coriolis force and direction of incoming flow	M b-7
c1	quasi 2d building	Advection, turbulence	W c1, A c1
c2	quasi 2d building	Advection, turbulence	M a1-2, A c2
c3	1 building	Advection, turbulence	W c3
c4	1 building	Direction of incoming flow	W c4
c5	1 building	Width of building	W c5
c6	several buildings	Flow interaction between buildings	W c6

Table 5.2: Test cases for model evaluation and comparison data sets from VDI Guideline (2005). (M: model results, A: analytic solution or plausibility check, W: wind tunnel data from the CEDVAL data base (Leitl, 2000), http://www.mi.uni-hamburg.de/CEDVAL_Validation_Data.427.0.html.

5.4.13.3 Field experiments and data

There is a general concern that the evaluation of model quality based solely on physical modelling is an unwise practice and that the use of full scale field data and experiments is an essential part of the evaluation process. This stems from two principal concerns:

- Several aspects of the problem may be extremely difficult or impossible to be physically modelled in a wind tunnel or water flume. This is a particular difficulty when there is need to model several physical phenomena simultaneously. A short list of examples of difficulty would include
    1. non-neutral atmospheric stability
    2. thermal effects due to heating or cooling of building surfaces; both for aerodynamically smooth and rough surfaces
    3. chemical reactions
    4. deposition and resuspension
- Physical modelling is based on skilled judgement as to what the important or dominant physical processes are, and this judgement may be wrong or not lead to a definitive solution.

There are two broad types of field data. There is extensive monitoring and associated data sets from permanent field stations both for meteorological and air quality concerns and this is available for model evaluation purposes. Additionally there have been many field experiment campaigns (though until recently these were not at the city or smaller scales) to obtain data under controlled conditions and these are also generally available for evaluation purposes.

It is of interest that in the USA, the regulatory authority for air quality is the US Environmental Protection Agency and they normally require evaluation studies to be based on field experiments with no role for physical modelling in the regulatory process.

Of course field experiments have their own difficulties and limitations; cost and resource provision not being the least. The wise approach, if resources permit, is to use both approaches within the regions of applicability of each approach.

It is definitely the case that evaluation studies cannot omit the use of field scale experiments and data; this is the reason for the investment in Europe, US and elsewhere in very expensive field studies.

Several European activities were mentioned in Chapter 4. To elaborate on this we note that there have recently been several field experiments that have provided extensive meteorological and dispersion data available for model evaluation. These include:

- Birmingham 1999
- Salt Lake City 2000
- San Diego
- Los Angeles

- London (DAPPLE)
- Hanover (VALIUM)
- Oklahoma City (Joint Urban 2003)
- Basel (BUBBLE)
- New York
- And others

One of the most extensive of these that has undergone considerable analysis of the meteorological and dispersion processes is that from Salt Lake City. A limited model evaluation study of the dispersion processes was recently undertaken with 6 microscale urban dispersion models applied to this large scale experiment (Hanna et al, 2004). The raw data were manipulated into a series of arcs ranging from 200m to 6 km from the source. Comparison was made based on the maximum concentration on an arc. Figure 5.2 shows the scatter plots. The previously shown Figure 5.5 is the MG-VG plot from this model evaluation. On the basis of this limited evaluation it can be concluded that some models could be deemed to be of correct quality and some not to be of correct quality.

The novel approach to be used in COST 732 is to give high priority to data sets for which both field and laboratory modelled data sets are available. This reduces the number of high priority data sets to London, Oklahoma City, Hanover, Basel and the MUST data set (see 6.2.5).

5.5 Model Sensitivity Analysis and Model Intercomparison

5.5.1 Model Sensitivity Analysis

Model sensitivity analysis is related to model evaluation and was initially considered in Chapter 2.4.3.3. Model sensitivity is the variability of the model results due to variations in model characteristics such as the as model inputs (initial conditions, external forcing), internal model “constants”, grid resolution and numerical solution parameters. The difference between model sensitivity and model uncertainty is that model uncertainty makes reference to differences between model results and measurements but model sensitivity is only about the differences among model results arising from variation in model characteristics. However, the knowledge of the model sensitivities is essential for the user of the model e.g. in order to be aware of the scale of possible uncertainties caused by inaccurate model characteristics.

A set of model simulations can be performed in order to asses the sensitivity of the model output to the uncertainty of each model characteristic. For any set of simulations different aspects of the model (physics, numerics, inputs etc.) can be tested for their influence on model outputs.

Model sensitivity can be assessed by

- single parameter sensitivity analysis: parallel model runs with variation of only one parameter at a time
- regional sensitivity analysis: simultaneous variation of (usually) correlated parameters.

Irwin (2000) points out that as the model formulation increases in complexity to explicitly treat more physical processes, the number of input variables is increased and thus there is the the likelihood of degrading the model's performance due to data representativeness uncertainty.

5.5.2 Model Intercomparison

Model intercomparison is an additional tool for model evaluation and it provides the necessary input for achieving improvement and optimization of models even in the absence of experimental data. To use the same domain and grid resolution, the same inflow and boundary conditions in order to produce compatible output is very important point for a successful intercomparison. A structured intercomparison among models will show up whether a general consensus exists among the models or whether there are outliers. By considering the results from a range of test cases the reasons for the consensus and the outliers can be determined. This will typically lead back to a fruitful discussion among the model developers and users concerning the scientific evaluation. Questions such as "are there weaknesses caused by absent processes that are important or by inclusion but with poor parameterizations" can be directly addressed..

For CFD models it is better if all models use the same turbulent closure and grid type in order to ignore additional inconsistency. A study of applicability of different models to specific tasks was done within a number of EU projects (SATURN, TRAPOS, SMEDIS, QNET-CDF, DAPPLE and EMU) as indicated in Chapter 4.

5.6 User-Oriented Evaluation

User-Oriented evaluation is a process whereby the users, those with the most collective experience of the model, are able to assist in the model evaluation process and to aid future development of the model. In a sense it is a structured feedback mechanism directed towards the user requirements introduced in section 5.1. This aspect of the evaluation can assist prospective users in deciding whether a model is appropriate to their intended use. Feedback to model developers as to how well the code can simulate the specific scientific problem of interest and as to how well model satisfies the users practical requirements is very important for further code development. User-Oriented model evaluation was dealt with successfully in the SMEDIS (Daish et al, 2000) project and that approach will be adopted here.

Information on models is conveniently gathered by means of questionnaires. This summary information is very useful to gauge the range of capabilities of the models and the attributes and difficulties faced in using the model within that range. Such a questionnaire would seek information on

- Current model usage
    - Type of user
        - background (engineer; consultant; regulator; academic; other)
        - type of experience (dispersion; fluid dynamics; thermodynamics; numerical methods; programming; consequence modelling; risk analysis)
        - length of experience (hours; days; weeks; months; years)
    - Model distribution - location outside model developer (worldwide;

  - industry; consultancies; regulatory authorities; universities)
  - Model performance and limitations (a summary of typical previous applications of the model and identification of the shortcomings and restrictions of the model)
  - The extent of the model tests (enough information about definition, design and analysis to allow reproduction of the validation experiment)
- Hardware and software requirements
  - Computer platforms (PC, Workstation; Vector/Parallel machine)
  - Memory
  - Disc space
  - Operating system (DOS; Windows; UNIX; Linux; VMS)
  - Additional software (Compiler; Graphics package; GIS; other
- Evaluation documentation
  - Was the evaluation documentation of use in determining whether the model was of an appropriate quality for the intended purpose?
  - Was the evaluation in-house or from a third party?
  - Was the evaluation active (that is, actually using the model) or passive (based on pre-existing and documented evaluations?
- Availability and Cost

5.7 Scale interaction and model nesting/downscaling

Most problems at the microscale do not exist in isolation. Atmospheric processes on the micrometeorological scale depend not only on the local features, but also on larger scale processes, e.g. those of the meso-meteorological or even regional scales. Similarly it is the micrometeorological scale flows and transport processes that determine the boundary conditions (such as the surface stress or the near surface velocity profile) for the larger scale processes.

Micrometeorological and dispersion models for inhomogeneous areas, like urban domains, are sensitive to the choice of boundary conditions. In many research models and test studies the boundary conditions are simplified or artificial, mostly based on the assumptions of horizontal homogeneities in corresponding directions on the inlet and outlet boundaries of the considered domain.

However, in most of urban simulations for real conditions only a small part of the urban area is considered in a micrometeorological model and urban heterogeneities outside the simulation domain affect the microscale processes. Therefore, it is important to build a chain of models of different scales with nesting of high resolution models into larger scale lower resolution models.

Different requirements should be considered for the main key parameters and levels of parameterisations for urban models of different scales (see Table 5.3). Usually, the microscale (street canyon) models are obstacle resolved and consider detailed geometry of the buildings and urban canopy, whereas the up-scaled city-scale (sub-meso) or mesoscale models consider parameterisations of urban effects or statistical descriptions of the urban building geometry. One example of such model downscaling for urban meteorology and air pollution modelling, based on the FUMAPEX methodology (Baklanov et al., 2005), is demonstrated in Figure 5.8.

Mesoscale models	**Sub-meso scale models**	**Street canyon scale models**
z_0, z_{0T}	$z_0(x)$, d(x)	
h_{UBL}	L_c, L_a, z_*	Detailed geometry
'Surface' fluxes (effective)	u_*^{IS}, H^{IS}, general: x_*^{IS}	$\bar{u}(h)$ second velocity scale for **horizontal** transport
Anthropogenic heat flux (non-surface) at some representative height	Dispersive fluxes	Heat exchange at vertical and horizontal building surfaces
Profiles of turbulent fluxes	Profiles of turbulent fluxes	Characteristic velocity variance in street canyon
Higher order moments?	Higher order moments (skewness, ...)	Higher order moments?
Synoptic forcing, average albedo	Mesoscale stability, albedo(x)	

Table 5.3: Key parameters for urban models of different scales (*COST 715, 2003*)

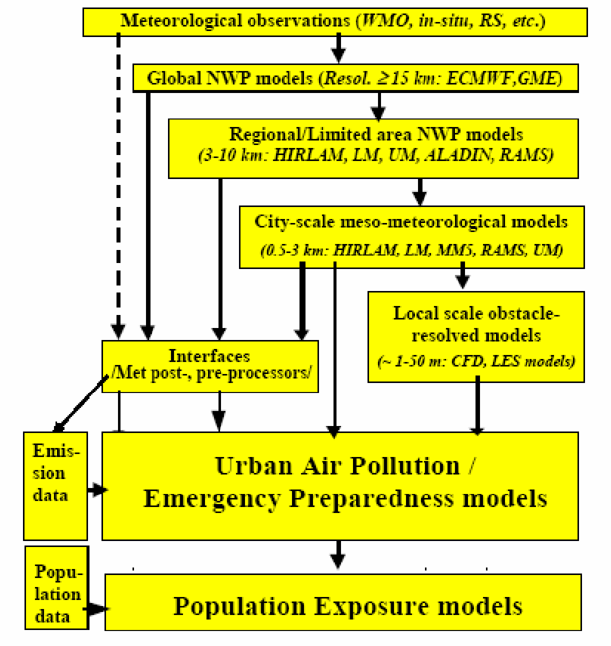


Figure 5.8. Current regulatory (dash line) and suggested (solid line) ways for forecasting systems of urban meteorology within Urban Air Quality Information and Forcast Systems (UAQIFSs) by downscaling from the adequate meteorological or numerical weather prediction (NWP) models to the urban/microscale obstacle-resolved models.

In a general sense, the scale interaction can play an important role in both directions: not only from a larger scale to the smaller microscale, but also from the urban/microscale to larger scale processes (e.g. atmospheric transport of harmful pollutants, initially released and dispersed in a street canyon; urban climate and wind climatology, etc.).

Therefore, two main types of the nesting techniques for the model downscaling can be chosen:

(i) one-way nesting, when the coarse-resolution model runs independently of the nest, just providing the initial and boundary conditions, and effects/feedbacks of the local/microscale run on the larger scale run are not considered, and

(ii) two-way nesting (Zhang et al., 1986), when the scale effects/feedbacks in both directions (from the mesoscale on the microscale and from the microscale on the mesoscale) are considered. In this case both the domains are run simultaneously to enable the feedbacks, and the terrain in the overlapping areas must be comparative to avoid mass inconsistency and generation of numerical noise.

The second way is not always reasonable to consider, because the two-way nesting approach is computationally more expensive in comparison with the one-way nesting. Therefore, for the considered specific problems it is recommended to do in advance (before suggesting the modelling system for end-users) a sensitivity study of the possible feedbacks from the microscale to larger scale processes.

One of the most important aspects in the model nesting is the necessary scale ratio (between the grid resolutions of the main and the nested models) to keep the numerical stability, suitable approximation and accuracy of the models. Long-term experience of many modellers shows that the ration should not be higher than 3 times (e.g. CMAQ, 1999).

Another important issue is the choice of the boundary and initial conditions for the inner nested model. For the boundaries a kind of Dirichlet conditions usually are chosen to provide the values with an interpolation from the coarse to the fine grid. However, it is necessary to be very careful in such approach to keep the mass/energy consistence, in other case a possible inconsistence can generate ‘parasite’ waves or explode the model in long simulations. As one of possible ways to diminish this problem could be the choice of a more soft conditions on one of the outlet boundaries (however, remember that the Neumann type of boundary conditions will increase the computation time).

In some cases, when we switch from the roughness length approach to the obstacle-resolved approach in the nested model, the interpolation procedure can be difficult for the meteorological fields within the urban canopy. This will require increasing the computation time due to the necessary additional iterations. In such cases the perturbation approach can be used. In this approach the main meteorological variables are considered as a sum of two components: background (large-scale) values, described by the coarse-resolution model, and perturbations due to microscale features, described by the nested fine resolution model.

Initial conditions for the inner domain can be obtained via sequential interpolation of values from the encompassing coarser domain. The four-dimensional data assimilation (FDDA) technique, based on Newtonian relaxation (or nudging) can be used. Nudging is a continuous form of FDDA that “relaxes” model results toward the "correct solution" by artificial tendency terms in the prognostic equations (see e.g. Seaman et al. 1995; Boucouvala et al., 2003).

Other following aspects of the scale-interaction and model down-scaling in the microscale modelling of atmospheric processes in urban areas are also important to consider in the quality assurance and improvement of micrometeorological models:

- Remoteness of the boundaries from considered obstacles in the modelling domain
- choice of initial conditions, boundary conditions or interpolated 3D fields: different interpolation procedures on each time step
- Data assimilation for nested domain runs
- Perturbation approach: advantages and disadvantages
- LES and DNS: further downscaling, periodical conditions vs. nested
- Thickening of the model grid in specific areas (e.g. in the method of finite elements)
- 'Hard' boundary conditions: problems with parasite waves and instabilities

The nesting of fine-resolution models into larger-scale lower resolution models is an important issue in micrometeorological and pollution modelling for urban areas, especially for model validations vs. real measurement data, for operational models, and for other real practical applications of the microscale models.

5.8 Aspects of model evaluation specific to non-CFD models

We have categorized models in decreasing order of complexity; LES, RANS, mass-consistent models and porosity and equivalent models. In the first three categories the buildings are resolved whereas in the last the models accommodate flow and dispersion aspects of the urban canopy in a statistical manner rather than resolved.

Porosity models (Cionco, 1965, Martilli et. al. 2002, Coceal and Belcher, 2004) use statistics of the urban surface to provide mass, momentum, turbulence, energy etc. modifications to the modelled equations to provide flow solutions within and above the modelled urban canopy. Similar and simpler, but more empirical models (Bentham and Britter, 2003, Britter and Hanna, 2003) have also been developed. These models may provide a description of the flow, turbulence and exchange processes in the urban canopy adequate for many purposes.

A variant of this approach is when statistics of the urban surface are used to produce a surface parameterisation of the effects of the urban canopy, and this is to be incorporated into the modelling as a condition at the surface, e.g. the displacement height and the roughness length. An extension of this approach is the development of "wall functions" that enable CFD codes to not have excessive gridding close to the surface. A still further extension would be the use of linearised perturbation models to model the effects of changes in surface roughness.

These models are immediately useful when incorporated into larger scale, such as mesoscale, models. The subsequent near surface solution will provide an appropriate boundary condition for the use of a building resolved RANS or LES model.

We are uncertain whether there has been any formal model evaluation study on these types of models. Any evaluation would likely require the use of spatially averaged variables; horizontally averaged or possibly volume averaged, such as velocity, speed, turbulence, shear stress, recirculating regions, temperature, mass, moisture, momentum and heat transfer etc. The requirement of spatially averaged variables may preclude the use of real field data though mock field data may be acceptable together with wind tunnel data. Microscale CFD results, that had been evaluated, would also seem to be an appropriate data source.

Models for dispersion in and near urban canopies are available (Hanna et. al., 2003). Model evaluation studies for these types of models would appear to be no different to previous evaluation studies for non-CFD dispersion models except that some degree of spatial averaging or estimation of maximum concentrations and plume statistics may be required.

A further class of model that must be included in this category is the "street canyon" model (Berkowicz, 2000) and its variants that are directly relevant to urban air quality studies, particularly those to do with traffic emissions. There have been extensive evaluation studies of this class of model using monitoring data and with dedicated experimental campaigns. Within the EU TRAPOS project a structured comparison of many models was undertaken.

There is substantial experience and background in the formal evaluation of model quality for non-CFD codes. This is particularly so in the dispersion and air quality areas and this has been driven by the regulatory context in Europe, the US and elsewhere. There is less evidence in Europe of formal evaluation of meteorological (non-CFD models) though this may be due to the small number of such models (linearised models, porosity models). However the surface parameterizations incorporated into CFD models can be interpreted as microscale non-CFD models. It would be useful to evaluate these parameterizations.

Experience has shown that there may be some generally expected values for the metrics for "state of the art/science" models when applied to particular data sets subject to a specified protocol. If the data set, the protocol and the expected values of the metrics are available it is possible for anyone to evaluate the quality of a model themselves.

Finally guidance needs to be given as to what an evaluated model could be used for, and not used for. These include application to general atmospheric dispersion using the Model Validation Toolkit (Olesen, 1995; Olesen, 2005), mesoscale transport and dispersion models (Chang et. al., 2003), dense gas dispersion models (Hanna et. al., 1993) and the EU SMEDIS project and more.

5.9 Aspects specific to CFD model evaluation

The CFD models are able to address problems of considerable complexity. They are also subject to substantial user options as to how the solution will be obtained. Consequently there will be several issues of specific relevance to the evaluation of CFD models that require consideration and choice. These include:

- **Neutral evaluator:** A choice must be made as to whether the evaluation will be undertaken by the intended user or by a neutral evaluator or whether all the models would be run by one individual.
- **Level of Commonality:** A choice would need to be made as to whether there would be an enforced commonality of sections of the

model evaluation e.g. common turbulence closure model or common gridding structure. This is often considered to be a “lowest common denominator” approach that penalises more advanced models or those that have more options.

- **Gross Features:** A choice must be made as to whether an objective assessment of the gross features of the flow (location of stagnation, separation and reattachment points or over drag forces etc) is needed for the evaluation.
- **Point positions for comparison:** The selection of spatial grid points and temporal positions for data for model evaluation will need to be made.
- **Point comparison:** A decision must be made whether a point comparison is to be made paired in space and time, paired in space but not in time etc
- **Point, spatial or volume comparison:** A choice of point, space or volume averaging of data for comparison will be required.
- **Time averaging:** A decision will be required as to what time average the model results correspond to.

6 Validation Data Requirements and Criteria for Compiling Validation Data Sets

6.1 Motivation

The validation of mathematical flow and dispersion models relies markedly on the availability of experimental data for model testing and for the evaluation of model quality. There has been no clear definition of what should be called 'validation data' and which quality requirements should be met by validation data. Consequently, various sources of atmospheric flow and dispersion data have been used, and in some cases misused, for validation purposes.

There is more to “validation data” than just measured flow and dispersion data available from monitoring stations in the field. Atmospheric flow and dispersion phenomena are primarily controlled by a set of complex boundary conditions and, depending on the type and scales of the problem, also driven by a particular set of initial conditions. As a consequence, a specific flow and dispersion situation can be characterized physically for validation purposes only, if all physical (and even chemical) boundary conditions are known and documented completely.

The validation data requirements are essentially based on the type of problem that is being considered and the type of model that is being used to model the problem. Each of these will lead to particular data requirements. From a field experimentalist’s or physical modeller's perspective, the data that can be provided can, at best , be what is possible with the 'state of the art' in instrumentation, what is practically possible and what financial resources are available to obtain the data; these may not necessarily be what is needed for the validation of mathematical models. Consequently compromises, involving some assumptions or approximations, are required. It is here that considerable care, skill and experience is often required.

However it is clear that the availability of comprehensive and appropriate reference or test data together with information on its quality significantly influences the outcome of model validation procedures; the results of a model validation are as good or as uncertain as the reference or test data.

In section 6.2, the validation data requirements, irrespective of what can actually be delivered are outlined together with the constraints of the numerical modelling and the experimental data. Sections 6.3 and 6.4 focus on how to fulfil the demanding data requirements with respect to data availability and data quality.

6.2 Validation Data Requirements and Constraints

6.2.1 Validation Data Requirements

Validation data should provide *test cases of different complexity*, ranging from very simple isolated structures to complex urban cases. The simpler test configurations would allow the testing of codes which are still under development or codes based on simple models which are not designed to be applied to complex urban cases, whereas the complex test configurations would allow the quantification of the model performance, accuracy and applicability in practice of more sophisticated models.

A second requirement is to provide validation data sets for *systematic testing* of numerical codes. The ability of a numerical model to replicate one particular test case representing one particular configuration is not adequate to determine the quality of the model. Consequently, systematic validation data sets are needed, in which just one governing boundary condition is varied systematically in reasonably small steps. Mean wind direction, model size, size of the modelled area, emission source locations or emission source types can be gradually varied to provide a near- ideal validation data set so as to fully document the model quality. Such a data set also allows the determination of the sensitivity of the problem itself, and of the mathematical model, to changes or uncertainties in the input parameters.

A third important requirement is of course a *complete validation data set* with information:

- with respect to the boundary conditions necessary for a physical classification of a test case,
- with respect to the model input data required to setup a model run without further assumptions,
- with respect to a documentation of the uncertainty of the reference data.

For the documentation of the physical boundary conditions of a flow and dispersion scenario, it is necessary to provide at least information on:

- the geometry of the problem with sufficient detail,
- the mean wind profile at the entrance of the domain,
- statistical and spectral turbulence parameters at the entrance of the domain (turbulence intensity profiles, turbulence spectra, length scales etc.),
- the mean temperature profile at the entrance of the domain,
- the source strength and source position,
- the background concentrations.

The input parameters at the domain entrance can be derived from the documentation of the physical boundary conditions for most of the models discussed here. An exception must be made for LES models because dynamic forcing at the entrance of the computational domain requires additional information on the temporal behaviour of the boundary conditions and their temporal and spatial correlation.

Complete documentation of the uncertainty of the measured validation data is required. This should contain the instrumentation accuracy, the uncertainty of the measured values with respect to their spatial and temporal representativeness and the repeatability of results under similar boundary conditions. For field and laboratory data it must be clearly documented as to how data have been acquired and what averaging, if any, has been performed.

In terms of data quality, validation data must fulfil higher quality criteria than standard field or laboratory experiments. The use of any experimental data as a reference requires them to be at least as precise as the maximum allowable model uncertainty. Some test data is derived from a *combination* of measured values and error propagation will introduce an uncertainty larger than the actual accuracy of any individual measurement. As one example, the non-dimensional concentration value C* is used to facilitate the comparison of measured and modelled pollutant concentrations by combining wind speed, measured concentrations, a characteristic length scale and the emission source strength. Consequently, the uncertainty of C* combines the individual uncertainty of the emission source strength , the wind speed measurement and the concentration measurement and these can easily sum to more than 20% uncertainty for typical laboratory conditions. A much higher uncertainty must be assumed for standard field data where frequent calibration of the instrumentation is often impractical and the uncertainty of individual measurements like wind speed, pollutant concentration and emission source strength derived from traffic counts is much higher. In many existing cases, important boundary conditions like background concentrations or wind directions are not even measured locally and must be derived from more global meteorological boundary conditions using unproven assumptions.

6.2.2 Numerical Modelling Constraints

For the characterization of urban flow and dispersion processes, the spatial and temporal distribution of mass, momentum and energy at the edges of the domain as well as the geometry of the problem should be known with appropriate accuracy. Whereas the geometry of a problem can generally be provided as accurately as required by most of the numerical flow and dispersion models, it is practically impossible to specify physical (and chemical) boundary conditions with sufficient resolution in space and time simultaneously for all bounds of a domain. As a consequence, simplifications and assumptions must be introduced and this directly affects the quality of model validation. One of the most prominent examples is the assumption of quasi-stationary boundary conditions or the corresponding clustering of so-called 'typical urban dispersion situations'. From a strict physical point of view, quasi-stationary conditions can be assumed only if all terms of the conservation equations are in equilibrium and the treatment of the problem allows for achieving a quasi-stationary state.

Another problem concerning the basic physical constraints of validation data arises from the fact that numerical flow and dispersion modelling is increasingly capable of dealing with dispersion problems from instantaneous sources with the flow driven by large scale turbulence. Both with respect to the release of a pollutant and with respect to the exposure/immission, dispersion modelling is now commonly required to predict short-term release scenarios and transient dispersion phenomena as they occur during accidental releases or under short-term transient atmospheric conditions. It is obvious, that a quasi-stationary approach will not work in this case for model validation because the 'history' of the boundary conditions is important for defining the physical state of a dispersion problem unequivocally. Depending on the actual release conditions and the local dispersion process along the path of a dispersing plume or puff, the governing time scales can change from a fraction of a second up to several minutes or even hours. However, a proper physical characterization of transient flow and dispersion phenomena requires the boundary conditions of a problem to be defined at all time scales inherently present and any simplification of boundary conditions will necessarily introduce a mostly not quantifiable uncertainty to model validation, even if the reference data are of high measurement quality.

Another set of criteria applied to reference data is defined by the numerical models to be validated and the input data required for setting up a model run. In order to ensure a direct compatibility of model results and validation data, all input data required for a certain model type ideally would be derivable directly from a complete set of boundary conditions to be documented in a validation data set. This means that the relevant physical requirements and constraints listed above are also required from a numerical modelling point of view. In order to run a model and to compare model results with reference data, a physically complete set of boundary conditions is needed in addition to the actual reference data. The actual requirements strongly depend on the type of model to be validated, and more or less significant simplifications can be applied (depending on the kind of model to be evaluated because not all models would accept or make use of the entire set of physically sound boundary conditions as mentioned above).

Stepping back in time, the most easily to validate model type are statistical atmospheric flow and dispersion models. The 'only' requirement arising from this type of model was that the reference data and the according boundary conditions had to be statistically safe and representative. By repeating dispersion experiments sufficiently often under similar conditions or running monitoring stations over sufficiently long periods of time, statistically representative 'mean validation data' could be collected for a set of representative 'mean boundary conditions'. The situation is still similar for more sophisticated models based on the 'mean' transport equations for mass, momentum and energy. The so-called Reynolds-Averaged-Navier-Stokes- or RANS-type of models is based on averaged equations restricting the model input and the model output to mean flow and dispersion fields. Despite the big improvement of model quality due to the physics incorporated, for validating a basic RANS model still quasi-stationary 'mean' boundary conditions must be defined for setting up a model run, and representative 'mean' reference flow and dispersion fields for a given set of boundary conditions are needed for model evaluation.

Concerning the validation data sets, so-called time resolving RANS codes are more demanding. With the entire turbulent phenomena still being parameterized such models intend to provide model results for non-stationary boundary conditions as a set of quasi-stationary solutions. The approach was first introduced and applied in technical CFD applications for which 'transient' flow and dispersion phenomena could clearly separated into a high frequency turbulent part which is completely parameterized and a low frequency quasi-stationary part which is simulated by non-stationary boundary conditions. Irrespective of the question whether this approach can safely be applied to atmospheric flow and dispersion models, suitable validation data sets would require time-resolved boundary conditions and the corresponding flow and dispersion fields for a true validation of time-resolved model results.

With the Large-Eddy-Simulation (LES) models becoming readily available, the temporal behaviour of quantities at the edges of a domain is becoming a substantial part of a validation data set. LES is resolving at least the large-scale fraction of turbulent flow and dispersion phenomena directly, not relying on a complete parameterization of turbulence. However, in order to validate the temporal progression/development of large-scale turbulent fluctuations, test data sets must provide the corresponding inflow/input data at the boundaries of the domain with sufficient temporal and spatial resolution and the reference data within the domain must be available in a temporally and spatially correlated form in order to evaluate the model quality. Up to now, there is no common view on how to validate the performance of an LES model except that the model should reproduce quasi-stationary flow and dispersion fields properly before it is tested for the modelling of transient flow and dispersion phenomena.

6.2.3 Field and Laboratory Data Constraints

It is obvious, that practical flow and dispersion measurements neither in the field nor under controlled laboratory conditions in a boundary layer wind tunnel can deliver validation data sets which fulfil entirely the demands of sophisticated numerical models. Field data, for instance, cannot provide the long-term mean quantities required for validation of a RANS-based numerical model because of the constantly changing 'meteorological' boundary conditions at full scale. The time scales involved in turbulent flow and dispersion at full scale range from seconds or minutes up to hours but within an hour, the meteorological boundary conditions will change significantly. Averaging situations with 'physically identical' boundary conditions is possible only, if the governing boundary conditions are measured and documented completely with sufficient accuracy and resolution. For example, in order to extract and average situations with a given stratification only, representative vertical temperature readings must be available from a sufficiently long field data record. Otherwise, averaging of field data – perhaps for a given wind direction only - will mix different stratification regimes, which is leading to reference data representative for a more global flow and dispersion regime only, like an annual mean value.

A second problem regarding the use of field data for model validation is, that individual field results represent unique dispersion situations with a significant scatter inherent for instance in the results of 'instantaneously' measured half-hourly mean values. In most field cases, the boundary conditions during a half-hour sampling period cannot be captured precisely enough in order to define the physical conditions representative for the corresponding mean values. Because of the different time scales driving full-scale flow and dispersion phenomena, a proper definition of the boundary conditions representative for a certain sampling period must consider also time periods directly before and after the sampling. Additionally, it is still an open question whether model validation requires individual dispersion situations to be replicated by numerical modelling or whether it would be more reliable to test the 'physical behaviour' of a flow and dispersion model by applying sets of different test cases with systematically changing boundary conditions.

The accuracy or reliability of field data cannot easily be quantified based on field data alone. It is not just the accuracy of the instrumentation used for field measurements which defines the reliability of data compiled from field measurements. In addition, the repeatability of field measurements for similar boundary conditions as well as the spatial representativeness of individual measurement locations with respect to a particular flow and dispersion problem must be evaluated and quantified with respect to the measured quantities before corresponding data can be used safely for model validation purposes.

Schatzmann and Leitl (2002) give an impression how severe that not yet widely recognised problem really is. They analysed data presented in form of half-hourly averages from an urban monitoring station for a full year. The large number of data points allowed carrying out a statistical analysis. After some filtering and removal of cases originating from low traffic and low wind speed situations, the data were grouped into classes with roughly the same meteorological conditions. Figure 6.2 shows half-hourly mean concentrations of NO_x presented in non-dimensional form. The data points (small dots) scatter significantly and show a variability of roughly a factor of 2.

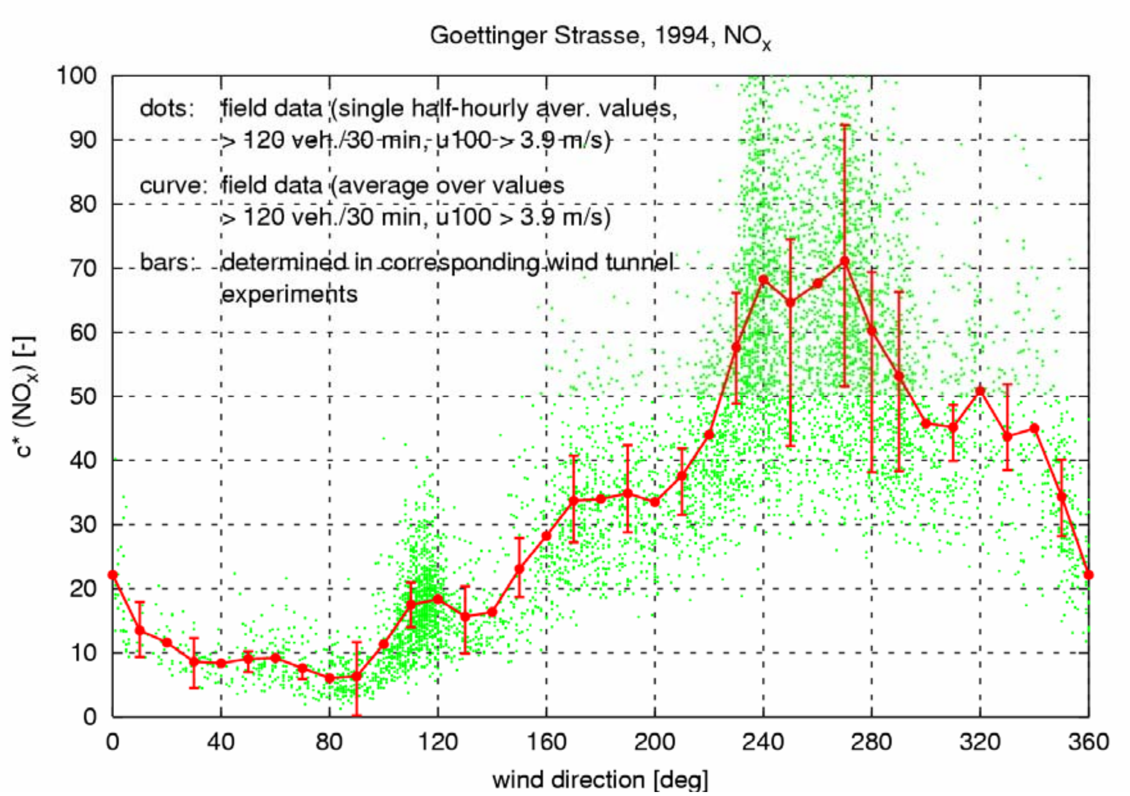


Fig. 6.2 : Normalized half-hourly mean concentration values as a function of wind direction measured over the period of one year at the street monitoring station Goettinger Strasse in Hanover/Germany (from Schatzmann and Leitl, 2002). Each individual dot represents a 30 min mean value. The "error-bars" (see text) were determined in corresponding wind tunnel experiments.

The large scatter of data points shown in Figure 6.2 supports the statement that the common 30 min mean concentrations measured inside the urban canopy layer have the character of random samples only. Depending on the wind direction, the variability between seemingly identical cases can be large. To simply increase the sampling time would not solve but worsen the problem since over periods longer than 30 min a systematic trend in meteorological conditions has to be expected. When the street canyon case is replicated in a boundary layer wind tunnel, long concentration versus time traces can be monitored. Such long time series, collected under carefully controlled constant ambient conditions, can then be split into many 30 min intervals. Depending on the complexity of the flow at the measurement point, which in a given geometry corresponds to the wind direction, 30 min intervals with high and low concentration means are found similar to those found in the atmosphere but with a smaller range. This is shown in Fig. 6.2 by the bars which indicate the minimum and maximum of the 30 min averages. We generally observe more variability in the field than in the wind tunnel even after other perturbing influences have been considered (such as non-local concentration contributions, variations in the emission rates for any particular traffic count etc). They result from the fact that clouds of pollutants are subject to low frequency disturbances in the atmospheric boundary layer which have time scales that are not small compared to the averaging time. These may not be present or not resolved in the wind tunnel. This is why we are using wind tunnel and field measurements in our evaluation process.

Laboratory data, as they can be compiled under precisely controlled boundary conditions in a boundary layer wind tunnel, can provide valuable information to extend field data and to quantify the uncertainty of reference data compiled from field measurements. The gap of information between field data sets and the needs of numerical models of different complexity can be closed at least partially by supplementing laboratory experiments. The biggest advantages of laboratory data are that the complexity of the problem can be chosen and that the physical boundary conditions can be controlled at will and can be kept constant over a sufficiently long period of time. Quasi-stationary conditions, as they are postulated for most of the model types can be achieved in a wind tunnel easily and statistically representative measurements can be realized even for systematically varying boundary conditions which are directly comparable with corresponding numerical results. Quantifying the accuracy or uncertainty of wind tunnel results is normally possible based on an error analysis as well as based on repetitive measurements under similar boundary conditions.

Despite the benefits laboratory data have, there are also severe restrictions in physical modelling, as for stratification modelling or the effects of gravity-driven flow and dispersion phenomena. Although the physical boundary conditions of a dispersion experiment can be recorded more precisely in a wind tunnel than in the field, most of the wind tunnel data sets also suffer from a general lack of temporal and spatial correlated measurements of boundary conditions as they are needed for instance for the validation of time dependent RANS or LES applications.

6.3 Classification and Mapping of Validation Data Sets

In order to classify validation data sets as well as to adjust a structured data set with respect to specific needs of a certain type of flow and dispersion model, the check list given in Table X was developed. The table is based on the general demands on reference data sets as stated above but is also providing check marks for specific information to be provided in an ideal validation data set. Since not all the information requested in the table will be used by all types of flow and dispersion models subjected, the data set is directly mapped to the model requirements. The extent and the quality of information needed, of course, also depend on the experience and skills of the user of a data set. Consequently, the amount of information required for a 'sufficiently' documented reference data set can vary significantly. For classification and structuring the documentation of data sets can be grouped into four major parts. Perhaps the most important but often least extensive part of a validation data set is a detailed description of the test(s) in general inclusive information on how the data was acquired, which post-processing was applied and which confidence interval can be assigned to individual quantities measured/documented in the data set. Further parts of a 'complete' data set must provide information on the geometry of the 'test site'/experiment and the emission source(s) involved. Finally, the actual reference data must be provided, containing flow and dispersion data and the associated physical boundary conditions.

It is certainly difficult to categorize and map reference data from field and laboratory measurements since the priority of model requirements can differ even for one model category. In this context, the check list given in Table 6.1 can be considered as a very basic evaluation tool for test data only. If essential (strongly required) information is missing, then the according data should not be used for a rigorous validation of a numerical flow and dispersion model. If, for example, a field test provides well-defined dispersion measurements for a completely documented source configuration but no reference data for the driving wind field is available, the test case cannot be used for validation of a RANS/LES-based dispersion model because it is not possible to distinguish between uncertainties in wind field calculation and dispersion simulation.

6.4 Potential Validation Data Sets

Facing the typical shortcomings that both field and laboratory data have with respect to validation data requirements and data quality, validation data sets should combine information from field experiments and laboratory campaigns wherever it is possible. A combination of physical modelling and field measurements can improve data quality and reliability of test data sets significantly. For example, systematic testing of an intended field site in a physical model can deliver information on the spatial and temporal representativeness of field measurements even before expensive instrumentation is set up in the field. Measurement locations can be optimized with respect to the accuracy and reliability of the results by specifying well-sited locations for measuring boundary conditions and the actual field data. In addition, basic information on the time scales inherent in the measured signals can be provided and adequate instrumentation can be chosen.

For the evaluation of the proper physical behaviour, systematic laboratory data can be counted as the source of choice for validation data. Systematic changes in test case geometry and complexity as well as in the boundary conditions within desired ranges are possible only under controllable laboratory conditions. The complexity of test cases can range from very simple single-obstacle-situations up to complex urban dispersion situation tested under precisely controlled, simplified and completely documented boundary conditions.

In a hierarchy of validation data sets with gradually increasing complexity the ultimate test cases would be necessarily field data of known and documented accuracy and reliability. In order to close the gaps inherent in field data sets, laboratory tests should go along with the field experiments. Combined field and laboratory tests allow for a systematic extension of field data as well as for the quantification of uncertainty/scatter of temporarily and spatially limited field data.

Following the strict data evaluation concept developed above, only a very limited set of validation data is currently available for testing and validation of 'state-of-the-art' microscale flow and dispersion models. Combined field and laboratory data are available from the MUST experiment (Biltoft, 2001), the VALIUM study (Schatzmann et al, 2006), the DAPPLE project (Arnold et al, 2004) as well as from the BUBBLE project (Rotach et al, 2005) and the Joint Urban 2003 experiment (Allwine, 2004). The data sets cover a wide range of geometrical complexity, ranging from a rather simple array of containers to the in-homogeneously structured roughness of the central business district in a modern city. Whereas the field data mainly focus on point-wise, local flow and dispersion measurements the corresponding laboratory data provide quasi-stationary flow and dispersion fields (MUST, VALIUM, JU2003) and instantaneous flow and dispersion data for well-defined boundary conditions. Test data for systematically changing boundary conditions are available from field data (MUST, VALIUM, JU2003) and, to a larger extent, also from physical modelling in boundary layer wind tunnels.

Table 6.2 documents the availability of information, without distinguishing between information from wind tunnel experiments and field tests. Currently, the table is mainly based on a literature survey. The content will be updated after approaching the originators of different data directly. In a next step it is intended to homogenize the existing field and laboratory data in order to compile consistent validation data sets. Upon approval of the authors of particular data, selected results of corresponding field and laboratory experiments will be merged to more or less complete validation data sets.

Legend
not required / not used <empty cell>
strongly required ■
required ▥
helpful information ▒

		Model Requirements						
	Data Set Check List	Semi-Empirical Models	Gaussian Type Dispersion Model	Gradient Transport Models (K-Models)	Analytical Solutions (Berljand/Huang)	RANS Flow and Dispersion Models	U-RANS Flow and Dispersion Models	LES Flow and Dispersion Models
Description / Documentation of Test Data								
sufficient/detailed information on								
location of the test site	☐	■	■	■	■	■	■	■
size of the area	☐	▒	▒	▒	▒	■	■	■
general topographical description	☐	▥	▥	▥	▥	■	■	■
general meteorological characteristics	☐					■	■	■
material(s) of the site	☐					■	■	■
experimental set-up / methodology and instrumentation								
measurement grid / measurement locations	☐	■	■	■	■	■	■	■
reference coordinate system	☐	■	■	■	■	■	■	■
instrumentation used for measurements	☐	▒	▒	▒	▒	■	■	■
pollutant chemistry involved	☐	▒	▒	▒	▒	■	■	■
data quality								
spatial and temporal resolution of the test results	☐	▥	▥	▥	▥	■	■	■
post-processing routines applied to raw measurement data	☐	▥	▥	▥	▥	■	■	■
expected/estimated variablility of the results	☐	▥	▥	▥	▥	■	■	■
confidence of calculated statistical results	☐	▥	▥	▥	▥	■	■	■
previous use of data for validation purposes / experiences								
reference list (if applicable)	☐	▒	▒	▒	▒	▥	▥	▥
Geometry of the 'test site'								
sufficient/detailed information on								
buildings/obstacles								
outer dimensions & location	☐	▒	▒	▥	▒	■	■	■
roof shape/type	☐	▒	▒	▥	▒	■	■	■
surface characterization	☐	■	■	■	■	■	■	■
topography								
height profile(s) / height contours of the ground	☐	■	■	■	■	■	■	■
surroundings								
geometrical characterization of the structures surrounding the test site / model area	☐	■	■	■	■	■	■	■
accuracy/reliability of geometry data								
accuracy of building dimensions	☐	■	■	■	■	■	■	■
accuracy of topography data	☐	■	■	■	■	■	■	■
statement on sufficiency of the documented model area	☐	▒	▒	▒	▒	▒	▒	▒
Emission source(s) specification								
sufficient/detailed information on								
emission source(s)								
temporal and spatial classification of tracer/pollutant source(s)	☐	■	■	■	■	■	■	■
number/location of sources	☐	■	■	■	■	■	■	■
emission data (source flow rate, duration of release(s) etc.)	☐	■	■	■	■	■	■	■
emission data type (statistically modeled, numerically modelled, measured, etc.)	☐	▥	▥	▥	▥	■	■	■
if required - post-processing of emission data (emissions from car traffic)								
composition of car fleet	☐	▒	▒	▒	▒	■	■	■
specification of driving patterns (cold start fraction, traffic lights, traffic jams etc.)	☐	▒	▒	▒	▒	■	■	■
recommended/site-specific emission model	☐	▒	▒	▒	▒	■	■	■
accuracy of source specification								
estimated/determined accuracy of emission data	☐	■	■	■	■	■	■	■
confidence level of emission data	☐	■	■	■	■	■	■	■

Table 6.1a: Validation Data Check List – Part 1/2

Legend

not required / not used <empty cell>

strongly required ■

required ▥

helpful information ░

Model Requirements

	Data Set Check List	Semi-Empirical Models	Gaussian Type Dispersion Model	Gradient Transport Models (K-Models)	Analytical Solutions (Berljand/Huang)	RANS Flow and Dispersion Models	U-RANS Flow and Dispersion Models	LES Flow and Dispersion Models
Physical Requirements								
boundary conditions								
inflow conditions								
representative mean vertical and lateral approach flow profile(s)								
wind speed (component resolved)	☐	▥	▥	▥	▥	■	■	■
wind direction	☐	▥	▥	▥	▥	■	■	■
turbulence intensities	☐	░	░	░	░	■	■	■
temperature/radiation	☐	▥	▥	▥	▥	▥	▥	▥
turbulent fluxes	☐	░	░	░	░	▥	■	■
pollutant/tracer concentrations (background concentrations)	☐	▥	▥	▥	▥	■	■	■
additional parameters of the atmospheric boundary layer								
mixing height	☐	▥	▥	▥	▥	░	░	░
turbulence length scales (macro-scales)	☐	░	░	░	░	░	░	░
representative time series of								
wind velocity components	☐					■	■	■
turbulent fluxes	☐					░	░	■
pollutant/tracer concentrations (background time series)	☐					■	■	■
test data								
flow field								
statistically representative wind flow data at a sufficent number of locations	☐					■	■	■
time resolved / time dependent flow measurements	☐					■	■	■
concentration field								
statistically representative concentration data	☐	▥	▥	▥	▥			
time resolved / tim dependent concentration measurements	☐							
source of data / data provider / contact person	☐	■	■	■	■	■	■	■

Table 6.1b: Validation Data Check List – Part 2/2

	Data Set Check List	JU2003	BUBBLE	MUST	DAPPLE	VALIUM
Description / Documentation of Test Data						
sufficient/detailed information on						
location of the test site	☐	☒	☒	☒	☒	☒
size of the area	☐	☒	☒	☒	☒	☒
general topographical description	☐	☒	☒	☒	☒	☒
general meteorological characteristics	☐	☒	☒	☒	☒	☒
material(s) of the site	☐	☒	☒	☒	☒	☒
experimental set-up / methodology and instrumentation						
measurement grid / measurement locations	☐	☒	☒	☒	☒	☒
reference coordinate system	☐	☒	☒	☒	☒	☒
instrumentation used for measurements	☐	☒	☒	☒	☒	☒
pollutant chemistry involved	☐	☒	☒	☒	☒	☒
data quality						
spatial and temporal resolution of the test results	☐	☒	☒	☒	☒	☒
post-processing routines applied to raw measurement data	☐	?	?	?		?
expected/estimated variablility of the results	☐	y/n	y/n	y/n	y/n	y/n
confidence of calculated statistical results	☐	y/n	y/n	y/n	y/n	y/n
previous use of data for validation purposes / experiences						
reference list (if applicable)	☐					
Geometry of the 'test site'						
sufficient/detailed information on						
buildings/obstacles						
outer dimensions & location	☐	☒	☒	☒	?	☒
roof shape/type	☐	☒	☒	☒	☒	☒
surface characterization	☐	☒	☒	☒	☒	☒
topography						
height profile(s) / height contours of the ground	☐	☒	☒	☒	☒	☒
surroundings						
geometrical characterization of the structures surrounding the test site / model area	☐	☒	☒	☒	☒	☒
accuracy/reliability of geometry data						
accuracy of building dimensions	☐	☒	☒	☒	☒	☒
accuracy of topography data	☐	☒	☒	☒	☒	☒
statement on sufficiency of the documented model area	☐	☒	☒	☒	☒	☒
Emission source(s) specification						
sufficient/detailed information on						
emission source(s)						
temporal and spatial classification of tracer/pollutant source(s)	☐	☒	☒	☒	☒	☒
number/location of sources	☐	☒	☒	☒	☒	☒
emission data (source flow rate, duration of release(s) etc.)	☐	☒	☒	☒	☒	☒
emission data type (statistically modeled, numerically modelled, measured, etc.)	☐	☒	☒	☒	☒	☒
if required - post-processing of emission data (emissions from car traffic)						
composition of car fleet	☐				?	☒
specification of driving patterns (cold start fraction, traffic lights, traffic jams etc.)	☐				?	☒
recommended/site-specific emission model	☐				☒	☒
accuracy of source specification						
estimated/determined accuracy of emission data	☐	☒	☒	☒	☒	☒
confidence level of emission data	☐	☒	☒	☒	☒	☒

Table 6.2a: Potential Validation Data – Part 1/2

	Data Set Check List	JU2003	BUBBLE	MUST	DAPPLE	VALIUM
Physical Requirements						
boundary conditions						
inflow conditions						
representative mean vertical and lateral approach flow profile(s)						
wind speed (component resolved)	☐	☒	☒	☒	?	☒
wind direction	☐	☒	☒	☒	?	☒
turbulence intensities	☐	☒	☒	☒		☒
temperature/radiation	☐	☒	☒	☒	☒	☒
turbulent fluxes	☐	☒	☒	☒	?	☒
pollutant/tracer concentrations (background concentrations)	☐	☒	☒	☒	?	☒
additional parameters of the atmospheric boundary layer						
mixing height	☐	☒	☒	☒	?	☒
turbulence length scales (macro-scales)	☐	☒	☒	☒	?	☒
representative time series of						
wind velocity components	☐	☒	☒	☒	?	☒
turbulent fluxes	☐	☒	☒	☒	?	☒
pollutant/tracer concentrations (background time series)	☐	☒		☒	?	
test data						
flow field						
statistically representative wind flow data at a sufficent number of locations	☐	☒	☒	☒	?	☒
time resolved / time dependent flow measurements	☐	☒	☒	☒	?	☒
concentration field						
statistically representative concentration data	☐	☒	☒	☒	?	☒
time resolved / tim dependent concentration measurements	☐	☒		☒	?	☒
source of data / data provider / contact person	☐	☒	☒	☒	☒	☒

Table 6.2b: Potential Validation Data – Part 2/2

7 Conclusions and Recommendations

It was concluded that there was a need for the development of a structured and documented protocol for the determination and communication of the quality of models used for microscale meteorological modelling, and associated pollutant dispersion modelling, for use in urban areas. The protocol has several distinct elements:

- A Scientific Evaluation Process
- A Verification Process that addressed both the code and the solution procedure
- The provision of appropriate and quality assured Validation Data Sets; in particular it was preferred that both field data and associated physical modelling data were utilised
- A Model Validation Process in which model results are compared with the experimental data sets.
- An Operational Evaluation Process that reflects the needs and responsibilities of the Model User

It was recommended that such a protocol be developed and documented in a short and clear format describing the actions required.

The protocol should then be tested, and if necessary, revised by the COST 732 community.

Simultaneously an implementation strategy should be developed to provide for wider dissemination throughout Europe and possibly further afield.

8 References

Adrian, R. J., Meneveau, C., Moser, R.D., and Riley, J.J. (1999) Workshop ‘Turbulence Measurements for LES’, Final Report, available on the World Wide Web at: http://davinci.tam.uiuc.edu/data/moser/les/workshop.pdf

AIAA (1998) Guide for the Verification and Validation of Computational Fluid Dynamics Simulations”, American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA.

Aldama, A. A. (1990) Filtering Techniques for Turbulent Flow Simulation, Springer Verlag, Berlin Heidelberg, New York.

Allwine K. J. (2004) Overview of JOINT URBAN 2003 - an atmospheric dispersion study in Oklahoma City, Proceedings of the AMS Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone, January 11-15, Seattle, Washington, USA.

Arnold, S. et al, (2004) Introduction to the DAPPLE Air Pollution Project. Science of the Total Environment. 332, p139-153.

Baetke, F. and Werner, H. (1990) Numerical Simulation of Turbulent Flow over Surface Mounted Obstacles with Sharp Edges and Corners. Journal of Wind Engineering and Industrial Aerodynamics, Vol. 35, 129-147.

Baklanov, A., Mestayer, P., Clappier, A., Zilitinkevich, S., Joffre, S., Mahura, A. (2005) On the parameterisation of the urban atmospheric sublayer in meteorological models. Atmos. Chem. Phys., 5, 12119–12176.

Bentham J.T. and Britter R.E. (2003) Spatially averaged flow within obstacle arrays. Atm. Env. 37, 2037-2043

Bartzis, J.G., Vlachogiannis, D. and Sfetsos, A., (2004) Thematic area 5: Best practice advice for environmental flows. The QNET-CFD Network Newsletter, Vol. 2, No. 4, pp. 34-39.

Berkowicz R. (2000) OSPM- a parameterised street pollution model. J. Env. Mon. Assess. , 65,323-331

Bezzo, F., Macchietto, S., Pantelides, C. C. (2000) A General Framework for the Integration of Computational Fluid Dynamics and Process Simulation. Comp. & Chem. Eng., 24: 653-658.

Biltoft CA (2001) Customer Report for Mock Urban Setting Test, DPG Document No. WDTC-FR-01-121, West Desert Test Center, U.S. Army Dugway Proving Ground, Dugway, Utah, 58pp.

Blocken, B., Roels, S. and Carmeliet, J.(2004) Modification of pedestrian wind comfort in the Silvertop Tower passages by an automatic control system, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 92, pp. 849-873.

Blocken, B., Roels, S. and Carmeliet, J. (2003) Pedestrian wind conditions in passages through buildings – Part 1. Numerical modelling, sensitivity analysis and experimental verification, Research report, Laboratory of Building Physics, Catholic University of Leuven, 99p.

Borrego, C.; Monteiro, A.; Ferreira, J.; Miranda, A.I.; Costa, A.M.; Sousa, M. (2005) Modelling uncertainty estimation procedures for air quality assessment. In 3rd International Symposium on Air Quality Management at Urban, Regional and Global Scales (AQM), 26-30 September 2005; Istanbul, Turkey - Proceedings of the 3rd International Symposium on Air Quality Management at Urban, Regional and Global Scales. Eds. S. Topçu, M.F. Yardim, A. Bayram, T. Elbir and C. Kahya, Vol. I, pp. 210-219.

Borrego, C., Schatzmann, M. and Galmarini, S. (2003) Quality Assurance of Air Pollution Models. In: Moussiopoulos, N. (Ed.): Air Quality in Cities. Springer-Verlag, ISBN 3-540-00842-x.

Boucouvala, D., Bornstein,R., Wilkinson, J., Miller, D. (2003) MM5 Simulations of a 1997 Southern California Ozone Study (SCOS97) Episode. Special SCOS97 issue of Atmospheric Environment.

Breuer, M., Jovičić, N., Mazaev, K.(2003) Comparison of DES, RANS and LES for the seperated flow around a flat plate at high incidence, Int. J. Numer. Meth. Fluids, Vol. 41, pp. 357-388.

Britter, R.E. (1993) The evaluation of technical models used for major-accident hazard installations. Report to the Commission of the European Communities, DG XII/E-1. Report No. EUR-14774-EN.

Britter R.E. and Hanna S., (2003) Flow and dispersion in urban areas. Ann .Rev. Fl. Mech., 35, 469-496.

Britter, R., and Schatzmann, M. (Eds.) (2007) Model Evaluation Guidance and Protocol Document. COST 732 report. COST Office Brussels, ISBN 3-00-018312-4.

Beychok, M.R. (2005) Fundamentals of Stack Gas Dispersion (Fourth Edition), published by author, Irvine, California, USA.

Cadafalch, J., Pérez-Segarra, C. D., Cònsul, R., Oliva, A. (2002) Verification of Finite Volume Computations on Steady-State Fluid Flow and Heat Transfer, ASME Journal of Fluids Engineering, 124:11-21.

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., Hall, R. (2001): The SMEDIS Database and Validation Exercise. Int. Journal of Environment and Pollution, 16, pp. 614-629.

Casey, M. and Wintergerste, T., (Eds.) (2000) ERCOFTAC SIG "Quality and Trust in Industrial CFD": Best Practice Guidelines.

Castro, I.P (2003): CFD for external aerodynamics in the built environment, The QNET-CFD Network Newsletter, Vol. 2, No. 2, 4-7. See also http://www.qnet-cfd.net.

Celik, I., Cehreli, Z. N. and Yavuz, I. (2005) Index of resolution quality for Large Eddy Simulations, Journal of Fluids Engineering, Vol. 127, pp. 949-958.

Chang, J.C. and Hanna, S.R. (2004) Air quality model performance evaluation. Meteo. Atmos. Phys. 87, 167-196.

Chang, J.C and Hanna, S.R. (2005) Technical Descriptions and User's Guide for the BOOT Statistical Model evaluation Software Package. Available through www.harmo.org/kit

CMAQ (1999) Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modelling System, Edited by Byun, D.W. and Ching, J.K.S., EPA/600/R-99/030, Washington, DC.

Cionco, R. M. (1965) A mathematical model for flow in a vegetative canopy. J. App. Met., 4, 517-522

Coceal O. and Belcher S.E. (2004) A canopy model of mean winds through an urban area. Q.J.R. Met.Soc. ,130, 1349-1372

Coleman, H. W. and Stern, F. (1997) Uncertainties and CFD Code Validation, Journal of Fluids Engineering, Vol. 119, pp. 795-803.

Coleman, H., Stern, F., Di Mascio, A., and Campagna, E. (2001) The Problem with Oscillatory Behaviour in Grid Convergence Studies, ASME Journal of Fluids Engineering, 123:438-439.

Cowan, I.R., Castro, I.P., and Robins, A.G. (1997) Numerical considerations for simulations of flow and dispersion around buildings, Journal of Wind Engineering and Industrial Aerodynamics, 67 & 68, 535-545.

Daish N.C., Britter R.E., Linden P.F., Jagger S.F., Carissimo B. (2000) SMEDIS: Scientific model evaluation of dense gas dispersion models, International Journal of Environment and Pollution, 14 , 39-51

Di Sabatino, S., Britter, R.E., Carruthers, D.J.Towards a formilised procedure for the evaluation of urban dispersion model. Personal communication, unpublished document.

ERCOFTAC (2000) Quality and Trust in Industrial CFD: Best Practice Guidelines. Casey, M. and Wintergerste, T. (Eds.).

Feddersen B, Leitl B, Rotach MW, Schatzmann M (2003) Wind tunnel investigation of the spatial variability of turbulence characteristics in the urban area of Basel City, Switzerland. Proceedings PHYSMOD2003, September 3–5, 2003, Prato, Italy, Firenze University Press: 23–25

Feddersen B, Leitl B, Rotach MW, Schatzmann M (2004) Wind tunnel modeling of urban turbulence and dispersion over the City of Basel (Switzerland) within the BUBBLE project. Fifth Symposium on the Urban Environment (AMS), Vancouver, Canada, 23–28 August 2004

Ferziger, J.H. and Perić, M. (2002) Computational Methods for Fluid Dynamics, Springer-Verlag Berlin, Heidelberg, New York, 3rd edition.

Fothergill, C.E., Roberts, P.T. and Packwood, A.R.(2002) Flow and dispersion around storage tanks - A comparison between numerical and wind tunnel simulations. Wind & Structures, Vol. 5, No. 2-4, 89-100.

Franke, J., C. Hirsch, A.G. Jensen, H.W. Krüs, M. Schatzmann, P.S. Westbury, S.D. Miles, J.A. Wisse, and N.G. Wright. Recommendations on the Use of CFD in Wind Engineering (2004). In J.P.A.J van Beeck (Ed.), Proceedings of the International Conference on Urban Wind Engineering and Building Aerodynamics: COST Action C14 - Impact of Wind and Storm on City Life and Built Environment, May 5 - 7, pages C.1.1 - C.1.11, Rhode-Saint-Genèse, Belgium, ISBN 2-930389-11-7.

Franke, J., and Frank, W. (2005) Numerical simulation of the flow across an asymmetric street intersection. In J. Náprstek and M. Pirner (Eds.) Proceedings of the 4th European-African Conference on Wind Engineering (EACWE4), 11 - 15 July, Czech Technical University, Prague, Paper #138.

Franke, J., Hellsten, A. Schlunzen, H., Carissimo, B. (Eds.) (2007) Best Practice Guideline for the CFD Simulation of Flows in the Urban Environment. COST 732 report. COST Office Brussels, ISBN 3-00-018312-4.

Geurts, B. J. and Fröhlich, J. (2002) A framework for predicting accuracy limitations in large-eddy simulations, Physics of Fluids, Vol. 14, pp. L41-L44

Gryning, S.-E., and Batchvarova, E., (2005) Strengths/weaknesses of laboratory/field and numerical data. In: Schatzmann, M., and Britter, R. (2005) (Eds.): Quality assurance of microscale meteorological models. COST 732 report. European Science Foundation, ISBN 3-00-018312-4.

Hall, R.C. (Ed.) (1997) Evaluation of modelling uncertainty. CFD modelling of near-field atmospheric dispersion. Project EMU final report, European Commission Directorate–General XII Science, Research and Development Contract EV5V-CT94-0531, WS Atkins Consultants Ltd., Surrey.

Hanjalić, K.. (1999) Second-Moment Closures for CFD: Needs and Prospects, International Journal of Computational Fluid Dynamics, Vol. 12, pp. 67-97.

Hanna S.R., Chang J., Britter R., and Neophytou, M., (2003) Overview of Model Evaluation History in the Atmospheric Air Quality Area. QNET-CFD Network Newsletter, 2, 1-5

Hanna S.R., Britter R.E. and Franzese P. (2003) A baseline urban dispersion model evaluated with Salt lake City and los Angeles tracer data. Atm. Env, 37, 5069-5082

Hanna, S., Fabian, P., Chang, J., Venkatram, A., Britter, R.E., Neophytou, M., Brook, D. (2004) Use of Urban 2000 field data to determine whether there are significant differences between the performance measures of several urban dispersion models. Paper 7.3, Fifth Conference on Urban Environment, American Meteorological Society, Vancouver.

Hanna S.R., Brown M.J., Camalli F.E, Chan S., Courier W.J., Huber A.H., Kim S. and Reynolds R.M. (2006) Detailed simulations of atmospheric flow and dispersion in downtown Manhattan: An application of five computational fluid dynamics models. Bulletin of American Meteorological Society, 87, 1713-1726.

Hirsch C., Bouffioux, V. and Wilquem F (2002) CFD simulation of the impact of new buildings on wind comfort in an urban area. In: G. Augusti, C. Borri, and C. Sacré (Eds.) Impact of Wind and Storm on City Life and Built Environment, Proceedings of the Workshop, pages 164-171, CSTB, Nantes.

Hoffman, R.N., Liu, Z., Louis, J.F., and Grassotti, C. (1995) Distortion representation of forecast errors. Monthly Weather Review, 123, 2758-2770.

Hutton, A.G. (2005) Evaluation of CFD codes – The European Project, QNET-CFD. In: Schatzmann, M., and Britter, R. (2005) (Eds.): Quality assurance of microscale meteorological models. COST 732 report. European Science Foundation, ISBN 3-00-018312-4.

Isukapalli, S.S. (1999) Uncertainty Analysis of Transport-Transformation Models, PhD Thesis, The State University of New Jersey, http://www.ccl.rutgers.edu/~ssi/thesis/ .

Irwin J.S., (2000): Modeling Air Quality Pollutant Impacts. EURASAP Newsletter, ISSN-1026-2172, 40. http://www.meteo.bg/EURASAP/Newsletter.html

Kastner-Klein, P., Leitl, B., Pascheke, F., and Schatzmann, M. (2004) Wind tunnel simulation of the Joint Urban 2003 Tracer Experiment. 84th AMS Annual Meeting, Seattle, USA, June 11-15.

Klein, P., Leitl, B., Schatzmann, M. (2007) Driving Physical Mechanisms of Flow and Dispersion in Urban Canopies. International Journal of Climatology (accepted).

Klein, M. (2005) An attempt to assess the quality of Large Eddy Simulations in the context of implicit filtering, Flow, Turbulence and Combustion, Vol. 75, pp. 131-147.

Knupp, P., and Salari, K., (2003) Verification of Computer Codes in Computational Science and Engineering, Chapman and Hall/CRC, Boca Raton.

Leitl, B. (2000): Validation Data for Microscale Dispersion Modelling. EUROTRAC Newsletter, 22, pp. 28-32.

Leitl B., Kastner-Klein, P., and Schatzmann, M. (2004) Wind Tunnel Modeling of Complex Flow and Dispersion Phenomena in support of the Joint Urban Atmospheric Dispersion Study 2003. Fifth Symposium on the Urban Environment (AMS), Vancouver, Canada, August 23-28.

Leitl, B., Schatzmann, M., and Kastner-Klein, P. (2005) Flow and Dispersion in Complex Urban Areas; Wind Tunnel Modeling in Support of Joint Urban 2003. Proceedings International Workshop on Physical Modelling of Flow and Dispersion Phenomena, London, Ontario.

Lohmeyer, A., Müller, W.J., and Bächlin, W. (2002) A comparison of street canyon concentration predictions by different modellers. Final results from the Podbi-Exercise. Atmospheric Environment, 36, pp. 157-158.

Martilli A., Clappier A. and Rotach M. (2002) An urban surface parameterisation for mesoscale models. Boundary Layer Meteorology, 164, 261-304

Menter, F., B. Hemstrom, M. Henrikkson, R. Karlsson, A. Latrobe, A. Martin, P. Muhlbauer, M. Scheuerer, B. Smith, T. Takacs, and S. Willemsen (2002) CFD Best Practice Guidelines for CFD Code Validation for Reactor-Safety Applications, Report EVOL-ECORA-D01, Contract No. FIKS-CT-2001-00154 , 46 pp.

Miles, S. and Westbury, P. (2003) Practical Tools for Wind Engineering in the Built Environment, The QNET-CFD Network Newsletter, Vol. 2, No. 2 , pages 11-14.

Mochida, A., Tominaga, Y., Murakami, S., Yoshie, R., Ishihara, T. and R. Ooka (2002) Comparison of various k-epsilon models and DSM to flow around a high rise building - report of AIJ cooperative project for CFD prediction of wind environment, Wind & Structures, Vol. 5, No. 2-4, 227-244.

Model Evaluation Group (1994) Guideline for model developers and Model Evaluation Protocol. European Community, DG XII, Major Technological Hazards Programme, Brussels, Belgium

Morgan M.J. and Henrion, M. (1990) A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. New York, Cambridge Univ. Press.

Moussiopoulos, N. (Ed.) (2003): Air Quality in Cities, Springer-Verlag, ISBN 3-540-00842-x.

Murphy, A.H., (1995) The Finlay Affair: A Signal Event in the History of Forecast Verification, Weather and Forecasting, 11/1, pp. 3-20.

Oberkampf, W. L., Trucano, T. G., and Hirsch, C. (2004) Verification, validation, and predictive capability in computational engineering and physics. Appl. Mech. Rev, 57(5):345 - 384.

Olesen, H.R. (1995) Data sets and protocol for model validation. Workshop on Operational Short-range Atmospheric Dispersion Models for Environmental Impact Assessment in Europe, Mol, Belgium, Nov. 1994, Int. J. Environment and Pollution, Vol. 5, Nos. 4-6, 693-701.

Olesen, H.R. (2005) User's Guide to the Model Validation Kit. National Environmental Research Institute, Denmark. 72pp. Research Notes from NERI no. 226. Available through http://research-notes.dmu.dk and www.harmo.org/kit.

Peltier L. J., Haupt, S.E., Wyngaard, J.C., Stauffer, D., Deng, A., Kredensor, F. (2006) Meteorological uncertainty effects in atmospheric transport and dispersion modelling: A demonstration. AMS Forum: Environmental Risk and Impacts on Society: Successes and Challenges, part of the AMS Annual Meeting, Atlanta, GA, 29 Jan- 2 Feb, (extended abstracts available from the web at: http://ams.confex.com/ams/Annual2006/techprogram/meeting_Annual2006.htm)

QNET-CFD Network newsletter (2003) Overview of Model Evaluation History and Procedures in the Atmospheric Air Quality Area. Volume 2, No. 5.

Richards, P.J. and Hoxey, R.P. (1993) Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 46 & 47, 145-153.

Richardson, L.F. (1910) The approximate arithmetical solution by finite differences of physical problems involving differential equations with an application to the stresses in a masonry dam, Trans. Royal Society London, Ser. A, Vol. 210, pp. 307-357.

Richardson, L.F (1927) The deferred approach to the limit, Trans. Royal Society London, Ser. A, Vol. 226, pp. 229-361.

Roache, P. J. (1994) Perspective: a method for uniform reporting of grid refinement studies, 116:405-413.

Roache, P. J. (1998) Verification and Validation in Computational Science and Engineering, Hermosa Publishers, New Mexico.

Roache, P. J. (2002) Code verification by the method of manufactured solutions, ASME Journal of Fluids Engineering, 124:4-10.

Roache, P. J. (2003) Criticisms of the “Correction Factor” Verification Method, ASME Journal of Fluids Engineering, 125:732-733.

Roache P. J., (1997) Quantification of uncertainty in Computational Fluid Dynamics, Ann. Rev. Fluid. Mech, 29: 123-160.

Rotach, M., Fisher, B., Piringer, M (Eds.) (2003) COST715 Workshop on urban boundary layer parameterisations, Report EUR 20355, ISBN: 92-894-4143-7.

Rotach, M., Vogt, R., Bernhofer, C., Batchvarova, E., Christen, A., Clappier, A., Feddersen, B., Gryning, S.E., Martucci, G, Mayer, H., Mitev, V., Oke, T.R., Parlow, E., Richner, H., Roth, M., Roulet, Y.A., Ruffieux, D., Salmond, J.A., Schatzmann, M., Voogt, J.A. (2005) BUBBLE- An urban boundary layer meteorology project. Theoretical and Applied Climatology, 81, pp. 231-261.

Roy, C. J. (2005) Review of code and solution verification procedures for computational simulation. Journal of Computational Physics, 205:131-156.

Royal Meteorological Society (1995) Policy Statement of the Royal Meteorological Society. Atmospheric Dispersion Modelling: Guidelines on the justification of choice and use of models, and the communication and reporting of results. Royal Meteorological Society.

Rubbert P. E (1997) The use of CFD in Airplane Design, CFD 97 Fifth Annual Conference of the CFD Society of Canada, Victoria, B.C.

Saltelli A. et al., (2000) Sensitivity Analysis. John Wiley & Sons. Chichester, New York

Scaperdas, A. and Gilham, S. (2004) Thematic Area 4: Best practice advice for civil construction and HVAC, The QNET-CFD Network Newsletter, Vol. 2, No. 4, pages 28-33.

Schatzmann M., Leitl B. (2002) Validation and Application of Obstacle Resolving Urban Dispersion Models. Atmospheric Environment 36, 4811-4821.

Schatzmann, M., Grawe, D., Leitl, B., and Müller, W.J. (2003) Data from an urban street monitoring station and its application in model validation. Proceedings, 26th International Technical Meeting on Air Pollution Modelling and its Application. Istanbul, Turkey, May 26-30.

Schatzmann, M., and Britter, R. (2005) (Eds.): Quality assurance of microscale meteorological models. COST 732 report. European Science Foundation, ISBN 3-00-018312-4.

Schatzmann, M., Bächlin, W., Emeis, S., Kühlwein, J., Leitl, B., Müller, W.J., Schäfer, K., Schlünzen, H. (2006) Development and Validation of Tools for the Implementation of European Air Quality Policy in Germany (Project VALIUM). Atmospheric Chemistry and Physics, 6, 3077-3083.

Schlesinger, S. (1979) Terminology for Model Credibility. Simulation, 32(3): 103-104

Schlünzen K.H. (1997): On the validation of high-resolution atmospheric mesoscale models , J. Wind Engineering and Industrial Aerodynamics, 67 & 68 , 479-492.

Seaman, N. L., Stauffer, D., Lario, A. (1995) A multiscale four-dimensional data assimilation system applied in the San Joaquin Valley during SARMAP. Part I: Modeling design and basic performance characteristics. J. Appl. Meteor., 34, 1739-1761.

Spalart, P., and Allmaras, S. (1992) A one-equation turbulence model for aerodynamic flows. Technical Report AIAA-92-0439, American Institute of Aeronautics and Astronautics.

Spalart, P., Jou, W.-H., Strelets, M., Allmaras, S. (1997) Comments on the feasibilty of LES for wings and on the hybrid RANS/LES approach, Advances in DNS/LES, 1st AFOSR Int. Conf. on DNS/LES, Greden Press, 1997.

Stern, F., Wilson, R. V., Coleman, H. W., and Paterson, E. G. (2001) Comprehensive Approach to Verification and Validation of CFD Simulations - Part 1: Methodology and Procedures. ASME Journal of Fluids Engineering, 123:793-802.

Stern, F., Wilson, R.V., Coleman, H., Paterson, E. (1999) Verification and Validation of CFD Simulations, 3th ASME/JSME Joint Fluids Engineering Conference, San Francisco, CA, 18-23 July.

Stern, J.; Flemming, R., (2004) Formulation of criteria to be used for the determination of the accuracy of model calculations according to the requirements of the EU Directives for air quality – Examples using the chemical transport model REM-CALGRID, Freie Universität Berlin, Institut für Meteorologie

Tominaga, Y., Mochida, A., Shirasawa, T., Yoshie, R., Kataoka, H., Harimot, K. and Nozu, T. (2004) "Cross Comparisons of CFD Results of Wind Environment at Pedestrian Level around a High-rise Building and within a Building Complex", Journal of Asian Architecture and Building Engineering, Vol. 3, No. 1.

VDI (2005): Umweltmeteorologie - Prognostische mikroskalige Windfeldmodelle - Evaluierung für Gebäude- und Hindernisumströmun. VDI 3783 Blatt 9. Available in English as VDI Guideline on Environmental meteorology - Prognostic microscale windfield models - Evaluation for flow around buildings and obstacles.

Wilks, D.S. (2006) Statistical Methods in the Atmospheric Sciences, 2nd edition, Vol. 91, International Geophysics Series, 630 pp, Elsevier Inc.

Wilson, R., Shao, J., and Stern, F. (2004) Discussion: Criticisms of the "Correction Factor" Verification Method (Roache, 2003), ASME Journal of Fluids Engineering, 126:704-706.

Wright, N.G. and Easom, G.J. (1999) Comparsion of several computational turbulence models with full-scale measurements of flow around a building. Wind and Structures, Vol. 2, No. 4, pp. 305-323.

Yoshie, R., A. Mochida, Y. Tominaga, H. Kataoka, K. Harimoto, T. Nozu, T. (2005) Shirasawa.Cooperative project for CFD prediction of pedestrian wind environment in the Architectural Institute of Japan. In J. Náprstek and M. Pirner (Eds.) Proceedings of the 4th European-African Conference on Wind Engineering (EACWE4), 11 - 15 July, Prague, Paper #292.

Zhang, D.-L., H.-R. Chang, N.G. Seaman, T.T. Warner, J.M. Fritsch (1986) A two-way interactive nesting procedure with variable terrain resolution. Mon. Wea. Rev., 114: 1.

Zitney, S. E., Syamlal, M. (2002) Integrated Process Simulation and CFD for Improved Process Engineering. In Proc. of the European Symposium on Computer Aided Process Engineering *12, ESCAPE-12. (J. Grievink and J. van Schijndel, Eds.) The Hague, The Netherlands: 397-402.

Appendix Glossary of terms, List of acronyms and abbreviations

A. Glossary of Terms

Accuracy		Measure of correctness	Oberkampf et al. (2002)
Error		Recognisable deficiency in any phase or activity of modelling and simulation that is not due to lack of knowledge	AIAA (1998)
	Acknowledged error	Knowledge of divergence from an approach or ideal condition that is considered to be a baseline for accuracy, e.g. - finite precision arithmetic - conversion of PDEs in discrete equations => *solution verification*	Oberkampf et al. (2002)
	Unacknowledged error	Blunders or mistakes, e.g. - programming errors - input data errors - compiler errors => *code verification*, SQA	Oberkampf et al. (2002)
Evaluation (scientific evaluation)		Examination of a model consisting of three parts: - *scientific review* or *assessment* - *verification* - *validation* Documented in the evaluation protocol.	Britter (1993), Royal Meteorological Society (1995), Daish et al. (2000)
Model	Conceptual model	Produced by analysing and observing the physical system. In Computational Physics dominated by the PDEs for conservation of mass, momentum and energy. Also includes auxiliary equations like: - turbulence models - constitutive equations - initial and boundary conditions	Oberkampf et al. (2002)
	Computerised model (computer or computational model, code)	Operational computer program that implements a *conceptual model*	Oberkampf et al. (2002)

Prediction		Use of a computational model to foretell the state of a physical system under conditions for which the computational model has not been validated	AIAA (1998)
Qualification		Determination of adequacy of the conceptual model to provide an acceptable level of agreement for the domain of intended application	Schlesinger (1979)
Scientific review or assessment		Examination of the models' scientific basis, e.g. - physical processes included, - how they are modelled, - approximations employed, - solution techniques, and the user-oriented features, e.g. - user interfaces - resources required	Britter (1993), Royal Meteorological Society (1995), Daish et al. (2000)
Uncertainty		Recognisable deficiency in any phase or activity of modelling and simulation that is not due to lack of knowledge	AIAA (1998)
	Aleatory uncertainty (variability, irreducible or inherent or stochastic or random uncertainty)	Inherent variation associated with the physical system or environment being considered. Mathematical representation as probability distribution	Oberkampf et al. (2002)
	Epistemic uncertainty (reducible, subjective, model form uncertainty)	Cause of nondeterministic behaviour derived from some level of ignorance or lack of knowledge about the system or the environment. Sources: - limited understanding of complex physical processes - insufficient knowledge concerning initial and boundary conditions => can be reduced by an increase in knowledge	Oberkampf et al. (2002)

Validation		Process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended use of the model	Britter (1993), Royal Meteorological Society (1995), AIAA (1998)
Verification		Process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model	Britter (1993), Royal Meteorological Society (1995), AIAA (1998)
	Code verification (numerical algorithm verification)	Addresses the software reliability of the implementation of all of the numerical algorithms that affect the numerical accuracy and efficiency	Oberkampf et al. (2002)
	Solution verification (numerical error estimation)	Deals with the quantitative estimation of the numerical accuracy of a given solution to the PDEs	Oberkampf et al. (2002)

B. Acronyms and Abbreviations

Acronyms and abbreviations	**Meaning**	**Link / Reference**
ACCENT	Atmospheric Composition Change. European Network of Excellence funded by EC - FP6	http://www.accent-network.org/
AEROHEX	Advanced Exhaust Gas Recuperator Technology for Aero-Engine Applications, funded by CEC, Growth Programme (2000-2003)	http://aix.meng.auth.gr/lhtee/projects/aerohex/aerohex.html
AIAA	American Institute of Aerodynamics and Astronautics	
ATREUS	Advanced Tools for Rational Energy Use towards Sustainability with emphasis on microclimatic issues in urban applications, funded by EC - FP5	http://aix.meng.auth.gr/atreus/index.htm

BOOT	Software package for model evaluation	www.harmo.org/kit
BPG	Best practice guideline	http://www.mi.uni-hamburg.de/Home.484.0.html
BUBBLE	The Basel UrBan Boundary Layer Experiment	http://pages.unibas.ch/geo/mcr/Projects/BUBBLE/
CAFÈ	Clean Air for Europe project funded by EC	http://ec.europa.eu/environment/air/cafe/
CEC	Commission for Environmental Cooperation	http://www.cec.org
CEDVAL	Database: Compilation of Experimental Data for Validation of Microscale Dispersion Models	http://www.mi.uni-hamburg.de/CEDVAL_Validation_Data.427.0.html
CFD	Computational Fluid Dynamics	
CFL	Courant-Friedrichs-Lewy number	
CITAIR	Science and research for better air in European Cities	http://www.harmo.org/harmoni/LinksTo.asp
City-Delta	European Modelling Exercise. Project organised by the Joint Research Centre of the EC, in support of the CAFE Programme	http://aqm.jrc.it/citydelta/
CHENSI	Microscale CFD model name	http://www.ec-nantes.fr/DAH/0/fiche_LLMF__structure/
CLEAR	Cluster of European Air Quality Research	http://www.nilu.no/clear/
COST action C14	Impact of Wind and Storms on City Life and Built Environment	http://www.costc14.bham.ac.uk/CostC14/workshop.pdf
COST action 715	Meteorology Applied to Urban Air Pollution Problems	http://www2.dmu.dk/atmosphericenvironment/cost715.htm
DAPPLE	Dispersion of Air Pollution and its Penetration into the Local Environment, funded by the UK EPSRC	http://www.dapple.org.uk/
DNS	Direct Numerical Simulation	
DOAP	Design or assessment parameters	
EC	European Commission	
ECORA	Evaluation of Computational Fluid Dynamic Methods for	http://domino.grs.de/ecora/ecora.nsf

	Reactor Safety Analysis	
EEA	European Environmental Agency	
EMU	Evaluation of modelling uncertainty, funded by EC	Hall, R.C. (ed.), 1997
ENV-e-City	ENvironmentally Viable electronic CITY, EU project	http://www.fit.fraunhofer.de/projekte/envecity
EPA	US - Environmental Protection Agency	http://www.epa.gov/
EPSRC	Engineering and Physical Sciences Research Council. The UK Government's leading funding agency for research and training in engineering and the physical sciences	http://www.epsrc.ac.uk/default.htm
ERCOFTAC	European Research Community On Flow, Turbulence And Combustion. Scientific Evaluation issues are addressed in the various Best Practice Guidelines	http://www.ercoftac.org/
EU	European Union	
EUREKA	A political initiative to encourage the trans-national development of technological research and development in Europe	http://www.eureka.be/about.do
EUROAIRNET	European Air Quality Monitoring Network	http://air-climate.eionet.europa.eu/databases/EuroAirnet/index_html
EUROTRAC-2	Transport and Chemical Transformation of Environmentally Relevant Trace Constituents in the Troposphere over Europe. EUREKA propject	http://www.gsf.de/eurotrac/
FAC2	Fraction of predictions within a factor of 2	
FB	Fractional bias	
FoM	Figure of merit	
FP5	Fifth Framework Programme of EC	
FP6	Sixth Framework	

	Programme of EC	
FUMAPEX	Integrated systems for Forecasting Urban Meteorology, Air Pollution and Population Exposure, funded by EC - FP5	http://fumapex.dmi.dk/
FWD	The Air Quality Framework Directive – Council Directive 96/62/EC of 27. September 1996 on ambient air quality assessment and management	http://ec.europa.eu/environment/air/ambient.htm
GUI	Graphical User Interface	
INTERGAIRE	Integrated urban Governance and Air Quality Management In Europe, funded by EC - FP5	http://www.nilu.no/clear/proj2.htm
LES	Large Eddy Simulation	
LHTEE	Laboratory of Heat Transfer and Environmental Engineering at the Aristotle University Thessaloniki	http://aix.meng.auth.gr/lhtee/index.html
MERCURE	Microscale model name	http://www.enpc.fr/cerea/en/axes.html
MG	Geometric mean	
MIMO	Microscale model name	http://aix.meng.auth.gr/lhtee/projects/
MISKAM	Microscale model name	http://www.lohmeyer.de/Software/winmiskam.htm
MITRAS	Microscale model name	http://www.mi.uni-hamburg.de/mitras
MMS	Method of Manufactured Solutions	
MoE	Measure of effectivness	
MSE	Model Scientific Evaluation	
NMSE	Normalised mean square error	
NoE	Network of Excellence	
OSCAR	Optimised Expert System for Conducting Environmental Assessment of Urban Road Traffic, funded by EC - FP5	http://www.eu-oscar.org
PICADA	Photocatalytic Innovative Coverings Applications for Depollution Assessment, funded by EC	http://www.picada-project.com/domino/SitePicada/Picada.nsf?OpenDataBase
QNET-CFD	Thematic Network for Quality and Trust in the Industrial Application of CFD	http://www.qnet-cfd.net
RANS	Reynolds Averaged Navier-	

	Stokes	
RME	Relative mean error	
RMS	Root mean square (error)	
RPE	Relative Percentile Error	Stern and Flemming (2004)
RSM	Reynolds Stress Modelling	
PDE	Partial Differential Equation	
SAPPHIRE	Source Apportionment of Airborne Particulate Matter and Polycyclic Aromatic Hydrocarbons in Urban Regions of Europe, funded by EC - FP5	http://www.gees.bham.ac.uk/research/sapphire/
SATURN	Studying Atmospheric Pollution in Urban Areas (EUROTRAC-2 subproject)	http://aix.meng.auth.gr/saturn/
SMC	Second Moment Closure	
SMEDIS	Scientific Model Evaluation of Dense Gas Dispersion Models	Daish N.C. et al, 1999; Di Sabatino S. et al. 1999, Cambridge Environmental Research Consultants Ltd, 2002
SQA	Software quality assurance	
TASCflow	CFD model	http://www.cfx-germany.com/cfxtf.html
TRAPOS	Optimisation of Modelling Methods for Traffic Pollution in Streets, funded by EC – coordinated to EUROTRAC-2	http://www2.dmu.dk/AtmosphericEnvironment/trapos/
URANS	Unsteady RANS	
VADIS	Microscale model name	http://www.dao.ua.pt/gemac/
VALIUM	Development and Validation of Tools for the Implementation Of European Air Quality Policy in Germany - Supported by the German Federal Ministry of Education and Research	http://www.mi.uni-hamburg.de/Valium.46.0.html
VDI	The German Association of Engineers	http://www.vdi.de/vdi/vrp/richtlinien/suche/index.php
VG	Geometric variance	

See also other related acronym lists e.g.:
http://www.gsf.de/eurotrac/what_is/a-cronym.htm

COST is an intergovernmental European framework for international cooperation between nationally funded research activities. COST creates scientific networks and enables scientists to collaborate in a wide spectrum of activities in research and technology. COST activities are administered by the COST Office.
Website: www.cost.esf.org

COST action 732 “Quality Assurance of Microscale Meteorological Models” has been set up to improve and assure the quality of micro-scale meteorological models that are applied for predicting flow and transport processes in urban or industrial environments. In particular it is intended

- to develop a coherent and structured quality assurance procedure for these type of models,
- to provide a systematically compiled set of appropriate and sufficiently detailed data for model validation work in a convenient and generally accessible form,
- to build a consensus within the community of micro-scale 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 contribute to the proper use of models by disseminating information on the range of applicability, the potential and the limitations of such models,
- to identify the current weaknesses of the models and data bases,
- to give recommendations for focussed experimental programmes in order to improve the data base and
- to give recommendations for the improvement of present models and, if necessary, for new model parameterisations or even new model developments.

EUROPEAN SCIENCE FOUNDATION ESF provides the COST Office through an EC contract

COST is supported by the EU RTD Framework programme

Price (excluding VAT) in Germany: € 9.50 **ISBN 3-00-018312-4**