THE MUST MODEL EVALUATION EXERCISE: PATTERNS IN MODEL PERFORMANCE

H. R. Olesen[1], A. Baklanov[2], J. Bartzis[3], F. Barmpas[4], R. Berkowicz[1], K. Brzozowski[5], R. Buccolieri[6], B. Carissimo[7], A. Costa[8], S. Di Sabatino[6], G. Efthimiou[3], J. Franke[9], I. Goricsan[10], A. Hellsten[11], M. Ketzel[1], B. Leitl[12], R. Nuterman[13], E. Polreich[14], J. Santiago[15], R. Tavares[8]

[1]NERI, Univ. of Aarhus, Denmark
[2]DMI, Denmark; [3]Univ. of West Macedonia, Greece; [4]Aristotle Univ., Greece; [5]Univ. of Bielsko-Biala, Poland; [6]Univ. of Salento, Italy; [7]CEREA, France; [8]Univ. of Aveiro, Portugal; [9]Univ. of Siegen, Germany; [10]Budapest Univ. of Technology and Economics, Hungary; [11]Helsinki Univ. of Technology, Finland; [12]Univ. of Hamburg, Germany; [13]Tomsk State Univ., Russia; [14]ZAMG, Austria; [15]CIEMAT, Spain

Abstract: As part of the COST 732 action more than a dozen different research groups have modelled the MUST experiment, as simulated in a wind tunnel. The model evaluation guidance developed within COST 732 recommends 'exploratory data analysis' as one of the elements in model validation. Experience has shown that such exploratory analysis is crucial to reveal shortcomings of models that might otherwise pass unnoticed. Conditions are best for detecting common patterns and anomalies if you have a situation where *several models are put into a common framework* – like the case at hand. The available material provides a unique opportunity to identify and explore patterns within model performance.

***Key words:** COST 732, MUST, CFD, atmospheric dispersion, model evaluation, model validation, exploratory data analysis.*

1. INTRODUCTION

The COST action 732 on *Quality Assurance and Improvement of Micro-Scale Meteorological Models* has fostered a comprehensive modelling exercise. More than a dozen different research groups have modelled the MUST (Mock Urban Setting Test) experiment, as simulated in a wind tunnel. Companion papers describe the model evaluation protocol and the metrics studied, while the current paper focuses on exploratory analysis and on the results of the exercise. The available material provides a unique opportunity to identify and explore patterns within model performance for a multitude of models and setups. A few central findings from the data exploration are reported here. However, due to size limitations the present paper can yield only a tiny glimpse into the universe of information that can be retrieved from the exercise.

The model evaluation guidance developed within the COST 732 action recommends 'exploratory data analysis' as one of the elements in model validation. Experience has shown that such exploratory analysis is crucial to reveal shortcomings of models that might otherwise pass unnoticed. Thus, exploratory analysis is a vital element in quality assurance of models. The following sections will illustrate this by several examples.

The present paper is limited to the study of CFD models. The performance of a number of non-CFD models has also been studied within the COST 732 action, but information on these studies should be sought elsewhere. A comprehensive report which provides a deeper description of model results will become available on the COST 732 web site (URL 1).

2. EXPERIMENTAL SETUP AND TOOLS FOR ANALYSIS

The experimental data for the MUST case allow the study of flow as well as of dispersion. The present paper gives only rudimentary information on the experimental setup and the cases treated. For more details the reader is referred to Leitl et al. (2007). Three cases are simulated by the models: Two representing flow, and one representing dispersion. They are referred to as, respectively, the 0 degree flow case, the -45 degree flow case, and the -45 degree dispersion case. *Figure 1* shows the geometry of the cases.

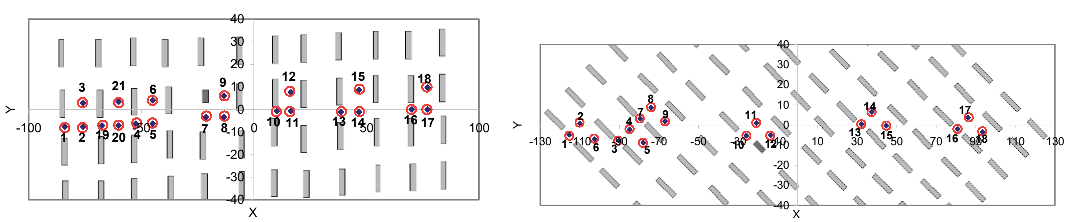

Figure 1. Layout of measurements and the buildings. The vertical profiles of wind and turbulence components are measured at the locations shown. The wind blows from the left. Left: the 0 degree case. Point 3 represents a "Wide Street". Right: the -45 degree case. Point 1 represents a "Narrow Street".

Tools in the form of Excel workbooks have been developed to allow exploratory analyses of model performance. There is a basic group of Excel files that allows *easy graphical inspection of the details of every case for all of the models*. Based on these 'full workbooks', another group of Excel workbooks have been derived, containing essential information extracted from the full workbooks. This extract is presented as summary plots and metrics. The figures here are mainly retrieved from the latter set of workbooks. There are a few additional files, in particular one – referring to the case with dispersion – that uses Gaussian fitting to describe the pollution cloud in terms of Gaussian plume parameters. This makes it easy to obtain an impression of main features of the plume, as modelled and as observed.

All of these files are available on a public web site (URL 1). The publicly available files may differ in some of their contents from the files used for the present study. The files can be used in future by other modelling groups who wish to put their results into the same framework as that used by the COST 732 group. Experiences can be reported on a Wiki (URL 2) - this makes it possible to continue pooling of experiences even after the COST action has finished.

3. RESULTS

Flow modelling, -45 degree case

Figure 1 shows the setup of the measuring positions where vertical profiles of u and w were measured. The locations can be classified into three groups according to the geometry: "Narrow streets" as exemplified by tower 1, "Crossings" exemplified by point 2, and "Wide streets" by tower 6.

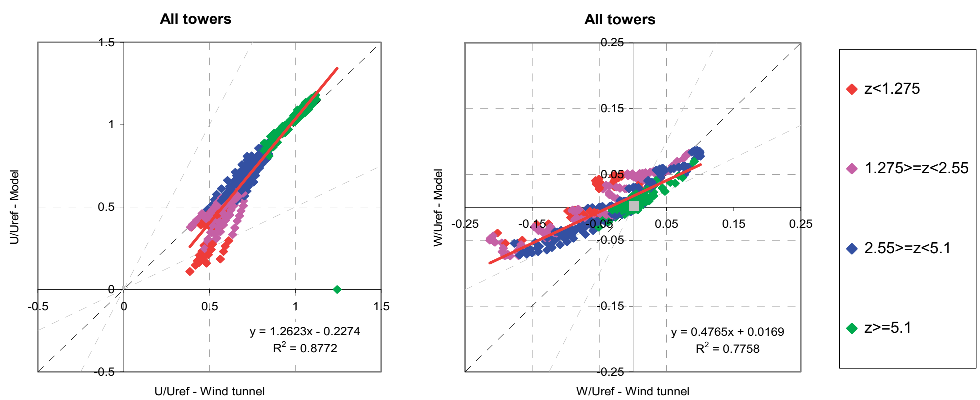

Figure 2. Flow, -45 degree case. Scatter plots for velocity components *u* (left) and *w* (right) for one particular model (Fluent). The points are colour-coded, with red indicating the lowest heights (below 1.275 m).

Figure 2 shows for one model a scatter plot of measured versus modelled values for the horizontal, along-wind velocity component *u*, and the vertical component *w*. All data points (498) in the vertical profiles are included. The buildings generally have a height of 2.5 m (full scale units, as used throughout the present paper). There are measurement positions both below and above roof height. It is important to note that the appearance of scatter plots like Figure 2 will depend very much on the proportion of data at various heights, and on how well the various measuring positions are represented. A similar statement applies to metrics such as Hit Rate. A value of a certain metrics tells nothing unless it appears in a context.

Therefore the results underlying the figure should be further examined. For the u component Figure 3 shows the same data as Figure 2, but split into three classes of data: "Wide streets", "Crossings", "Narrow Streets". The model appears much less successful in performance of u for *Narrow Streets* than for the other locations. Close to the ground the modelled u comes close to zero, whereas this is not the case for the measured values. However, one should also note, that due to the technical set-up of the measurements, only very few measuring points below 1.275 m exist in the case of *Wide Streets* and *Crossings*.

Concerning the vertical velocity *w*, Figure 2 shows that the range of modelled values is smaller than the range of measured values. In particular, predicted downward velocities are numerically much smaller than measured. The available material makes it easy to determine whether this behaviour of w is peculiar to the model examined or whether it is a common feature for many models.

Figure 4 shows the behaviour of w for a variety of models and model setups, performed by 16 different modelling groups. There are 3 sets of Miskam runs, 5 of Fluent, and a variety of other models. The plots are shown anonymously in order to avoid misleading interpretation of the results, which should be accompanied by a more thorough explanation than can be given here. Certain characteristic patterns can be detected, but it is also evident that for some models the pattern deviates from the general. Such deviations deserve exploration. During the process of the MUST exercise within COST 732, it was often found that deviations were a symptom of trivial errors – misplaced buildings, data extraction errors, interpolation procedures etc. But besides highlighting more or less trivial errors, the visualisation of performance patterns can also reveal more subtle questions that deserve to be pursued.

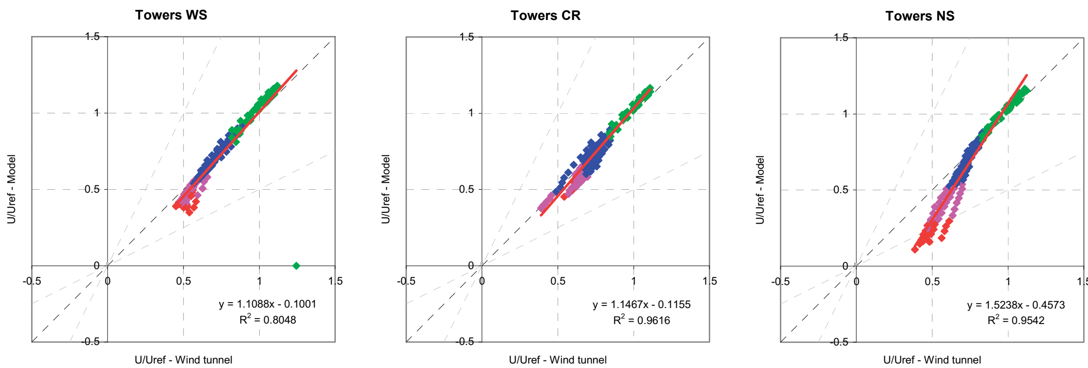

Figure 3. Flow, -45 degree case. Same data as Figure 2 for the *u* component, but split into 3 classes of data: "Wide streets" (WS), "Crossings" (CR), "Narrow Streets" (NS).

Figure 4 reveals that it is a common feature for nearly all models that for downward vertical velocities, the numerical value of w is underpredicted. We may further note - although this is not illustrated - that the behaviour of w is quite distinct for the three groups of measurement positions.

The models' ability to predict turbulent kinetic energy is in general poor.

Flow, 0 degree case

Figure 1 shows the setup for the 0 degree flow case. Patterns of model behaviour in terms of prediction of velocity components and turbulent kinetic energy can be studied in a manner similar to the example discussed for the -45 degree case. One general conclusion on the state of the art is that all models have some difficulty in predicting u for *Wide Streets* – they generally predict too low a wind speed close to the roof top and slightly above it. Otherwise, prediction of *u* is generally good. Concerning w, the range of predicted w values is smaller than the range observed – similarly to the -45 degree case.

Dispersion, -45 degree case

Figure 5a shows the setup of the -45 degree dispersion case. In the wind tunnel the source was placed at ground level, while the concentration measurements were performed at a height of 1.275 m. Although the net of monitors at first sight looks dense, the sparseness of monitors is a problem when results are interpreted.

Figure 5b illustrates the difficulties in comparing measurements and model results. It shows a cross-section relatively close to the source (18 m from it). The modelled values are zero within a certain interval where there is a building. It is seen that the measurements lack a sufficiently detailed resolution to properly evaluate model performance. The presence of buildings complicates matters.

Although difficult to interpret, point-by-point comparisons do carry some information. For each measurement point, the measured value is compared to the modelled value at the closest point in the modelling grid. This will normally be a point outside buildings, unless the grid has been chosen with too poor resolution. However, for some of the models the data were provided interpolated exactly to the measuring points; in those cases interpolated values were used. Due to the discontinuity at the building walls (see e.g. Figure 5b), such interpolation is dangerous and requires care if it shall not produce misleading values. Figure 6 shows a rather typical scatter plot of modelled versus measured concentrations for all measuring points (256). Despite the obvious mispredictions for some high concentrations (close to the source), the Hit Rate is as high as 0.74 because there are many points with low concentration values where the model fits fairly well (further away from the source).

Figure 7 is a map showing the plume trajectory according to measurements and 7 models. The trajectory has been determined by defining a series of cross-sections and fitting a Gaussian curve to each of these. Points within buildings were excluded from the fitting. This method was applied for both measurements and model results. A Gaussian fit can describe only some main features of the pollution cloud, but it does provide useful information.

It is seen that most models predict the plume trajectory well. One model (A) has a severe problem. Another (B) has a problem in the first part of the trajectory, close to the release point. Inspection of individual cross-sections confirms that this is not an artefact of the fitting procedure. It is interesting to note that in terms of Hit Rate, model B has the best score of all the 25 available model runs. This is so because the Hit Rate is based on 256 data pairs, while the displaced part of plume only affects around 20 of these.

This fact illustrates the point that *metrics alone do not assure quality and is no substitute for looking at data*. The Model A is the only model based on a diagnostic flow field model, which does not properly predict the deflection of the plume trajectory by the buildings.

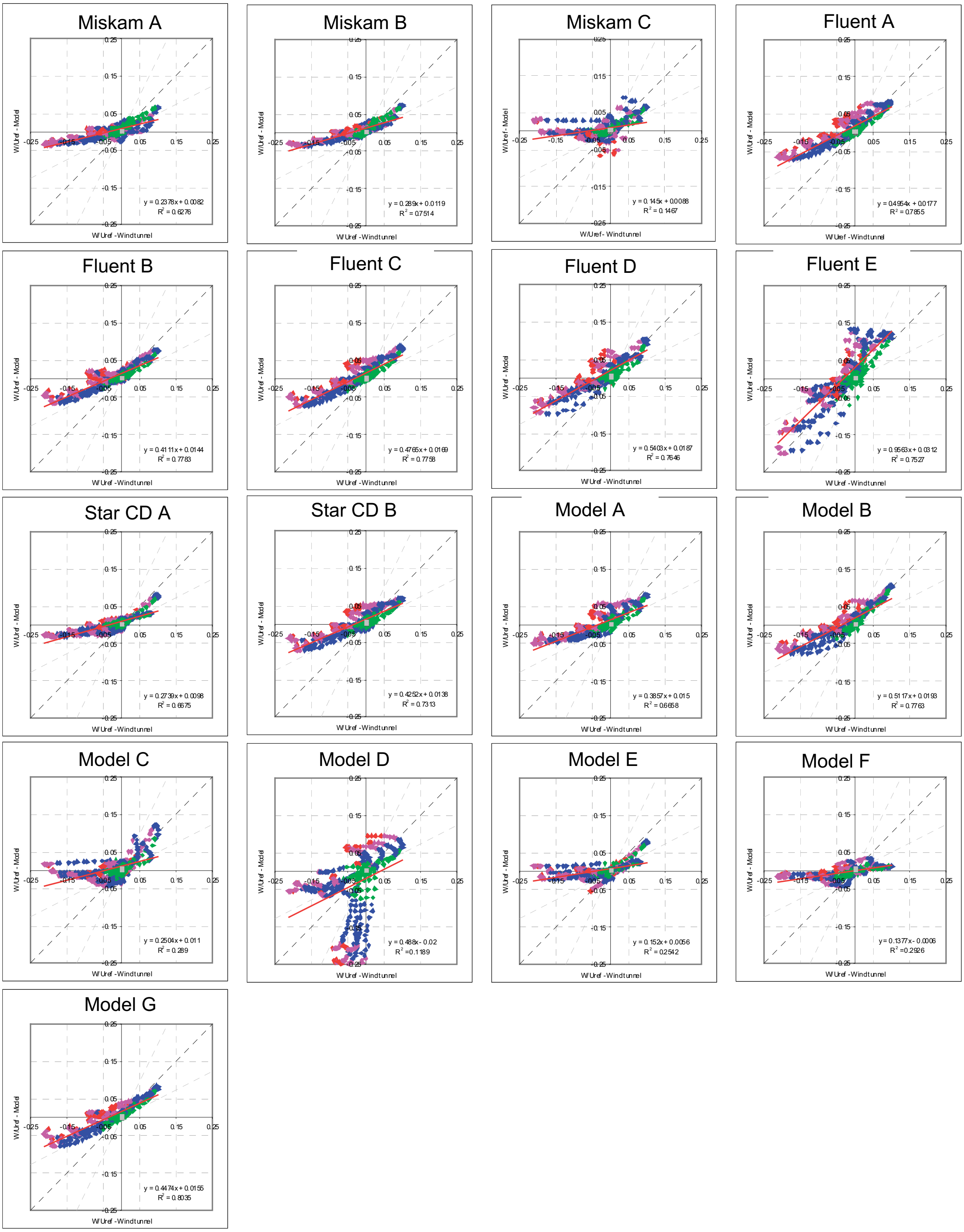

Figure 4. Scatter plot of w for 17 models. Each model is only represented by one run by the same group, except that Fluent D and Fluent E are by the same group.

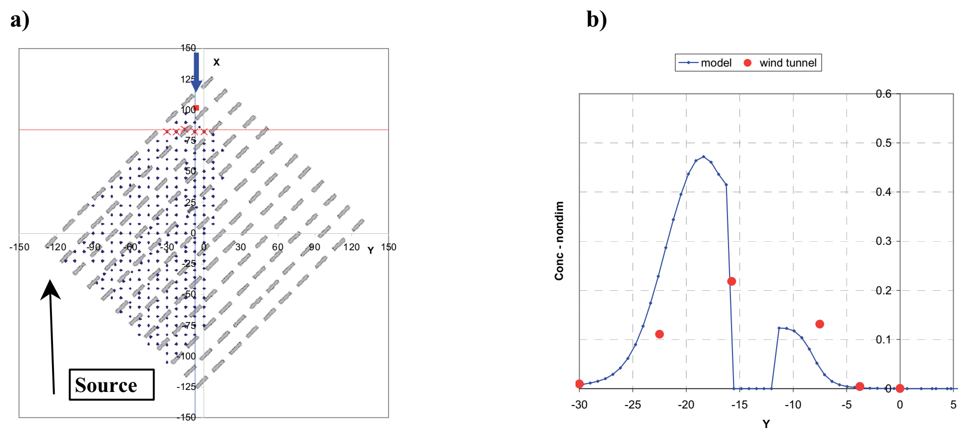

Figure 5. (a): Layout of dispersion experiment. Monitors are located at blue dots. (b): Profile of concentrations in cross-section at 18 m from the source, along the red line in (a). The modelled concentration is zero where there is a building.

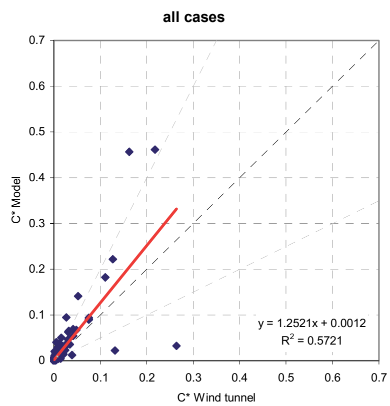

Figure 6. Scatter plots of concentrations for one of the Fluent runs.

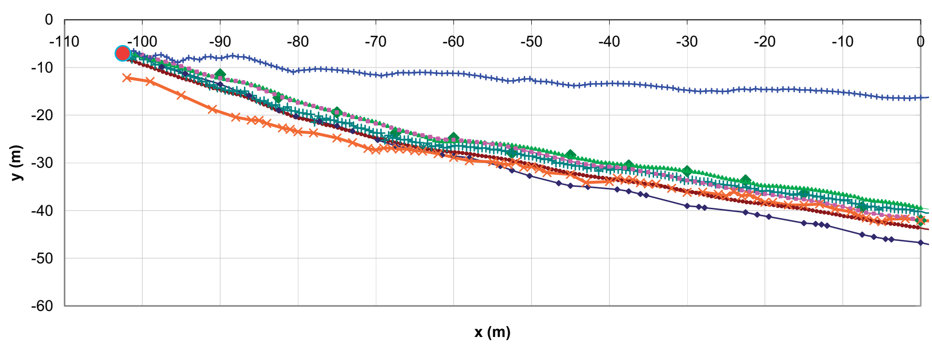

Figure 7. Map indicating plume trajectory according to measurements (diamonds) and 7 models (curves).

Figure 8 shows the estimated maximum concentration in the plume, according to measurements and 7 models. The estimation of the maximum is obtained by Gaussian fitting. The estimate is rather uncertain at the two cross-sections closest to the release point where the plume is too narrow to be adequately resolved by measurements. Anyhow, at greater distances the figure shows a general trend which can also be discerned when looking at data in other ways: There is a tendency for the models to predict too high values along the centre line.

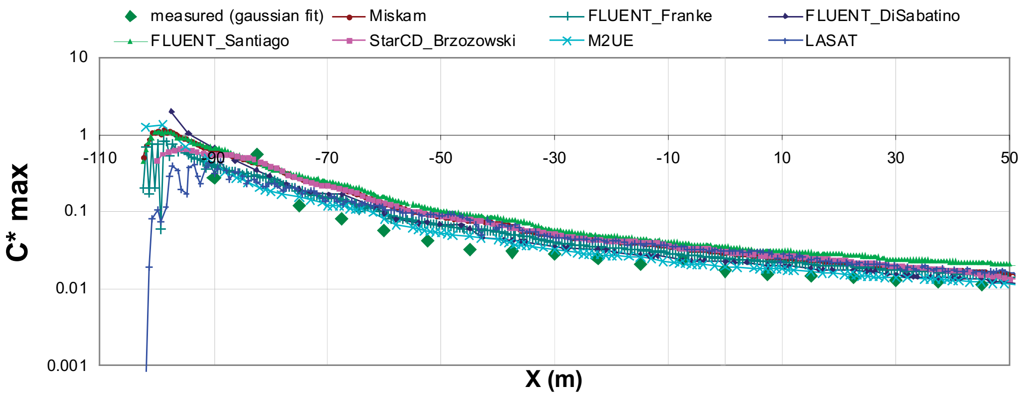

Figure 8. Estimated maximum concentration in plume, using Gaussian fitting. Note that the scale is logarithmic.

4. CONCLUSIONS

The power of exploratory analysis has been illustrated by examples. Exploratory analysis should never be omitted when performing quality assurance of models. It is our experience that even experienced modellers do sometimes perform errors, and that these can be overlooked unless one inspects data, presented in graphical form. Conditions are best for detecting anomalies if you have a situation where *several models are put into a common framework*. Such a framework makes it also possible to identify features common to several models, and thereby derive an indication of the state of art. The MUST exercise provides one such, excellent framework. Here, space permits to show only a few results. Readers who wish to look seriously into the state of the art should consult the references found through URL 1, including the Excel tools themselves.

REFERENCES

Leitl, B., K. Bezpalcova, F. Harms, 2007: Wind Tunnel Modelling Of The MUST Experiment, 11th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Cambridge, July 2-5, 2007. http://www.harmo.org/Conferences/Proceedings/_Cambridge/downloadItem.asp?CID=P435

URL 1: Official web site of COST 732: http://www.mi.uni-hamburg.de/index.php?id=464

URL 2: Wiki on Atmospheric Dispersion Modelling which has a particular forum for COST 732: http://atmosphericdispersion.wikia.com/wiki/COST_732_forum