17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

ASSESSING THE PERFORMANCE OF ATMOSPHERIC DISPERSION MODELS

Steven Herring[1] and Pablo Huq[2]

[1]CBR Advice Group, Defence Science and Technology Laboratory, Porton Down, Salisbury, England
[2]College of Earth, Ocean and Environment, University of Delaware, USA

Abstract: There are a large number of models and tools available for predicting the dispersion of material released into the atmosphere. These take many forms, ranging from simple models that execute in a fraction of a second to complex models requiring weeks of computational time. The primary purpose of many of these models is to predict the hazard resulting from releases of hazardous materials, and so it is critical to understand the level of predictive accuracy that might be expected from them. However, this is not easily determined from comparisons with experimental measurements for a number of reasons which include: the limited number of releases typically made, limited range of meteorological conditions, limited range of test environments and the arrangement and number of sensors used. In addition, the results of a comparison may be significantly influenced by practical decisions relating to what data to include and exclude, and the approach adopted to temporal and spatial averaging of the data. Given these factors, it is difficult to understand how the performance of one model relates to another.

The paper presents the issues involved in comparing model predictive performance to experimental data by reference to a number of comparison exercises against experimental data sets. It then presents an approach which is designed to make the performance of atmospheric dispersion models (of whatever type) more transparent. This involves clearly defining the basis on which the comparison is made, the use of the widely accepted BOOT software for generating quantitative measures and comparing the performance of the chosen model to the performance of a simple analytic model for the same experimental data. A simple analytical model is adopted as it is easily understood, and provides a clear baseline performance against which the performance benefits of more sophisticated models may be easily assessed.

Key words: *dispersion modelling, validation*

INTRODUCTION

A wide range of models have been developed for predicting the transport and dispersion of material released into the atmosphere. These range from simple analytic or empirical Gaussian plume models defined by a single equation, to models based on complex Computational Fluid Dynamics (CFD) simulations. Whereas the former will execute in a fraction of a second on standard desktop computer, the latter may require weeks of computational time on a High Performance Computing (HPC) facility. Independent of the method that is used to predict the dispersion, there is a need for the end user to be able to readily assess how good any particular model is, and what its limitations are. Thus there is considerable interest in methodologies for comparing the performance of different dispersion models (Warner et al., 2001), and also the development of automated of tools to facilitate such comparisons (Andronopoulos et al., 2015).

It is very difficult to describe the performance of a dispersion model in terms of a single metric. Although most studies utilise a number of generally recognised statistical parameters, researchers and organisations have developed their own favoured metrics on which they place particular emphasis. For example, some researchers favour the 2-D Measure of Effectiveness (2-D MOE) and Normalised Absolute Difference (NAD) parameters (Warner et al., 2001), while others look at the Cumulative Factor (Tull and Suden, 2014). However, when authors wish to compare their results to those of others they will generally relate them to the bounds for an acceptable model suggested by Hanna and Chang (2012). While these criteria may provide a reasonable basis for evaluating the performance of a dispersion model, examination of the process reveals that the values of the metrics are highly dependent on the way in which the comparison is

made. This being so, it may be difficult for a third party to fully appreciate the performance of a model without access to full details of the comparison process. The purpose of this paper is to propose a way by which the performance of models may be made more transparent.

ISSUES IN VALIDATING DISPERSION MODELS

The acquisition of sufficient dispersion data from field trials to validate models involves the deployment of a large amount of instrumentation and a large number of releases. This is complex and costly even for open terrain experiments, and the complexity and costs increase substantially for urban experiments. It is important to appreciate that even the biggest datasets consist of only a limited number of releases at a limited range of conditions. A consequence of this is that they invariably represent statistically small samples which limit the degree of confidence that can be obtained from comparing model predictions against a single data set.

Although field trials may involve releases made at a range of different times during the day and at night, they are generally conducted over a few days or weeks. This inevitably limits the range of meteorological conditions covered by the dataset. The environment is obviously fixed by the location of the trial, while the amount of data obtained is governed by the numbers of sensors that are available, their spatial distribution and sampling periods. To illustrate the difficulties we note that even though a large number of samplers (600 in all) were used in the Prairie Grass open terrain experiment (Haugen and Barad, 1958), the number that recorded data was limited to a fraction of these. The complexity of the urban wind field suggests that an even larger number of samplers are required to obtain good coverage, however, this is prohibitedly expensive, and only 130 samplers were deployed in the Joint Urban 2003 experiment (Allwine and Flaherty, 2006). In addition, although the dispersion of a plume is three dimensional, most sampling is restricted to a horizontal plane close to the ground, with only small numbers of measurements in the vertical dimension. This means that even in large experiments, such as those referred to above, the data is likely to be sparse, and of low fidelity.

The accuracy of any dispersion prediction is fundamentally linked to the accuracy of the meteorological data that is input to the model. The need for accurate meteorological data is typically addressed in research experiments by deploying a range of instrumentation at a number of locations, and at a range of heights. Nevertheless, it not obvious how this data should be consolidated to derive the most appropriate wind field information for input into a dispersion model. In addition to the meteorological uncertainties above, there are also typically smaller uncertainties relating to the definition of the source and release conditions which may introduce systematic errors that are important in assessing accuracy in the near-field

Some of the limitations associated with conducting field trials can be overcome by the use of wind and water tunnel experiments, which provide well-defined, constant, meteorological conditions that support repeatable measurements. In addition, the constant conditions and reduction in scale permit the acquisition of large statistical samples (Harms et al., 2011). However, the walls of the tunnel physically limit the maximum dimensions of the turbulence scales, and although Reynolds number independent flows can be achieved, there is a limit to the detail that can be represented (e.g. roof top geometries are generally simplified and trees omitted). It is also difficult to vary stability conditions, and most data from wind and water tunnel experiments are restricted to neutral buoyancy conditions, although stability effects are of great importance.

APPROACHES TO MODEL VALIDATION

A wide range of statistical metrics have been used by researchers in conducting dispersion model validation exercises. These include the Normalised Mean Square Error (NMSE), geometric mean, coefficient of correlation and percentage of observations and predictions within factors of two or ten (FAC2 and FAC10). Many are incorporated in the widely used BOOT statistical package (Chang and Hanna, 2005), but other measures and tools exist. In addition to the choice of metrics, the results of a validation exercise depends upon decisions made by the analyst regarding the basis on which the comparison is made. Consider the design of field studies involving arc maximum concentration values. This depends on the samplers being deployed on arcs, but even if there are a large number of samplers,

such an approach is problematic. This is because the plume is only likely to cover a small number of samplers on a given arc and not provide a well-defined distribution (particularly in urban areas) which makes evaluation of the arc maximum difficult.

Once the performance metrics have been chosen, the resulting values are highly dependent on the criterion used for determining the data that should be included in the comparison. After defining the threshold at which a sampler reading is taken to be zero, there are then three options for filtering the observed and predicted data pairs (Boubert and Herring, 2015): 1) accept all, 2) accept if both above threshold 3) accept if one above threshold. It is important to note that the application of strategy (3) to a spatially and temporally correlated comparison will likely lead to significantly poorer performance metrics than if strategies (1) or (2) are adopted. In addition, the performance metrics can generally be improved by adopting longer averaging times if a series of samples are taken at each location.

MAKING MODEL PERFORMANCE TRANSPARENT

The combination of factors described in the preceding sections, mean that if the results of a model comparison are presented in isolation without a detailed description of the analysis process, then it is difficult for a third party to evaluate how good the model is. It may perform well for a particular data set, or may appear to perform well because of the particular validation strategy adopted. In order to address the problem of making make the results of model comparisons more transparent, we propose that as well as presenting the results for the chosen model further effort should also be made to include a comparison with a standard reference model whose details are freely available. The utility and acceptance of such a standard model is greatly increased if it is based on well-founded understanding and physical principles, and can account for a range of conditions.

A REFERENCE MODEL

One possible candidate for a standard reference model for continuous ground level releases is the analytical model developed by Franzese and Huq (2011) which has been validated against data from urban experiments conducted in Oklahoma City, Salt Lake City, London, and St. Louis.

The model is based on a standard Gaussian formulation, in which the mean concentration c is predicted by equation (1), in which y indicates the crosswind direction, z the vertical direction, σ_y and σ_z are the standard deviations of the crosswind and vertical distributions of concentration, respectively and U is the reference wind speed and Q the mass release rate.

$$c = \frac{Q}{\pi U \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right) \quad (1)$$

In contrast to other simple urban dispersion models which are based solely on implementing empirical relationships derived from particular experiments, this model is based on fundamental dispersion theory. This means that the horizontal and vertical diffusion coefficients are determined according to the theories of Taylor (1921) and Hunt and Weber (1979) respectively as discussed in Franzese and Huq (2011).

The validation conducted by Franzese and Huq (2011) showed that model predicted the existence of near and far field urban dispersion regimes, and suggested that urban dispersion was governed by the characteristic length scales of atmospheric boundary layer turbulence, rather than urban canopy length scales which were more likely to affect dispersion only in the vicinity of the source. The model predictions demonstrated a convincing collapse of data for both daytime and nighttime conditions as shown in Figure 1.

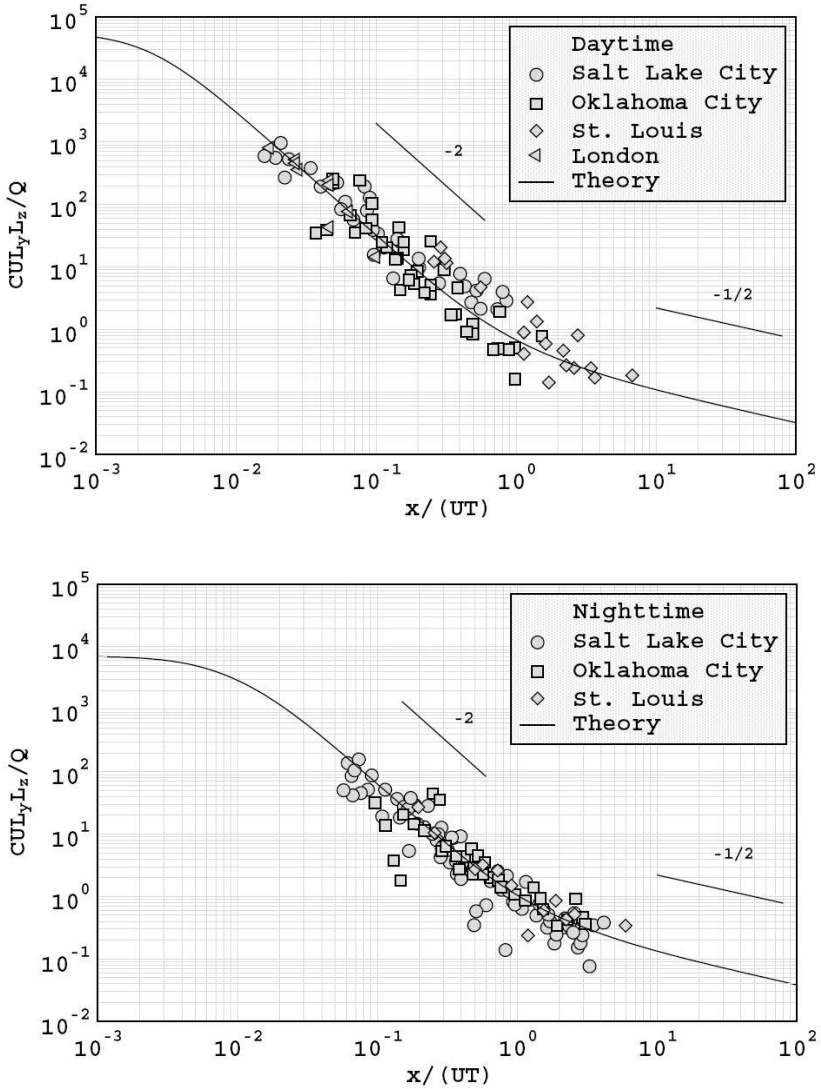

Figure 2. Comparisons between observed daytime and nighttime data and analytical model predictions.

The results therefore suggest that the model may be used to obtain realistic predictions of atmospheric dispersion in urban areas in both daytime and nighttime conditions, but has the advantage of being simple in formulation and applicable to a wide range of conditions. The authors therefore propose to undertake work to demonstrate how this model may be used as a reference model for quantifying the performance benefits of more sophisticated methods.

CONCLUSIONS

An examination of the literature shows that there is a lack of consistency in the approaches adopted to validating dispersion models in terms of the metrics used to assess performance. In addition, the wide variety of experimental arrangements and relatively limited data typically available, coupled with the large effect of subjective decisions on the inclusion or exclusion of data, mean that it is often difficult to establish the true performance of a model. One way of alleviating this is to always include a comparison against a reference model. It is suggested that the simple analytic plume model for ground-based releases developed by Franzese and Huq (2011) provides a suitable starting point for such an approach which the authors intend to develop.

REFERENCES

Allwine, K.J. and J.E. Flaherty 2006: Joint Urban 2003: Study Overview and Instrument Locations, Technical Report PNNL-15967, Pacific Northwest National Laboratory.

Andronopoulos, S. and F. Barmpas, J.G. Bartzis, K. Baumann-Stanzer, E. Berbekar, G. Efthimiou, C. Gariazzo, F. Harms, A. Hellsten, S. Herring, K. Jurcakova, B. Leitl, S. Trini-Castelli 2015: COST ES1006 Model Evaluation Protocol. ISBN 987-3-9817334-1-9.

Boubert, B. and S. Herring 2015: Validation of UDM Against Data from the MUST Experiment, Dstl/84143.

Chang, J.C. and Hanna, S.R. 2005: Technical Descriptions and User's Guide for the BOOT Statistical Model Evaluation Software Package, Version 2.0, Technical report.

Franzese, P. and P. Huq, 2011: Urban dispersion modeling and experiments in daytime and nighttime atmosphere. *Boundary Layer Meteorology*, **139**, 395-409.

Hanna, S.R. and J. Chang (2012): Acceptance criteria for urban dispersion model evaluation, Meteorology and Atmospheric Physics, 116 (3-4), 133-146.

Harms, F. and Leitl, B., Schatzman, M., Patnaik, G. 2011: Validating LES-based flow and dispersion models, Journal of Wind Engineering and Industrial Aerodynamics, pp 289-295.

Haugen, D. and M.L. Barad, 1958: Project Prairie Grass, a field program in diffusion, Volume 1. Air Force Cambridge Research Center.

Haugen, D. and M.L. Barad, 1958: Project Prairie Grass, a field program in diffusion, Volume 2. Air Force Cambridge Research Center.

Hunt J.C.R. and Weber A.H. 1979: A Lagrangian statistical analysis of diffusion from a ground-level source in a turbulent boundary layer. Quart J Roy Meteor Soc 105(444):423–443.

Taylor, G.I. 1921: Diffusion by continuous movements. Proc Lond Math Soc 20:196–211.

Tull, B. and P. Suden 2014: Urban Dispersion Model Evaluation of the QUIC and HPAC Models Using the DAPPLE Dataset. 16th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes 8-11 September 2014, Varna, Bulgaria.

Storwold, D.P., 2007: Detailed Test Plan for the Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT 07), Meteorology Division, West Desert Test Center, U.S. Army Dugway Proving Ground.

Warner, S., and N. Platt, J. F. Heagy, S. Bradley, G Bieberbach G. Sugiyama, J. S. Nasstrom, K.T. Foster, D. Larson User-Oriented Measures of Effectiveness for the Evaluation of Transport and Dispersion Models, Institute for Defense Analyses Alexandria, Virginia Paper P-3554, Jan 2001.