2 Key to the Model Validation Kit

The kit has been used at the Harmonisation conferences

The *Model Validation Kit* has been used for a series of workshops and conferences on *Harmonisation within Atmospheric Dispersion Modelling for Regulatory purposes* (see www.harmo.org). During the series of Harmonisation conferences, many papers have used the Kit, which was introduced in 1993. The present Guide describes the material after a revision in autumn 2005.

This chapter serves as a key to the entire material. Its purpose is to give you a background, so you can assess how well the kit fulfils your needs, and give you a qualified background to decide which parts of the Kit you will work with.

2.1 Some basic recommendations

Exploratory data analysis is important!

It is recommended that any model evaluation exercise start with clear definitions of the evaluation goal and the variables to be considered, followed by exploratory data analysis, and then statistical performance evaluation. The implications of this are discussed more closely in the User's Guide to BOOT, which is part of the material at hand (Chang and Hanna, 2005).

Thus, statistical model performance evaluation should not be a stand-alone exercise. It is highly recommended to be coupled with exploratory data analysis, which can reveal model errors, and errors and inconsistencies in data. The Model Validation Kit offers tools for this.

2.2 The Model Validation Kit

A common frame of reference

The Model Validation Kit is intended to be used for evaluation of atmospheric dispersion models. It is a collection of four field data sets as well as software for model evaluation. The Kit is a practical tool intended to serve as a common frame of reference for model performance evaluation. It is, however, limited in scope, as described in subsequent discussions.

History

The Kit has been used for the series of Harmonisation workshops and conferences. A preliminary version of the Kit was used for the workshop in 1993, while a subsequent version was used essentially unchanged throughout the period 1994 - 2005 (in 1997, a supplement was added). It has been distributed in hardcopy (diskette/CD and paper) to more than 250 research groups during that period.

The package was updated to Version 2.0 in October 2005. The new version allows the same studies to be carried out as the previous version, but has been revised in several respects. New software and computing environments have made it necessary to update the package. Furthermore, the documentation is significantly improved and brought up to date. The package can be downloaded from the Internet at www.harmo.org/kit.

Elements of the package

The package contains the following main elements:

- Field data sets from Kincaid, Indianapolis, Copenhagen and Lillestrom;
- The BOOT statistical model evaluation software package;
- Tools for exploratory data analysis, useful for diagnostic model evaluation;
- A recommended procedure (protocol) for model performance evaluation. The approach is explained in the Chapter ***Step by step instructions***. This procedure is relatively simple and thus has some limitations.

For the Kincaid experiment there is also supporting material that can be useful (video clips and a *Dispersion Visualisation Tool*) - see Chapters 8 and 10.

Note that although the emphasis of the Model Validation Kit is on the protocol, some tools included in the Kit - in particular the BOOT software - are general and can be applied for problems beyond the scope of the protocol.

When the Model Validation Kit is distributed on CD, the material is organised in folders as described in Chapter 4 on *Package contents*. Here, in the documentation we use the folder names of the CD.

The material can also be downloaded from the Web in a number of packages (self-extracting zipped files).

2.3 Data sets

The Model Validation Kit addresses the classic problem of a single stack emitting a non-reactive gas. The Kit comprises data from the following four field experiments:

- The Kincaid experiment (1980-81) with tracer releases from a 187-m stack. There are 171 hours of tracer data from monitoring arcs at distances from 0.5 to 50 km. In the Model Validation Kit, the emphasis is on arc-wise maximum concentrations.
- Data from an experiment in Copenhagen, Denmark in 1978-79 with releases from a non-buoyant elevated source (115 m) in neutral and unstable conditions. Nine hours of tracer data are available on arcs from 2 to 6 km. Both arc-wise maxima and crosswind-integrated concentrations are considered reliable.
- Data from an experiment in Lillestrøm, Norway (1987) with tracer releases from a non-buoyant source at 36 m in stable (winter) conditions. Sampling took place during 8 15-minute periods, not during an entire hour. Therefore, when comparing observations with models yielding one-hour averages, crosswind integrated

concentrations can be compared without problems, whereas it is not straightforward to compare arc-wise maxima.

- The Indianapolis experiment (1985) with tracer releases from an 84-m power plant stack in the city of Indianapolis, USA. There are 170 hours of tracer data from monitoring arcs at distances from 0.25 to 12 km. The emphasis is on arc-wise maxima.

Quality indicator

One experience from the past work – an experience that has been repeatedly confirmed – is the usefulness of assigning a quality indicator to experimental data, indicating how reliable a particular set of observations is. Such a quality indicator can be assigned by subjective methods (e.g., inspection of graphs), or assigned by a computer code according to certain objective criteria. The use of a quality indicator is valuable, because subsets of data can be selected in a well-defined manner. This can be utilised to discard data that would have been misleading if they were blindly included in an analysis. For two of the experiments, Kincaid and Indianapolis, the tracer data have been flagged by a manually assigned quality indicator assessing the quality of arc-wise maximum concentrations.

The quality index has values of 0, 1, 2 and 3, with 2 and 3 representing the most reliable data. Comparison studies of observed data with model results should in general be conducted with a quality indicator of 2 or 3.

The data sets are described in the chapter ***Field data***.

2.4 The BOOT software

BOOT is a general tool

The main tool for statistical performance evaluation is the BOOT software package. The BOOT program has been improved and is now available in version 2.0 with a comprehensive, rewritten 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. Although it has been primarily used to evaluate the performance of air dispersion models, the same procedures and approaches implemented in BOOT also apply to other types of models.

Performance measures considered in BOOT

Compared to the previous version of BOOT, the program now includes some additional performance measures, and an implementation of the ASTM statistical model evaluation procedure (see later). 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.) With the new software version, FB and MG can be separated into overpredicting and underpredicting components. Bootstrap resampling is used to estimate the confidence limits of a performance measure - hence the name BOOT of the package.

Files related to BOOT

On the distribution CD, the `BOOT` folder contains the BOOT program, a comprehensive User's Guide and various sample files. The `Tools` folder contains additional utilities for use in the present context, as described in Chapter 6 on ***Step by step instructions*** .

2.5 Tools for exploratory data analysis

When performing model evaluation, it is not sufficient to consider just statistical evaluation that produces some performance metrics. Rather, it is recommended that exploratory data analysis also be performed using graphical techniques.

The SIGPLOT graphical package: features and drawbacks

The Model Validation Kit includes some tools for such graphical analyses in the form of the SIGPLOT graphical package and the RESIDUAL utility. The SIGPLOT package is offered as an option that is specifically tailored for model performance evaluation. It must be mentioned that the SIGPLOT program, as well as a number of associated utility programs included in the Model Validation Kit only function in a DOS environment. The package can produce residual plots, where model residuals are depicted as a function of independent variables such as the downwind distance and time of day. Examples are shown in *Figure 7* (in Chapter 6).

It is recognised that the somewhat archaic SIGPLOT package is only one of the many ways of performing exploratory data analysis. More modern and interactive tools than the SIGPLOT package can certainly be used to achieve the same goals. For example, a potential alternative is to use Microsoft Excel for data handling and graphical analyses. Excel offers some very powerful tools for interactive data analysis. In particular, its *Autofilter* feature is useful for investigation of model behaviour. Nevertheless, Excel does not offer the specialised plots that SIGPLOT produces. The advantages of using SIGPLOT are that you will be able to produce residual and other types of specialised plots with data in a relatively standardised format, which has been used by others. Furthermore, the required utilities are already prepared, and the procedures for using the software are described in detail. The drawback is that you will have to work in a DOS environment (Section 6.2.5 provides some hints on this).

More details on Sigplot can be found in the chapter ***Step by step instructions*** as well as in the chapter ***SIGPLOT software***.

2.6 Limitations

It must be recognised that model evaluation studies performed on the basis of the Model Validation Kit are limited in scope. These limitations can be summarised as follows:

- Only four experimental data sets are considered.
- The emphasis is on operational short-range models.
- The problem of interest is relatively simple, namely a point source emitting a non-reactive gas over flat terrain, due to the fact that

this is the scenario represented by the four field experiments. On the other hand, much of the software included in the Kit is general and applicable to many different release scenarios.

- Further, the emphasis is primarily on a) *arc-wise maximum concentrations*, and to some extent b) *cross-wind integrated concentrations*.

- The Kit does not explicitly account for the stochastic nature of dispersion problems.

The tools in the Kit can be used to diagnose strengths and weaknesses of the models, but as a consequence of the above limitations, you should be careful in interpreting the results.

Quantile-quantile plots cannot be expected to show one-to-one correspondence

To further elaborate the last bullet in the above list, atmospheric dispersion processes are stochastic, whereas models in general predict only ensemble averages - not individual realisations. This means that there is a basic conceptual problem with the procedure of directly comparing model predictions to observations, as they cannot be expected to have the same statistical distribution. One consequence is that if the monitoring network is sufficiently dense and if the data represent a sufficient number of scenarios, then a "perfect model" is likely to underpredict the highest observed concentrations (this issue is elaborated by *Olesen*, 1997).

Note further that the so-called quantile-quantile plots from an entire experimental database should not stand alone as the result from a model evaluation. A very useful supplement is residual plots, which provide more insight into model behaviour.

Despite its limitations the Model Validation Kit has the advantage of being straightforward to apply and practically oriented. It also provides a common framework where the results of different studies can be intercompared.

2.7 An alternative: The ASTM methodology

A separate "ASTM package" exists

As noted, there is a concern that direct comparison of model predictions against observations could cause misleading results. Therefore, an alternative approach has been proposed by John Irwin, and has resulted in ASTM Standard Guide D6589. This procedure has also been incorporated in the latest version of the BOOT software as an option. The procedure is *not* treated in depth in the present compendium. However, there exists also a separate package (software and data sets), specifically devised as an implementation of the ASTM procedure - here referred to as the *ASTM package*. It was prepared by John Irwin and is available on the Internet (www.harmo.org/astm). This is not part of the Model Validation Kit, but it can be used as a supplement or an alternative to the Model Validation Kit.

The chapter ***Notes on the "ASTM package"*** of the present Compendium outlines the main principles of the ASTM methodology. Further, it explains some features that distinguish the two packages and lists certain issues of concern.

2.8 Structure of the User's Guide

In order to become acquainted with the Model Validation Kit, the two subsequent chapters are recommended reading. They concern, respectively, ***Pitfalls and FAQ***, and ***Package contents***.

Then follows a long chapter on ***Field data***, yielding an overview of the four field experiments and of the data included in the kit.

Chapter 6 ***Step by step instructions*** explains in detail how the tools of the kit can be used. You may choose not to use all of the tools, as some of them - especially those related to the SIGPLOT package - may seem unfamiliar to today's computer users

After Chapter 6 several short chapters with optional information follow, concerning:

- The SIGPLOT software
- The Dispersion Visualisation Tool
- Tools for Grapher
- Video clips from Kincaid
- Notes on the "ASTM package"
- Changes since the previous version of the Model Validation Kit.

Details on the BOOT software are not included here, as there is a separate User's Guide in the `BOOT` folder of the CD. The User's Guide also contains a general discussion on model evaluation.