Experiences using the Delta Tool for model quality assurance

Helge R. Olesen and Matthias Ketzel
A broad perspective

› The classic single stack problem.
› The ”EU Directive problem” where the Delta Tool comes in.
Classic single stack problem

› You have a single source surrounded by arcs of monitors.
› How do you evaluate model performance?
› “Model Validation Kit”
› ASTM standard guide D6589 on statistical evaluation of dispersion models

John Irwin
Video...

Dispersion experiments

Borex 1992 & 1994
Inherent incertaintly

› The atmosphere is turbulent with stochastic variations
› Nature provides us with individual realisations.
› With our models we can only hope to predict ensemble averages.
The graph shows the concentration (ppt) of a compound plotted against crosswind distance (km). The concentration peaks are observed at specific crosswind distances, indicating the dispersion pattern of the compound in the environment.
Classic stack problem

Conclusion already established in the eighties:
- Pairing model results and observations with respect to time and space is not useful when dealing with a single stack. It yields a correlation close to zero.
Parameters we may consider

- Arc-wise maxima (Model Validation Kit)
- Near centerline concentrations (ASTM Standard Guide)
- Crosswind integrated concentrations (Cy) and sigma’s (Sy) (John Irwin’s most recent recommendations)
John Irwins latest recommendations

› Compare group geometric mean values of observed Cy and Sy values.
John states: "We place too much emphasis on the importance of Cmax in our model evaluations; Cmax is dependent upon Cy and Sy (not the other way around)."

› More on John’s work at his website

   www.jsirwin.com

› (you may also go though the URL mentioned in my abstract)

May 6, 2013
Helge R. Olesen.
The "EU Directive problem"

▸ Models are used to assess compliance with limit values at numerous locations in Europe.

▸ How should requirements for model performance be defined? In other words: What "Model Quality Objectives" will it make sense to put into a legal framework?

▸ The Delta Tool is work in progress to address this problem. JRC Ispra leads the work.

▸ Philippe Thunis has a presentation tomorrow on the Delta Tool.
Differences between the two problems

› The classic single stack problem: We try to model a highly fluctuating signal. If we compare observations and model results in space and time correlation will be close to zero.

› The ”EU Directive problem”: For the majority of monitoring sites multiple sources interact. Background pollution is important. For the majority of sites we deal with a relatively smooth signal.
Further differences between problems

› The classic single stack problem: The emission source strength is known.

› The ”EU Directive problem”: The emission inventory is part of the problem. The quality of the total system is an issue: emission modelling plus atmospheric transport and chemistry modelling.
FAIRMODE Meeting – draft agenda first day

Wednesday, 10th of April 2013
Elzenveld Conference Center in Antwerp, Belgium

<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-09:45</td>
<td>Registration and welcome coffee</td>
</tr>
<tr>
<td>09:45-10:15</td>
<td>Welcome by the host</td>
</tr>
<tr>
<td>10:15-11:00</td>
<td>FAIRMODE – where we are within working group 2</td>
</tr>
<tr>
<td>11:00 -11:20</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>11:20-12:30</td>
<td>WG2 – Quality assurance and benchmarking</td>
</tr>
</tbody>
</table>

SG4: Benchmarking and Model Quality Objectives

- Experiences/comments from participants regarding Model quality objectives and discussion
- Discussion
- DELTA tool: Overview of updates
- DELTA tool: MS experiences & feedback
The Delta Tool

- A set of software. The user prepares
  1. Data for one year of observations for a number of stations.
  2. Model results at the corresponding locations.
- The system provides a common frame of reference for evaluating model performance.
- It is possible to make some exploratory analyses within the system. However, the core is a Benchmarking report.
Notes on the Benchmarking tool

› It is in a testing fase.
› The tool operates with performance criteria, but these are provisional and subject to discussion within the Fairmode community.
› There have been several versions of the tool with some major changes.
› For normalisation the plots now use measurement uncertainty (Delta Tool version 3.x).
Benchmarking report

Target plot

Summary of statistics: Do results comply with performance criteria?
Results within the inner circle are within the measurement uncertainty.

Centered Root Mean Square Error axis

Bias axis

HCAB (2009) – Danish traffic station, NO$_2$
Summary of statistics: Do results comply with performance criteria?

<table>
<thead>
<tr>
<th>INDICATOR</th>
<th>SUMMARY STATISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>O B S</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Exceed,</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>T I M E</td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Corr</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>StdDev</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>S P A C E</td>
<td></td>
</tr>
<tr>
<td>Corr</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>StdDeV</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>A Q D</td>
<td></td>
</tr>
<tr>
<td>RDE</td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

Nb of stations/groups: 1 valid / 1 selected

HCAB
Delta Tool test data set: POMI (2005), Chimere – challenging Italian data set. NO\textsubscript{2}
<table>
<thead>
<tr>
<th>INDICATOR</th>
<th>OBS</th>
<th>Time</th>
<th>Space</th>
<th>AQD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Blue dots</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
</tr>
<tr>
<td>Exceed,</td>
<td>Blue dots</td>
<td>&lt;2, 2</td>
<td>&lt;2, 2</td>
<td>&lt;2, 2</td>
</tr>
<tr>
<td>Bias</td>
<td>Green dots</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
</tr>
<tr>
<td>Corr</td>
<td>Red dots</td>
<td>&lt;2, 2</td>
<td>&lt;2, 2</td>
<td>&lt;2, 2</td>
</tr>
<tr>
<td>StdDev</td>
<td>Green dots</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
</tr>
<tr>
<td>RDE</td>
<td>Orange dots</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
<td>&lt;2, 0</td>
</tr>
</tbody>
</table>

**Summary Statistics**

Nb of stations/groups: 60 valid / 60 selected

**Indicators**

- **Mean**: Average value across all stations.
- **Exceed**: Proportion of days exceeding a certain threshold.
- **Bias**: Bias relative to a reference value.
- **Corr**: Correlation with a reference variable.
- **StdDev**: Standard deviation of values.
- **RDE**: Relative difference from expected value.

**Units**
- ug/m³
- days
- %
Some remarks on the Delta Tool

› Keep in mind: What do the underlying data represent?

› Establishing a common frame of reference is a good thing. It allows us to make a proper comparison of comparable results.

› However, care should be taken, especially when it comes to policy aspects.

› The data underlying two different benchmarking reports may represent two different challenges.
Remarks - continued

As an extreme, imagine the classic single stack problem put into the Delta Tool framework.

Poor performance in respect to the target plot is not necessarily an indicator of unacceptable model performance – it depends on the challenge you pose to a model. Further, you should consider fitness for purpose.
Examples where high correlation is hard to achieve

- A site under influence from a major point source is difficult to model correctly.
- A traffic site where traffic data are inaccurate, possibly just slightly shifted in time.
Delta Tool: Practical aspects

Not everything runs smoothly when you are a first time user. However, feedback to the developers can improve this.

A number of minor issues were brought up at the Fairmode plenary in April, regarding such things as better explanations of graphics.

There is so far not a set of explanatory notes, which could ease interpretation of the Benchmarking reports.
Conclusion

› Those of you working with modelling in the context of the European Air Quality Directive should sooner or later acquaint yourself with the Delta Tool.

› Report any issues you may find to the developers.

› We should step carefully as to how the Delta Tool eventually should be used in conjunction with the Air Quality Directive. We should not impose unreasonable requirements on ourselves.