Evaluating consequence models

15th May 2018

Simon Coldrick, Fluid Dynamics Team
Introduction

• Background
• Consequence modelling: why do it?
• The need for model evaluation
• A trivial example
• Stages of model evaluation
• How will the evaluation be used?
• The SAPHEDRA project
• The question
Background

• In consequence assessment for major hazards, decisions are made based on the output of predictive models

• Model developers, users and decision makers need to be assured of the robustness of the predictions

• Techniques for evaluating models have therefore existed since the early development of models

• There have been numerous model evaluation initiatives
Consequence modelling: why do it?

• To predict outcomes, instead of:
  – Doing experiments
    • Expensive
    • Limited range
    • Practicality (scale etc.)
  – Gathering data from accidents
    • Unplanned
    • Limited range
The need for model evaluation

• Predictions rely on models of reality, rather than reality itself

• An early mistrust signalled by “garbage in – garbage out”
The need for model evaluation

- Developments in computing and software and its accessibility mean that highly realistic looking results can be obtained...
The need for model evaluation

- There is a need to examine the correspondence between the model physics/results and the real world:
  - to assess the credibility or fitness-for-purpose of a model
  - US GAO (1979): “to inform a decision maker, who may be far removed from the modelling process, of the quality of the model results”
The need for model evaluation

• How much confidence do we have in model predictions?

Evaluation is a process by which interested parties, who were not involved in a model’s origins, development and implementation, can assess the model’s results in terms of its structure and data inputs so as to determine, with some level of confidence, whether or not the results can be used in decision making. Our concerns in advocating evaluation for complex models are threefold: (1) for
Model evaluation – other benefits

• An indication of the applicability of a model to new problem areas
  – e.g. new materials or scenarios
• To encourage model improvement through management of model quality
• To identify areas for improvement in models as well as shortcomings in experimental datasets
Recent initiatives
Model development process

1) Construct a set of equations which one can argue represents the process

2) Provide a method to solve those equations (analytically or numerically)

3) Test the solutions against all known relevant observations to check that they do represent those observations

Satisfactory?

Yes

4) Make new predictions for what will be observed in experiments which have not yet been done

No

Revise the model

Do the new experiments
Model evaluation process

• The underlying model must have a sound physical basis (Scientific assessment)

• The computer implementation of the model must be consistent with its mathematical basis (Verification)

• The model must represent reality to an acceptable degree (Validation)

• The interaction between the user and the model must be considered (User-oriented assessment)

• The level of uncertainty when evaluating and using the model and its effect on output must be considered (Sensitivity and uncertainty analysis)
A trivial example

• Walking from the hotel to the railway station
• A computer model to predict arrival time at the station, based on the equation:

\[ t_a = t_0 + \frac{S}{U} \]
A trivial example

- It has been validated so it must be ok to use

Some Statistical Performance Measures (SPM):

<table>
<thead>
<tr>
<th></th>
<th>MRB</th>
<th>MRSE</th>
<th>MG</th>
<th>VG</th>
<th>FAC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>100%</td>
</tr>
<tr>
<td>Actual</td>
<td>-0.00034</td>
<td>0.00000</td>
<td>0.99966</td>
<td>0.99932</td>
<td>100%</td>
</tr>
</tbody>
</table>

Acceptable: ?
Comparable: Yes

- Based on these, the model could be deemed acceptable
- Its performance could be compared with similar models
- The results are slightly conservative i.e. overestimates travel time – not a bad thing
A trivial example

- I try the model out and miss the train by 3 minutes
- What went wrong?
  - In a **scientific assessment** we discover that the model was developed for a rural location and I used it in an urban setting (**documentation/user oriented assessment**)
  - The **scientific assessment** shows that it is missing essential physics sub-models for delay e.g. a waiting-at-traffic-lights sub-model
  - The model was used outside its **limits of applicability**
- Model evaluation should help to identify these shortcomings, among others
The challenge of model evaluation in consequence modelling

• Model evaluation is not new and is an established part of product design processes
  – Emphasis can be placed on validation because the models are often applied in a narrow region of operation

• In consequence modelling, important decisions are based on modelling results for complex, sometimes completely new scenarios, for which there may be little or no experimental data
  – The same reliance cannot be placed on model validation and other tests are needed to provide assurance that the predictions are reliable
Scientific assessment

• Scientific assessment may be one of the most important aspects of model evaluation
  – validation can be used to assess whether a model is “right,” but it must be “right for the right reason”

• A good model should incorporate a realistic description of the physical processes being modelled as this provides the confidence necessary to apply the model beyond the range of observations used to test the model

• The problem for an evaluator is to determine if this is the case...

Venkatram (1988) “it has now become fashionable to compute all sorts of statistical performance measures that are supposed to quantify performance. In my opinion, the reason for this is that it is easy to compute statistics and compare numbers (which are often meaningless). On the other hand, it is much more difficult to examine the science that governs models.”
**Verification**

- Verification is checking that the computer implementation of a model is consistent with its mathematical basis

  “Mr. Herschel ... brought with him the calculations of the computers, and we commenced the tedious process of verification”
  - Charles Babbage c. 1839

  “It is emphasised that the undertaking of model verification is an extremely tedious task”
Verification in practice

- The responsibility of the model developer?
- Verification of CFD models (vendor/user)
- Sense checks (mass balance etc.)
- Analytical solutions
- Test cases
- Manufactured solutions
- Demonstration that verification has been done
Validation

• The comparison of model predictions against physical experiments. Straightforward?
  – Lack of data
  – Independent datasets
  – Old data
  – Data processing
  – Creating and maintaining datasets
  – Scale of test data (field scale/lab scale)
  – How to compare models with experiments (comparison metrics)
  – Consistent basis for model comparison
User-oriented aspects

• Users:

On two occasions I have been asked, "Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?" ... I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

-Charles Babbage

• Software user interface
• Documentation
• Training
User-oriented aspects

• Blind model comparison exercises have shown many times that different users of the same model can produce very different results for the same scenarios (e.g. Dalmarnock fire tests)

• A single user of the same model can produce different results
User-oriented aspects

• Different users of different models can produce different results

• Best Practice Guidelines (BPG) and documentation can help to address user aspects
Sensitivity/uncertainty analysis

• Which model inputs affect the output?
• What ranges may the model inputs take (including the composition of the model) and how do they affect the output?
• To indicate the variability in the output
• To inform future use of the model
• To identify model faults
How will the results of the evaluation be used?

• Acceptance criteria?
  – Science
  – Validation results
    • Based on previous experience

• “Safety” factor?

• Model comparison?

• Ongoing process of model development
A few slides on SAPHEDRA
Developing a protocol in SAPHEDRA

• Objectives:
  – To review previous protocols for the evaluation of dispersion, fire and explosion models with identification of the capabilities and limitations of different evaluation approaches, and lessons learnt from evaluation exercises
  – To identify best practice that could be used to develop a new model evaluation protocol for SAPHEDRA
  – To make recommendations for the model evaluation protocol structure
Review of model evaluation protocols and standards

- 11 protocols and standards reviewed covering:
  - Model evaluation in general
  - Atmospheric dispersion in general
  - Dense gas dispersion
  - Consequence modelling
  - Gas explosions
  - CFD modelling in general
  - CFD for hydrogen safety assessment
  - Fire modelling
Lessons learnt summary

• Previous applications of model evaluation protocols and standards were reviewed

• There were relatively few (published) applications

• Most reports on model evaluation focus on the model, not the evaluation method

• The largest number of applications were for the LNG dispersion MEP and the ASTM E1355-12 Standard because they are required by US regulation
Recommendations

• A number of recommendations have been extracted from:
  – The review of protocols and lessons learned
  – Previous applications
  – Experience of model evaluation
  – Experience of constructing protocols
Can ADMLC contribute over and above the information already available on this topic?

- The processes of model evaluation are well understood and there is sufficient information to go about evaluating models in different fields of application.
- The way in which models are compared with experiments differs according to the application and can be quite simple e.g. peak overpressure, or quite complex e.g. statistical goodness-of-fit.
- A review of these approaches would be informative for those who need to evaluate models, but are unsure of the different techniques and their relative merits.
Finally

Lesson 5: Sophisticated comparison metrics may be more trouble than they are worth.

Fire Model Validation – Eight Lessons Learned

KEVIN McGRATTAN, RICHARD PEACOCK, and KRISTOPHER OVERHOLT
Fire Research Division
National Institute of Standards and Technology
Gaithersburg, Maryland, USA
Thanks for listening

HSL is the commercial arm of the Health and Safety Executive, HSE. Our commercial work delivers high quality science to meet the needs of industry and government in the UK and overseas. Our commercial customers can commission services and research using our state-of-the-art scientific laboratory in Buxton, as well as analytical expertise from other parts of HSE’s science base.