Description of work requested by ADMLC

Topic for which funding is sought
A Review of Methods used to assess the Performance of Atmospheric Dispersion Models

ADMLC outline technical annex

An important aspect of using atmospheric dispersion models is to understand how a model performs against measured data. Many commercially available models are validated using widely available datasets, applying various statistical metrics (e.g. correlation, Root Mean Square Error etc), and/or graphics (e.g. scatter, Q-Q plots, statistical distributions). It is important that the underlying model has sound physics and mathematics (see attached* 'Framework for Model Evaluation in HSE', Coldrick 2018; also RR1099), and it is imperative that model validation does not mis-represent the true performance of the model, by cherry-picking datasets to make a model look good, or by not considering the quality of the monitoring data being used. Validation carried out to a high standard can provide confidence that the model is suitable for use in other locations and is especially advantageous where monitoring data is unavailable.

However, it is important that when monitoring data is available, model users are able to test and verify the model performance for their specific project and understand the uncertainties (including both model inputs and inherent model uncertainties) in the specific case they are undertaking. For this study, we are assuming the model is based on sound science and has been verified and the focus should be on the validation stage of the HSE model evaluation framework.

Many factors may influence the type of model performance tests carried out such as model inter-comparison tests, model type (e.g. Numerical/Gaussian/CFD), the quality of monitoring data and how this data is used in the evaluation, environments (e.g. urban/rural), source types, monitoring methods or different averaging periods (e.g. short term v long term) and the type of datasets being compared (e.g.model v observations or model v model). In each case the performance tests required, or the acceptability criteria for 'good performance', may differ.

There are several published methods of quantifying model performance, including statistical metrics (Chang and Hanna (2004); Venekatram (2008); Liu et al (2011)) and downloadable tools (CERC Model Evaluation Toolkit, FAIRMODE tools, or the open-source R library 'OpenAir') which assess model performance, and which publish model acceptability criteria. It has been noted that these tools use different statistical metrics for assessing model performance, and that the published model 'acceptability criteria' can vary. Methodologies to interpret the data are constantly evolving, such as the analysis of sensors in an arc around a source (Hanna, Chang and Strimaitis (1993); https://www.nfpa.org/News-and-Research/Data-research-and-tools/Hazardous-Materials/LNG-model-evaluation-protocol-and-validation-database-update).

ADMLC is interested in seeking tenders that will review the different published methods of comparing models against observed data. Observed data can include sources other than ground level monitors (e.g. satellites). We are not looking to develop new statistical metrics, or formalise an approach, however we are looking for the review to consider all the published methods, highlight the strengths and weaknesses that these methods offer (e.g. do different model evaluations lead to the same conclusion) and provide recommendations and guidance on when to use particular statistical tests.

Stage 1
The study should firstly include a literature review of how model performance is currently evaluated in commercially available packages (validation documents) and in research papers. This should include published model performance, or how a particular dataset has been used in different studies, across the range of different model types (e.g. Gaussian or numerical) and model backgrounds (e.g. commercially available, operational, research models and ensemble modelling systems). The performance metrics used in the evaluation studies should be evaluated. This should consider the combination of metrics used and how this can be used to diagnose problems or biases within a model. A single metric can be misrepresentative and misleading but can be very informative when placed alongside other metrics (graphical and numerical).
It is also important to consider the data requirements to support the calculation of different metrics, such as quality and quantity of the data (e.g. spatial distribution, sampling frequencies and uncertainties), and if the usefulness and robustness of the metrics can vary depending on the datasets used to verify a model.

Stage 2
Having identified performance metrics in the literature review, the performance metric should be assessed in further detail by outlining the advantages and disadvantages of using a particular method for different model types in case studies. These case studies should use datasets and model types in agreement with the Committee. The performance metrics considered should not necessarily be restricted to statistical values but should also include other approaches such as graphical analysis of the data or how the dataset is used in the analysis (e.g. would the outcome change if particular parts of a dataset were selected). This should consider how performance metrics may relate to long term (e.g. annual means) and short term (hourly values or percentiles) standards and, in particular, Air Quality Standards (e.g. 99.79th percentile of hourly mean).

Stage 3
The study should summarise the findings (identified in Stages 1 and 2) and provide recommendations of which model evaluation methods should be used for different situations (e.g. Gaussian/Numerical, Urban/Rural environment, Point source or non-point source, long or short term), guidance on using datasets for model evaluation and if any further work is needed on this topic.

* Also available on request to admlc@ukhsa.gov.uk

Timescales		
	Item or deliverable	**Date**
	Start date	
	Intermediate stages or deliverables	Add rows as needed
	Draft report for ADMLC comment	
	Final report	

Costs
Indicate points at which intermediate payments, if any, are required.
Note that ADMLC will only make intermediate payments on receipt of identified deliverables or the draft report

CVs of Project Staff
Provide CVs of 2 staff involved in the project.
This section should be no more than 1 page in total.