Environment Agency

Evidence

Comparison of simple and advanced regional models (CREMO)

Outcomes for the Environment Agency

Report – SC060037/R

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It is our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency.

Published by:
Environment Agency, Horizon House, Deanery Road, Bristol, BS1 5AH

www.environment-agency.gov.uk

ISBN:

Further copies of this report are available from our publications catalogue: http://publications.environment-agency.gov.uk or our National Customer Contact Centre: T: 08708 506506
E: enquiries@environment-agency.gov.uk.

Author(s):
Bernard Fisher

Dissemination Status:
Publicly available

Keywords:
Intercomparison, regional air quality models, simple, advanced, evaluation, diagnostic, acceptance criteria, policy support, regulation

Research Contractor:
Centre for Atmospheric and Instrumentation Research
University of Hertfordshire College Lane, Hatfield, AL10 9AB, UK. Tel: +44 (0) 1707 284520
Email: r.s.sokhi@herts.ac.uk

Environment Agency's Project Manager:
Bernard Fisher, Evidence Directorate

Collaborator(s):
Ranjeet Sokhi and Charles Chemel (University of Hertfordshire)
Garry Hayman (Hayman Atmospheric Consulting)
Stephen Griffiths (E.ON New Build and Technology)
Keith Vincent (AEA Technology)
Tony Dore (Centre for Ecology and Hydrology)
Paul Sutton and Raymond Wright (RWE npower)

Project Number:
SC060037/R

Product Code:
LIT

Evidence at the Environment Agency

Evidence underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us, helps us to develop tools and techniques to monitor and manage our environment as efficiently and effectively as possible. It also helps us to understand how the environment is changing and to identify what the future pressures may be.

The work of the Environment Agency's Evidence Directorate is a key ingredient in the partnership between research, guidance and operations that enables the Environment Agency to protect and restore our environment.

This report was produced by the Research, Monitoring and Innovation team within Evidence. The team focuses on four main areas of activity:

- **Setting the agenda**, by providing the evidence for decisions;
- **Maintaining scientific credibility**, by ensuring that our programmes and projects are fit for purpose and executed according to international standards;
- **Carrying out research**, either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available.

Miranda Kavanagh

Director of Evidence

Executive summary

CREMO (Comparison of Simple and Advanced Regional Models) is a five-year Environment Agency project that finished in early 2012. The results from the CREMO project and other recent UK and international evaluation studies can be used to assess the policy and decision support provided by various kinds of regional scale air quality models. The models considered within these studies include those with realistic meso-scale meteorology and those which assume simplified atmospheric transport conditions. A focus of this report on outcomes is on the acceptance criteria for the use of models in regulatory situations, especially under the EU Air Quality Directive.

Evaluation studies involve selecting appropriate metrics or diagnostics (parameters summarising key aspects of the behaviour of a model) and then deciding on the way in which they may be applied in regulatory and policy decisions. It is argued that the evaluation of models should consider how a model will be used (that is, what decisions will be made, whether for policy and decision support, or to develop new models).

Considerable progress has been made and the advanced models are essential for assessing secondary pollutants, such as ozone and particulate matter. However conclusions from evaluation studies completed both as part of CREMO and internationally, remain provisional and further work is required.

In the absence of an objective basis for setting acceptance criteria for models, the underlying pragmatic principle should be to use whatever has comparable accuracy with best existing international practice. In regulatory applications, the error expected in current types of models should be accepted and this uncertainty built into any decisions made on the basis of models.

This report presents the outcomes of the CREMO project by providing 11 key questions and their answers representing a statement of the main conclusions. Each question and answer is followed by supporting evidence.

Five additional Environment Agency reports providing evidence supporting this main report have been produced within the CREMO project and are listed below.

HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, R., 2012a. *Comparison of simple and advanced regional models (CREMO): Model evaluation protocol*. Report SC060037a/R. Bristol: Environment Agency.

HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, D.R., 2012b. *Comparison of simple and advanced regional models (CREMO): Model evaluation report.* Report SC060037b/R. Bristol: Environment Agency.

HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, R., 2012c. *Comparison of simple and advanced regional models (CREMO): Ozone diagnostics*. Report SC060037c/R. Bristol: Environment Agency.

HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, D.R., 2012d. *Comparison of simple and advanced regional models (CREMO): Model evaluation: Ground-level ozone*. Report SC060037d/R. Bristol: Environment Agency.

DERWENT, R.D., 2012. *Chemical mechanism choice: impacts of ozone precursor emissions reductions in the United Kingdom on episodic peak ozone in the United Kingdom*. Report SC060037e/R. Bristol: Environment Agency.

Additional external publications have been produced as a result of CREMO and appear in the reference section marked with an asterisk.

Acknowledgements

The author acknowledges valuable discussions held with colleagues within the context of CREMO and under the letter of agreement between Defra, the Environment Agency and the US Environmental Protection Agency establishing cooperation on exposure science.

The Environment Agency acknowledges the role of its partners in the CREMO project, some of whom made significant contributions in kind to the work programme. The partners included the University of Hertfordshire, AEA Technology, the Joint Environment Programme of the power industry, CEH Edinburgh, RD Scientific and Hayman Atmospheric Consulting.

Contents

1 Introduction

This report summarises the main outcomes of value to the Environment Agency from the 'Comparison of simple and advanced regional models' (CREMO) project. The overall aim of the CREMO project was to enable the Environment Agency to make an informed decision on the use of advanced regional-scale atmospheric chemical transport models as one of its assessment tools. In particular, the project evaluated the performance characteristics of the Community Multiscale Air Quality (CMAQ) modelling system[1] for real regulatory applications through comparison of CMAQ with existing methods. The project applied CMAQ to a series of assessments (including acid deposition, particulate matter and ozone) and tested its capabilities through targeted comparisons with 'simpler' models and with measurements according to agreed model acceptance criteria.

As part of the modelling within CREMO, calculations were made using emissions from UK and European sources in 2003. The stationary sources devised for demonstration purposes were designed to be representative i.e. they were devised to reflect the kinds of sources that the Environment Agency regulates and the broad regions where such sources may occur. However, in every other respect the stationary sources were hypothetical i.e. they were not intended to indicate any particular actual source or site. The representative, but hypothetical, nature of the devised stationary sources is best indicated by the descriptor: "representative/hypothetical". There were 2 representative/hypothetical sources: a large coal-fired power station in southern England, and a large oil refinery on the south coast of England. The representativeness of these sources was established by reference to the Pollution Inventory for England and Wales.[2]

1.1 Advances in air pollution assessment

The emissions of sulphur dioxide (SO_2), oxides of nitrogen (NOx), volatile organic compounds (VOCs) and ammonia contribute to a number of environmental impacts that affect human health and/or ecosystems:

- acid deposition/eutrophication;
- ground-level ozone;
- particulate matter.

In recent years air pollution assessments have focused on secondary pollutants, such as nitrogen dioxide (NO_2), ozone (O_3) and particulate matter (PM).

A number of different models have been used by Defra and the Environment Agency to cover specific impacts and spatial scales such as:

- the Fine Resolution Atmosphere Multi-Pollutant Exchange (FRAME) for acid deposition;
- the Trajectory model with Atmospheric Chemical Kinetics–Atmospheric Dispersion Modelling System (TRACK-ADMS) for annual audits;

[1] When the project was commissioned, MODELS-3 was the operational version of this community air quality model. MODELS-3 comprised the CMAQ modelling system and the MM5 mesoscale meteorological model. The MM5 model has since been replaced by the Weather Research and Forecasting (WRF) numerical weather prediction model. All references to MODELS-3 have been updated to CMAQ to avoid confusion.

[2] Information on annual emissions from individual sources is available from the Pollution Inventory for England and Wales (http://www.environment-agency.gov.uk/business/topics/pollution/32254.aspx).

- the Ozone Source–Receptor Model (OSRM) for ozone.

The last decade has seen an enormous increase in the sophistication of computer programs, most notably for this project, the US Environmental Protection Agency (USEPA) supported CMAQ (Community Multi-scale Air Quality) modelling system, and the development of cheap, fast and accessible computing platforms. Regulators therefore need to be able to assess the capabilities of different types of air quality models with different levels of sophistication.

There have also been a number of studies comparing the performance of regional models using measurements, such as Phase 1 of the Model Evaluation Exercise for the Department for Environment, Food and Rural Affairs (Defra) (Carslaw 2011a, b and c) and AQMEII, the Air Quality Modelling Evaluation International Initiative, (http://aqmeii.jrc.ec.europa.eu/). However there are no universally accepted, objectively based acceptance criteria for air quality models, though there are commonly applied metrics.

The CREMO project has developed criteria based on footprints, which are of particular value for industrial sources, to assess the performance of different models when calculating $PM_{2.5}$. Metrics for assessing ozone, for which many criteria have been proposed, are also discussed in this report (AQEG 2009).

Comparison of a model against observations is not the only, or a sufficient, way of judging a model. This project also focused on diagnostic assessment, which involves ensuring that complex modelling is undertaken with understanding. In other words the reasons for the behaviour and response of a system are fully understood. The use of models is strongly influenced by their use within regulatory frameworks, such as the EU Air Quality Directive, but are not confined to this. The main use envisaged under the Environment Agency supported CREMO project is within air quality management to demonstrate that the certain actions will lead to changes in concentration, resulting in compliance with regulatory standards or Directive limit values.

1.2 Project aims and objectives

The CREMO project had two main objectives:

1. To provide a technique for assessing the contribution of industrial emissions of NOx and VOCs, under realistic meteorological conditions, to ambient levels of ozone based on the CMAQ model, involving comparison with simpler methods and observations;
2. To provide a technique for assessing the contribution of industrial emissions under realistic meteorological conditions to ambient levels of size-speciated particulate matter, PM_{10} and $PM_{2.5}$, based on CMAQ, involving comparisons with simpler methods, and with observations.

These above aims were met by the following specific objectives:

- To compare the performance of CMAQ with the simpler models FRAME and TRACK-ADMS, and to produce footprints of deposition and concentrations resulting from industrial emissions regulated by the Environment Agency.
- To assess the capabilities of CMAQ as a practical tool for modelling acid deposition, ozone and size-speciated particulate matter.
- To evaluate the capabilities of CMAQ to predict regional ozone concentrations and the ozone response to changes in emissions of NOx and VOCs.
- To assess the capabilities of CMAQ to calculate the contribution of regulated industrial emissions to size-speciated particulate matter concentrations and associated chemical species.

- To identify the main operational applications of CMAQ by examining the variability and uncertainty resulting from changes to input parameters through sensitivity analysis.
- To synthesise and integrate the outcomes of the previous tasks and make recommendations on how and under what circumstances CMAQ should be used by the Environment Agency for regulatory applications.

1.3 Structure of this report

The rest of this report takes the form of a series of questions and answers. Eleven questions are posed followed by an answer (in italics), which is supported by evidence given in the following text.

2 Questions and answers with supporting evidence

Q1 Why do the Environment Agency or other regulators require complex sophisticated models?

Answer

Changes in $PM_{2.5}$ concentration, which are a small fraction of an air quality objective (for example, 1 µg/m^3) can have significant health implications. Thus small changes in ambient $PM_{2.5}$ concentrations need to be modelled accurately. Because $PM_{2.5}$ is generated on a regional scale, is not dominated by near-field effects and is a mixture of components, models are the only way of assessing the impact of individual sources and of estimating the population exposure, as they provide detailed spatially resolved concentrations.

A recent report by the Committee on the Medical Effects of Air Pollutants (COMEAP) quantified the effects of particulate air pollution ($PM_{2.5}$) on human health (COMEAP 2010). The report attempts to clarify the way in which the benefits of a reduction in average $PM_{2.5}$ concentration is expressed (using metrics such as life years saved or increased life expectancy).[3] Even what might be regarded as a small, but sustained, reduction of 1 µg/m^3 of $PM_{2.5}$ is estimated to produce a large health benefit, equivalent to roughly half the gain produced by the removal of all motor vehicle accidents. Even though processes which the Environment Agency regulate are only partly responsible for $PM_{2.5}$ concentrations, this large estimated health burden means that it is important to be able to quantify accurately potential changes in the concentration of $PM_{2.5}$.

A review of the current position of the status of $PM_{2.5}$ in the UK is given in a report published by the Scotland and Northern Ireland Forum for Environmental Research (SNIFFER 2010). Data analysis in the report suggests that $PM_{2.5}$ is unlikely to breach the annual limit value in the Air Quality Directive, but that meeting the exposure reduction target will be more difficult. To meet this target, a 1.5 to 2 µg/m^3 reduction in $PM_{2.5}$ concentrations is likely to be required.

The SNIFFER report points out that the extent of the Environment Agency's role in meeting the target is not clear as there are many sources of $PM_{2.5}$ that are not within the Environment Agency's regulatory remit. With the exception of London, emissions from 'industrial' sources are broadly as important as emissions from vehicle exhausts, but the Environment Agency regulates only a fraction of these 'industrial' sources. Data do not suggest a single dominant contribution from an industrial sector regulated by the Environment Agency. So there are unlikely to be any simple direct measures to significantly reduce $PM_{2.5}$ which the Environment Agency could implement. The report mentions models but does not draw conclusions about them, acknowledging that the Environment Agency is currently evaluating through the CREMO project the use of models to assess the industrial contribution to $PM_{2.5}$.

[3] The report states that removing all anthropogenic ('man-made') particulate matter air pollution (measured as $PM_{2.5}$) could save the UK population approximately 36.5 million life years over the next 100 years and would be associated with an increase in UK life expectancy from birth (that is on average across new births) of six months. This demonstrates the public health importance of adopting measures to reduce air pollution.

The SNIFFER report cites Derwent *et al.*. (2009), who applied the UK Photochemical Trajectory Model (UK PTM) in a study of the non-linear response of secondary particulate matter to changes in precursor emissions. For a single receptor in rural southern UK it was concluded that, for $PM_{2.5}$ concentrations, the largest reduction in $PM_{2.5}$ was derived from a reduction in ammonia. This modelling study illustrates a useful methodology to a policy question and answer. However, given that other models have yet not been used to address this question, this cannot be considered the definitive answer to the question of the change in $PM_{2.5}$ to a change in emissions. The result cannot be subject to a dynamic evaluation (Dennis *et al.* 2010), as large changes in ammonia in Europe are unlikely to occur in the medium term, preventing the response of the model to emission changes to be tested.

The COMEAP study quantified the risk from $PM_{2.5}$ using the Pollution Climate Mapping (PCM) model, an air quality model used exclusively in the UK. According to COMEAP (2010):

> *PCM is a pragmatic attempt to get close to understanding and accounting for all of the measured mass concentration for ambient concentrations of $PM_{2.5}$.*

PCM consists of a mixture of empirical and simple sub-models, though in practice its implementation is quite complex. PCM was not part of the urban evaluation analysis within Phase 1 of the Defra Model Evaluation Exercise, comparing air quality model performance (Carslaw 2011a). PCM's performance for calculating the pattern of annual average $PM_{2.5}$ across the country would probably be regarded as acceptable, but its use in the apportionment[4] of $PM_{2.5}$ to sectors or the attribution[5] of $PM_{2.5}$ to individual sources has not been evaluated. This must be regarded as one of the weaknesses of the COMEAP report.

The CMAQ model has been used internationally and is undergoing extensive international comparisons.[6] Its performance with regard to source attribution cannot be said to be fully evaluated, but its performance over a much wider range of conditions and data sets, involving the use of advanced observational networks, suggests that:

- it is a better tool for evaluating source apportionment and source attribution than PCM;
- results from the CMAQ model will have more international standing than the application of PCM alone.

Previous Environment Agency studies (Abbott and Vincent 2006, Abbott *et al*. 2006, Vincent and Abbott 2008) have assessed the benefits of its regulation in the areas of air quality and acid deposition using the semi-empirical TRACK-ADMS model. One example of the model's use is an estimate of the long-term benefits of a reduction in exposure to PM_{10}. Results suggest that:

- the average PM_{10} and $PM_{2.5}$ concentrations in the UK would have been 3μg/m^3 higher in 2005 without the emission regulations implemented between 1990 and 2005;

[4] Apportionment refers to that fraction of $PM_{2.5}$ which arises from the whole of a source sector or category (for example, power stations or road transport emissions).
[5] Attribution is taken to mean how much of the ambient $PM_{2.5}$ a specific source (for example, a particular power station or oil refinery) is responsible for.
[6] For examples of CMAQ applications, see the AQMEII intercomparison and presentations associated with the USEPA/Defra/Environment Agency Workshop: 'Improving and Applying Local, Regional and Global Air Quality Models' held in London, 6–10 December 2010. The presentations from this meeting are available on the Institute for Air Quality Management web site (http://www.iaqm.co.uk/resources.html).

- Environment Agency regulatory actions between 2010 and 2020 are expected to yield a further reduction in life years gained of approximately 230,000 life years.

These studies show that relative simple models can be used to estimate the outcomes of policy measures, but for greater confidence and for consistency, there is a need to establish that other more internationally recognised models would have obtained broadly similar results.

Q2 How should the Environment Agency address regional trends in air quality under changes in emissions?

Answer
The CMAQ model is at appropriate tool for the Environment Agency to investigate changes in ambient concentration and deposition under emission conditions markedly different from current emissions.

The decision to use an air quality model for policy and planning must be regarded as partly subjective. Both Defra (Derwent *et al*. 2010), and informally the USEPA (Dennis *et al*. 2010), have published model evaluation protocols. Benchmarking procedures (Thunis *et al*. 2011a) have been produced, but air quality models, not least because of the limitations on observational data, have not generally been subject to rigorous analysis. The collaborative project FAIRMODE (Forum for Air Quality Modelling in Europe 2011) has attempted to interpret the required performance of air quality models in the Air Quality Directive in its draft guidance. It was therefore not possible to state which performance measures should be met in the CREMO project. Hence a pragmatic view was taken, that modelling groups which contributed to CREMO should be able to demonstrate that their results meet the acceptance criteria in the published Defra Air Quality Model Evaluation Protocol (Derwent *et al.* 2010), most of the time.

Preliminary results from the ongoing Defra Model Intercomparison Exercise suggest that most of the models commonly used in the UK, when evaluated against airborne gas concentrations and acid deposition on local and regional scales, meet the acceptance criteria in its protocol. This would suggest that the criteria cannot be used to discriminate between the quality of models. Hence the usefulness of more complex models, such as CMAQ, possibly lies in the other roles that they can play, such as greater scientific understanding, emission sector apportionment and source attribution.

The performance of the CMAQ model in calculating background concentrations and deposition for a recent year, when measurements were available, was evaluated in Phase 1 of Defra's Model Evaluation Analysis using its Model Evaluation Protocol (Derwent *et al*. 2010). However a comparison of $PM_{2.5}$ predictions with observations was not included. The Defra Model Evaluation has attempted to follow the advice in the report by the Defra Science Advisory Council[7] on environmental modelling (Ferguson and Harrison 2010) which addressed desirable, behavioural issues related to environmental models (strategy, good practice and review) rather than technical ones. It is fair to say that other organisations, such as the USEPA, invest more heavily in evaluating the performance of models than organisations in the UK. Full evaluation involves considerable resource and effort, which rarely achieves the ambitions outlined

[7] It is not clear whether the Science Advisory Group was familiar with complex fundamental deterministic models such as CMAQ, and their use and evaluation, although the report reviewed climate models.

in the Science Advisory Council report. Development of more complex modelling, as a consequence of high performance computing, may make full evaluation increasingly difficult.

Preliminary results from a number of versions of CMAQ suggest that it:

- generally meets acceptance criteria in the Defra Model Evaluation Protocol;
- has comparable performance to other models in common use within the UK.

CMAQ's main advantage is that source apportionment and attribution for $PM_{2.5}$ components can be calculated directly within the model. Its main disadvantage is that, as a fundamental model, to run CMAQ requires considerable computer resource and extensive input data sets. Within Phase 1 of the Defra Model Evaluation Exercise a number of different configurations of the CMAQ model were successfully run by four groups in the UK and the results submitted for analysis**.**

The optimal CMAQ configuration for assessing $PM_{2.5}$ apportionment and attribution is an important consideration before CMAQ calculations are carried out. Any consultant chosen by the Environment Agency to run CMAQ should be able to demonstrate that its chosen configuration allows CMAQ runs to be made in an efficient way in terms of run time, set-up costs, *etc*. It is necessary to consider optimised formulations of the CMAQ model to make it as practical as possible. Specific decisions about the number of vertical layers in CMAQ and the link from global scale to regional scale have to be made.

CMAQ can also be used to investigate changes in emission categories or sectors, and individual source scenarios, in order to form air quality policy decisions relating to $PM_{2.5}$ for processes regulated by the Environment Agency. Due to CMAQ's fundamental treatment of processes one should have greater justification for using the CMAQ model to assess concentrations and depositions under future emission scenarios than other more highly parameterised air quality models. By 2020 there could be significant changes to global background concentrations, so that source-receptor relationships may be different from those in recent years. However the study of future scenarios has not yet been undertaken and is a topic for future investigations.

Q3 Is there an accepted configuration of the CMAQ model which balances the need for fast efficient running against the requirement to produce accurate detailed concentrations in space and time?

Answer
Complex chemical transport models such as CMAQ usually require more time to run than is usual in Environment Agency assessments. The optimal configuration for running CMAQ (for UK regulatory purposes) is not known, but considerable experience towards developing this goal has been obtained during the CREMO project.

Groups collaborating with the Environment Agency already have considerable experience in running CMAQ over long periods, typically annual runs, under a number of different conditions, but the optimal configuration to give adequate performance at minimum computer cost is not known.

The team at the University of Hertfordshire has compared the conditions applied in the various meso-scale models, such as the CMAQ modelling system and the European

Monitoring and Evaluation Programme (EMEP) model. For example, the team was able to compare 2006 CMAQ runs for AQMEII, which were at a 18 km resolution, with previous 2003 runs at 5 km resolution.

There is a wide choice of configurations, with variations in the horizontal and vertical size of grid cells, apart from differences in meteorology, emissions, boundary conditions and chemical schemes. One could easily visualise 3^n, where n is 4 or more different versions of the same calculation! The differences between the different versions of the same CMAQ calculation are therefore not surprising.

The optimal set-up (the set-up requiring the lowest computer resource and least preparation) for running CMAQ efficiently to answer a policy question or a regulatory issue has not yet been selected.

Q4 What process should the Environment Agency follow to determine whether a model is suitable for use as a regulatory tool?

Answer
The Environment Agency needs to decide on (a) an agreed evaluation protocol, (b) a metric which should be used in the assessment and (c) which models are acceptable for the use envisaged.

These are complex issues. There are a number of stages in the evaluation procedure. Stage one is evaluation by the developer or model reviewer. For CMAQ, the developer is the USEPA. Extensive evaluation has been undertaken in collaboration with the USEPA through Phase 1 of the Defra Model Evaluation Exercise and AQMEII.[8] Figure 2.1 illustrates the stages in an ideal scientific model evaluation.

The AQMEII evaluation protocol based on Dennis *et al*. (2010) has four stages or types of evaluation. For Environment Agency purposes, there should be a fifth stage in the evaluation which relates to policy. This concerns how the other stages of an evaluation should be used to make decisions and regulate air pollution.

However the type of a model used at the policy stage may not be the same as that used in the scientific stages of model evaluation. Instead it could be a model emulation.[9] It is vital not to miss out the fourth stage, the diagnostic evaluation, to get

[8] The Air Quality Modelling Evaluation International Inter-comparison. See http://aqmeii.jrc.ec.europa.eu for more details. The AQMEII model evaluation framework is described in Dennis *et al.* (2010).

[9] For explanation of an emulator, see http://mucm.aston.ac.uk/MUCM/MUCMToolkit/index.php?page=MetaOverviewEmulators.html. If a computer simulation is computationally expensive, so that evaluating a simulation model for a choice of input data takes a significant amount of computing time, one may be limited to evaluating the model at a small number of sets of different input data. This can make various analyses difficult, as one may want to know model predictions at a large number of different input values. One can deal with this problem by building an emulator: a statistical model of the model, constructed from a fairly small number of runs of the simulation model. The emulator will both predict output values, and report uncertainty in any prediction. In the Managing Uncertainty in Computer Models (MUCM) project, a toolkit for constructing Gaussian process emulators is described in which all parameter values and their interactions have uncertainty described by Gaussian functions. Gaussian functions possess convenient properties, making emulator

directly to the policy stage. The diagnostic evaluation should lead to an **understanding** of:

- how the model is working;
- its underlying principles;
- which processes are most significant;
- how they interact.

This process is not easy to specify in a formal way. Crucial to understanding is the interaction between the forcing terms in the system, primarily emissions, and non-linearities and feedbacks.

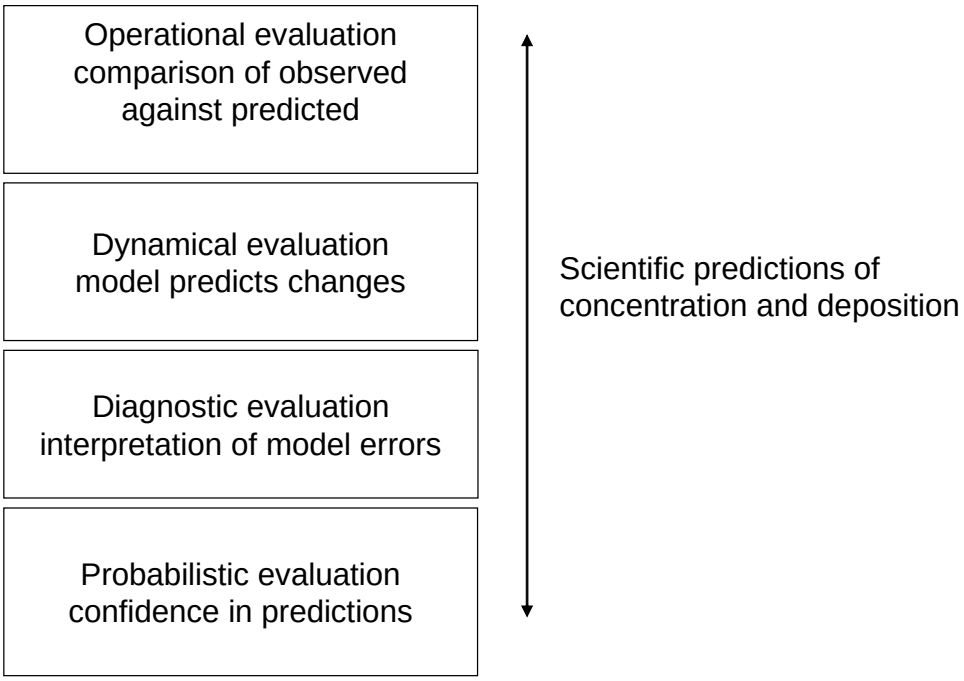


Figure 2.1 Scientific model evaluation or assessment framework for a model developer or reviewer

Figure 2.2 shows an ideal evaluation framework of a model which a policy-maker should apply when using an air quality model. Scientific application of the model or integrated assessment is the final part of the process, but this does not imply that this stage should necessarily be done by the policy-maker.

If a model is computationally expensive, so that evaluating a model for a single choice of input parameters takes a considerable amount of computing time, the evaluation of the model may be limited to a small number of different input values. This can make some analyses difficult, as one may want to know model predictions at a large number of different input values. It is possible to deal with this problem by building an emulator – a statistical model of the original model, constructed from a fairly small number of runs of the original model. The emulator will both predict model output values and report uncertainty in any prediction. The role of emulators should be considered by

---

approximation easy. Other approximate models, which could be considered emulations of full simulation models, are described later in this report.

those faced with computationally expensive, three-dimensional regional chemical transport models, such as CMAQ.

As an example of the emulation methodology, a statistical emulation of the moderately complex meso-scale model, The Air Pollution Model (TAPM)[10] was used to approximate the consequences on health of a notional power station emission (Fisher *et al.* 2010). The health outcome was expressed as the number of life years lost as a consequence of one year's operation of the power station source under a range of emission conditions, without the need for a large number of runs of TAPM.

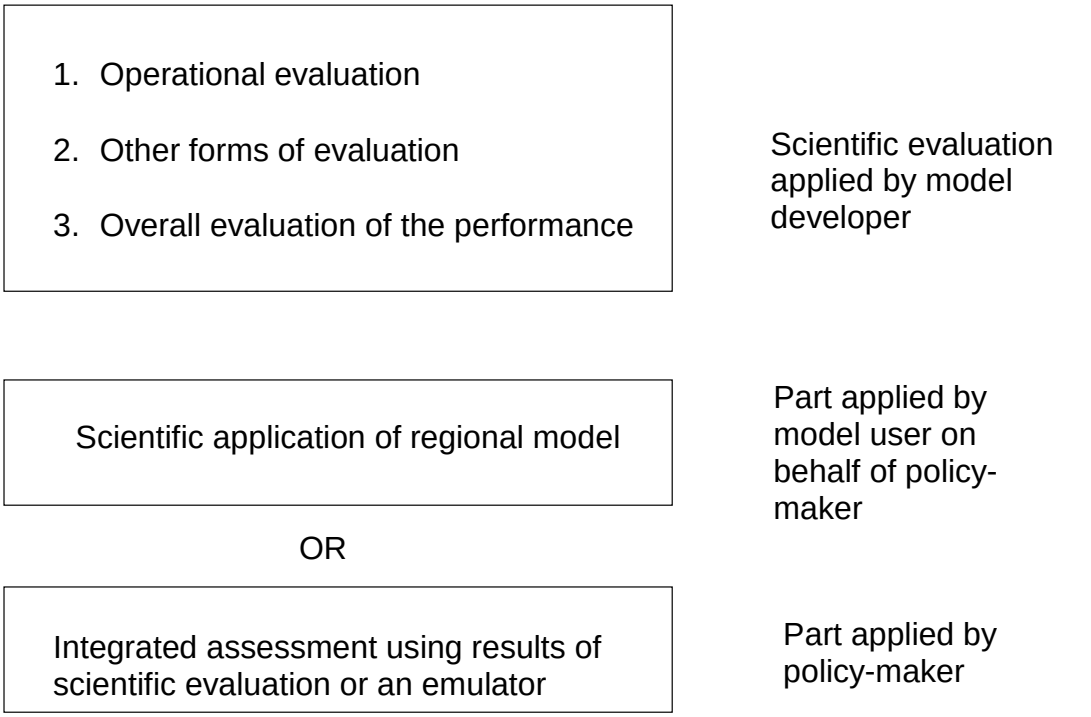


Figure 2.2 A model evaluation framework for a policy-maker showing the relationship to the model user

A similar kind of emulation was used in connection with the chemical transport model CMAQ. The USEPA has developed a Response Surface Model (RSM), which is a representation of the response of ozone and $PM_{2.5}$ to changes in emission. In this model, numerous pre-specified individual air quality modelling simulations are aggregated into a multi-dimensional air quality 'response surface'. RSM incorporates statistical relationships between model inputs and outputs that can be used to develop emissions control scenarios. The model required hourly multi-pollutant emissions for a base year and future years on a 36-km national grid of the following pollutants:

- carbon monoxide (CO);
- nitrogen oxides (NOx);
- volatile organic compounds (VOCs);
- sulphur dioxide (SO_2);

[10] The moderately complex meso-scale model, TAPM, was applied with UK emissions and meteorology in 2003 to illustrate the approach. The model was used as an efficient way to generate a set of results distributed through parameter space. It would have been difficult at the time of the calculation to have generated enough CMAQ results to have been able to build an emulator.

- ammonia (NH_3);
- particulate matter less than or equal to 10 microns (PM_{10});
- particulate matter less than or equal to 2.5 microns ($PM_{2.5}$).

Originally RSM was developed from several hundred runs of the air quality model CMAQ with a 36 km horizontal domain, which encompassed the contiguous USA for the base year 2001 and more recently for the base year 2008 using the versions of CMAQ available at the time the runs were made.

The RSM experimental design covers a change in the baseline emissions of zero to 120 per cent using likely emissions projections to provide annual $PM_{2.5}$ concentrations, visibility and deposition estimates. Once RSM has been generated, it can be used to simulate the functions of a more computationally expensive atmospheric chemistry model.

RSM can be used to derive analytical representations of model sensitivities to changes in model inputs. It is designed to show how CMAQ air quality model predictions of the atmosphere respond to emission reductions for selected sources and pollutants. With RSM it is possible to evaluate the air quality changes that result from adjusting emissions control on the precursors NOx, SO_X, NH_3, elemental carbon (EC), organic carbon (OC) and VOCs for a variety of source sectors, such as electricity generating units, point sources, area sources and mobile sources.

An analysis concluded that ambient $PM_{2.5}$ in each of nine urban areas in the USA is largely independent of the precursor emissions in all the other urban areas (USEPA 2006). Thus RSM allows the analysis of air quality changes in these urban areas and associated counties independently of one another. The limitations of the approach are that, as a statistical model, RSM does not reveal understanding of the relationship within the model, and its application is confined to the conditions and emissions scenarios described by variations around the base case.

In addition to the four well-established phases (operational, dynamic, probabilistic and diagnostic), the policy and decision support phase is a fifth phase within the model evaluation framework which broadly encompasses the model evaluation framework for policy makers outlined in Figure 2.2. The CREMO project has dealt extensively with operational evaluation and to some extent with diagnostic evaluation, but has not addressed dynamic and probabilistic evaluation. These forms of evaluation are constrained by limits to the observational data base for testing models and the computational resource required to undertake many runs of a complex model. One should not ignore the varying complexity and the differing number of parameters associated with different kinds of models.

This is illustrated schematically in Figure 2.3, which envisages that models can be classified into generalised types of model such as Eulerian, statistical and empirical models. Empirical models rely on incorporating observed data into the model from an extensive monitoring network. EMEP refers to the three-dimensional regional chemical transport model applied within the EMEP programme which has similarities to CMAQ. It can be run over a European or a UK domain. An outline of the computational resources needed to perform a run of the CMAQ model is given in the Appendix.

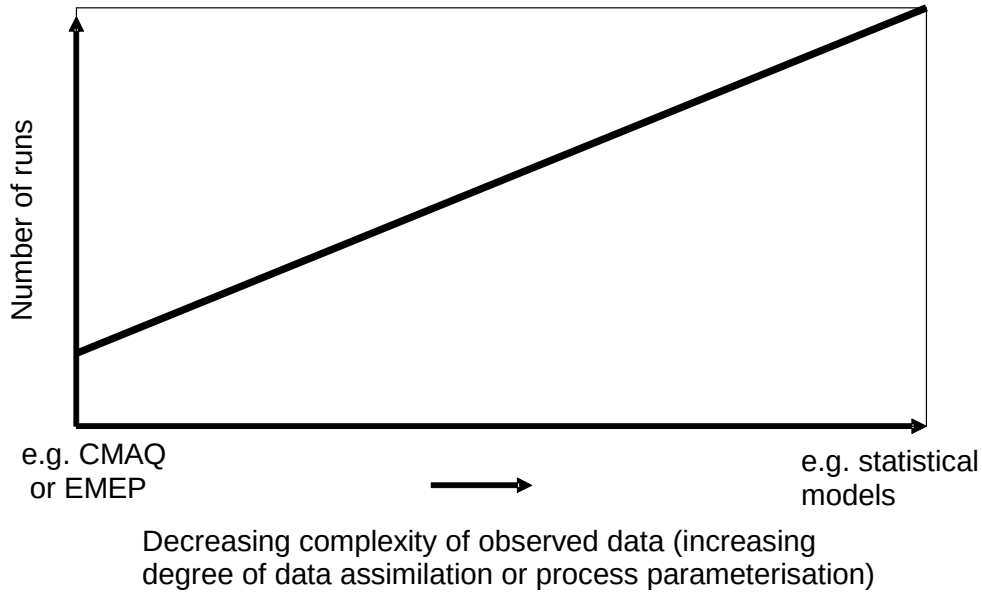


Figure 2.3 Relationship of the number of model runs able to be performed to degree to which model depends on observed data for a fixed computing resource

Q5 Has a formal evaluation procedure been carried out within CREMO?

Answer
The formal procedural approach to model evaluation, outlined in the answer to question 4, has been tested in CREMO. It turns out that there is no simple, single acceptance criterion for judging whether an air quality model is suitable for use.

The CREMO project evaluated the performance characteristics of CMAQ for possible regulatory application so that the Environment Agency could make an informed decision as to whether to incorporate advanced modelling as one of its assessment tools.

The different versions of the CMAQ model were compared with three existing air quality models:

- FRAME for acid deposition;
- TRACK-ADMS for annual audits;
- OSRM (related to IDOP[11]) for ozone.

[11] IDOP stands for integrated downwind ozone production. IDOP enables a comparison of VOC species, or emission sources, in terms of how much ozone is produced in the downwind environment under ideal conditions for producing ozone. It is based on selected runs of the Photochemical Trajectory Model, of which the Ozone Source–Receptor Model (OSRM) is a generalisation. IDOP was not used in the intercomparison but is mentioned here as it is a convenient, simplified method for regulatory purposes when screening VOC emissions. It is described by Derwent and Nelson (2003). In cases when the IDOP approach suggests that an emission is not acceptable, more sophisticated approaches such as OSRM or CMAQ should be applied to test the conclusion.

In CREMO, CMAQ was applied in a series of assessments (including acid deposition, particulate matter and ozone) to test its capabilities through targeted comparisons with simpler models and with measurements, according to agreed model acceptance criteria (Hayman *et al.* 2012a). The report (Hayman *et al*. 2012a) also lists the sites, measurements and methods available in the UK for evaluating models against observations, as well as summarising the models. The aim of CREMO was not to establish which model had the best performance judged against observations, but to establish whether the models meet acceptance criteria, meaning that they would be suitable for use in developing air quality management plans or deciding on future air quality strategies or programmes.

An evaluation protocol involves proposing suitable metrics or concentration /deposition diagnostics. If a model complied with the evaluation protocol, then the model is suitable for use in making decisions. An evaluation protocol is thus a formulation of model acceptance criteria, a number of which have been published in the literature. The criterion given in the Air Quality Directive (EC 2008) is of great importance, because of the compliance requirement imposed on EU countries. The Directive accepts that models can be used for air quality assessment to:

- reduce the number of sampling sites;
- prepare plans and abatement measures;
- determine where pollution is coming from (source apportionment) as a component of a plan.

Table A of Annex I of the Air Quality Directive contains data quality objectives for ambient air quality assessment. It states that:

> *The uncertainty for modelling is defined as the maximum deviation of the measured and calculated concentration levels for 90% of individual monitoring points, over the period considered by the limit value (or target value in the case of ozone), without taking into account the timing of the events. The uncertainty for modelling shall be interpreted as being applicable in the region of the appropriate limit value (or target value in the case of ozone). The fixed measurements that have to be selected for comparison with modelling results shall be representative of the scale covered by the model.*

This criterion, which a model should meet, is not easy to implement. The guidance published by FAIRMODE (2011) and guidance on NO_2 (Denby 2011) attempt to interpret the criterion, so that it can be implemented. The FAIRMODE view is that the modelling community is so broad and the applications of models are so varied, that there is not yet a clear and common understanding of what can be regarded as ‘best practice’ in modelling and refers extensively to ‘good practice’ rather than ‘best’. The interpretation of the Directive text is difficult. Basically the criterion should be expressed as a mathematical formula.

The term ‘model uncertainty’ also remains open to interpretation. FAIRMODE suggests the following interpretation, which it calls the ‘Relative Directive Error’ (*RDE*), defined mathematically at a single station as follows:

$$RDE = \frac{\left|O_{LV} - M_{LV}\right|}{LV}$$

where O_{LV} is the closest observed concentration to the limit value and M_{LV} is the corresponding ranked modelled concentration.

This is not an easy concept to explain in words. If PM_{10} is considered, the limit value *LV* is the daily average concentration of 50 µg/m^3 (not to be exceeded 35 times). The daily concentrations of the cumulative probability distribution of measured daily observed concentrations can be plotted against the cumulative distribution of the daily modelled concentrations in a quantile–quantile plot in which concentrations at given quantiles (probabilities at selected intervals between 0 and 1) are plotted against one another. M_{LV} is then the (ranked) modelled concentration which corresponds to the observed concentration O_{LV} closest to the limit value *LV*.

Figure 2.4 illustrates the concept of the Relative Directive Error using a specific example of the probability density functions of the model predictions and observations at a sampling site. The observations indicate that there are, say for the purposes of this example, 62 daily readings above a limit value of 50. The 62nd highest observed reading is O_{LV} close to the limit value 50 amongst the 365 daily readings. From the probability distribution of modelled concentrations, one finds the corresponding predicted daily concentration M_{LV} which is the 62nd highest concentration. Then *RDE* is defined by the formula given above.

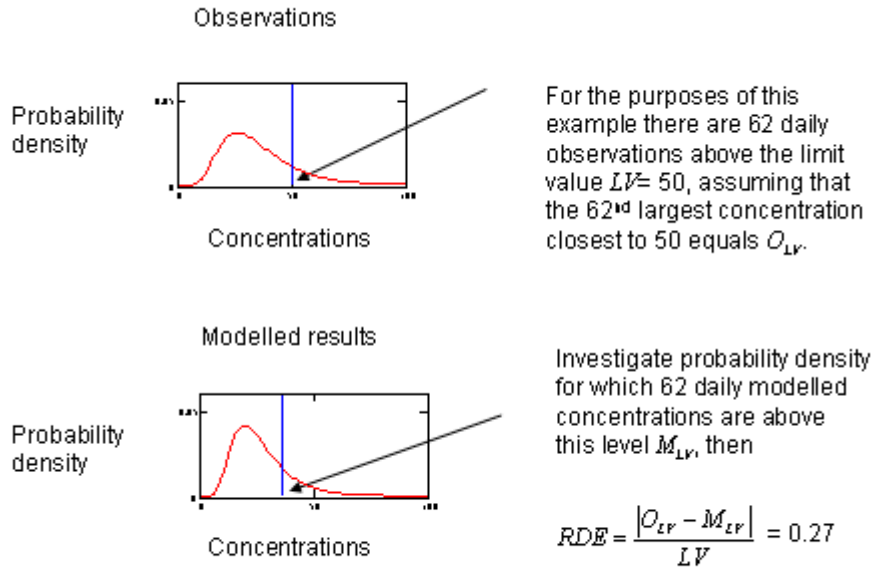


Figure 2.4 Illustration of the concept of *RDE*

For the annual average PM_{10} limit value, the annual average observation is compared with the annual average prediction and the concept of an observation close to the limit value does not apply. Then *RDE* equals the difference between the observed and predicted annual average concentrations divided by the limit value. In this case, *RDE* is not purely a measure of model performance. Its value will also depend on whether the limit value is being achieved. This illustrates the care needed when considering the normalisation of metrics and that no single metric is always appropriate in every situation.

FAIRMODE refer to an alternative formulation known as the Relative Percentile Error (*RPE*), which is defined at a single station, as:

$$RPE = \frac{|O_P - M_P|}{O_P}$$

where O_p and M_p are the observed and modelled concentrations at the quantile p or percentile ($P=100p$) level, used to define exceedance of the limit value.

In the example illustrated in Figure 2.4 of daily PM_{10} concentrations, p equals 35/365 and P equals 35/365x100 per cent. The definition of *RPE* is illustrated graphically in Figure 2.5 in which the air quality standard specifies that 35 daily readings are allowed to exceed the limit value out of 365 readings. In the probability density distribution of modelled concentrations, 35 daily modelled concentrations exceed a value of M_p. Then *RPE* is defined by the formula given above.

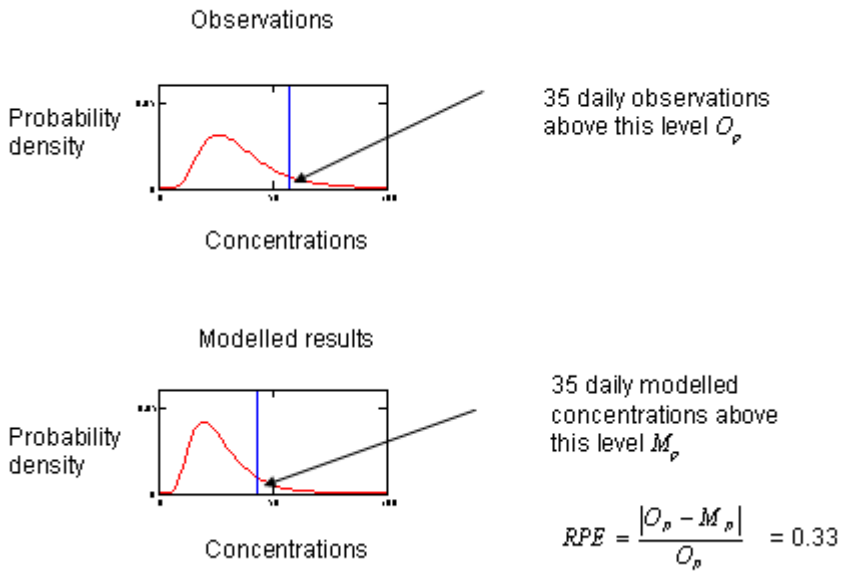


Figure 2.5 Illustration of *RPE*

The Air Quality Directive states that:

- for daily and annual average NO_2 concentrations, the model uncertainty (interpreted in FAIRMODE as either the *RDE* or the *RPE*, in order to specify a number) has to be less than 50 per cent;
- for the annual average NO_2 concentration, the model uncertainty (taken here to be either the *RDE,* or the *RPE*) has to be less than 30 per cent;
- for daily average particles $PM_{2.5}$ and PM_{10}, the model uncertainty has yet to be defined;
- for the annual average particle limit value, the model uncertainty (taken here to be either *RDE* or *RPE*) has to be less than 50 per cent.

FAIRMODE (2011) points out a further complication. This is that the Air Quality Directive states that the uncertainty should be determined from the maximum of 90 per cent of the available monitoring stations. This is interpreted to be a clause that allows any outliers (that is 10 per cent of the monitoring stations) to be excluded from the uncertainty calculation. If taken literally, this would mean that any model domain with less than 10 stations would not be able to exclude any outliers. Because there are many urban areas where there are less than 10 stations, the FAIRMODE report argues that the Air Quality Directive must mean that all the stations should be used to assess model uncertainty when the number of suitable stations is less than 10. It is worth noting that only stations representative of the same spatial scale (urban, background, kerbside or rural) as the air quality model are to be applied in the uncertainty assessment.

The specific measure of uncertainty set by the Air Quality Directive can be criticised on operational grounds. While setting a specific performance limit as a regulatory requirement for air quality models is an attractive way to ensure models are used fairly and consistently, model uncertainty cannot be managed in this way because models are diverse, are subject to continuous development, and cannot be subject to managerial control. It would be better if the Air Quality Directive set out the principles by which models are judged and that specific acceptability criteria were the subject of informal guidance, which could be regularly updated.

The Defra Model Evaluation Protocol (Derwent *et al*. 2010) sets the following criteria. The predictions of the model should be accepted if:

- the percentage of model predictions within a factor of two (*FAC2*) of the observations is greater than 50 per cent;
- the magnitude of the normalised mean bias (*NMB*) is less than 0.2.

The normalised mean bias (*NMB*) is defined as:

$$NMB = \frac{\sum_{i=1}^{N} M_i - O_i}{\sum_{i=1}^{N} O_i}$$

where N is the number of observations, M_i are the calculated values, O_i are the observed values. *NMB* should satisfy $-0.2 \leq NMB \leq 0.2$.

The *NMB* will artificially put a higher weighting on higher concentrations and so the mean normalised bias, *MNB,* is a better metric for judging model performance. *MNB* is defined as:

$$MNB = \frac{1}{N} \sum_{i} \frac{M_i - O_i}{O_i}.$$

However the normalised mean bias, *NMB*, has been used in a number of recent model evaluations and the air quality model performance using this metric is illustrated in later sections of this report. It is also important to consider carefully the quality of the observations and the size of the sample *N*, the number of observed–calculated pairs. Care must be taken when applying criteria. (For example, see the discussion of the *AOT40* and *SOMO35* metrics under Q9. Any criterion that includes a cut-off can have a very large bias because of the normalisation factor. It would be better to normalise such metrics by the limit value, or some other fixed value.)

Phase 1 of the Defra Model Evaluation Exercise showed that air quality models generally comply with the *FAC2* criterion (that is 50 per cent within a factor of two). An earlier comparison of simple regional transport models by the Environment Agency (Abbott *et al.* 2001) for acid deposition also suggested that all regional transport models could be expected to meet the *FAC2* criterion. The 95th percentiles of the predicted deposition rates for the simple models available at the time of the study, TRACK, Hull Acid Rain Model (HARM) and FRAME, were within a factor of two of the annual average value and the 5th percentiles were within half the annual average value, so the uncertainty bounds found in the recent Defra Model Evaluation Exercise are to be expected. Only the *NMB* criterion provides a way of discriminating between the performance of air quality models. The advantage of the more complex operational models, such as CMAQ, is their ability to test their performance for short-term average concentrations (daily and hourly) and to be able to investigate complex interactions and feedbacks between processes more fully. The performance of the CMAQ model published as part of the CREMO study (Chemel *et al*. 2010) suggests only moderate

performance when calculating PM_{10}. Particulate matter was not considered in Phase 1 of the Defra Model Evaluation Exercise, so definite conclusions cannot be drawn about this component.

The set of possible statistical measures for evaluating model performance listed by Chemel *et al.* (2010) includes the percentage within a factor of two (*FA2*) and the normalised mean bias (*NMB*). For each performance measure, a value could be specified that could be regarded as 'good', based on previous evaluation studies. This process would lead to a generally accepted standard based on experience, though there may be no rigorous basis for such a standard.

Published results from Defra model evaluations are part of this process. They are one step towards performance measures, which all models subsequently used for assessment purposes would be expected to meet. The results from Phase 1 of the Defra Model Evaluation Exercise suggest that most of the models commonly used in the UK, when evaluated against airborne gas concentrations and acid deposition on local and regional scales, meet the Defra Model Evaluation Protocol *FAC2* and *NMB* acceptance criteria most of the time, and would be regarded as acceptable. A comparison between models and observations for $PM_{2.5}$ has not yet been completed, so to set an acceptance criterion for $PM_{2.5}$ would be premature.

As part of a model intercomparison within CREMO (Chemel *et al*. 2011), all the models considered (CMAQ v4.6, CMAQ v4.7, TRACK-ADMS and FRAME) satisfy the *FAC2* criterion that 50 per cent of the modelled results should be within a factor of two for the annual mean concentrations of the species considered. In contrast, none of the models satisfy the criterion that the normalised mean bias *NMB* should be in the range -20 to 20 per cent for all the species considered, which includes PM_{10}. This variable performance requires further investigation. The *NMB* criterion may not be easily satisfied by all outputs (all chemical species) from a model.

Another metric explored in CREMO is the national budget, such as the fraction of the national emission deposited over the country or the spatial average of the concentration over the country (possibly expressed as the population weighted mean) as illustrated in Chemel *et al.* (2011). For model evaluation, an empirically derived estimate of the predicted quantity is therefore required. The Concentration Based Estimated Deposition (CBED) model (Smith *et al.* 2000, Smith and Fowler 2001) fulfils this role. Broadly speaking, it is an empirical model based on interpolating observations. It is used to produce spatial patterns and budgets from observations. CBED contains assumptions regarding the way wet deposition and concentrations are interpolated between monitoring sites, so that its predictions should not be regarded as purely empirical estimates. It is not possible to assess the accuracy of CBED estimates directly, but a Monte-Carlo analysis, in which key parameters are varied over plausible ranges, would give an indication of its accuracy.

The acceptance criteria for long-term average concentrations or depositions predicted by models should be simple, pragmatic and applicable to the policy or regulatory goals. The criteria based on *FAC2* and *NMB* can be readily applied in operational evaluation when there are some observational data sets.

Q6 What types of diagnostic metrics have been used in CREMO?

Answer

A footprint of $PM_{2.5}$ concentration is a useful diagnostic metric for testing complex systems and for use in regulatory applications. It is of particular interest to the Environment Agency because of its requirement to regulate major stationary sources. It is one example of sensitivity analysis for complex models, which is an area of active research.

A footprint metric is a response function, showing how concentrations or deposition are influenced spatially by emissions from a single specified source, such as a power station. Figure 2.6 provides an example.

Footprints are obtained from the difference between the concentration or deposition, when all sources are included, and the concentration or deposition, when all sources, except for the specified source under consideration are included. In simple models the individual footprints can be calculated directly.

The footprint metric has two purposes:

- The footprint metric is a diagnostic showing how a system that includes very complex processes changes as a result of a change in emission at a specified position.
- It is useful for regulatory purposes as it shows how emission reductions may change the concentrations.

In principle, a reduction strategy should follow a sequential change in emissions, tracking which emission reductions are most effective.

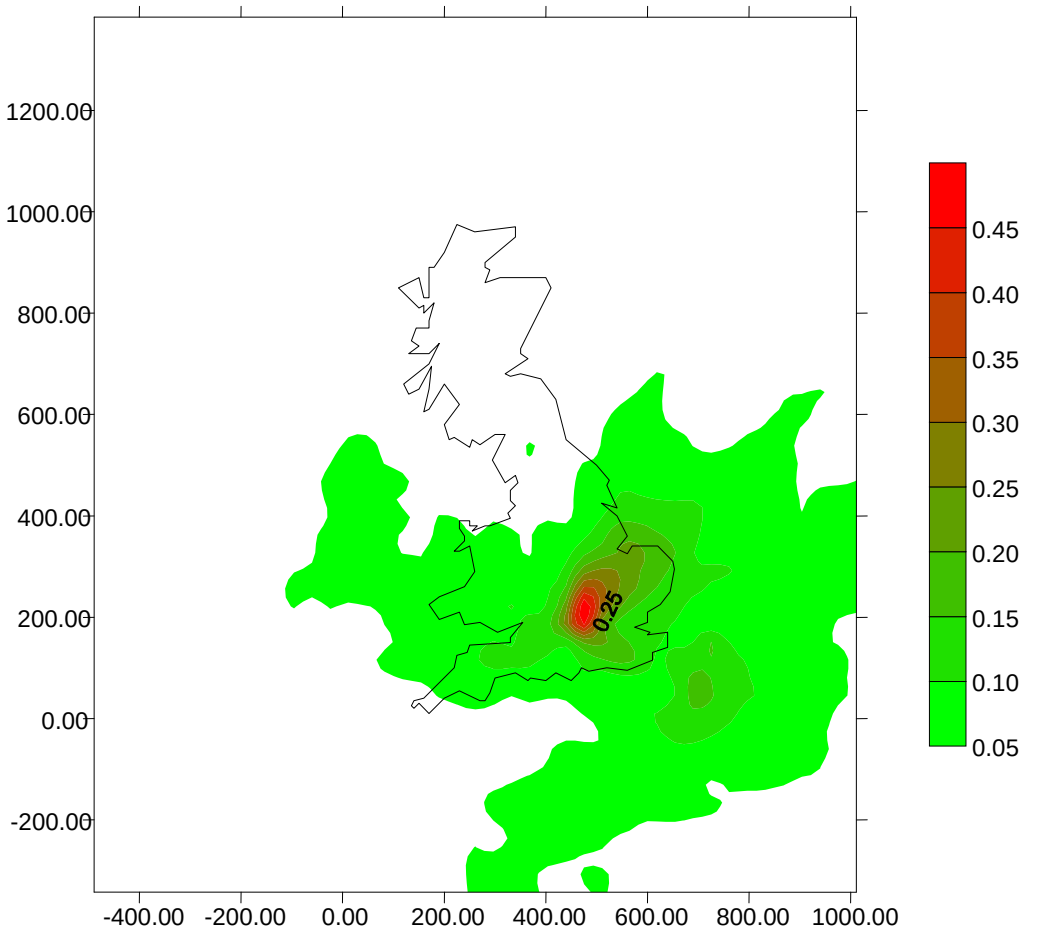


Figure 2.6 PM_{10} concentrations in 2003 from power station in southern England using the CMAQ model with a 15 km grid resolution

When making policy decisions, it is important to understand the underlying structure of the complex system under investigation. This means a systematic approach to the way the system responds to change. It may be possible to scale and multiply several simple footprints in order to assess a proposed emission reduction strategy with sufficient accuracy.

Thus if Δq_i is the reduction in the source strength of the i^{th} source under a reduction strategy scenario, an approximate estimate of the reduction in concentration may be given by ΔC, where:

$$\Delta C = \sum_i \Delta q_i G_i$$

and G_i is the footprint of the i^{th} source of unit source strength. (Such an approach would have to be evaluated first using a three-dimensional chemical transport model, such as CMAQ). The outcome from the application of footprint metrics should be some kind of integrated assessment, meta-model or emulation. This is the way a footprint can be applied to regulatory decision-making.

A step towards investigating source attribution and source apportionment for categories of sources is provided by the footprints from different types of industrial installations. By scaling emissions from representative sources within source sectors, one would hope to be able to extrapolate footprint results to obtain results for the change in emission of a whole source sector or for the overall change resulting from an emission control policy.

For policy applications, the EMEP source-receptor tables (NMI 2010) provide an example of national footprints (that is the influence of all the sources in one country on the whole area of a surrounding country. See discussion of Figure 2.14).

Using representative footprints, one can estimate approximately the contribution from a range of sources and it should be possible to develop other simple approaches. The word ‘simple’ in this context means that few equations are solved. However to compare the footprint shape of different models, it is necessary to apply some form of pattern recognition. It is not straightforward to describe the influence of a source on surrounding areas in terms of a (single) numerical parameter.

The footprint metric has been applied in a quantitative way in CREMO using the spatial correlation coefficient (r) and the coefficient of variation of the root mean square differences (CVRMSE) (see discussion of Table 2.6). CMAQ v4.6 was chosen as the reference model. With these quantitative measures (see Chemel *et al*.. 2011), differences between the footprints of the SO_2, NOx and PM_{10} concentrations, and the total deposition of non-sea salt sulphur and nitrogen have been compared for the different models.

Another quantitative approach to evaluating footprints is to consider the distance dependent structure of the footprint of a single source (Fisher *et al*. 2011). The weighted average concentration given by the average concentration along a typical trajectory, excluding approximately any dilution arising from dispersion, is obtained by multiplying concentration by distance. This footprint metric is defined by:

$$\frac{r}{2\pi}\int_0^{2\pi} C(r,\theta)d\theta$$

where $C(r, \theta)$ is the concentration at distance r from the source in a direction θ.

The footprint metric is scaled by a value of the concentration near to the source and is illustrated in Figure 2.7. The dependence of the PM_{10} concentration along a radial trajectory does not decrease rapidly with distance. This can be interpreted to arise because of the gradual production of secondary aerosol in the atmosphere. A single numerical diagnostic is the distance between the source and the point at which the weighted secondary PM_{10} concentration is a maximum. This diagnostic then summarises the influence of the source on secondary aerosol formation. The differences between different regional models can be compared by considering the dependence of the weighted average concentration dependence on distance and determining the distance at which this is a maximum.

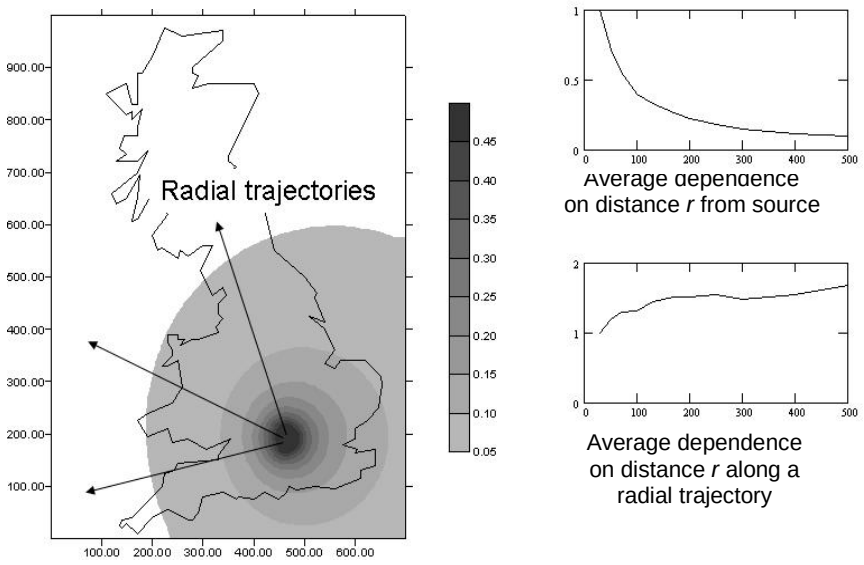


Figure 2.7 Illustrative example of the annual average PM_{10} concentration footprint in µg/m^3 from a major stationary point source such as a power station

The variation in concentration along a south to north or west to east transect through the source, as illustrated later by Hayman *et al*. (2012d), would decrease much more strongly with distance. This is because the spread of air mass trajectory ensures that a transect contains a factor proportional to the inverse of the distance from the source (1/*r*).

Q7 How should the Environment Agency make use of air quality models?

Answer

Environment Agency decisions regarding the purposes for which different types of models are used should be based on evaluations in CREMO and similar international evaluation studies.

Progress over the past decades has led to the development of air quality models, which in turn has raised issues, such as which model is most appropriate for a given application or whether model results should be combined with observations? A decision by the Environment Agency regarding which purpose the different types of models should be used for, ought be based on evaluations in CREMO results and other similar international studies. A different model might be applied in an operational situation to that applied in a policy situation depending on the context. The application could be to assess current and future budgets, to look at future regulation or to

undertake source apportionment. The decision-maker may use a specific model for purely practical reasons or convenience.

A pragmatic view should be taken about the differences in model predictions. If two models give similar results within an acceptable margin or error, then either may be used. Their value in determining practical outcomes should be the guiding principle. Adopting this approach in CREMO means that one should take a broad view of how models are used.

The driving force in Europe is the application of models to demonstrate compliance with Air Quality Directives and whether measures to meet limit values at future dates will be regarded as sufficient, if based on the models considered. Account needs to be taken of ongoing evaluation studies, such as Phase 2 of the Defra Model Evaluation Exercise, the planned Defra CMAQ evaluation study and further Environment Agency studies of $PM_{2.5}$.

Regulators face the problem of applying increasingly complex models. In principle it is desirable that the results of a calculation can be replicated. However this becomes increasingly difficult as models become complex. The detailed configuration may be difficult to set up on different computer platforms and decisions about which input data sets and options to apply become complicated to document.

There is also a case for using an ensemble of models, assuming that the models are as independent as possible. This has been considered extensively in the AQMEII model intercomparison (Solazzo *et al*. 2012a), in which the results of seven North America models and 11 European models were evaluated. The average of a model ensemble can give better results than individual models when there is cancelling of errors. The ensemble can also provide an estimate of the uncertainty in complex model predictions, an estimate not easily made available by other methods.

In AQMEII, for the long-term mean ozone concentrations, the average of a suitably selected *subset* of the models – typical bias about 2 parts per billion (ppb) – performed better than the ensemble average of all the models (typical bias about 5 ppb). In the AQMEII intercomparison, all the models were of a similar type, namely three-dimensional chemical transport models using a mesoscale meteorological model to drive the calculations. It is possible to identify models that were not independent of each other because they incorporated almost identical model components. This is important, otherwise the ensemble spread will not give a good indication of uncertainty. It was also recognised that the boundary conditions at the edge of the North American and European model domains can give rise to uncertainty in the long-term mean ozone. This could explain the biases in long-term mean ozone found, which no amount of averaging of individual models would eliminate. Within the CREMO intercomparison there are only a limited number of models of a different type, so one cannot develop an ensemble estimate.

One can also combine model results with observations to produce assimilated predictions. However any kind of assimilation requires knowledge of uncertainties in the model and in the observations: the bigger the uncertainty in a model, the lower the weighting on the model prediction. No conclusions have yet been made on the advantages of aggregating different models and observations, but results from the various model intercomparison exercises under way should eventually lead to decisions on the correct way to use these approaches.

Much attention has also been paid to computationally simpler ways of assessing source apportionment and source attribution to avoid re-running complex models many times. Most widely used is the decoupled direct method (DDM), which has been

implemented in three-dimensional air quality models, such as CMAQ, to calculate first-order sensitivities with respect to emissions, and initial and boundary concentrations. Each computer run of the model includes auxiliary equations evaluating the sensitivity coefficients (the change in concentration to a change in input parameter value, such as the impact of a major power station source.) This is similar to the footprint approach discussed above. However the use of DDM avoids the brute force approach of needing to repeat complex calculations many times.

Implementing DDM in any model requires locating those modules (chemistry, advection, aerosol formation) where a non-linear response may arise, because in such steps the equations for the sensitivities have a different form than the equations for the concentrations. In the case developed by Dunker *et al*. (2002) for the three-dimensional chemical transport model CAMx, this required deriving new equations for the sensitivities from the equations in the chemistry solver and the advection scheme in the model. At each time step, integration of the sensitivity coefficients is carried out separately from, and only after, concentrations are integrated as part of the model. Napelenok *et al.*. (2006) applied the DDM method to CMAQ version 4.3, using the same numerical algorithms in DDM for the transport-related processes: advection, diffusion, deposition *etc*., as for concentrations in the base model. Attention was paid in this version of the DDM method to particulate formation, a process that is not linear. Aerosol formation and cloud dynamics were treated differently in the DDM version of the model, essentially by deriving additional equations for the derivatives of the parameters associated with these processes.

For secondary air pollutants, precursor emissions affect concentrations in non-linear and inter-dependent ways. First-order sensitivities describe the linear response of the model to a change in input parameters and higher-order sensitivities describe higher order responses. Second-order sensitivity coefficients are computed by differentiating the governing equations of the first-order sensitivities with respect to the parameter of interest. Sensitivity coefficients represent the responsiveness to infinitesimal perturbations. To asses larger perturbations around a base case (for example, to describe significant emissions reductions), second-order sensitivity coefficients should be included as part of a Taylor series expansion. Cohan *et al.* (2006) characterised the non-linearity of ozone response to NOx emitted from source regions using second-order sensitivities. Non-linearity was found to increase with the magnitude and emission density of a source region.

A recent alternative approach to assessing source apportionment and source attribution in a simple way has been the development of the adjoint model of CMAQ. This is a linear model describing the sensitivity of *input* values to changes in *output*, the reverse of the usual sensitivity analysis. Hence adjoint models correspond to running the base model backwards in time. This is advantageous as there may be one output, but numerous inputs. A typical example of the application of the adjoint method would be the assessment of the contribution of black carbon to health effects. Since the three-dimensional chemical transport models discussed in this report are very complex, deriving the adjoint model is usually done by applying automatic computer methods to differentiate each step in the sequence of numerical operations which make up the model. The latest version of CMAQ v5.0 is not available with a DDM option, nor is there an adjoint version yet, but work is going on to develop these tools.

A type of meta-model or emulation for the Sydney Greater Metropolitan Region area has been developed using sensitivity coefficients. It was developed using the three-dimensional chemical transport model CAMx, to evaluate the impact of new sources (Yarwood *et al*. 2011). The screening model relies on the higher order sensitivities of ozone to NOx and VOC emissions. By using the high-order decoupled direct method

(HDDM) on the model CAMx, it is possible to develop a parametric model of the impact of new sources, of the form:

O_3 impact = function (NOx emissions, VOC emissions, source location)

This equation is derived from a Taylor series expansion of the ozone response to extra sources of NOx (Q_{NOx}) and VOC (Q_{VOC}) for five typical locations in the Greater Sidney Metropolitan area, given by

$$\Delta O_3 = Q_{NOx}\frac{\partial O_3}{\partial Q_{NOx}} + Q_{VOC}\frac{\partial O_3}{\partial Q_{VOC}} + \tfrac{1}{2}Q_{NOx}\frac{\partial {O_3}^2}{\partial {Q_{NOx}}^2} + \tfrac{1}{2}{Q_{VOC}}^2\frac{\partial^2 O_3}{\partial {Q_{VOC}}^2} + Q_{NOx}Q_{VOC}\frac{\partial^2 O_3}{\partial Q_{NOx}\partial Q_{VOC}}$$

where the sensitivity coefficients $\frac{\partial O_3}{\partial Q_{NOx}}$, $\frac{\partial O_3}{\partial Q_{VOC}}$ and second order terms are calculated within one run of the model CAMx. Using the HDDM approach, adjustments for the different species of VOC can also be introduced using VOC reactivity factors.

These methods do not fully explore the response to large changes in emissions. Hence further runs of the full model are required taking one beyond the region of parameter space for which the model has been evaluated. To have confidence, it is necessary to understand the structural behaviour of the model response to changes in input, not just formal numerical results!

Q8 Can footprints of $PM_{2.5}$ concentration be used to regulate stationary sources?

Answer

There is a need to investigate footprints under different background atmospheric concentrations, such as those expected in 2020, before this approach can be adopted.

Sufficient CMAQ runs to investigate footprints under significantly different background atmospheric concentrations have not yet been completed, and are a high priority for $PM_{2.5}$ assessment.

The situation is illustrated diagrammatically in Figure 2.8. If the input emissions change (shown as arrows in Figure 2.8), the two models' predicted responses, shown as the solid and dashed trajectories in future years under time varying meteorology, may change. The two models may respond identically to the addition of a stationary source under current conditions. The ellipses drawn in grey denote agreement after a specified future time; good if background conditions do change over time but not so good under a change of background conditions.

In CREMO, the model responses shown by stationary source footprints agree quite well in 2003 background conditions. The predicted model response under future background conditions in 2020 has not been investigated. Further work is planned by the Environment Agency to investigate changes in $PM_{2.5}$ concentrations over the UK between current conditions and future emissions scenarios representing possible conditions in 2020.

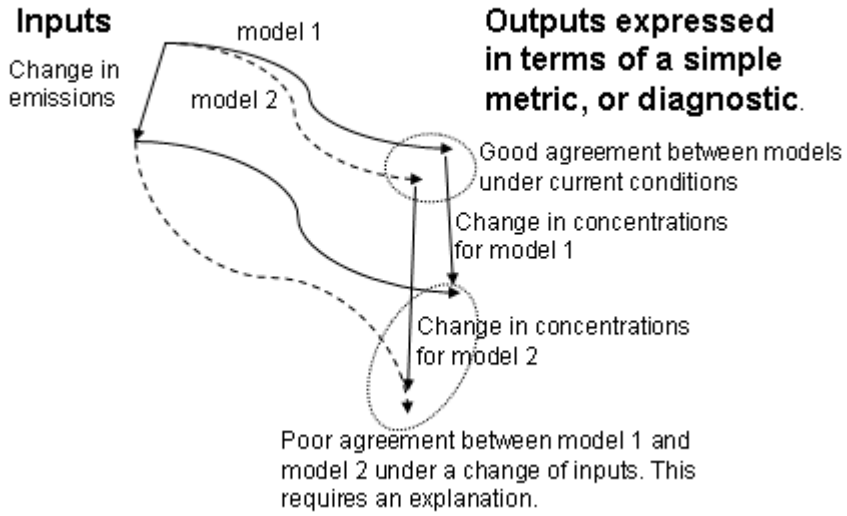


Figure 2.8 Illustration of responses of two models to a change in emissions under different background conditions

Such studies are similar to work undertaken by the USEPA to investigate the impact in a future year of new vehicle standards. Air quality modelling was performed (USEPA 2011a), using CMAQ v4.7 for three emissions cases:

- a 2005 base year;
- a 2030 reference case projection without additional vehicle standards;
- a 2030 control case projection with additional vehicle standards.

The projected controls primarily affect NOx emissions (USEPA 2011b). Concentrations of the annual and 24-hour average $PM_{2.5}$ concentrations, daily maximum eight-hour ozone concentrations, annual nitrogen and sulphur deposition levels, and selected annual and seasonal air toxic concentrations, as well as visibility impairment, were calculated. Measurements in 2005 were used to evaluate performance.

Q9 Can the Environment Agency adopt a formal, prescriptive approach to the acceptance or rejection of air quality models?

Answer

Acceptance criteria for regulatory air quality models used by the Environment Agency should be flexible and pragmatic. This does not discount the systematic use of benchmarking tools to describe the performance of models and this should be encouraged.

Acceptance criteria for the use of regional models in CREMO have been tested (Hayman *et al.* 2012b), based on the operational approach adopted in Phase 1 of the Defra Model Evaluation Exercise. However, when considering more than one pollutant species and more than one acceptance criteria, the appropriate way to aggregate the acceptance criteria is not obvious. In Defra's Model Evaluation Exercise the implied aggregated criterion is that all acceptance criteria for all pollutant species should be satisfied.

An acceptance criterion for ozone depends on whether the annual average ozone concentration or peak ozone level during episodes is being considered. Annual

average ozone depends largely on domain boundary conditions and removal within the domain, while episodic ozone concentrations rely on regional generation within the domain. It is therefore necessary to distinguish between an acceptance criterion for shorter periods, such as episodes, and an acceptance criterion for annual average ozone, which will depend in part on, but will not be dominated by, episodic ozone concentrations.

Operational evaluation criteria can be applied to the highest of the daily maximum rolling eight-hour mean ozone concentrations in a year and to the annual average ozone using measurements from a network. These definitions require a detailed explanation (Hayman *et al.* 2012c). The running eight-hour mean ozone concentrations are calculated from the hourly average ozone concentrations over fixed periods from 00.00 to 00.59 onwards. These hourly averages are then taken consecutively in groups of eight, and the eight-hour averages for the periods 00.00–07.59, 01.00–08.59 *etc.* are then calculated. By convention, the running eight-hour mean concentration is assigned the time of the last hour of the period. The maximum of these 24 mean concentrations in a day is the daily maximum eight-hour running mean. The performance of the CMAQ model, when calculating the daily maximum ozone concentrations, is illustrated later in Table 2.8 using various metrics including *NMB*.

The paper by Francis *et al.* (2011) demonstrates that the CMAQ model shows agreement with observations during an episode of high ozone in 2003. The simpler ozone trajectory model OSRM shows good agreement with CMAQ and with observations, for monthly mean daily maximum ozone concentrations at Harwell (Vincent *et al*. 2010) and for annual mean ozone and NO_2 along transects (Hayman *et al*. 2012d). This suggests that both OSRM and CMAQ are acceptable for long-term averages.

There is no one clearly preferred health related ozone metric and taking a threshold of 35ppb is somewhat arbitrary. If $A_{8,i}$ is the maximum eight-hourly average ozone on day i in ppb, during a year with N days (N = 365 or 366), then the metric *SOMO35* (the sum of maximum ozone above 35 ppb) is defined as:

$$SOMO35 = \sum_{i=1}^{i=N} \max((A_{8,i} - 35),0)$$

The units of *SOMO35* are ppb.days or ppm.days. Typically one expects *SOMO35* to take a value of a few thousand ppb.days in southern England.

Another kind of ozone metric is used for the protection of ecosystems. The metric *AOT40* (accumulated hourly mean ozone above a threshold of 40ppb, expressed in $\mu g/m^3$.hours) is the sum of the differences between the hourly ozone concentration greater than 80 $\mu g/m^3$ (= 40 ppb) and 80 $\mu g/m^3$, over a given period, using the one-hour values measured (or modelled) between 08:00 and 20:00 Central European Time each day, following the Air Quality Directive definition. Units are $\mu g/m^3$.hours or ppb.hours. The threshold sets a critical level above which harmful effects are expected.

For criteria which include a cut-off, such as *AOT40* and *SOMO35*, the normalised mean bias *NMB* should not be used in an acceptance criterion, as *NMB* will take very large values when the normalisation factor is small. It would be better to normalise by some fixed value such as a target value. So, for example, the normalised bias criterion for *AOT40* would become

$$\frac{1}{N}\sum_{i}\frac{AOT40Modelled_{i} - AOT40Observed_{i}}{TV},$$

where *TV* is the target value for *AOT40*. The Air Quality Directive sets an *AOT40* target value, *TV*, of 18,000 μg/m^3.hours to protect vegetation.

There is no commonly accepted agreement over the best choice of ozone metric. There are examples of diagnostic ozone metrics in the review by Middleton *et al.* (2007), while Rao *et al*. (2011) argue in favour of the seasonal average as the most suitable metric. The Integrated Process Rate analysis of the CMAQ model undertaken by Francis *et al*. (2011) to understand the causes of an ozone episode over south-east England considered the contributions of cloud processes, chemical processes, advection, diffusion, vertical advection, vertical diffusion, horizontal advection and dry deposition, at different model heights. For this episode, in a local south-east England domain, meteorological processes were shown to have the greatest influence. The analysis demonstrates how the build-up of an episode at a location can be interpreted, but does not immediately generalise to an indicator for managing ozone.

The occurrence of photochemical episodes, when high ozone concentrations are generated, suggests the need to be aware of a variety of possible behaviours. Under steady conditions for a range of VOC emission strengths (Figure 2.9), ozone would be expected to tend towards a limit, which is the long-term solution of a system of ordinary differential equations.

Under some conditions, corresponding to lower emission rates, the limit value would be expected to be close to the initial condition, similar to the background ozone concentration. This represents an approximate balance between production and destruction of ozone over the domain. Under other conditions, corresponding to higher emission rates, the limit value is expected to be far from the initial state, corresponding to the build-up of an ozone episode. The large difference in the limiting ozone concentration between these two emission situations corresponds to a bifurcation in the steady-state behaviour of the differential equations describing the ozone system. Near a bifurcation, a large change in the limiting ozone concentration occurs for a small change in emissions. A sensitivity analysis of a numerical chemical transport model would break down around the bifurcation point.

As an illustration, in Figure 2.9 the atmospheric chemistry within an anticyclone over central Europe under increases in VOC emissions of up to a factor 10 is shown in the right hand upper figure (with a limit value of high, episodic ozone very different from the initial concentration). The atmospheric chemistry within an anticyclone over central Europe under decreases in VOC emissions of up to a factor 10 is shown in the left hand lower figure (with a limit value not much different from the initial concentration).

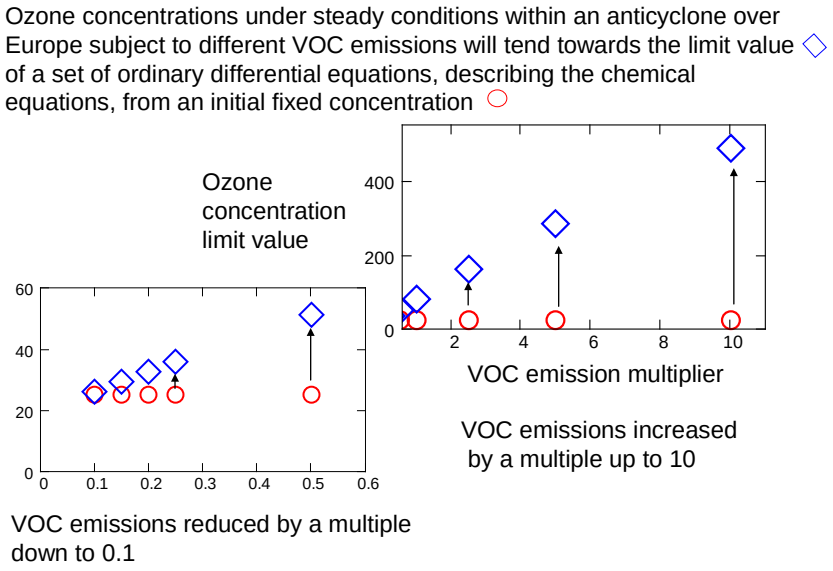


Figure 2.9 How the background concentration of ozone ○, taken to have a fixed initial value, might tend towards a limit value ◊ which is the solution of a system of ordinary differential equations after a long period of time

The occurrence of ozone episodes will to some extent influence the long-term average ozone concentration, although their occurrence may vary considerably from year to year depending on weather conditions. The build-up of ozone during an episode is likely to depend approximately on the density of precursor emissions (NOx and VOCs) averaged over the region of the anticyclone.

Ozone formation can be attributed to specific sources during an ozone episode. This has been carried out using the CMAQ model (Yu *et al*. 2008) by investigating ozone formation when the emissions from a specified source are changed. The long-term average production of ozone can be attributed to a specific source by taking the difference in annual ozone concentration with and without the source in question. This is shown in Figure 2.10, using the CMAQ model to calculate the footprint of an oil refinery on the south coast of England in 2003. In this case, the annual average ozone attributed to the refinery is seen to be small and negative.

It is important to distinguish between the situation that leads to the build-up of ozone, depending on the regional emissions of NOx and VOCs, and the situation that depends largely on the input of ozone entering into a region from sources on a hemispherical scale. The long-term average effect of individual stationary sources is seen to be small (negative and of magnitude < 0.2 μg/m^3 in the example shown in Figure 2.10). Hayman *et al.* (2012d) illustrate the result shown in Figure 2.10 in another way. They plot the influence of the refinery on the annual average ozone concentration as a function of distance along a south to north transect through the refinery location.

Hayman *et al.* (2012c) recommend a number of metrics to describe the impact of ozone including *SOMO35* and *AOT40*. The only one of the proposed metrics which focuses explicitly on peak ozone concentrations is the maximum running 8-hour mean concentration. Peak concentrations in all the other metrics are attenuated to some extent by the presence of lower ozone concentrations within the averaging, so that for these metrics the build up of ozone during an episode illustrated in Figure 2.9 is not

seen. Hayman *et al.* (2012d) discuss comparisons of the calculated response in peak ozone concentrations to changes in precursor emissions in a number of different models. There appear to be differences between the model predictions, suggesting that further understanding of the behaviour of ozone during episodes is required. It is interesting to note that Derwent (2012) showed that the choice of chemical mechanism made little difference to the response in peak ozone concentrations to changes in precursor emissions during one ozone episode over southern England in July 2006, when concentrations reached over 100ppb. This suggests that the difference in the details of chemical reactions is not the single most important factor in explaining differences between predictions of ozone during episodes.

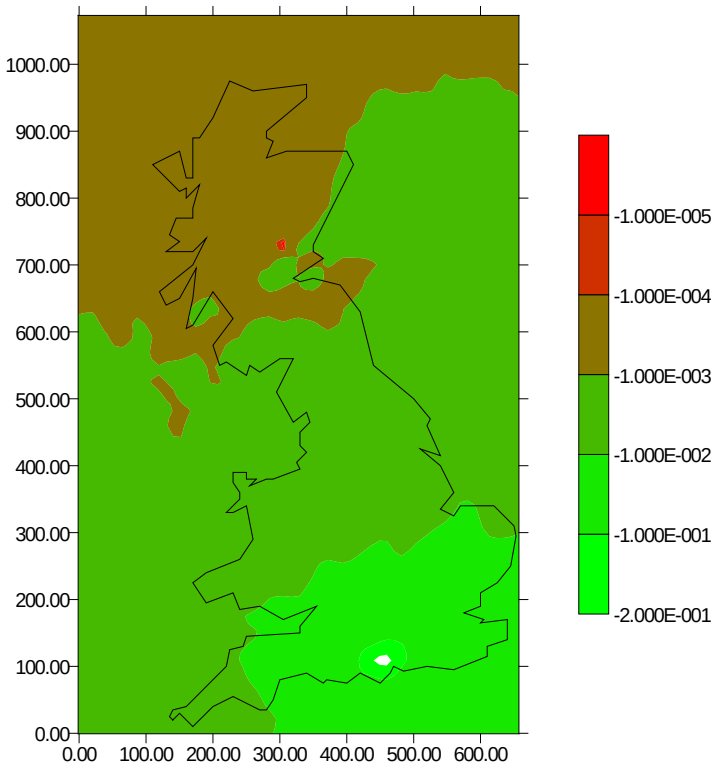


Figure 2.10 Annual average ozone concentration in µg/m^3 from an oil refinery in central southern England in 2003 (VOC emissions ~0.2 kg/s, NOx emissions ~0.2 kg/s) using the CMAQ v4.6 model

As one moves away from situations for which there are observational data, acceptance criteria for the use of models become more difficult to apply. Some dynamic evaluations are possible, where one can assess model performance using observations to determine whether the model responds as expected. However as one generally considers situations different from current or historical regimes, it is not possible to test the models against measurements. Extrapolation far from the current or past situations could become very uncertain.

A model is a system of inputs and outputs, and the objective is to try to establish whether the response of the system to a change of inputs is reliable even when observations on which to test the response are not available. The acceptance of a model in these situations relies more on understanding of the system (in other words on expert judgement). In contrast, for operational evaluation, quantitative acceptance criteria have been set. For the other types of evaluation (probabilistic, diagnostic, *etc.*), air quality models should be subject to continuous development and challenge. Any policy or regulatory outcome should be subject to challenge from a variety of alternative modelling approaches, all of which should be tested against commonly applied

operational acceptance criteria (such as *NMB*<0.2 and *FA2*>50 per cent). This process of challenge is similar to the process of testing scientific theories. If a variety of models show a similar degree of agreement, then the simplest or most practical should be used.

Simpler models can be considered to be emulations of more complex models provided they describe approximately the same response to input changes as more complex, fundamental models. Then there is justification for using them on the grounds of efficiency. This does not mean that the more complex fundamental models should be neglected, since they provide the essential test bed for exploring understanding of the response of a model to changes in input over a wider range of conditions, covering broader 'what if' questions. Within CREMO the testing of emulations of complex models has not developed far enough for their use to be recommended for regulation.

If one model shows a very different result from existing models designed for the same purpose, then the cause should be investigated to try to explain the reasons. The model developer should have an understanding of the principles underlying the model. Key to this is selecting the diagnostics or metrics, the summary parameters describing key aspects of model behaviour, such as mass balances, chemical species lifetimes *etc*. Although prescriptive advice cannot be given, it is expected that model developers understand what their models do and can propose a series of suitable diagnostic or metrics.

The CREMO Model Evaluation Protocol (Hayman *et al*. 2012a) is a development of an earlier protocol developed by Yu *et al*. (2007) and contains a brief description of each of the models used within the project. A number of the models in CREMO have also been included within Phase 1 of the Model Evaluation Exercise conducted by Defra during 2010–2011. The larger set of models used in the Defra evaluation are summarised in a report by the Air Quality Modelling Review Steering Group (Williams *et al*. 2011) and in a series of questionnaires available on the King's College London website.[12]

The evaluation of models is therefore not a wholly quantitative procedure. It can be made more systematic by the use of checklists (Risbey *et al*. 2001) and by benchmarking models. A draft procedure for benchmarking models (Thunis *et al*. 2011a) is under discussion in Europe, as well as a benchmarking tool (Thunis *et al*. 2011b), and some applications are available (Chemel *et al*. 2010, Pederzoli *et al*. 2011). The purpose of benchmarking is to evaluate the performance of air quality models and indicate ways for making improvements. The draft benchmarking procedure discusses acceptance criteria but does not set numerical acceptance criteria. Thus benchmarking tools should be used to describe the performance of models and should be encouraged.

Q10 What kind of tools can assist in making expert judgement, and what judgements can be made at the present time?

Answer

Both statistical metrics and diagnostic metrics should be considered when making a model evaluation as part of expert judgement, which can be made more systematic by

[12] The supporting questionnaires for the Defra model evaluation reports are available to download from the web portal http://www.erg.kcl.ac.uk/downloads/Policy_Download/ using the log in details: User name = policy, Password = AirMen.

using check lists and benchmarking tools. At the present time, the Environment Agency could set a model performance criterion for 'daytime' ozone predictions, but could not yet set one for particulate matter predictions.

This section contains examples of the use of metrics to judge regional air quality model performance. Dore *et al.* (2012) presented results based on Phase 1 of the Defra Model Evaluation Exercise which encompass a wider set of models than that considered here. If one wishes to express an acceptance criterion, this should be in numerical form (that is, it should be expressed as a number). The examples shown below are in the form of summary tables, focusing on particulate matter, the secondary pollutant with the greatest health effects.

The statistical metrics are not independent. For example, *FAC2* is expected to be related to *NMB*. In addition, when judging model performance, it is necessary to decide how to aggregate the independent statistical metrics together in order to make a decision. An objective method for doing this is not readily apparent. Summaries of the performance of models within CREMO and from Phase 1 of the Defra Model Evaluation Exercise (Carslaw 2011c) are given in tables below.

The model TRACK-ADMS was developed to enable the Environment Agency to assess contributions from major industrial sources (Abbott and Vincent 2006, Abbott *et al*. 2006, Vincent and Abbott 2008). Although of similar complexity, FRAME model results could not be included in the intercomparison, as FRAME output only contains concentrations of the inorganic components of particulate matter (sulphate, nitrate and ammonium) and not total particulate. The reduced complexity models, TRACK-ADMS and FRAME, both assume simplified meteorology to calculate the long-term average atmospheric concentration. They contain some degree of data assimilation which improves predictions. (TRACK-ADMS includes a bias correction.) They are suited to calculating the contributions from large industrial sources and have been subject to uncertainty analysis by undertaking a Monte Carlo analysis of the variation in output over plausible ranges of input parameters.[13] Their limitation is that one cannot be sure that the calibrated choice of parameter input values, used to evaluate the models, is applicable under future emission scenario conditions when emissions and boundary conditions over the model domain may be very different from the ones used to test the models. This was explained more fully in the answer to question 8 and is the reason why more fundamental models, such as CMAQ, have been developed.

Operational Evaluation using statistical metrics

Table 2.1 shows an example of calculating PM_{10} concentrations across the UK, comparing the performance of the CMAQ model against a simpler model TRACK-ADMS. Observations from the UK Automatic Urban and Rural Network (AURN) are used in the evaluation.

Note that the performance measures of the simple models in Table 2.1 to 2.4 are in italics.

[13] Only the uncertainty in wet deposition was analysed (airborne concentrations were not considered in the study, but wet deposition is a secondary component with similarities in production to particulate matter). For the studies by Abbott and Vincent (2006), Abbott *et al.* (2006) and Vincent and Abbott (2008), the uncertainty analyses suggest that 90 per cent or more values of the annual average predictions of secondary quantities lie within a factor of two of observations. Thus it is expected that the factor of two metric *FAC2* considered in the model evaluations will nearly always take values greater than about 0.9.

Table 2.1 Comparison of performance in predicting annual average PM_{10} concentration at AURN sites in the UK in 2003 for two versions of an advanced model and a simpler, reduced model (in italics)

Model metric PM_{10} for 2003	CMAQ v4.6	CMAQ v4.7	*TRACK-ADMS*
FAC2 (%)	88.2	100	*100*
*r (*correlation coefficient)	0.09	0.0	*0.45*
NMB	-0.33	-0.09	*-0.20*
Single power station contribution (%)	0.34	0.28	*0.28*

Source: Chemel *et al.* (2011)

Tables 2.2 to 2.4 show examples of summary tables illustrating the performance of regional models for calculating the three major secondary components of PM_{10} (sulphate, nitrate and ammonium, respectively) from Phase 1 of the Defra Model Evaluation Exercise (Carslaw 2011c).

Table 2.2 Comparison of performance in predicting annual average airborne sulphate over the UK in 2003 for various advanced models and simpler, reduced models (in italics)

Metric SO_4 in air annual mean model	Number of sites	*FAC2* (fraction)	*NMB*	*r* (correlation coefficient)
CMAQ JEP	12	0.89	0.10	0.80
CMAQ UH	12	0.97	-0.26	0.84
EMEP UK	12	1.00	-0.12	0.83
EMEP Unified	12	1.00	0.03	0.91
NAME	12	0.57	-0.45	0.33
FRAME	*12*	*0.84*	*0.28*	*0.75*
HARM	*12*	*1.00*	*-0.16*	*0.86*

Table 2.3 Comparison of performance in predicting annual average airborne nitrate over the UK in 2003 for various advanced models and simpler, reduced models (in italics)

Metric NO_3 in air annual mean model	Number of sites	*FAC2* (fraction)	*NMB*	*r (*correlation coefficient)
CMAQ JEP	12	1.00	0.29	0.95
CMAQ UH	12	1.00	-0.03	0.97
EMEP UK	12	1.00	-0.00	0.97
EMEP Unified	12	1.00	0.12	0.93
NAME	12	1.00	0.2	0.88
FRAME	*12*	*0.92*	*0.14*	*0.92*
HARM	*12*	*0.92*	*-0.33*	*0.71*

Table 2.4 Comparison of performance in predicting annual average airborne ammonium over the UK for various advanced models and reduced models (in italics)

Metric NH_4 in air annual mean model	Number of sites	*FAC2* (fraction)	*NMB*	*r* (correlation coefficient)
CMAQ JEP	12	1.0	-0.03	0.94
CMAQ UH	12	0.92	-0.38	0.95
EMEP UK	12	1.00	-0.17	0.96
EMEP Unified	12	1.00	-0.23	0.93
NAME	12	1.00	-0.22	0.95
FRAME	*12*	*0.58*	*-0.36*	*0.96*
HARM	*12*	*0.08*	*-0.63*	*0.96*

Based on comparing *NMB,* no systematic difference between the performance of the more complex models and that of the simpler models can be seen. The level of performance appears to be better for single species (sulphate, nitrate and ammonium) rather than for the multi-component secondary species, particulate matter. This should be expected as the connection between emissions and airborne concentrations is more direct for single species.

AQMEII is a model evaluation initiative involving research groups from North America and Europe with the goal of advancing the methods for evaluating regional-scale air quality modelling systems. Within AQMEII, predicted concentrations of ozone and fine particulate matter and the deposition of sulphate and nitrogen from various regional-scale models were submitted for analysis to the European Union Joint Research Centre's ENSEMBLE model evaluation platform (Galmarini and Rao 2011). Data from some thousands of surface monitoring stations (chemical and meteorological), ozone profiles and MOZAIC aircraft measurements from the two continents have been gathered and harmonised in the ENSEMBLE system for direct comparison with model data.

The CMAQ model was applied to simulate air quality over North America and Europe for the year 2006 (Appel *et al.* 2012) as one part of the AQMEII project. Table 2.5 shows the seasonal, domain averaged normalised mean biases (*NMB*) of daily average $PM_{2.5}$ concentrations from the CMAQ model for the North American Air Quality System network and the European AirBase network in the year 2006. As far as possible, the CMAQ model configurations were similar for North America and Europe, with both simulations utilising version 4.7.1 of the model. The North American simulation used 34 vertical layers and a 12 km horizontal grid spacing, while the European simulation used 34 vertical layers and an 18 km horizontal grid spacing covering most of Europe. The overestimate of $PM_{2.5}$ at North American sites is thought to be due to an overestimate in the unspeciated $PM_{2.5}$ mass (Appel *et al.* 2012), which makes up a significant proportion of the $PM_{2.5}$ mass in the CMAQ model version 4.7.1 used in the AQMEII intercomparison. Improvements to treating this component are incorporated in later versions of CMAQ. If one considers all the models for which predictions were available in AQMEII, then generally there appears to be significant underprediction of both PM_{10} and $PM_{2.5}$ (Schere *et al.* 2012). Some of the large underprediction of PM_{10} is thought to arise from a lack of information about the source strength of wind-blown dust.

Table 2.5 *NMB* of daily average $PM_{2.5}$ comparisons between predictions and observations in different seasons of 2006 over North America and Europe made under AQMEII using the CMAQ model

Season and domain	Approximate number of sites	*NMB*
Winter, North America	958	0.304
Winter, Europe	160	-0.55
Spring, North America	958	0.189
Spring, Europe	160	-0.369
Summer, North America	958	-0.046
Summer, Europe	160	-0.372
Autumn, North America	958	0.363
Autumn, Europe	160	-0.242

Table 2.6 shows the seasonal, domain-wide normalised mean biases of daily average PM_{10} concentrations for the North American Air Quality System and European AirBase networks.

Table 2.6 *NMB* of daily average PM_{10} comparisons between predictions and observations for different seasons in 2006 over North America and Europe made under AQMEII using the CMAQ model

Season and domain	Approximate number of sites	*NMB*
Winter, North America	956	-0.479
Winter, Europe	1000	-0.648
Spring, North America	956	-0.565
Spring, Europe	1000	-0.562
Summer, North America	956	-0.574
Summer, Europe	1000	-0.612
Autumn, North America	956	-0.465
Autumn, Europe	1000	-0.468

The model performance for $PM_{2.5}$ and PM_{10}, especially for the European domain, shows large under-prediction and is not as good as in some of the UK model comparisons. The large negative biases, representing under-prediction, are understandable, if some components of the particulate matter have not been treated properly due to limited knowledge of emissions of these components and their behaviour in the atmosphere. The difference in the performance of the models for $PM_{2.5}$ and PM_{10} suggest that some components of the coarse fraction of particulate (difference between PM_{10} and $PM_{2.5}$) have not been included. This is understandable as dust, such as windblown soil in rural areas or dust re-suspended by vehicles in urban areas, is of primary local origin, lies in this size range and cannot be treated accurately in CMAQ. This interpretation of model performance for particulate matter applies to any Eulerian model with a grid resolution of some kilometres or more, not just to CMAQ, as demonstrated in the study by Solazzo *et al.* (2012b). The modelling of a footprint from a power station is likely to be better, as there is little locally derived coarse particulate, the source strength, which consists largely of inorganic compounds, is better known and the chemistry better understood.

The normalised mean bias is the common metric in all these evaluations. The performance is better for the smaller region (the U.K.) and for individual species (sulphate, nitrate and ammonium) than for the multi-component secondary species, particulate matter, and for whole continents. The more scientifically based, fundamental models do not perform better than the simpler models with a few ‘tuned’ input

parameters. This may seem a disappointing result for fundamental science, but is perhaps understandable given the opportunities for errors to arise in the input data and in the observations used for testing complex models.

The FAIRMODE report (2011) on the application of models under the European Union's Air Quality Directive notes that, if observed data are included in a model, evaluation has to be carried out differently from the normal model evaluation shown in Tables 2.1 to 2.6. Inclusion of observed data (data assimilation) is the basis of purely statistical models. When data assimilation is included in a model, evaluation is done by either running the assimilation procedure a number of times with the exclusion of different sets of observation sites for each of the runs, or through separating monitoring data into assimilation and evaluation data sets. The models considered in Tables 2.1 to 2.6 were generally developed purely on the basis of understanding atmospheric processes.

The PCM model is a statistical model used extensively for policy application in the UK and has been used as the basis for UK air quality modelling when reporting ambient air quality assessments under the European Air Quality Directive (Grice *et al*. 2010). The Defra intercomparison does not contain a model evaluation based on the PCM model, presumably because a comparison of model performance on the same basis could not be done with an assimilated model. The PCM model is the 'official' model used for mapping annual average concentrations over the U.K. land area and for estimating future trends (Grice *et al.* 2010). However some demonstration of consistency with process models would strengthen confidence in the PCM approach. PCM is simpler to apply than process models. It requires a detailed source inventory broken down by grid square (1 km $\times$ 1 km for the UK) and by source sector. This detail is not available for every year of interest. Furthermore, broadly speaking, an assimilated model does not provide diagnostic information about the influence of one atmospheric process upon another. Such relationships provide understanding of the importance of individual processes, leading to diagnostic evaluation. However PCM shows the direct relationship between spatial distributed concentrations and emissions, which is an advantage when making policy decisions.

As argued above, the statistical metrics listed in Tables 2.1 to 2.6 cannot be used to evaluate PCM, as PCM already incorporates observations at national monitoring sites. So how can predictions from a model using assimilated data be compared with predictions from the process models listed in Tables 2.1 to 2.6, which are largely derived from first principles? (Some regional process models use measurements to specify the background concentrations entering the modelling domain.) If observational data sets were divided up between those used for assimilation and those used for testing, it might be possible to define a more generalised statistical metric based on information theory, such as the Akaike Information Criterion (Burnham and Anderson 2003), but could this be justified rigorously? Loosely speaking, the lower degree of freedom in setting parameters in the assimilated model should be taken into account in the way the performance metric is defined. Rather than go down this route, it is suggested that a pragmatic view is adopted that compares the differences between the PCM model and other process model predictions, and diagnoses the causes of differences.

Ways of directly intercomparing model predictions (rather than comparing models' performance with measurements) were suggested by Chemel *et al*. (2011). An example is given in Table 2.7 using the results of CMAQ v4.6 as the baseline model. The two statistical metrics *r* and CVRMSE readily separate the model CMAQ v4.7 linked to the baseline model and the model TRACK-ADMS, derived independently. The models tested by Chemel *et al*. (2011) were also used to compare the annual average PM_{10} footprint from a power station source. Statistical metrics were used to evaluate

the similarity between the footprints calculated by the regional models. In this case no observational data base is available, as there is generally no direct way of attributing concentrations to a specified source on a regional scale.

Table 2.7 Model intercomparison of the footprint of annual average air concentrations of PM_{10} with the footprint of a power station in central southern England in 2003[a,b,c]

	Spatial correlation coefficient *r*		*CVRMSE* (percentage)	
	CMAQ v4.7	TRACK-ADMS	CMAQ v4.7	TRACK-ADMS
Annual average PM_{10}	0.82	0.61	39.8	105.8

Notes [a] See Figures 2.6 and 2.7 for illustrations of footprints.
[b] Over an approximately square region with sides of length 350 km covering southern England using the CMAQ model v4.6 calculation as the baseline.
[c] The statistical metrics applied are the spatial correlation coefficient *r* and the coefficient of variation of the root mean square error CVRMSE (square root of the variance divided by the mean), expressed as a percentage.

The tables above illustrate the use of statistical metrics for evaluating predictions of PM_{10} and $PM_{2.5}$ from air quality models. Model performance depends on the ability to predict concentrations of all species of interest, although the two species of importance for regional models are ozone and PM. When using CMAQ, it is also possible to calculate short-term average concentrations, such as daily concentrations. In Table 2.8, the performance of CMAQ v4.6 is shown for the short-term variation of the two secondary species, ozone and PM_{10}. Model performance for 'daytime' ozone concentrations over a year is seen to be superior to that of PM_{10}.

Table 2.8 Comparison of performance in predicting the daily maximum ozone and daily mean PM_{10} concentrations at AURN sites in the UK in 2003 for CMAQ v4.6

CMAQ v4.6 metric	Maximum daily running eight-hour mean ozone	Daily mean PM_{10}
NMB	0.05	-0.34
r (correlation coefficient)	0.69	0.47
FAC2 (%)	76.7	26.8
Number of sites	~40	~40

Source: Chemel *et al*. (2010)

The AQMEII project also provided CMAQ performance statistics for ozone for the many hundreds of ozone monitoring sites in Europe and North America. The statistics by season and domain of the normalised mean bias for the daytime average ozone concentration are shown in Table 2.9. The normalised mean ozone bias is generally within ±10 per cent apart from during the summer season in Europe. This discrepancy cannot be readily explained. The CMAQ ozone predictions in Table 2.8 for the UK in 2003 also satisfy this performance criterion. In addition the ozone predictions in 2003 from the OSRM Lagrangian trajectory model for the UK in 2003, described by Hayman *et al.* (2010), comply with this performance measure.

Table 2.9 Comparison between predictions and observations made under AQMEII using the CMAQ model showing the normalised mean bias for daytime[a] average ozone for different seasons over North America and Europe in 2006

Season and domain	*NMB*
Winter, North America	-0.134
Winter, Europe	0.084
Spring, North America	-0.041
Spring, Europe	-0.048
Summer, North America	0.098
Summer, Europe	0.016
Autumn, North America	0.084
Autumn, Europe	0.323

Note: [a] daytime = 8 am to 8 pm.

The ozone conditions at the boundary of the North American and European domains may be a major factor influencing ozone concentrations within the domains in the winter. A different factor, episodic ozone generation, may be a major factor influencing ozone concentrations in the summer. Given the similar level of model performance from diverse models encompassing the main factors affecting ozone concentrations, one may provisionally propose that a normalised mean bias of within ±10% is currently the achievable performance of the current generation of photochemical models for the normalised 'daytime' average ozone concentration.

A comparison was made of the impact of an oil refinery in southern England on annual ozone concentrations in 2003 using the OSRM and CMAQ models. The predicted change in the annual average ozone concentration along a horizontal transect through the refinery shows that the emissions from the refinery lead to a decrease in ozone on average in both models. The decrease is thought to be caused by the reaction of ozone with NO releases. The two models produce similar results, giving confidence in the use of the OSRM model developed from the Photochemical Trajectory Model, which was specifically designed for use in developing national policy for regulating ozone and is not a fully comprehensive chemical transport model. The annual average ozone footprint from the CMAQ model (see Figure 2.10) also shows negative ozone concentrations out to distances of a few hundred kilometres.[14]

An acceptance criterion based on a normalised mean bias for 'daytime' average ozone concentrations of less than ±10% is currently achievable for the current generation of photochemical models. The CMAQ model would satisfy this criterion. No acceptance criterion can yet be set for models used to predict concentrations of particulate matter.

Diagnostic Evaluation using a footprint metric

Although statistical metrics provide useful summaries of model performance and illustrate the degree of agreement, or disagreement, they do not lead directly to conclusions regarding which model or which type of model should be used in air quality assessment. Thus operational evaluation does not provide information about the

[14] An ozone footprint cannot be calculated from the output from the IDOP method mentioned in the discussion for question 5. Instead it gives an upper bound on ozone concentration production during episodic conditions. The IDOP approach should be regarded as an initial filter, providing a very approximate upper bound on ozone increments for major individual sources if results from the OSRM or CMAQ models are not available.

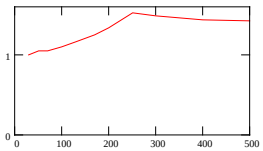
influence of one process upon another, leading to understanding of why a model behaves in a certain way.

Instead it is necessary to consider diagnostic metrics, which describe the behaviour of model systems in simple ways. As an illustration, examples of a diagnostic metric based on the footprint of PM and sulphur and nitrogen deposition from a power station source are given below in Figures 2.11, 2.12 and 2.13. In Fig. 2.11 the distance weighted footprint, illustrated in Figure 2.7, is used to compare the predictions of the annual average PM_{10} concentrations from the models, previously compared in operational evaluations using statistical metrics. All the plots have been normalised to the concentration in the near field and are expressed as the ratio of the average concentration along a radial trajectory to the near-field concentration.

Footprints used in diagnostic evaluations cannot be compared directly with observations. However they are useful for diagnosing whether different models are treating processes, in this case the distance dependence, in different ways.

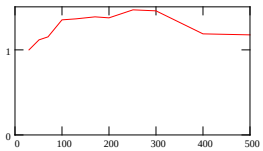
In Figure 2.11, the distance-dependent weightings of the annual average PM_{10} in the footprint of a power station are compared for three types of models. Broadly, the models show little dependence on distance suggesting that the removal is largely in balance with production, but the detailed behaviour is different even for the two models whose structure is similar (CMAQ v4.6 and CMAQ v4.7). Sutton *et al*. (2012) showed that apparently small differences in the temporal profile of ammonia emissions over the year in 2003 (even if the spatial distribution of the annual total is identical) can make differences to the prediction of the annual spatial distribution of components of $PM_{2.5}$ and of acid deposition using CMAQ v4.6 and v4.7. This suggests that differences in the set-up of model runs can make differences to the spatial distribution, even if the models themselves (CMAQ v4.6 and CMAQ v4.7) are formulated in a similar way.

In Figure 2.12, the distance-dependent weighting of the sulphur and nitrogen deposition from a power station are seen to be different from that of the annual average PM_{10}. This demonstrates that the processes of sulphur and nitrogen formation and sulphur and nitrogen loss are not approximately in balance, and have distinctly different length scales.

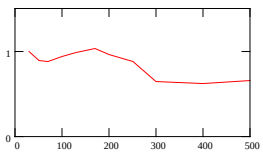


Annual average PM10 concentration from
power station in µg m-3
CMAQ v4.6 at 5km resolution

Average dependence on distance *r*
along radial trajectory



Annual average PM10 concentration from
power station in µg m-3
CMAQ v4.7 at 5km resolution

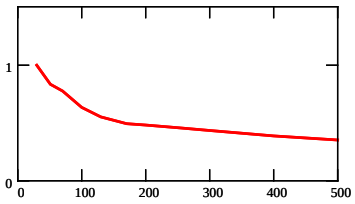


Annual average PM10 concentration from
power station in µg m-3
TRACK-ADMS at 1km resolution

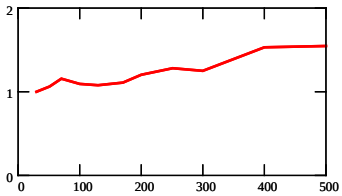
Figure 2.11 Examples of dependence of PM_{10} concentration on distance along a radial trajectory, derived from the footprint of a power station source in central southern England, scaled by the value of the concentration near the source, for

two versions of an advanced model CMAQ (v4.6 and v4.7) and the simpler, reduced model TRACK-ADMS

Distance dependence of deposition fields



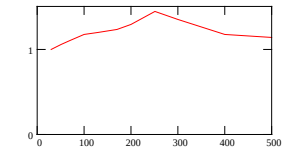
Distance dependence S deposition CMAQ 4.6
Decays because it is dominated by primary S species



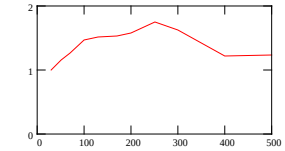
Distance dependence N deposition CMAQ 4.6
Increase because it is dominated by secondary formation of N species

Figure 2.12 Examples of dependence of sulphur and nitrogen deposition on distance along a radial trajectory, derived from the footprint of a power station source in central southern England from the CMAQ v4.6 model, scaled by the value of the deposition near the source

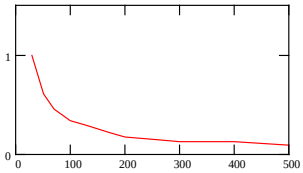
In Figure 2.13, the distance-dependent weightings of $PM_{2.5}$ and NO_2 for various different kinds of sources are shown, illustrating different kinds of distance dependencies. Small-scale features of the plots may be the result of edge effects in the domain considered and may also arise as a result of averaging over wind direction.



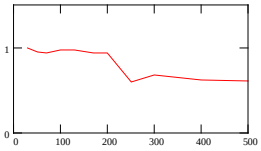
CMAQ v4.6 PM2.5
interpolated at 1km



CMAQ v4.7
PM2.5
interpolated at 1km



Annual average NO2 concentration
from Heathrow in µg m-3
CMAQv4.6 at 1km resolution



Annual average PM2.5
concentration from Fawley
Refinery for 2003 using
CMAQ v4.6 model in µgm-3.

Figure 2.13 Examples of dependence of $PM_{2.5}$ concentration on distance along a radial trajectory, derived from the footprint of a power station source in central southern England, scaled by the value of the concentration near to the source, for two versions of the advanced model CMAQ (v4.6 and v4.7), and for the NO_2 concentration around Heathrow Airport in 2003 and the $PM_{2.5}$ concentration around an oil refinery in southern England using CMAQ v4.6

Footprints can be used to make more subtle comparisons between models. The EMEP model has been used to derive the contributions from individual countries to the regional concentration of $PM_{2.5}$ over Europe. The individual country contributions represent national footprints, and in Figure 2.14, the EMEP footprint for the UK is compared with the footprint from a typical power station. It is seen that, out to distances of 500 km, the balance between production and loss is approximately maintained in both calculations. As expected, loss mechanisms begin to dominate beyond this distance. The footprints are expressed in concentration units. The maximum $PM_{2.5}$ concentrations in the two cases are 0.8 and 0.2 µg/m^3, respectively.

Although the species profiles (SO_2 : NOx: NH_3) of the UK emissions and a single coal-fired power station source are different, an approximate estimate can be made of the dominant precursor emissions, taken to be the sum of primary sulphur and nitrogen species (SO_2 + NOx). Although only rough estimates, there is consistency in magnitudes. Six times the precursor source strength gives roughly four times the maximum $PM_{2.5}$ concentration. The distance-dependent weighting appears to be different in detail. This may be a result of the different spatial resolution of the two models. It could also be the result of detailed differences in the treatment of particle formation in the two models. The distance-dependent weighting is again seen to be a valuable diagnostic metric for investigating model behaviour.

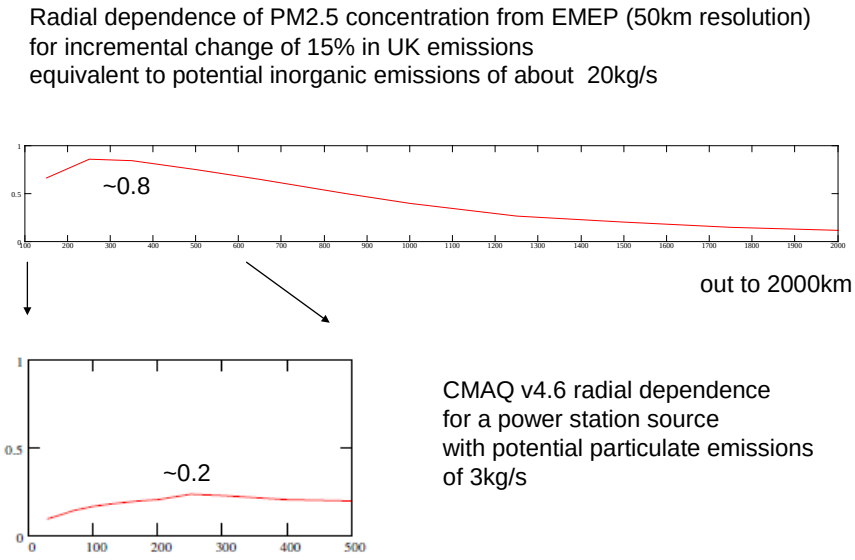


Figure 2.14 *Top:* Dependence of $PM_{2.5}$ concentration on distance in 2003, along a radial trajectory starting in central England, based on a 15% reduction in UK emissions, based on the EMEP model (Klein *et al*. 2011). *Bottom:* Dependence of $PM_{2.5}$ concentration on distance in 2003, along a radial trajectory, from the footprint of a stationary power station source. Notional potential source strengths for the two types of sources and maximum values of $PM_{2.5}$ concentrations (0.8 and 0.2 µg/m^3) are shown.

Another kind of diagnostic evaluation is to look at the behaviour of individual processes within the model. An example of this has been demonstrated by Derwent (2012) who compared the chemical schemes commonly applied in regional air quality models. This involved taking the chemical mechanisms out of models and applying them within a single model, the Photochemical Trajectory Model (PTM) (Derwent *et al*. 2009). Hence the application of the mechanisms in the inter-comparison is not exactly the same as the way the mechanisms are applied within the chemical transport model, but this harmonisation is inevitable if the behaviour of a single process is to be compared.

Six chemical mechanisms (CRI, CB-05, CBM-4, SAPRC-99, SAPRC-07 and OSRM) were evaluated and a Monte Carlo analysis was applied to assess uncertainty. Only runs from the Monte Carlo analysis with acceptable parameters were retained. From these, mid-afternoon ozone concentrations during July 2006 were compared with observations at Harwell, Oxfordshire. The mean fractional bias from the chemical schemes were very low, in the range -0.04 to -0.09. There was some indication of underestimation of peak ozone but it was concluded that there was little difference in the output of any of the 6 mechanisms.

An policy response metric can be set in terms of whether ozone reductions are greater for a 30% reduction in NOx emissions, or a 30% reduction in VOC emissions. The analysis also showed that there was little difference in the policy response metric between the chemical mechanisms, when applied within the PTM model framework.

There are examples where diagnostic evaluation has not been carried out, leaving the outcome of a model intercomparison uncertain. For example, model results from the large European project MEGAPOLI, on regional pollution in cities, showed generally under-prediction of particulate matter concentrations compared with observations (Schlünzen and Haller 2011). The observed values contained measurement data at up to 20 stations situated in the Rhine-Ruhr area, which is close to the centre of the

European model domain for most of the regional chemical transport models applied. For $PM_{2.5}$ and PM_{10} the mean values of the chemical transport models CHIMERE and SILAM agreed quite well with measured data. However the predictions from one model FARM (the Flexible Air quality Regional Model (Finardi *et al*. 2012)) were too high for PM_{10} and $PM_{2.5}$. In the case of the CHIMERE and SILAM chemical transport models the bias was negative, which is what is generally expected. The FARM model was found to frequently predict concentrations greater than observed values and not to predict the frequently occurring lower concentration values. However its behaviour for other pollutants (SO_2, NO_2 and ozone), which include primary and secondary species, was not exceptional. The FARM model appears to be an outlier for PM_{10} concentrations and over-predicted particulate matter, especially in winter, but not SO_2, NO_2 or ozone. The performance of the FARM model is shown in Table 2.10, in which PM_{10} concentrations (Silibello *et al*. 2011) are compared with observed data across Europe available from the EMEP Chemical Coordinating Centre at NILU.

Table 2.10 Comparison between predictions and observations of PM_{10} concentrations from the FARM model expressed in terms of the metrics *FAC2* and *NMB* for summer and winter months of 2005

Metric	June 2005	December 2005
FAC2(%)	86	48
NMB(%)	-1	69

This behaviour could not be explained. It could be based on an overestimate of residential heating emissions, but with a grid resolution of 25 km normally leading underestimates this seems unlikely. A physical explanation could be that exchange through the boundary layer and convection was lower than in other models, but this factor did not show up in calculations of the other pollutant species. The difficulty in explaining differences in model behaviour in complex models, where many factors may contribute in complicated ways, may be regarded as a weakness of large computer models. It is hard to pinpoint the influence of specific processes.

The use of diagnostic metrics can therefore reveal aspects of model behaviour. However there is no systematic way of choosing them. Diagnostic evaluation must be considered to be at an exploratory stage. The same can be said of dynamic and probabilistic evaluation, which have also not received a great deal of attention (see Galliland *et al.*. (2008), for an example of dynamic evaluation). Future Environment Agency studies and Phase 2 of the Defra Model Evaluation Exercise may provide more information on these aspects of model evaluation.

Q11 How does one move from scientific evaluation to policy and decision support?

Answer

Moving from scientific evaluation to policy and decision support requires an understanding of model behaviour. A formal procedure cannot be specified, but has to rely on expert judgement. This can be made more systematic by the use of benchmarking tools and check lists.

The European Monitoring and Evaluation Programme (EMEP) contains examples of scientific evaluation and policy metrics. Making the step to policy support depends on understanding the system, which means recognising suitable diagnostics. In EMEP, the policy metric is the set of national source–receptor matrices. The diagnostic metric is the set of individual country footprints for 15 per cent changes in emissions (annual

country-to-grid source–receptor matrices; see http://www.emep.int/SR_data/index_sr.html). An example of the radial dependence for the UK was given in Figure 2.14, showing the circumferential average footprint of the UK centred on the middle of a grid cell in central England, using grid cells of dimension 50 km $\times$ 50 km.

In the USA the CMAQ model and equivalent advanced chemical transport models, such as CAMx, have been accepted for policy applications. An example of their potential use arises when conducting reviews of new sources of $PM_{2.5}$ under the inter-pollutant trading policy for $PM_{2.5}$ set up by the USEPA in 2008. The principle is that primary $PM_{2.5}$ can be regarded as equivalent to some multiple of $PM_{2.5}$ precursor. Any proposed choice of ratios of $PM_{2.5}$ precursors for use in an inter-pollutant offset programme in $PM_{2.5}$ non-attainment areas must be accompanied by a technical demonstration showing the net air quality benefits of adopting the chosen ratios. This technical demonstration has to include a series of sensitivity runs with advanced chemical transport models, in order to compare modelled $PM_{2.5}$ concentration changes with reductions of direct $PM_{2.5}$ emissions and $PM_{2.5}$ precursor emissions (for example SO_2 and NOx) from anthropogenic point sources within the area of interest.

A chemical transport model, such as CMAQ or CAMx, at a grid resolution of 12 km or less, is the recommended way of predicting secondary $PM_{2.5}$ concentrations from changes in precursor emissions from a specified source. A dispersion model, such as AERMOD, or a chemical transport model, such as CMAQ or CAMx, at a grid resolution of 4 km or less, is the recommended way to predict changes in primary $PM_{2.5}$ concentrations. The inter-pollutant trading ratios for $PM_{2.5}$, between directly emitted $PM_{2.5}$ emissions and precursor emissions, can thus be determined. The technical demonstration requires making use of:

- the chemical transport model;
- the emissions inventory data;
- observed ambient data for $PM_{2.5}$ and its component species.

3 Conclusions

The conclusions from the CREMO project are summarised in this report as answers to questions 1 to 11. Each answer is followed by supporting evidence. Overall it is concluded that there is currently no completely objective basis for setting acceptance criteria for models. The underlying principle has been to use whatever is comparable with existing practice. The error expected in current types of models should be accepted and this uncertainty built into any decisions made on the basis of models.

This conclusion is consistent with the recent technical guidance from the Forum for Air Quality Modelling in Europe (FAIRMODE 2011), which accepts that it is not possible to provide specific recommendations regarding a model's acceptability criteria. The guidance is not a comprehensive review of the performance expected from current air quality models. While the draft procedure for benchmarking models (Thunis *et al*. 2011a) does discuss acceptance criteria, it does not propose numerical values for acceptance criteria.

3.1 Future work

Further runs of CMAQ are planned by the Environment Agency. Under future work, the industrial contributions to $PM_{2.5}$ in 2006 and 2020 will be estimated provided the model performance is shown to be equivalent to, or better than that, of other regional air quality models.

The under-prediction of particulate matter seen in the results in Tables 2.6 and 2.7 occurs in other chemical transport models. Although related to mass closure (calculating the correct mass of all the particle components measured in current instruments), the under-prediction is not completely understood and requires further investigation.

References

References marked with an asterisk are those produced directly as a result of the CREMO project.

ABBOTT, J. and VINCENT, K., 2006. *Auditing the contribution of Environment Agency regulated processes to pollution.* Science Report SC030172/SR3. Bristol: Environment Agency.

ABBOTT, J., HAYMAN, G, VINCENT, K., METCALFE, S., DORE, T., SKEFFINGTON, R., WHITEHEAD, P., WHYATT, D., PASSANT, N. and WOODFIELD, M., 2001. *Uncertainty in acid deposition modelling and critical load assessments*. R&D Technical Report P4-083(5)/1. Bristol: Environment Agency.

ABBOTT, J.A., VINCENT K.J. and STEDMAN J.R., 2006. *Methods for auditing the contribution of Environment Agency regulated processes to pollution*. Science Report SC030172/SR2. Bristol: Environment Agency.

APPEL, W., CHEMEL, C., ROSELLE, S.J., FRANCIS, X.V., HU, R.-M., SOKHI, R.S., RAO, S.T. and GALMARINI, S., 2012. Examination of the Community Multiscale Air Quality (CMAQ) model performance over the North American and European domains. *Atmospheric Environment*, 53,142-155.

AQEG, 2009. *Ozone in the United Kingdom.* Report of the UK Air Quality Expert Group. Prepared at the request of the Department for Environment Food and Rural Affairs, the Scottish Executive, Welsh Assembly Government and the Department of the Environment Northern Ireland. London: Defra. Available from: http://archive.defra.gov.uk/environment/quality/air/airquality/publications/ozone/documents/aqeg-ozone-report.pdf [Accessed 18 April 2012].

BURNHAM, K.P. and ANDERSON, D.R., 2003. *Model selection and multimode inference: A practical information-theoretic approach.* New York: Springer.

CARSLAW, D., 2011a. *Defra urban model evaluation analysis – Phase 1.* London: Defra. Available from: http://uk-air.defra.gov.uk/library/reports?report_id=654 [Accessed 20 April 2012].

CARSLAW, D., 2011b. *Defra regional and transboundary model evaluation analysis – Phase 1*. London: Defra. Available from: http://uk-air.defra.gov.uk/library/reports?report_id=653 [Accessed 20 April 2012].

CARSLAW, D., 2011c. *Defra deposition model evaluation analysis – Phase 1*. London: Defra. Available from: http://uk-air.defra.gov.uk/library/reports?report_id=652 [Accessed 20 April 2012].

COMEAP, 2010. *The mortality effects of long-term exposure to particulate air pollution in the United Kingdom.* A report by the Committee on the Medical Effects of Air Pollutants, Chairman: Professor J. G. Ayres, Chairman of the Subgroup on Quantification of Air Pollution Risks, J. F. Hurley. London: Health Protection Agency. Available from: http://comeap.org.uk/membership/128-the-mortality-effects-of-long-term-exposure-to-particulate-air-pollution-in-the-uk.html [Accessed 20 April 2012].

*CHEMEL, C., SOKHI, R.S., YU, Y., HAYMAN, G.D., VINCENT, K.J., DORE, A.J., TANG, Y.S., PRAIN ,H.D. and FISHER, B.E.A., 2010. Evaluation of a CMAQ simulation at high resolution over the UK for the calendar year 2003. *Atmospheric Environment,* 44, 2927-2939 [supplementary data available online].

*CHEMEL, C., SOKHI, R.S., DORE, A.J., SUTTON, P., VINCENT, K.J., GRIFFITHS, S.J., HAYMAN, G.D., WRIGHT, R., BAGGALEY, M., HALLSWORTH, S., PRAIN, H.D. and FISHER, B.E.A., 2011. Predictions of UK regulated power station contributions to

regional air pollution and deposition: a model comparison exercise. *Journal of Air and Waste Management Association*, 61, 1236-1245.

COHAN, D.S., HAKAMI, A., HU, Y. and RUSSELL, A.G., 2005. Non-linear response of ozone to emissions: Source apportionment and sensitivity analysis. *Environmental Science and Technology*, 39, 6739-6748.

DENBY B., 2011. *Guide on modelling nitrogen dioxide (NO_2) for air quality assessment and planning relevant to the European Air Quality Directive. Result of activities in the Forum for Air Quality Modelling in Europe*, FAIRMODE Working Group 1, Version 4.6. ETC/ACM Technical Paper 2011/15. Bilthoven, The Netherlands, European Topic Centre on Air Pollution and Climate Change Mitigation, RIVM.

DENNIS, R., FOX, T., FUENTES, M., GILLILAND, A., HANNA, S., HOGREFE, C., IRWIN, J., RAO, S.T., SCHEFFE, R., SCHERE, K., STEYN, D. and VENKATRAM, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. *Environmental Fluid Mechanics*, 10, 471-489.

DERWENT, R. and NELSON, N., 2003. *Development of a reactivity index for the control of the emissions of organic compounds*. R&D Technical Report P4-105 RC8309. Bristol: Environment Agency.

DERWENT, R., WITHAM, C., REDINGTON, A., JENKIN, M., STEDMAN, J., YARDLEY, R. and HAYMAN, G., 2009. Particulate matter at a rural location in southern England during 2006: Model sensitivities to precursor emissions. *Atmospheric Environment*, 43, 689-696.

DERWENT, D., FRASER, A., ABBOTT, J., JENKIN, M., WILLIS, P. and MURRELLS, T., 2010. *Evaluating the performance of air quality models*. Report prepared for the UK Department for Environment, Food and Rural Affairs, Issue 3/June 2010. London: Defra. Available from: http://uk-air.defra.gov.uk/reports/cat05/1006241607_100608_MIP_Final_Version.pdf

*DERWENT, R.D., 2012. *Chemical mechanism choice: impacts of ozone precursor emissions reductions in the United Kingdom on episodic peak ozone in the United Kingdom*. Report SC060037e/R. Bristol: Environment Agency.

DORE, A.J., CARSLAW, D., CHEMEL, C., DERWENT, R.G., FISHER, B.E.A., GRIFFITHS, S.J., LAWRENCE, S., METCALFE, S.E., REDINGTON, A., SIMPSON, D., SOKHI, R., SUTTON, P., VIENO, M. and WHYATT, J.D., 2012. Evaluation and inter-comparison of acid deposition models for the UK*. Paper to International Technical Meeting on Air Pollution Modelling and its Applications,* Utrecht, The Netherlands.

DUNKER, A.M., YARWOOD, G., ORTMAN, J.P. and WILSON, G.M., 2002. The decoupled direct method for sensitivity analysis in a three-dimensional air quality model's implementation, accuracy, and efficiency. *Environmental Science and Technology*, 36, 2965-2976.

EC, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient air quality and cleaner air for Europe. *Official Journal of the European Union*, L152, 11 June 2008, 1-44.

FAIRMODE (Forum for Air Quality Modelling in Europe), 2011. *The application of models under the European Union's Air Quality Directive: A technical reference guide*. Technical Report No 10/2011. Copenhagen: European Environment Agency. Available from: http://www.eea.europa.eu/publications/fairmode [Accessed 20 April 2012].

FERGUSON, N. and HARRISON, R., 2010. *SAC-MOD: Modelling, uncertainty and prediction*. Final report of the Science Advisory Council SAC-MOD sub-group. Paper SAC-MOD (10)15. London: Defra. Available from: http://webarchive.nationalarchives.gov.uk/20111030114052/http:/sac.defra.gov.uk/documents/ [Accessed 20 April 2012].

FINARDI, S., SILIBELLO, C., D'ALLURA, A.. and P. RADICE, P., 2012. Analysis of pollutants exchange between the Po Valley and the surrounding European region. *Paper to 8th International Conference on Air Quality - Science and Application.* 19-23 March 2012, Athens, Greece.

*FISHER, B., CHEMEL, C., HU, R.-M. and SOKHI, R., 2010. Emulating complex calculations for regulatory decisions. In *Air Pollution Modeling and its Application XX*, (ed. D.G. Steyn and S.T. Rao), pp. 195–199. Dordrecht: Springer.

*FISHER, B.E.A., CHEMEL, C., FRANCIS, X.V., HU, R.-M., SOKHI, R.S., HAYMAN, G.D., VINCENT, K.J., DORE, A.J., BAGGALEY, M., GRIFFITHS, S., SUTTON, P. and WRIGHT, R.D., 2011. Diagnostic evaluation of complex and simple atmospheric chemical transport models by considering single source impacts in the UK. In *Air Pollution Modeling and its Application XXI* (ed. D.G. Steyn and S. Trini Castelli), pp. 427-431. Dordrecht: Springer.

*FRANCIS, X., CHEMEL, C., SOKHI, R.S., NORTON, E.G., RICKETTS, H. and FISHER, B., 2011. Mechanisms responsible for the build-up of ozone over South East England during the August 2003 heat wave. *Atmospheric Environment*, 45, 6880-6890.

GALMARINI, S. and RAO, S.T., 2011. The AQMEII two-continent regional air quality model evaluation Study: Fueling ideas with unprecedented data. *Atmospheric Environment*, 45, 2464.

GALLILAND, A.B., HOGREFE, C., PINDER, R.W., GODOWITCH, J.M., FOLEY, K.L. and RAO, S.T., 2008. Dynamic evaluation of regional air quality models: assessing changes in O_3 stemming from changes in emissions and meteorology. *Atmospheric Environment*, 42, 5110-5123.

GRICE, S.E., LINGARD, J.J.N., STEDMAN, J.R., COOKE, S.L., YAP, F.W., KENT, A.J., BUSH, T.J., VINCENT, K.J. and ABBOTT, J., 2010. *UK air quality modelling for annual reporting 2008 on ambient air quality assessment under Council Directives 96/62/EC, 1999/30/EC and 2000/69/EC*. AEAT/ENV/R/2859 Issue 1. Harwell, Didcot: AEA Technology.

HAYMAN, G.D., ABBOTT, J., DAVIES, T., THOMSON, C.L., JENKIN, M.E., THETFORD, R. and FITZGERALD, P., 2010. The Ozone Source–Receptor Model: a UK ozone policy tool. *Atmospheric Environment*, 44, 4283-4297.

*HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, R., 2012a. *Comparison of simple and advanced regional models (CREMO): Model evaluation protocol*. Report SC060037a/R. Bristol: Environment Agency.

*HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, D.R., 2012b. *Comparison of simple and advanced regional models (CREMO): Model evaluation report.* Report SC060037b/R. Bristol: Environment Agency.

*HAYMAN, G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, R., 2012c. *Comparison of simple and advanced regional models (CREMO): Ozone diagnostics*. Report SC060037c/R. Bristol: Environment Agency.

*HAYMAN ,G., SOKHI, R., CHEMEL, C., GRIFFITHS, S., VINCENT, K., DORE, A.J., SUTTON, P. and WRIGHT, D.R., 2012d. *Comparison of simple and advanced regional models (CREMO): Model evaluation: Ground-level ozone*. Report SC060037d/R. Bristol: Environment Agency.

KLEIN, H., GAUSS, M., NYIRI, A. and STEENSEN, B.M., 2011. *Transboundary air pollution by main pollutants (S, N, O_3) and PM: The United Kingdom*. MSC-W Data Note 1/2011. Oslo: Norwegian Meteorological Institute.

MIDDLETON, D.R., LUHANA, L. and SOKHI, R.S., 2007. *Overview of indicators for estimating the contribution of Environment Agency regulated sources to regional ozone*. Science Report: SC030171/SR3. Bristol: Environment Agency.

NAPELENOK, S. L., COHAN, D.S., HU, Y. and RUSSELL, A.G., 2006. Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM). *Atmospheric Environment*, 40, 6112-6121.

NMI, 2010. *Transboundary acidification, eutrophication and ground-level ozone in Europe for 2008*. EMEP Status Report 2010. Oslo: Norwegian Meteorological Institute. Available from: http://emep.int/publ/reports/2010/status_report_1_2010.pdf [Accessed 20 April 2012].

PEDERZOLI, A., THUNIS, P., GEORGIEVA, E., BORGE, R. and CARRUTHERS, D., 2011. *Performance criteria for the benchmarking of air quality model regulatory applications: the 'target' approach*. Presented 14th Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, held 2–6 October 2011, Kos, Greece.

RAO, S.T., PORTER, P.S., MOBLEY, J.D. and HURLEY, F., 2011. Understanding the spatio-temporal variability in air pollution concentrations, *EM*, journal of the Air and Waste Management Association, November 2011, 42-48.

RISBEY, J., VAN DER SLUIJS, J.P., RAVETZ, J. and JANNSEN, P., 2001. *A check list for quality assurance in environmental modelling*. Report Number NW&S-E-2001-11. Utrecht: Department of Science, Technology and Society, Utrecht University. Available from: http://www.nusap.net/downloads/reports/E2001-11.pdf [Accessed 20 April 2012].

SCHERE, K., VAUTARD R., SOLAZZO, E., HOGREFE C. and GALMARINI S., 2012. Results and lessons learned from Phase 1 of the AQMEII. EM, the Air and Waste Management Association Magazine for Environmental Managers, July 2012, 30-37.

SCHLUENZEN, K.H. and HALLER M., Editors, 2011. *Evaluation of Integrated Tools*, MEGAPOLI Scientific Report 11-03, MEGAPOLI Deliverable 7.2. http://megapoli.dmi.dk/publ/MEGAPOLI_sr11-03.pdf [Accessed 30 April, 2012].

SILIBELLO, C., D'ALLURA, A., FINARDI, S. and RADICE, P., 2011. Comparison of computed and measured aerosol depth (AOD) over Europe for a year long chemical transport model simulation. *Paper to 14th Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes.* 2-6 October 2011, Kos, Greece.

SMITH, R.I. and FOWLER, D., 2001. Uncertainty in wet deposition of sulphur. *Water, Air and Soil Pollution: Focus*, 1, 341-354.

SMITH, R.I., FOWLER, D., SUTTON, M.A., FLECHARD, C. and COYLE, M., 2000. Regional estimation of pollutant gas deposition in the UK: model description, sensitivity analyses and outputs. *Atmospheric Environment,* 34, 3757-3777.

SNIFFER, 2010. *$PM_{2.5}$ in the UK*. Final Report Project ER12. Edinburgh: Scotland and Northern Ireland Forum for Environmental Research. Available from: http://www.sniffer.org.uk/Webcontrol/Secure/ClientSpecific/ResourceManagement/UploadedFiles/PM25%20Report%20Final%20%2820Dec10%29.pdf [Accessed 20 April 2012].

SOLAZZO, E., BIANCONI, R., VAUTARD, R., APPEL, K.W., BESSAGNET, B., BRANDT, J., CHRISTENSEN, J.H., CHEMEL, C., COLL, I., VAN DER GON, H.D., FERREIR, J., FORKEL, R., FRANCIS, X.V., GRELL, G., GROSSI, P., HANSEN, A., JERIČEVIĆ, A., KRALJEVIĆ, L., MIRANDA, A.I., MORAN, M.D., NOPMONGCOL, U., PIROVANO, G., PRANK, M., RICCIO, A., SARTELET, K.N., SCHAAP, M., SILVER, J.D., SOKHI, R.S., VIRA, J., WERHAHN, J., WOLKE, R., YARWOOD, G., ZHANG, J., RAO, S.T. and GALMARINI, S., 2012a. Model evaluation and ensemble modelling and

for surface-level ozone in Europe and North America. *Atmospheric Environment*, 53, 60-74.

SOLAZZO, E., BIANCONI R., PIROVANO G., MATTHIAS V., VAUTARD R., MORAN M.D., APPEL K.W., BESSAGNET B., BRANDT J., CHRISTENSEN J.H., CHEMEL C., COLL I., FERREIRA J., FORKEL R., FRANCIS X.V., GRELL G., GROSSI P., HANSEN A.B., MIRANDA A.I., NOPMONGCOL U., PRANK M., SARTELET K.N., SCHAAP M., SILVER J.D., SOKHI R.S., VIRA J., WERHAHN J., WOLKEM R., YARWOOD G., ZHANG J., RAO S.T. and GALMARINI S., 2012b. Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII. *Atmospheric Environment*, 53, 75-92.

SUTTON, P., CHEMEL, C., GRIFFITHS, S. and SOKHI, R.S., 2012. *Investigation, using CMAQ, of sensitivity of air quality modelling to input ammonia emissions*. Paper to International Technical Meeting on Air Pollution Modelling and its Applications, Utrecht, The Netherlands.

THUNIS, P., GEORGIEVA, E. and GALMARINI, S., 2011a. *A procedure for air quality models benchmarking. Version 2*. Ispra, Italy: European Commission, Joint Research Centre. Available from: http://fairmode.ew.eea.europa.eu/models-benchmarking-sg4 [Accessed 20 April 2012].

THUNIS P., GEORGIEVA, E. and PEDERZOLI, A., 2011b. *The DELTA tool and Benchmarking report template. Concepts and user guide. Version 2*. Ispra, Italy: European Commission Joint Research Centre. Available from: http://fairmode.ew.eea.europa.eu/models-benchmarking-sg4/delta_tool_concepts_userguide.pdf [Accessed 20 April 2012].

USEPA, 2006. *Interim regulatory impact analysis (RIA) for the proposed National Ambient Air Quality Standards for particulate matter*, Chapter 3, pp. 3-1 to 3-26. Washington DC, US Environmental Protection Agency, Office of Air Quality, Planning and Standards. Available from: http://epa.gov/pm/actions.html [Accessed 20 April 2012].

USEPA, 2011a. *Air quality modeling technical support document: Heavy-duty vehicle greenhouse gas emission standards final rule*. Report EPA-454-R-11-004. Washington DC: US Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division.

USEPA, 2011b. *Heavy-duty vehicle greenhouse gas (HDGHG) emissions inventory for air quality modeling technical support document*. EPA-420-R-11-008. Washington DC: US Environmental Protection Agency, Office of Air Quality Planning and Standards, Office of Air and Radiation Air Quality Assessment Division.

VINCENT, K. AND ABBOTT, J., 2008. *Air quality and deposition benefits from Environment Agency regulation*. Science Report SC060108. Bristol: Environment Agency.

VINCENT, K., COOKE, S. and MURRELLS, T., 2010. *Ozone Source–Receptor Model (OSRM): Assessment of Fawley ozone footprint*. Private communication.

WILLIAMS, M., BARROWCLIFFE, R., LAXEN, D. and MONKS, P., 2011. *Review of air quality modelling in Defra*. Report prepared for Defra by the Air Quality Modelling Review Steering Group. London: Defra. Available from: http://uk-air.defra.gov.uk/reports/cat20/1106290858_DefraModellingReviewFinalReport.pdf [Accessed 20 April 2012].

YARWOOD G., SCORGIE, Y., AGAPIDES, N., TAI, E., KARAMCHANDANI, P., HOLMAN, C., BAWDEN, K., SPENCER, J. and TRIEU, T., 2011. *A screening method for ozone impacts of new sources based on high-order sensitivity analysis of CAMx simulations for Sydney*. Presented at 14th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, held 2–6 October

2011, Kos, Greece. Available from: http://www.harmo.org/Conferences/Proceedings/_Kos/topicIndex.asp?topicID=1 [Accessed 20 April 2012].

*YU, Y., CHEMEL, C., SOKHI, R.S., HAYMAN, G., GRIFFITHS, S., VINCENT, K., DORE, A.J., DERWENT, D., MIDDLETON, D.R. and WRIGHT, R., 2007. Protocols for the CREMO regional model inter-comparison study. Produced for the CREMO project.

YU, Y., SOKHI, R.S., KITWIROON, N., MIDDLETON, D.R. and FISHER, B., 2008. Performance characteristics of MM5–SMOKE–CMAQ for a summer photochemical episode in southeast England, United Kingdom. *Atmospheric Environment*, 42, 4870-4883.

List of abbreviations

ADMS	Atmospheric Dispersion Modelling System
AQMEII	Air Quality Modelling Evaluation International Initiative
AURN	Automatic Urban and Rural Monitoring Network
CBED	Concentration Based Estimated Deposition
CEH	Centre for Ecology and Hydrology
CMAQ	Community Multiscale Air Quality
COMEAP	Committee on the Medical Effects of Air Pollutants
CREMO	Comparison of simple and advanced regional models
CVRMSE	coefficient of variation of the root mean square error
DDM	decoupled direct method
Defra	Department for Environment, Food and Rural Affairs
EC	elemental carbon
EMEP	European Monitoring and Evaluation Programme
FAC2	factor of two
FAIRMODE	Forum for Air Quality Modelling in Europe
FRAME	Fine Resolution Atmosphere Multi-Pollutant Exchange
HARM	Hull Acid Rain Model
HDDM	high-order decoupled direct method
IDOP	integrated downwind ozone production
JEP	Joint Environment Programme
MNB	mean normalised bias
NMB	normalised mean bias
NOx	oxides of nitrogen
OC	organic carbon
OSRM	Ozone Source–Receptor Model
PCM	Pollution Climate Mapping
PM	particulate matter
ppb	parts per billion
ppm	parts per billion
RDE	relative directive error
RSM	Response Surface Model
SMOKE	Sparse Matrix Operator Kernel Emissions
SNIFFER	Scotland and Northern Ireland Forum for Environmental Research
TAPM	The Air Pollution Model
TRACK	Trajectory model with Atmospheric Chemical Kinetics
UK PTM	UK Photochemical Trajectory Model

USEPA	US Environmental Protection Agency
VOC	volatile organic compound
WRF	Weather Research and Forecasting

Appendix: Computational statistics relating to CMAQ runs

The run times and computational resources needed to undertake calculations of the hourly concentrations over the UK for a year using a chemical transport model such as CMAQ will depend on:

- configuration and resolution used;
- computing resource available.

The latter will vary considerably as computational resources improve year by year. As a broad indication, the following run times have been provided by the University of Hertfordshire for the calculations used in CREMO, assuming that the calculation is performed on an 128 PC array:

- 7 days for an annual CMAQ calculation using a spatial grid resolution of 5 km;
- 36 hours for an annual calculation using a spatial grid resolution of 15 km;
- 24 hours for an annual calculation using a spatial grid resolution of 45 km.

However this does not take account of the time required to prepare the data sets needed. Some 1–2 weeks is required to process the emission data for a chosen year. Comparable times to those given above are required to produce the necessary meteorological data using a meteorological driver such as the Weather Research and Forecasting (WRF) model.

Of course these preparation times are not needed for every calculation. One could imagine that an archive of emission and meteorological data could be built up for a number of key years. These could be retrieved and reused according to the application required, reducing the processing time considerably compared with a completely new application of the model.

Furthermore, tasks need not be undertaken sequentially which would shorten the overall time for doing an annual calculation. For example, it might be possible to split the annual CMAQ calculation into four seasonal runs, which could be done in parallel. This would reduce the fine scale CMAQ calculation to as little as two days. Any organisation wishing to provide a routine service, given experienced staff and a suitable archive, could thus offer operational CMAQ output at a speed which could not have been envisaged a few years ago.

Would you like to find out more about us, or about your environment?

Then call us on
08708 506 506* (Mon-Fri 8-6)

email
enquiries@environment-agency.gov.uk

or visit our website
www.environment-agency.gov.uk

incident hotline 0800 80 70 60 (24hrs)
floodline 0845 988 1188

* Approximate call costs: 8p plus 6p per minute (standard landline).
Please note charges will vary across telephone providers

Environment first: This publication is printed on recycled paper.