Environment Agency

Evidence

Comparison of simple and advanced regional models (CREMO)

Model Evaluation Report

Report – SC060037/b

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):
Garry Hayman (Hayman Atmospheric Consulting)
Ranjeet Sokhi and Charles Chemel (University of Hertfordshire)
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)

Dissemination Status:
Publicly available

Keywords:
Regional air quality models, evaluation, comparison, diagnostic, acceptance, criteria, metrics, monitoring networks

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):
Hayman Atmospheric Consulting
22 The Old Pound
Wootton, Abingdon, Oxfordshire, OX13 6BQ
garry_hayman@btinternet.com

Project Number:
SC060037/b

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

The Environment Agency is responsible for the regulation of specific industrial sectors and has long used regional-scale atmospheric chemical transport models to assist in setting emission limits. The Department for Environment, Food and Rural Affairs (Defra) also makes use of such models to assist in the development of policy measures relating to the environmental impacts resulting from such emissions. Defra and the Environment Agency have used a number of different models to cover specific impacts and spatial scales. Since the late 1990s, a number of 'advanced' models have been developed with the capability to address multi-pollutant issues on multiple scales. One of these advanced models is the Community Multiscale Air Quality (CMAQ) modelling system developed originally by the US Environmental Protection Agency.

The overall aim of the 'Comparison of simple and advanced regional models' (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 an assessment tool. In particular, the project evaluated the performance characteristics of advanced regional air quality models 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.

This report describes the application of the model evaluation protocol developed for the CREMO project. It provides a summary of the protocol, discusses possible acceptance criteria, and describes applications of the protocol to operational and diagnostic evaluations of some of the models (CMAQ, TRACK-ADMS and FRAME) examined in the CREMO project. TRACK-ADMS (Trajectory model with Atmospheric Chemical Kinetics–Atmospheric Dispersion Modelling System) is used by the Environment Agency for annual audits and FRAME (Fine Resolution Atmosphere Multi-Pollutant Exchange) for acid deposition. The model evaluation protocol and the outcomes of the CREMO project are detailed in two further reports (Hayman *et al.* 2012 and Fisher *et al.* 2012).

A model evaluation is used to demonstrate that a model is suitable for its intended application. The evaluations undertaken to date do not show any clear difference between the different models. Thus, none of the models can be ruled out as unsuitable for the intended applications. To a certain extent this is to be expected as the evaluations have largely been operational. Further inter-comparisons are planned as a follow up to the CREMO project and in other initiatives.

Contents

1 Introduction

The overall aim of the 'Comparison of simple and advanced regional models' (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 test its capabilities through targeted comparisons with 'simpler' models and with measurements according to agreed model acceptance criteria.

1.1 Background

The emissions of sulphur dioxide (SO_2), oxides of nitrogen (NO_X), 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 (PM).

These impacts do not necessarily occur in the immediate vicinity of the emission source but often involve long-range transport to the affected areas, a result of the timescales for chemical processing of the emissions in the atmosphere.

The Environment Agency is responsible for the regulation of specific industrial sectors and has long used regional-scale atmospheric chemical transport models to assist in setting emission limits. The Department for Environment, Food and Rural Affairs (Defra) also makes use of such models to assist in the development of policy measures relating to the environmental impacts resulting from such emissions, including provision of input into the United Nations Economic Commission for Europe (UNECE) Convention on Long-range Transboundary Air Pollution (CLRTAP) protocols and EU Directives.

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

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

[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.

Since the late 1990s, a number of ‘advanced’ models have been developed with the capability to address multi-pollutant issues on multiple scales. These modelling systems include:

- Community Multiscale Air Quality (CMAQ) modelling system (Byun and Schere 2006);
- Unified European Monitoring and Evaluation Programme(EMEP) model (Simpson *et al*. 2003);
- CHIMERE model (Bessagnet *et a*l. 2009 and references therein).

These are all available for use by the air pollution research community.

In 2007, Defra commissioned a review of its ozone modelling tools (including OSRM) (Monks *et al*. 2007) as part of a wider review of its air pollution modelling activities.The review noted that the UK modelling approach differed from other countries in its use of boundary layer trajectory models. One of its key recommendations (R1.1) was to move to a Eulerian framework as used by advanced models, such as CMAQ and EMEP. Other recommendations were to:

- compare Eulerian model results with the results from observations and with those from comparative Lagrangian models to ensure continuity (R 1.2);
- conduct a model comparison exercise where two of the current Lagrangian-based models are compared to two (or more) regional air quality Eulerian-based models.

As part of the Joint Environment Programme (JEP), the power generators (E.ON and RWE npower) have being using CMAQ to investigate the contribution and significance of the power generation sector (see for example, Griffiths and Lennard 2006, Lennard *et al*. 2006, Sutton 2008).

1.2 Project aims and objectives

The CREMO project had two main aims:

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 CMAQ and 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 PM_{10} and $PM_{2.5}$, based on CMAQ and involving comparison with simpler methods and observations.

These aims were to be 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 their 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 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 report

This report describes the application of the model evaluation protocol developed for the CREMO project (Hayman *et al*. 2012). Section 2 provides a summary of the protocol while Section 3 describes applications of the protocol to operational and diagnostic evaluations of some of the models examined in the CREMO project. Conclusions are presented in Section 4.

2 CREMO model evaluation protocol

The main models participating in the CREMO project were:

- CMAQ modelling system (versions 4.6 and 4.7), operated by the University of Hertfordshire and the power generators (Griffiths and Lennard 2006, Lennard *et al*. 2006, Sutton, 2008, Chemel *et al*. 2010);
- FRAME model, developed and operated by the Centre for Ecology and Hydrology (CEH) (Dore *et al*. 2007, Matejko *et al*. 2009, Vieno *et al*. 2010);
- TRACK-ADMS, developed and operated by AEA Technology (Lee *et al*. 1999a, 1999b, Abbott and Vincent 2006);
- OSRM, developed and operated by AEA Technology (Hayman *et al*. 2010).

Model evaluation is a key part of the iterative cycle of model development, testing and confidence building. An important goal of the performance evaluation is to:

- determine a model's degree of acceptability and usefulness for a specified task;
- establish that the model is providing the results for the right reason.

The Model Evaluation Protocol for the CREMO project prepared by Hayman *et al*. (2012) sets out:

- how the evaluations (operational, dynamic and diagnostic) were to be undertaken;
- the outputs that were to be produced in the CREMO project.

2.1 Acceptance Criteria

As discussed by Hayman *et al*. (2012), there are no theoretical criteria for model acceptance. The criteria that exist have resulted from expert judgement based on an analysis of actual performance.

In the US, model performance goals and criteria have been set for ozone (USEPA 1991, Russell and Dennis 2000) and more recently for $PM_{2.5}$ (Boylan and Russell 2006). Boylan and Russell (2006) define performance 'goals' as the level of accuracy considered to be close to the best a model can be expected to achieve in that application. Performance 'criteria' are defined as the level of accuracy considered to be acceptable for standard modelling applications.

In Europe, the air quality daughter directives specify accuracy criteria for air pollution modelling of the pollutants which they cover. In the first daughter directive (covering sulphur dioxide, oxides of nitrogen, carbon monoxide, lead and particulate matter), the accuracy for modelling (and objective estimation) is defined:

> *as the maximum deviation of the measured and calculated concentration levels, over the period considered by the limit value, without taking into account the timing of the events* (EC 1999).

These criteria have generally been retained in the consolidated Air Quality Directive (EC 2008), although the modelling uncertainty is now based on:

the maximum deviation of the measured and calculated concentration levels for 90 per cent 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 maximum deviation (also called the maximum relative directive error) should not exceed the modelling objective for the specific pollutant and time period. Further work in this area is in progress through the FAIRMODE programme.[2] One of the FAIRMODE activities has been to develop a toolkit to benchmark air quality models.

Derwent *et al*.(2010) recommended two acceptance criteria for the Defra Air Quality Model Evaluation Protocol being used in the Defra Model Evaluation Exercise (Carslaw 2011a,b). These are as follows:

- the fraction of modelled concentrations that lie within ±50 per cent (that is a factor of two) of the observed value should be greater than 50 per cent.
- the normalised mean bias (*NMB*)[3] should be $-0.2 \leq NMB \leq 0.2$ where *NMB* is defined as:

$$NMB = \frac{\sum_{i=1}^{N} [P_i - O_i]}{\sum_{i=1}^{N} O_i}$$

where P_i are the calculated values, O_i are the observed values and N is the number of observed-calculated pairs.

Table 2.1 summarises these model acceptance criteria. If all these statistical measures are within the ranges shown, and the graphical performance procedures are also interpreted to yield acceptable results, the model is judged to be performing acceptably.

[2] http://fairmode.ew.eea.europa.eu/models-benchmarking-sg4
[3] Note that this is different from mean normalised bias.

Table 2.1 Specification of model acceptance criteria

Organisation	Species	Model acceptance criteria
USEPA	O_3	Unpaired highest prediction accuracy: ±15–20% Normalised bias: ±5–15% Gross error of all pairs >60 parts per billion (ppb): 30–35%
	$PM_{2.5}$	Mean fractional error (*MFE*) and the mean fractional bias (*MBF*) are less than or equal to approximately +50% and ±30%, respectively. Additionally, the model performance criteria for major components of $PM_{2.5}$ are met when both *MFE* and *MBF* are less than or equal to approximately +75% and ±60%, respectively.
EU	SO_2, NO_2, NO_X, CO	Hourly: Relative Directive Error (*RDE*) 50% Eight-hour averages: *RDE* (Hayman *et al.* 2012) 50% Daily averages: *RDE* 50% Annual averages: *RDE* 30%
	Benzene	Annual averages: *RDE* 50%
	PM	Daily averages: *RDE* not yet defined Annual averages: *RDE* 50%
	O_3	Hourly: *RDE* 50% Eight-hour averages: *RDE* 50%
UK/Defra	All	The fraction of modelled concentrations that lie within ±50% of the observed value should be greater than 50%. Normalised mean bias (*NMB*) should be $-0.2 \leq NMB \leq 0.2$.

3 Application of the protocol

The CREMO model evaluation protocol (Hayman *et al.* 2012) has been applied to evaluate the operational performance of one or more of the participating models (CMAQ, TRACK-ADMS and FRAME). Details of the applications have been published as follows:

- Operational evaluation: gas phase species (Chemel *et al*. 2010);
- Operational and diagnostic evaluation: deposition (Chemel *et al*. 2011).

The Defra Model Evaluation Exercise is also relevant to the CREMO project (Carslaw 2011a,b). These applications are briefly summarised below.

3.1 Operational evaluation: gas phase species

Chemel *et al*. (2010) undertook the first long-term 'operational' evaluation of the CMAQ model under UK conditions. The model was run on multiple grids using one-way nests down to a horizontal resolution as fine as 5 km over the whole of the UK. The simulation was conducted for the year 2003, which contained several pollution episodes throughout the year (for example, calm weather smogs in February and March, and heat waves in July and August).

The operational evaluation, defined in the CREMO model evaluation protocol (Hayman *et al*. 2011), involved a comparison of model outputs against UK surface measurements of the following air pollutants (namely CO, NO_2, O_3, PM_{10} and SO_2) and species contributing to acidic and nitrogen deposition (namely NH_3, SO_2,HNO_3, and HCl for gases, and SO_4^{2-}, NO_3^-, NH_4^+, Cl^-, and Na^+ for aerosols).

The main findings of this evaluation study were as follows:

- The performance characteristics of the modelling system were found to be variable according to acceptance criteria and to depend on the type (for example, urban, rural) and location of the sites, as well as time of the year (for example, for NH_3).
- The performance of the techniques used for 'operational' evaluation generally conformed to expected levels and ranged from good (for example, O_3, SO_4^{2-}) to moderate (for example PM_{10}, NO_3^-). The moderate performance for PM_{10} is reflected by the moderate performance for NO_3^- and NH_4^+. At a few sites, low correlations and large standard deviations for some species (for example SO_2) suggest that these sites are subject to sources that are not well described in the model. Overall, the model tends to over-predict O_3 and under-predict aerosol species (except SO_4^{2-}). Discrepancies between predicted and observed concentrations may be due to a variety of intertwined factors, which include inaccuracies in meteorological predictions, chemical boundary conditions, temporal variability in emissions, and uncertainties in the treatment of gas and aerosol chemistry. Reasons for these discrepancies have not been clearly identified yet. Further work is required to investigate the respective contributions of such factors on the predicted concentrations.

Chemel *et al*. (2010) noted a number of limitations in the approach adopted for model evaluation in their work. Evaluation techniques that aim at comparing predicted values of the modelled variables with measurements provide only an overall evaluation of model performance (Dennis *et al*. 2010). Indeed these comparisons do not examine whether the results of the model are correct for the right reasons nor how sensitive is the model performance to chemical and meteorological processes.

Such an evaluation (often referred to as ‘diagnostic’ evaluation) complements the ‘operational’ evaluation and is being considered for future work. In particular, further work is needed to evaluate the capabilities of the modelling system to:

- predict the response of regional ozone concentrations to changes in emissions of NOx and VOCs;
- calculate the contribution of regulated industrial emissions to size-speciated PM concentrations and associated chemical species.

In CREMO this ‘diagnostic’ evaluation involved comparison with simpler methods that are already adopted as policy tools in the UK such as:

- the TRACK-ADMS modelling system, combining TRACK and ADMS (Carruthers *et al*. 1994) for annual audits;
- the Photochemical Ozone Creation Potential (POCP) method (Derwent *et al*. 1998, Derwent and Nelson 2003, Derwent *et al*. 2007) and OSRM (Hayman *et al.* 2010) for O_3;
- FRAME for acid deposition.

3.2 Operational evaluation: deposition

Chemel *et al*. (2011) estimated the contributions of the emissions from a UK regulated fossil fuel power station to regional air pollution and deposition for the year 2003 using four air quality modelling systems. The modelling systems varied in complexity and emphasis in the way they treat atmospheric and chemical processes, and include the CMAQ modelling system (versions 4.6 and 4.7), TRACK-ADMS, and FRAME.

An evaluation of the baseline calculations against UK monitoring network data revealed that:

- all modelling systems tended to under-estimate the annual mean air concentrations of SO_2, NOx and PM_{10};
- there was a high variability in the response of the modelling systems for non-sea salt sulphur and nitrogen deposition.

No individual modelling system was found to provide the overall best performance. The CMAQ modelling system version 4.6 was selected as the ‘reference’ for the comparison of the modelled footprint, simply because it was the most comprehensive model used and this version was reasonably well understood. However the selection was arbitrary and should not be considered as implying that CMAQ version 4.6 was the best model used.

Maps of the annual mean air concentration and total deposition increments due to the power station were summarised for each modelling system and compared using a range of diagnostic metrics. The footprint derived by subtracting the footprints of the power station emissions can account for a significant fraction of the local impacts for some species for 2003 (for example, more than 50 per cent for air concentration and non-sea salt sulphur deposition close to the source).The spatial correlation and the coefficient of variation of the root mean square error (CVRMSE) were calculated for each model footprint and for those calculated by the CMAQ modelling system version 4.6. The correlation coefficient quantified the model agreement in terms of spatial patterns and the CVRMSE measures the magnitude of the difference between model footprints.

There were no clear reasons for the differences between model results in terms of the treatment of the key processes within the models. Such differences depend, inter alia, on the treatment of plume chemistry and emissions data processing. For instance, for the CMAQ modelling system, emissions from point sources were mixed instantaneously into the entire grid cell identified at the level of sources plume rise, while for TRACK-ADMS and FRAME point sources, plumes are tracked in a Lagrangian reference frame. In addition, the current theoretical understanding of the processes leading to acid deposition is limited. Detailed process-level studies are needed to pinpoint deficiencies in acid deposition modelling.

There are large uncertainties in the assessment of contributions of industrial sources to regional air pollution and deposition. A critical question that remains to be examined is whether uncertainties such as those reported in the present work still render such model footprints meaningful for policy applications. Quantifying the uncertainty associated with a single modelling system is extremely difficult given the range of inputs and process calculations. Hence an ensemble average of model calculations could be used to provide an estimate of the uncertainty associated with an industrial source footprint. However it has to be recognised that simple air quality modelling systems such as TRACK-ADMS and FRAME still have run times much faster than those of advanced systems such the CMAQ modelling system.

3.3 Defra model intercomparison

A further comparison of simple and complex models in the UK, including those participating in the CREMO project, is being undertaken as part of the Defra Model Evaluation Exercise (Carslaw 2011a,b). There are three separate inter-comparisons addressing different air pollution issues:

- (acid and nitrogen) deposition,
- regional and transboundary pollution;
- urban air quality.

The models in the CREMO project are contributing to one or both of the deposition and regional/transboundary air pollution inter-comparisons (marked with an asterisk in Table 3.1).

Table 3.1 Models participating in the Defra Model Intercomparison Exercise

Deposition	Regional/Transboundary Pollution
CMAQ (University of Hertfordshire)*	CMAQ (University of Hertfordshire) *
CMAQ (JEP)*	CMAQ (AEA Technology) *
EMEP	CMAQ (King's College)
EMEP4UK	EMEP
FRAME *	EMEP4UK
HARM	OSRM *
NAME	AQUM
	NAME

For the model intercomparison related to deposition and regional/transboundary air pollution, hourly concentrations of air pollutants and deposition terms were calculated for the 2006 calendar year; the CREMO project used 2003.

Model outputs were provided to King's College London for processing and calculation of performance metrics. Examples are given of the results for SO_2 and NO_2 in Table 3.2.The complete analysis can be found in Defra (2011a,b). All the models passed the acceptance criteria (Derwent *et al*. 2010).

Table 3.2 Summary Defra Model Evaluation Exercise statistics for annual mean SO_2 and NO_2 concentrations

(a) SO_2

Model	*FAC2*	*MB*	*MGE*	*NMB*	*NMGE*	*RMSE*	*r*
CMAQ.JEP	0.73	1.23	1.30	0.67	0.71	1.84	0.84
CMAQ.UH	0.84	1.18	1.26	0.64	0.68	1.81	0.82
EMEP4UK	0.86	0.41	0.71	0.22	0.39	1.00	0.66
EMEP.Unified	0.73	0.98	1.06	0.53	0.58	1.42	0.77
FRAME	0.96	0.12	0.57	0.07	0.31	0.79	0.77
HARM	0.37	2.05	2.07	1.11	1.12	2.62	0.80
NAME	0.04	4.38	4.38	2.38	2.38	5.45	0.84

(b) NO_2

Model	*FAC2*	*MB*	*MGE*	*NMB*	*NMGE*	*RMSE*	*r*
CMAQ.JEP	0.72	-2.24	2.79	-0.25	0.32	3.23	0.92
CMAQ.UH	0.84	-2.41	2.41	-0.27	0.27	2.74	0.97
EMEP4UK	1.00	-1.58	2.11	-0.18	0.24	2.52	0.94
EMEP.Unified	0.94	-1.13	1.80	-0.13	0.20	2.48	0.92
FRAME	0.97	-0.85	1.91	-0.10	0.22	2.60	0.92
HARM	0.12	-5.32	5.32	-0.60	0.60	6.47	0.94
NAME	0.94	0.02	2.05	0.00	0.23	2.95	0.91

3.4 OSRM: factor of two and normalised mean bias

Hayman *et al*. (2010) derived a range of performance statistics for OSRM from a comparison of modelled and observed concentrations (of ozone) at sites where ozone monitoring is undertaken. The performance statistics included:

- the fraction of modelled concentrations that lie within a factor of two of the observed value;
- the normalised mean bias (NMB), derived for years from 1999 to 2005.

Figure 3.1 shows plots of the percentage of modelled concentrations within a factor of two of the observed values against the normalised mean bias as derived using OSRM.

The upper panel of Figure 3.1 is for the year 2003. It can be seen that the points with normalised mean biases lying between -0.2 and 0.2 usually exceed the condition that

50 per cent of the modelled concentrations lie within a factor of 2 of the observed values. The lower panel shows that this pattern is replicated for other years (that is 1999 to 2005).

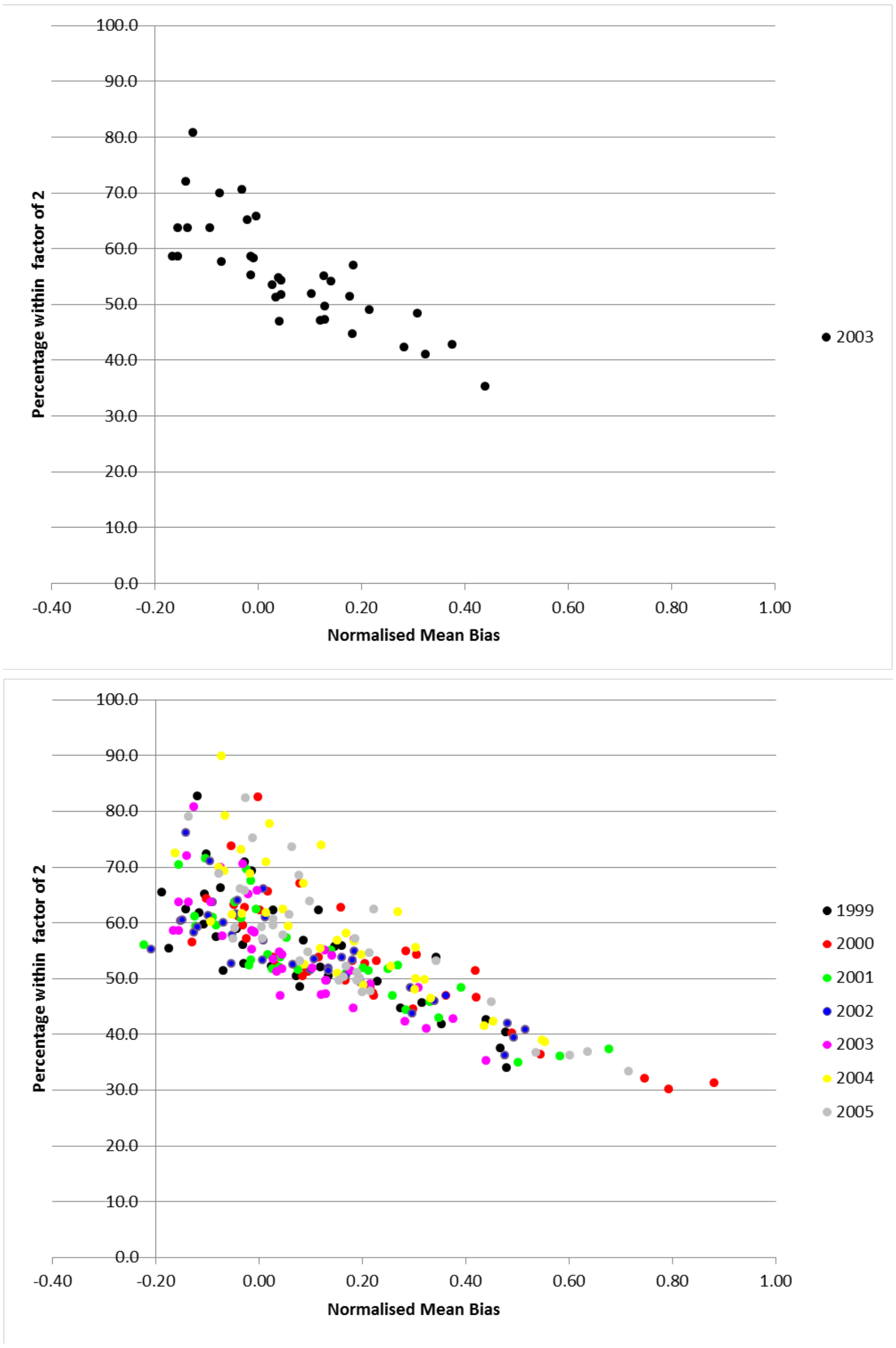

Figure 3.1 Plots of percentage of modelled hourly ozone concentrations within a factor of two of the observed values against *NMB* derived using OSRM for 41 UK sites for 2003 (upper panel) and for 1999–2005 (lower panel)

The maximum hourly ozone concentration in the day has been used to assess the performance of models such as EMEP Unified (Simpson *et al*. 2003). This provides a less stringent test as the need for exact co-incidence in the timing of events is relaxed. Figure 3.2 compares the same statistics derived using the peak daily ozone concentrations for 1999–2005. As expected, the model performance is improved on both measures using the peak daily ozone concentrations.

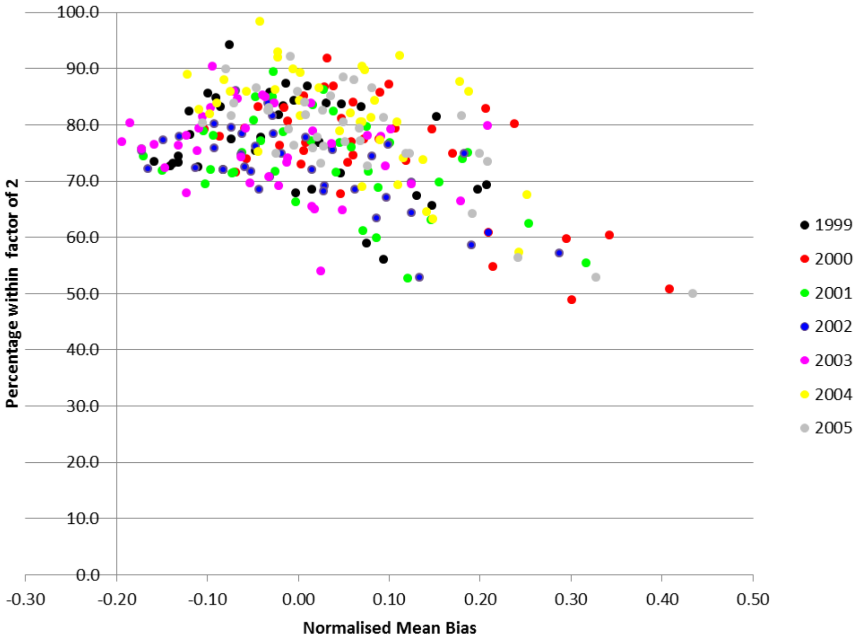

Figure 3.2 Plot of percentage of modelled maximum daily ozone concentrations within a factor of two of the observed values against the *NMB* derived using OSRM for 41 UK sites for 1999–2005

4 Conclusions

Operational and diagnostic evaluations have been undertaken according to the model evaluation protocol developed for the CREMO project (Hayman *et al*. 2012). Chemel *et al*. (2010) have described an operational evaluation of the CMAQ (v4.6) model. The first phase of the Defra Model Evaluation Exercise was also an operational evaluation. The evaluation described by Chemel *et al*. (2011) was a form of dynamic evaluation but was a purely model intercomparison. There were no measurements to evaluate the accuracy of the calculated footprints.

Hayman *et al*. (2012) considered whether there were criteria that could be used to determine if particular models (or model configurations) were acceptable for the specific application. They concluded that the parameters used in the Defra Model Evaluation Exercise (Derwent *et al*. 2010) represented the minimum acceptable level of model performance:

- the fraction of modelled concentrations that lie within ±50 per cent of the observed value should be greater than 50 per cent;
- the normalised mean bias (*NMB*) should be $-0.2 \leq NMB \leq 0.2$.

From the results presented here, the models generally comply with the first Defra criterion and so only the second provides a way of discriminating between models. In one sense, this is not surprising as a successful comparison with observations is a necessary step in a model's development and in its acceptance by the air quality modelling community. Furthermore, for the regional/transboundary pollution models, the comparison used only rural ozone monitoring sites. A more stringent test would have required the use of (sub)urban background sites, not only for ozone but also for oxides of nitrogen (NO and NO_2).

In its report on the requirements of environmental models for regulatory applications, the US National Research Council (NRC 2007) described the key elements of model evaluation. Model evaluation is an ongoing process and the key questions posed by Beck (2002) continue to provide a useful framework:

1. Is the model based on generally accepted science and computational methods?
2. Does it work, that is, does it fulfil its designated task or serve its intended purpose?
3. Does its behaviour approximate that observed in the system being modelled?

The evaluations undertaken to date do not show any clear difference between the different models. Thus, none of the models can be ruled out as unsuitable for the intended applications. To a certain extent, this is to be expected as the evaluations have largely been operational.

The evaluations have largely focused on the annual timescale. The Defra Model Evaluation Exercise has shown the value (for ozone) of considering different time periods.

An earlier comparison of simple regional transport models by the Environment Agency for acid deposition (Abbott *et al*. 2001) suggested that regional transport models could be expected to meet the 'factor of two' criterion. The 95th percentile of the predicted deposition rates for the simple models available at the time of the study: TRACK, HARM (Hull Acid Rain Model) and FRAME, was within a factor of two of the annual average value and the 5th percentile was within half the annual average value.

One of the advantages of the more complex chemical transport models available today is the ability to test their performance for short-term average concentrations (daily and hourly) and to be able to investigate more fully, complex interactions and feedbacks between processes.

5 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.

BECK, M.B., 2002.Model evaluation and performance. In *Encyclopedia of Environmetrics* (ed. A.H. El-Shaarawi and W.W. Piegorsch), pp. 1275-1279. Chichester: Wiley.

BESSAGNET, B., MENUT, L., CURCI, G., HODZIC, B., GUILLAUME, A., LIOUSSE, C., MOUKHTAR, S., PUN, B., SEIGNEUR, C. and SCHULZ, M., 2009. Regional modeling of carbonaceous aerosols over Europe –focus on secondary organic aerosols, *Journal of Atmospheric Chemistry,* 61, 175–202.

BOYLAN, J W. and RUSSELL, A.G., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmospheric Environment*,40, 4946-4959.

BYUN, D. and SCHERE, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modelling system. *Applied Mechanics Review*, 59, 51-77.

CARRUTHERS, D.J., HOLROYD, R.J., HUNT, J.C.R., WENG, W.-S., ROBINS, A.G., APSLEY, D.D., THOMPSON, D.J. and SMITH, F.B., 1994. UK-ADMS: a new approach to modelling dispersion in the earth's atmospheric boundary layer. *Journal of Wind Engineering and Industrial Aerodynamics*, 52, 139-153.

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 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].

CHEMEL, C., SOKHI, R.S., YU, Y., HAYMAN, G.D., VINCENT, K.J., DORE, A.J., 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.

CHEMEL, C., SOKHI, R.S., DORE, A.J., SUTTON, P., VINCENT, K.J., GRIFFITHS, S.J., , 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.

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.G., JENKIN, M.E., SAUNDERS, S.M. and PILLING, M.J., 1998. Photochemical ozone creation potentials for organic compounds in North West Europe calculated with a Master Chemical Mechanism. *Atmospheric Environment,* 32, 2419-2441.

DERWENT, R.G., JENKIN, M.E., PASSANT, N.R. and PILLING, M.J., 2007. Photochemical ozone creation potentials (POCPs) for different emission sources of organic compounds under European conditions estimated with a Master Chemical Mechanism. *Atmospheric Environment*, 41, 2570-2579.

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 [Accessed 18 April 2012].

DORE A.J., VIENO, M., TANG, Y.S., DRAGOSITS, U., DOSIO, A., WESTON, K.J. and SUTTON, M.A., 2007. Modelling the atmospheric transport and deposition of sulphur and nitrogen over the United Kingdom and assessment of the influence of SO_2 emissions from international shipping. *Atmospheric Environment*, 41, 2355-2367.

EC, 1999. Directive 1999/30/EC of the Council of the European Union of 22 April 1999 relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air. *Official Journal of the European Union*, L163, 29 June 2009, 41-60.

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.

*FISHER B.E.A., CHEMEL, C., FRANCIS, X.V., HU, R.-M., SOKHI, R.S., HAYMAN, G.D., VINCENT, K.J., DORE, A.J., GRIFFITHS, S., SUTTON, P. and WRIGHT, R.D., 2012. *Comparison of simple and advanced regional models (CREMO) Outcomes for the Environment Agency*. Bristol: Environment Agency.

GRIFFITHS, S.J. and LENNARD, R.J., 2006. *Simulation of UK air quality and deposition impacts in 2010 using Models-3*. Joint Environment Programme (JEP) Report PT/06/BE1368/R.

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.D., SOKHI, R.S., CHEMEL, C., GRIFFITHS, S.J., VINCENT, K.J., DORE, A.J., SUTTON, P. and WRIGHT, R., 2012. *Comparison of Simple and Advanced Regional Models (CREMO): Model evaluation protocol.* Report SC060037a/R. Bristol: Environment Agency.

LEE, D.S., KINGDON, R.D., JENKIN, M.E. and WEBSTER, T., 1999a. Modelling the contribution of different sources of sulphur to atmospheric deposition in the United Kingdom. *Environmental Modelling and Assessment*, 5, 105-118.

LEE, D.S., KINGDON, R.D., PACYNA, J.M., BOUWMAN, A.F. and TEGEN, I., 1999b. Modelling the atmospheric transport and deposition of calcium in the UK and northern Europe. *Atmospheric Environment*, 33, 2241-2256.

LENNARD, R.J., GRIFFITHS, S.J. and SUTTON, P., 2006. *Simulation of European emissions impacts on particulate matter concentrations in 2010 using Models-3*. Joint Environment Programme (JEP) Report PT/06/BE1376/R.

MATEJKO, M., DORE, A.J., HALL, J., DORE, C.J., BŁAŚ, M., KRYZA, M., SMITH, R. and FOWLER, D., 2009. The influence of long term trends in pollutant emissions on deposition of sulphur and nitrogen and exceedance of critical loads in the United Kingdom. *Environmental Science and Policy*, 12, 882-896.

MONKS, P.S., BLAKE, R.S. and BORRELL, P., 2007.*Review of tools for modelling tropospheric ozone formation and assessing impacts on human health and ecosystems.* A report prepared for the Department for Environment, Food and Rural Affairs. London: Defra. Available from: http://randd.defra.gov.uk/Document.aspx?Document=AQ0706_6733_FRP.pdf [Accessed 18 April 2012].

NRC, 2007. *Models in environmental regulatory decision-making*. Report prepared by the Committee on Models in the regulatory decision process for the National Research Council of the National Academies.Washington DC: The National Academies Press. Available from:http://www.nap.edu/catalog/11972.html [Accessed 18 April 2012].

RUSSELL, A. and DENNIS, R., 2000. NARSTO critical review of photochemical models and modeling. *Atmospheric Environment*, 34, 2283-2324.

SIMPSON, D., FAGERLI, H., JONSON, J.E., TSYRO, S., WIND, P. and TUOVINEN, J.-P., 2003.*Transboundary acidification, eutrophication and ground level ozone in Europe.Part I: Unified EMEP model description*.EMEP Report 1/2003.Oslo: Norwegian Meteorological Institute. Available from: http://emep.int/publ/reports/2003/emep_report_1_part1_2003.pdf [Accessed 18 April 2012].

SUTTON, P., 2008. *Use of long-range transport models for habitats assessment of power stations*. Presentation given at meeting,'Models-3/CMAQ: recent developments and applications', held Loughborough University, November 2008. Available from: http://ncasweb.leeds.ac.uk/mesomaq/images/Meetings/3_habitatsassess_psutton.pdf [Accessed 18 April 2012].

USEPA, 1991. *Guideline for regulatory application of the urban air shed model*. Research Triangle Park, NJ: US Environmental Protection Agency. Available from: www.epa.gov/scram001/guidance/guide/uamreg.pdf [Accessed 18 April 2012].

VIENO, M., DORE, A.J., BEALEY, W.J., STEVENSON, D.S. and SUTTON, M.A., 2010. The importance of source configuration in quantifying footprints of regional atmospheric sulphur deposition. *Science of the Total Environment*, 408, 985-995.

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.