KING'S College LONDON

Defra Phase 2 regional model evaluation

King's College London
Centre for Ecology and Hydrology
Met Office
rdscientific
Ricardo-AEA
University of Hertfordshire
University of Manchester

David Carslaw, Paul Agnew, Sean Beevers, Charles Chemel, Sally Cooke, Lucy Davis, Dick Derwent, Xavier Francis, Andrea Fraser, Nutthida Kitwiroon, Gordon McFiggans, Tim Murrells, Alison Redington, Nicholas Savage, Ranjeet Sokhi and Massimo Vieno

version: 10th October 2013

Summary key points

General points

1. Nine models or model variants have been considered in this report: AEA-CMAQ and OSRM from Ricardo-AEA; AQUM and NAME from the Met Office, EMEP-4-UK from CEH, KCL-CMAQ from King's College London, the PTM from rdscientific, Hertfordshire-CMAQ from the University of Hertfordshire and WRF-Chem from the University of Manchester. The models include three variants of the CMAQ model, set up and run according to practises at each institution. In addition to these models, an 'ensemble' model was constructed by taking the mean hourly values by measurement site of six of the models that predicted at every site. The ensemble model was evaluated in the same way as the main models to determine whether ensemble modelling improved the comparison with measurements. The ensemble model also conveniently shows the 'typical' response of models.

2. Almost all models provided predictions for 23 measurement sites across the UK and Mace Head in Ireland. These receptors include 17 sites in rural areas and 7 at urban background locations in London, Manchester and Birmingham. Hourly mean predictions of O_3, NO_x (NO and NO_2), NO_y, HNO_3, H_2O_2 were made for each receptor location for 2006. Most results have been considered separately for rural and urban locations.

3. The evaluation has focused on several key areas. These areas include an evaluation at a location where 'baseline' or 'background' O_3 dominates (Mace Head), the ability of the models to predict specific O_3 metrics of relevance to air quality policy, the effect of precursor (NO_x and VOC) emission reductions at a UK and European scale and the performance of the models in predicting common meteorological variables measured at Met Office surface stations in the UK.

Comparison with measurements

4. The models are shown to predict annual mean rural O_3 concentrations within 10%, with a tendency towards a slight positive bias.

5. The models tend to show a positive bias when predicting the annual mean O_3 concentration at urban receptors. The higher postive bias in urban areas (typically around +20%) is likely due to model grid resolution and the ability to adequately model urban NO_x concentrations. Nevertheless, KCL-CMAQ does not show as strong a positive bias and has a lower error in urban areas compared with other models. The estimation of urban NO_x emissions could also be important e.g. underestimates would tend to result in increased concentrations of urban O_3.

6. For higher concentration O_3 metrics the models give a wider range of predictions. For the number of days with a daily maximum of running 8-hour means >100 µg m^{-3} (50 ppb), the range of predictions across the models is large at about ± 25 days — compared with a typical observed value of about 30 days. The models predict most sites within a factor of two, but not within a mean bias of $\pm 20\%$.

7. For the maximum daily mean O_3 concentration almost all models predict within a $\pm$20% bias, with a tendency towards negative bias in both rural and urban locations.

8. The models do not tend to explain the variation between sites very well for the three metrics considered above, as most of the time the predictions are no better than taking the mean of all measurements.

9. As noted above, most models fail to capture higher O_3 concentrations well. Cluster analysis shows (using Lullington Heath as an example) that the 'missed' conditions arise when air-masses come from continental Europe in summertime. For AEA-CMAQ highest O_3 concentrations at Lullington Heath are associated with Atlantic air-mass origins, suggesting that it misses some of the important high concentration conditions when air masses are from Europe.

10. Consideration has also been given to predictions of NO_x and NO_2 in urban and rural areas. Broadly speaking the models tend to underestimate NO_x concentrations in urban areas and overestimate concentrations in rural areas. For all areas, typically about half the annual mean NO_x predictions are within a factor of two. While it would be expected that large grid models would tend to underestimate concentrations in urban areas, not all models show this characteristic. In particular, the KCL-CMAQ model predicts NO_x, NO_2 (and O_3) well in urban areas.

11. Brief comparisons have been made with the new daily air quality index for O_3, which is on a scale from 1–10 ('low' to 'very high') over June and July 2006. These results show that all models tend to overestimate the index (concentration) for low values of the index and underestimate the index when the index (concentration) is 'high'. The variation between the models is substantial. For example, when the observed O_3 concentration is at level 7 (High) the corresponding modelled O_3 is typically at level 4, the lowest level on the moderate scale (on average). However, there are large differences between the models. For the same conditions, AQUM for example would estimate the index to be 5 (middle of the moderate range), whereas WRF-Chem would be 3.

12. Overall it is difficult to identify model performance consistency. Models that do well for some higher concentration metrics do less well with others.

13. The predictions made by the models for threshold-type statistics such as the number of days where the maximum rolling 8-hour mean is >100 μg m^{-3} are very sensitive to small uncertainties in the predicted O_3 concentration. For example, for the EMEP-4-UK model at Harwell, the estimate varies from 20 to 78 days for a $\pm$10% uncertainty in predicted O_3 concentration. This variation is similar to the range seen across the different models. Care should be exercised when comparing model predictions for threshold-type statistics, even if they are important health-based standards or limits.

14. In urban areas there is a relationship between the mean bias of a NO_x (and O_3) concentration and model grid size and emission. Models with larger grid sizes tend to underestimate concentrations of NO_x and overestimate concentrations of O_3. In addition, there is an important effect of the emission estimate used. The KCL-CMAQ model tends to predict urban NO_x and O_3 concentrations well. This

model has a relatively small grid size and assumes higher urban NO_x emissions than other models. These results suggest that the models can predict urban O_3 and NO_x concentrations well provided they use a suitable grid size and representative NO_x emissions estimates.

Boundary conditions — analysis of Mace Head data

15. The model predictions have been compared at the Mace Head site, which is most strongly affected by the model boundary conditions assumed in each model to better understand the influence of the lateral boundary conditions used by the models.

16. Cluster analysis has been used to extract air-mass origins from the 'clean' Atlantic to understand how the models differ. It is found that EMEP-4-UK and Hertfordshire-CMAQ have the best overall agreement for these air-mass origins. Furthermore, the NAME model is shown to systematically underestimate O_3 concentration by $\approx$6 ppb and the KCL-CMAQ model overestimate concentrations by about 3 ppb. Very similar results are seen when using a more comprehensive air mass allocation approach to isolate 'baseline' conditions. It is difficult however to relate the boundary conditions for O_3 to predictions of O_3 at sites across the UK.

17. The model ensemble provided the best overall agreement with measurements across all the comparisons made.

Effect of precursor emission reductions

18. Four emission reduction scenarios have been considered to test how the model predictions of O_3 respond for different O_3 metrics. These are:

    Scenario S1 reduction of total anthropogenic NO_x and VOC by 30% across the UK + Europe;

    Scenario S2 reduction of total anthropogenic NO_x and VOC by 30% across the UK only;

    Scenario S3 reduction of total anthropogenic NO_x by 30% across UK + Europe and,

    Scenario S4 reduction of total anthropogenic VOC by 30% across UK + Europe.

    These emission reduction scenarios are primarily used to test how the models respond to emission changes rather than representing specific policy-relevant reduction scenarios.

19. For *annual mean O_3* concentrations For scenarios S1 to S3, all models show that an increase in O_3 concentrations can be expected at both urban and rural locations. The increase in O_3 is more pronounced in urban areas due to the reduction in emissions of NO_x. Only scenario S4 (30% reduction in VOCs) would lead to more general reductions in the concentration of O_3. There is, however, a very wide range in O_3 predictions at specific receptors. For example, for scenario S1 at Harwell, the annual mean O_3 is predicted to increase by between 1 to 6 $\mu g\ m^{-3}$. There is only a minor difference in the effect of Scenario S1 (NO_x/VOC reduction UK + Europe) and scenario S2 (NO_x/VOC reduction UK only).

20. For a *higher concentration metric* such as the number of days with a daily maximum of running 8-hour means >100 μg m^{-3}, the effect of the scenarios is as follows. For S1 there tends to be a small increase in O_3 concentrations in urban areas and small decreases in the most rural areas and little change predicted elsewhere. For S2 and S3 almost all locations show an increase in the number of days exceeding can be expected. Again, reducing only VOCs (S4) is the only scenario that consistently reduces exceedances days across all sites by between 1 and 6.

21. A general and important characteristic of all scenarios run is that the models give a very wide range in responses to O_3 predictions. The wide range of responses is important because Defra can expect the predicted efficacy of measures to vary widely depending on model choice. This is the case both for longer term average concentrations e.g. annual means and higher concentration O_3 metrics.

22. For the higher concentration O_3 metrics both the sign and the magnitude of the O_3 change can vary depending on the model used. For example, the number of days with a daily maximum of running 8-hour means >100 μg m^{-3} for scenario S1 in rural areas the models predict either a slight overall increase in the days exceeding the limit (AQUM, 1.6 days on average) to a relatively large decrease (EMEP-4-UK, 6.1 days on average). In urban areas there is an even larger range of responses e.g. a decrease of 2 days for EMEP-4-UK to an increase in 7 days for KCL-CMAQ. The increases for KCL-CMAQ is likely related to the generally better performance for NO_x concentration predictions in urban areas for this model, where reductions in NO_x emissions would be expected to increase urban O_3 concentrations.

NO_x and VOC sensitivity

23. The scenarios S3 and S4 (NO_x and VOC reduction UK + Europe) allows an investigation of NO_x and VOC-sensitive regimes i.e. whether reducing NO_x or VOCs would be most beneficial for O_3 reduction. These effects are analysed with respect to different 'indicator species' that help to show whether a model tends to be in a NO_x or VOC-sensitive regime. These results show the transition from VOC to NO_x-sensitive regimes is reasonably consistent across the models and consistent with the original findings of Sillman (1995). However, the results from the NAME model could highlight issues with the data and the Hertfordshire-CMAQ results show much lower sensitivity to changes in NO_x or VOCs compared with the other models. The latter results may explain why Hertfordshire-CMAQ tends to result in less change in O_3 concentrations for scenarios S3/S4 in comparison with the other models.

24. The results show that for peak (maximum daily of rolling 8-hour mean) O_3 concentrations at Harwell that the models range from almost entirely VOC-sensitive (AQUM and NAME) to almost entirely NO_x sensitive (Hertfordshire-CMAQ). It should be noted however that O_3-NO_x-VOC sensitivity at individual locations and for specific events is often very uncertain and the these results reflect that point. The sensitivity will be very dependent on both emission assumptions (e.g. adequacy of absolute emission estimates and VOC speciation) but also the whole chemistry-transport system. It is therefore perhaps not very surprising that the models do not agree well in this respect. However, from a policy perspective these

findings are important because it cannot be said with certainty whether reducing NO_x or VOCs is the most beneficial approach for reducing peak O_3 concentrations.

25. Considering a wider range of O_3 sites reveals more consistent model behaviour. For example, the models tend to show a south-east, north-west gradient, with the south east of England being more VOC-sensitive and the north-west of the UK and Ireland being more NO_x sensitive. AQUM, NAME and EMEP-4-UK tend to predict that the south-east of England is more VOC-sensitive than the other models. The models are therefore reasonably consistent in identifying the spatial variation in NO_x and VOC sensitivity, but as noted previously, in intermediate environments there can be a wide range of model responses.

26. There is more consistency in answer to the question whether action taken at a UK level is more effective than action at a UK+European level. The results show that for almost all models and locations it is better to take action at a UK+European level than only at a UK level. However, the NAME model tends to show a greater number of days at all sites where action taken at a UK level is more effective. Furthermore, there is some indication for a few sites in the south-east of England that this is also the case.

27. It is found that there is a strong relationship between the European VOC emission estimates used in the models and the extent to which the models are NO_x or VOC sensitive. In particular, the assumptions concerning European biogenic VOC (BVOC) estimates vary considerably between the models and influence the extent to which the models respond to NO_x or VOC control. For example, O_3 predictions using NAME (zero BVOC) are much more sensitive to VOC reductions compared with Hertfordshire-CMAQ for example that assumes much higher European BVOC emissions. Indeed, it can be shown that at a site such as Harwell the models vary from being predominately VOC sensitive to NO_x sensitive, which depends on the VOC emission assumptions.

Surface meteorological predictions

28. Most models have provided hourly predictions of common meteorological variables including wind speed, wind direction, ambient temperature and relative humidity. These predictions were made at 11 Met Office surface stations for 2011.

29. In general the models predict the surface variables above well with the best performing models overall being AQUM, KCL-CMAQ and NAME for wind speed. There was some evidence that AEA-CMAQ and Hertfordshire-CMAQ tended to show less diurnal variation in wind speeds than is present in the measurements. Considering the relationship between surface O_3 predictions and meteorology shows that there is no obvious relationship between poor O_3 predictive performance and simultaneously poor performance in meteorological predictions; except perhaps for the urban receptors.

30. Across all meteorological sites AEA-CMAQ and KCL-CMAQ show a similar positive bias of $\approx$20 degrees. AQUM and NAME show a slight positive bias of 7 to 12 degrees. Hertfordshire-CMAQ only has a bias of 4 degrees. The analysis also

shows that the spread in wind directions for the models tends to be quite narrow and the most important characteristic of wind direction predictions in the positive bias shown by several of the models.

31. Model predictions of boundary layer height (BLH) vary considerably between the models. However, these variations do not translate to clear differences in model performance for O_3. BLH itself may not be a sufficiently useful quantity in this respect and a deeper consideration of the vertical structure of the atmosphere may be necessary.

32. It should be stressed that the meteorological comparisons consider only surface observations and not profile measurements. The latter may reveal more useful model deficiencies. In particular, there is no clear relationship between the quality of the meteorological prediction and the concentration of O_3, that could perhaps be revealed by considering by other meteorological variables or derived quantities related to vertical mixing.

33. The generally good quality meteorological predictions suggests that modelled values could usefully be substituted for surface observations or help predict in locations without surface measurements.

Contents

1 Introduction

1.1 Background

This report summarises the main findings of Phase 2 of the Defra Model Evaluation Exercise. The report focuses on the use of models that consider transboundary air pollution, and specifically the prediction of surface O_3 concentrations. In Phase 1, the evaluations were focused on comparing hourly concentrations of O_3 across 11 measurement sites in the UK.[1] Phase 2 builds on this analysis and extends Phase 1 to consider a wider range of issues following a meeting with the model groups.

Model evaluations can cover an enormous range of issues and can involve very detailed types of analysis. The main aims of this report are:

1. To focus on important metrics for ambient O_3 concentrations. There are many O_3 metrics in use which are directly related to European Directives or wider health concerns and it is useful to know how well the models predict O_3 concentrations for a range of these metrics.

2. To consider model performance with respect to NO_x and NO_2 and how well the models predict common surface meteorological measurements including wind speed, wind direction, ambient temperature and relative humidity.

3. To understand the range in predicted concentrations that might be expected through using different modelling systems. Defra could in principle rely on any one of these models for policy purposes and it is important to know how sensitive predictions are to different models and model configurations.

4. To understand how O_3 concentrations respond to different precursor emission controls. In this report several scenarios at the European and UK scale are assessed that consider reductions in the emissions of NO_x and VOCs.

It should be stressed that while many useful comparisons can be made between the models, there are limits to the analyses that can be undertaken before specific model sensitivity runs are required e.g. to test the importance of emission assumptions on the predictions.

Similar to the Phase 1 report, the analysis of the models is based on the year 2006. This year was initially chosen because many of the groups involved had already configured their models for 2006 as part of other activities e.g. for the USA-European Air Quality Model Evaluation International Initiative (AQMEII) (Galmarini and Rao 2011; Rao et al. 2011; Appel et al. 2012). It is useful to put 2006 into perspective by comparing some of the key O_3 statistics with those since 2000 to 2012. At Harwell for example the annual mean O_3 concentration was 27 ppb and was the 5th highest over those 13 years. However, 2006 was a more important year for peak O_3 concentrations. The maximum rolling 8-hour mean O_3 concentration was 92 ppb and was the third highest over the 12 years. Finally, in terms of the days where the maximum rolling 8-hour mean is >50 ppb (100 μg m^{-3}) it can be shown 2006 was a high year with 31 days; second only to 2003 (66 days). In 2006 the high peak concentrations of O_3 were experienced during July, which is considered specifically in this report.

[1] The phase 1 report can be found at http://uk-air.defra.gov.uk/reports/cat20/1105091514_RegionalFinal.pdf

During 2006 there were also periods where biomass burning in Russia had an influence on air quality in the UK. For example Saarikoski et al. (2007) report periods of high $PM_{2.5}$ concentrations during the spring of 2006 in Helsinki. Witham and Manning (2007) report elevated PM_{10} concentrations in the UK during May and September 2006 due to Russian wild fires. It is possible given the relative increased frequency of these events during 2006 compared with other years that they may also have had an influence of O_3 concentrations. Such an influence is difficult to detect in a clear way in the observations but may nevertheless have had an influence. Considering another year would help to establish the importance of these effects.

1.2 The models evaluated

A wide range of models for predicting surface O_3 concentrations have been considered in this report. Not all models provided hourly mean results for all receptors for a full year. WRF-Chem provided results for mid June to the end of July for base case and Scenario 1. The PTM model provided results for Harwell only at a single time (3:00–3.15pm) each day. Where possible full data sets have been analysed i.e. the whole of 2006 because that makes it possible to compare the model outputs with a range of O_3 metrics e.g. UK/EU limits which generally require a full year of data. The lack of full data sets for all models means that it has not been possible to evaluate the performance of all models in an entirely consistent way.

The models made predictions at the receptor locations shown in Table 1.1.[2] While the types of model evaluated in this report (i.e. regional, large grid size models) would generally be used for predictions of O_3 in rural areas it is also important to consider their performance in urban areas. Urban areas will have higher populations exposed and O_3 concentrations can be expected to increase in urban areas in future as local NO_x emissions are increasingly controlled.

Meteorological predictions were provided by all models except OSRM and the PTM. The following text gives a brief overview of the models based on the responses to a questionnaire developed by Defra.

[2]Note that the NAME model did not provide predictions at all sites shown in Table 1.1.

Table 1.1: Receptor locations considered in this report.

site	code	latitude	longitude	site.type
Aston Hill	AH	52.50	-3.03	Rural Background
Birmingham Tyburn	BIR1	52.51	-1.83	Urban Background
Bottesford	BOT	52.93	-0.81	Rural Background
Bush Estate	BUSH	55.86	-3.21	Rural Background
London Bloomsbury	CLL2	51.52	-0.13	Urban Background
Eskdalemuir	ESK	55.32	-3.21	Rural Background
Glazebury	GLAZ	53.46	-2.47	Rural Background
Harwell	HAR	51.57	-1.33	Rural Background
High Muffles	HM	54.33	-0.81	Rural Background
London N. Kensington	KC1	51.52	-0.21	Urban Background
Ladybower	LB	53.40	-1.75	Rural Background
Lullington Heath	LH	50.79	0.18	Rural Background
Lough Navar	LN	54.44	-7.90	Rural Background
London Eltham	LON6	51.45	0.07	Suburban Background
Manchester Piccadilly	MAN3	53.48	-2.24	Urban Background
Manchester South	MAN4	53.37	-2.24	Suburban Background
Mace Head	MH	53.33	-9.90	Rural Background
Rochester Stoke	ROCH	51.46	0.63	Rural Background
Sibton	SIB	52.29	1.46	Rural Background
Strath Vaich	SV	57.73	-4.78	Rural Background
London Teddington	TED	51.42	-0.34	Urban Background
Wicken Fen	WFEN	52.30	0.29	Rural Background
Yarner Wood	YW	50.60	-3.72	Rural Background

1.2.1 AEA-CMAQ

In common with two other groups Ricardo-AEA use the EPA Community Multiscale Air Quality (CMAQ) modelling system http://www.cmaq-model.org. CMAQ is designed for applications ranging from regulatory and policy analysis to understanding the complex interactions of atmospheric chemistry and physics. It is a three-dimensional Eulerian atmospheric chemistry and transport modelling system that simulates ozone, particulate matter (PM), toxic airborne pollutants, visibility, and acidic and nutrient pollutant species throughout the troposphere. Designed as a 'one-atmosphere' model, CMAQ can address the complex couplings among several air quality issues simultaneously across spatial scales ranging from local to hemispheric.

Numerical weather data are produced using WRFv3 on the same scale as CMAQ. Boundary and forcing conditions are provided by ECMWF for 2006.

At Ricardo-AEA it has been run at horizontal resolutions of 48 km (Europe) and 12 km (UK) for this study and the AQ forecasting. The 48+12 km simulation uses a 26 layer vertical structure with 12 layers below 800m and a lowest layer of 9 m.

European emissions are based on the 2006 EMEP emissions UK emissions are based on the 2006 NAEI. Temporal profiles from Jenkin et al (2000) are used for the main emissions SNAP sectors. Natural emissions are based on the Biogenic Potential Inventory.

1.2.2 AQUM

AQUM is a limited area configuration of the Met Office Unified Model (MetUM) which uses the UKCA chemistry scheme (Savage et al. 2013). The MetUM is a sophisticated system capable of modelling regions from limited areas to globally and with timescales

from less than hourly to climate scales. UKCA development first began in 2003 as part of a joint project initially comprising the Met Office and the universities of Cambridge and Leeds, with the aim of building a chemistry and aerosols sub-model within the Met Office's Unified Model for use in climate modelling. Since 2005, AQUM (Air Quality in the Unified Model) has been developed by the Met Office as a configuration of UKCA for modelling regional air quality. AQUM is run online, as part of the Met Office Unified Model, which is an Eulerian meteorological model.

For modelling air quality in the United Kingdom, the following emissions data sets are typically used: NAEI emissions at 1 km × 1 km resolution over the UK, ENTEC — 5 km × 5 km emissions (2007) for shipping surrounding the UK and EMEP emissions at 0.5° × 0.5° over the remainder of Europe. AQUM uses the RAQ (Regional Air Quality) scheme, which is an updated version of the STOCHEM chemical mechanism. Dry deposition is based on a Wesely scheme.

1.2.3 EMEP-4-UK

The EMEP-4-UK model (Vieno et al. 2010) is based (and kept updated) on the EMEP MSC-W model (Simpson et al. 2012). The development of the EMEP-4-UK model first stared in 2006 by Massimo Vieno (University of Edinburgh, CEH Edinburgh), Peter Wind and David Simpson (Norwegian Meterological Institute).

Meteorology is based on WRF 2.2, 3.1.1, and 3.2. For emissions EMEP uses the NAEI for the UK and EMEP everywhere else. EMEP-4-UK was run at a spatial resolution of 50 km × 50 km to provide initial and boundary conditions. It was then run at a finer resolution of 5 km × 5 km for the main model results.

For chemistry the model uses the EMEP MSC-W model chemistry scheme, although a run has also been based on an alternative scheme (see below). For deposition use is made of the EMEP MSC-W model deposition scheme.

1.2.4 Hertfordshire-CMAQ

The University of Hertfordshire also use CMAQ (together with Ricardo-AEA and KCL). The version used in this study was 4.7.1 (released June 2010). Hertfordshire-CMAQ used a grid resolution of 18 km and WRF 3.2.1. Emissions were based on the TNO inventory and chemistry from CB05. The boundary conditions were based on data from Geos-Chem.

1.2.5 KCL-CMAQ

KCL also use the CMAQ model described above. In the present study, European emissions were based on EMEP (2005) and UK emissions on NAEI (2005) and (2006). Meteorological data were based on WRF3.1. The chemical mechanism used was Carbon Bond-05 with aerosol and aqueous chemistry. The dry deposition scheme was based on a surface exchange aerodynamic method of Pleim et al. (2001) which uses surface resistance, canopy resistance and stomatal resistance to compute dry deposition velocities. KCL-CMAQ was run at a 9 km grid resolution for the UK.

1.2.6 NAME

NAME is a 3 dimensional Lagrangian dispersion model. Pollutant emissions are represented by releasing millions of air parcels, each able to represent the released mass of many different species. The air parcels are carried by the three-dimensional wind field obtained from the Met Office's Unified Model (UM). Local turbulent motion is simulated using a random walk technique which requires a diffusion coefficient calculated from the local turbulent velocity variance and the local turbulent timescale. Above the boundary layer these two quantities are fixed, but within the boundary layer they are defined in terms of the local atmospheric stability and local surface quantities. A detailed description can be found in Morrison and Webster (2005). The UM provides direct output of boundary layer height for use in NAME. The model dry deposition scheme is based on a resistance analogy parameterisation to calculate a species dependent deposition velocity. It is applied to all air parcels within the boundary layer. The wet deposition method uses scavenging coefficients to model washout and rainout of pollutants.

1.2.7 OSRM

The development of the OSRM (Ozone Source Receptor Model) has been led by Ricardo-AEA working through an enduring consortium of leading UK experts under contract to Defra (and previous Departments) since 1999 (Hayman et al. 2010). Following the initial design of the model in a research and development stage, various features of the model were enhanced to improve model performance, to take account of further developments in the underlying science and to make the model more suitable for direct application to Defra air quality policy. Since around 2005, the emphasis has shifted from development to maintenance and application of the model as a policy tool for examining the response of the UK ground-level ozone climate to changes in precursor emissions in the UK and Europe.

OSRM is a trajectory model. The OSRM is used annually to model UK ground-level ozone concentrations at 10 km × 10 km resolution (at 3,000 specified receptors). The model can also be used to produce the annual metrics for 41 specific monitoring sites and in addition, hourly averages at each site. The UK emissions are based on the NAEI and the rest of Europe on the EMEP 50 km × 50 km inventory.

The UK Met Office provides meteorological datasets derived from the NAME model. 30 boundary layer meteorological parameters are provided at 6-hourly resolution over a year, covering a domain from 30˚W to 40˚E and 20˚ to 80˚N at 1˚spatial resolution. These data are used to derive 96-hour back trajectories to specified receptors. The OSRM now has met data in this form for each calendar year from 1999 to 2011.

The current version of the OSRM uses an updated version of the mechanism in STOCHEM: 70 chemical species involved in 195 thermal and photochemical reactions.

1.2.8 PTM

The PTM (Photochemical Trajectory Model) model is a moving air parcel trajectory model that is used to describe photochemical ozone and fine particle formation in north west Europe. The PTM model is used to quantify the contribution made by each VOC species and each VOC source category to the long-range transboundary formation and transport of ozone and PM across North West Europe. These contributions are defined in terms of Photochemical Ozone Creation Potentials POCPs and SOAPs.

The PTM uses SO_2, NO_x, NH_3, VOCs, CO and CH_4 emissions taken from 2010 version of NAEI for the UK and SO_2, NO_x, NH_3, VOCs, CO and CH_4 emissions for the rest of Europe were taken from the EMEP webdab (2010). Isoprene emissions were taken from EMEP. Terpene emissions were taken from Hope Stewart and Nick Hewitt for UK and GEIA for Europe.

4-day 3-D back-track trajectories from Met Office Unified model providing latitude, longitude, altitude, boundary layer depth, temperature were used to describe the meteorological processes. Between 30 and 1,000 equal probability trajectories arriving at each arrival point between 15:00 and 15:15z each day from Met Office NAME model were used in the present study. A Wesely dry deposition velocity scheme was used but no treatment was given for wet deposition. All model results were obtained using the CRIv2 chemical mechanism.

The model provides a probability distribution for its output instead of a single number. In the output provided, the uncertainties in all input parameters including meteorological processes, emissions, chemical mechanism, photolysis rates and deposition velocities were treated by means of Monte Carlo sampling and many thousands of model runs. Only those model runs that produced 'acceptable' ozone levels compared with observations were actually used in policy analyses. In the Defra model intercomparison exercise the PTM model provided 50%-percentile output for 15:00z for the rural Harwell, Oxfordshire site based on only the 'acceptable' model runs.

The PTM produced results for Harwell for July 2006.

1.2.9 WRF-Chem

The Weather Research and Forecasting/Chemistry model (WRF-Chem) is fully coupled 'online' model (Grell et al. 2005). The air quality component of the model is fully consistent with the meteorological component; both components use the same transport scheme (mass and scalar preserving), the same grid (horizontal and vertical components), and the same physics schemes for subgrid-scale transport.

1.2.10 Model ensemble

The number of models considered in this report provides the opportunity to assess the effect of using model *ensembles*. In model ensembles the simulation errors in different models are assumed to be independent and the mean of the ensemble can be expected to outperform individual models, thus providing an improved 'best estimate'. By sampling modelling uncertainties, ensembles of air quality models should provide an improved basis for probabilistic prediction. Note however, members of a multi-model ensemble share common systematic errors.

In this context a model ensemble refers to pooling all the model results in some way to represent the average (or median) of all models. Ensemble modelling has been shown many times to produce better results than any single model (Solazzo et al. 2012). In this study a new model ensemble was produced by taking the mean of the hourly predictions of O_3 and other species by site for six of models i.e. not including the PTM or WRF-Chem where full year data were not available. Although the mean was used in this study, it was found that taking the median produced largely similar results.

Table 1.2: Summary of model characteristics.

model	type	boundary conditions	meteorology	chemistry (km)	grid[a]	emissions[b]
AEA-CMAQ	Eulerian	STOCHEM	WRF3	CB05	12	NAEI/EMEP
AQUM	Eulerian	GEMS/MACC	UM	UKCA/RAQ	12	NAEI/EMEP ENTEC
EMEP-4-UK	Eulerian	EMEP MSC-W model	WRF3.1.1	EMEP MSC-W model	5	NAEI/EMEP
Hertfordshire-CMAQ	Eulerian	Geos-Chem	WRF3.2.1	CB05	18	TNO
KCL-CMAQ	Eulerian	STOCHEM	WRF3.1	CB05	9	NAEI/EMEP
NAME	Lagrangian	Mace Head[c]	UM	STOCHEM	≈10?	NAEI/EMEP ENTEC
OSRM	Lagrangian	STOCHEM with Mace Head	NAME (UM)	STOCHEM	10	NAEI/EMEP ENTEC
PTM	Lagrangian	Mace Head?	UM	MCM3.1	≈10?	NAEI/EMEP GEIA
WRF-Chem	Eulerian	?	?	?	?	?

[a] For UK grid.

[b] ENTEC emissions are used for shipping.

[c] Based on analysis of extracted 'baseline' air-masses.

1.3 NO_x and VOC emission assumptions

An important issue in the intercomparison are the assumptions related to NO_x and VOC emissions. The intercomparison exercise itself was based on model runs that had already been produced by the modelling teams and there was no opportunity to ensure the use of the same emission inventories. However, this is not necessarily a disadvantage because each group essentially makes its own judgement as to what emissions are most appropriate and carry out their own evaluations accordingly.

Nevertheless, knowing something about the emission assumptions can help with the interpretation. Table 1.3 summarises the emissions of NO_x and VOCs by model, split by the UK and Europe. Note that some caution is needed when considering the European emissions because each model uses different geographic definitions. For this reason, it is also useful to consider the *ratio* of NO_x to VOCs because a ratio will be less susceptible to the precise definition of the geographic area.

TABLE 1.3: Summary of emissions of NO_x and VOCs for UK and Europe assumed by the models.

model	location	NO_x (kt yr^{-1})	VOC			NO_x/VOC ratio
			Anthropogenic (kt yr^{-1})	Biogenic (kt yr^{-1})	Total (kt yr^{-1})	
KCL-CMAQ	UK	1590	1029	210	1239	1.28
AEA-CMAQ	UK	1632	997	NA	NA	NA
EMEP-4-UK	UK	1447	1019	30	1049	1.38
NAME	UK	1633	1013	0	1013	1.61
AQUM	UK	1459	926	93	1019	1.43
Hertfordshire-CMAQ	UK	1600	927	313	1240	1.29
OSRM [a]	UK	1582	939	NA	NA	NA
WRF-Chem [a]	UK	1119	1036	NA	NA	NA
KCL-CMAQ	Europe[b]	14352	11982	10864	22846	0.63
AEA-CMAQ	Europe	16202	9909	18435	28344	0.57
EMEP-4-UK [a]	Europe	21803	15341	NA	NA	NA
NAME	Europe	18964	11143	0	11143	1.70
AQUM	Europe	8439	4890	1122	6012	1.40
Hertfordshire-CMAQ	Europe	16566	14802	27142	41944	0.39
OSRM [a]	Europe	21019	11399	NA	NA	NA
WRF-Chem [a]	Europe	7134	4420	NA	NA	NA

[a] Not including biogenic emissions.

[b] Note the area defined as Europe in the models varies. Also, Europe is assumed to include the UK.

Considering Table 1.3 for UK emissions of NO_x, there is reasonable consistency between the models. However, the WRF-Chem model assumes substantially less NO_x compared with the other models. A similar picture emerges for UK anthropogenic VOC emissions where all models assume about 1000 kt yr^{-1}. Where the models differ more is in the assumption of biogenic emissions. These emissions are not necessarily easy to extract from models because they can depend on the prevailing ambient conditions — in particular surface temperature. Where specific estimates are available, they vary by a large amount. For example for the UK, NAME assumes zero contribution from biogenic VOCs and AQUM assumes only 93 kt yr^{-1}, whereas Hertfordshire-CMAQ assumes 313 kt yr^{-1}. The differences are also important for European emissions, where the ratio of NO_x/VOC can be considered. Hertfordshire-CMAQ tends to have a much lower ratio (0.39) compared with the other models for which emissions are available e.g. KCL-CMAQ (0.63), NAME (1.70) and AQUM (1.40). These differences in assumptions are likely to affect how the models respond to emissions control and whether they consider conditions to be NO_x or VOC-sensitive.

1.4 Analysis approaches

The analysis has been conducted using R software (R Core Team 2013) and an R package called *openair* (Carslaw and Ropkins 2012). The document itself embeds R code that generates all the tables, plots and statistics 'on the fly' using the approach of Xie (Xie 2013a; Xie 2013b). This is a *dynamic reporting* approach where the document must be compiled in order to be produced. All the code used to generate the plots, tables etc. is given in Appendix C, which itself is generated directly from the report automatically. There are several reasons for adopting this approach:

- All the analysis is entirely reproducible: every plot, table etc. can be reproduced by anybody with access to the input data. This approach ensures that the process

can be made as open and as transparent as possible as every aspect of the analysis can be scrutinised by others.

- All the tools used to produce this report are free and open-source allowing anyone to access them.
- The approach makes it much more efficient to take account of revised data e.g. if a group finds a problem with the model results, new data are re-imported and the document re-compiled and all the analysis is automatically regenerated.
- The model evaluation necessarily can only cover certain aspects of interest. By making all the code available the modellers can develop the analysis further and indeed use it for other situations.

Much of the analysis in this report uses various model evaluation statistics to compare the models with observations. No single statistical measure can encapsulate and characterise differences between models and observations and for this reason several statistics have been used based mostly on Derwent et al. (2010). These statistics broadly consider model bias, error and correlation when compared with measurements. The summary statistics are defined in Appendix B.

To help with interpretation of these statistics summary results have been ranked in a simple way in summary tables throughout this report to show the best performing model first and so on. The ranking is based on the overall performance for the *Coefficient of Efficiency*, *COE* (Legates and McCabe 2012; Legates and McCabe Jr 1999). The *COE* is a simple, good overall indicator of model performance and often other statistics follow the order e.g. the correlation coefficient, r or the Root Mean Square Error, *RMSE*. It is also easy to interpret — see Appendix B. In particular, a value of 1 indicates a perfect model and a value of zero shows that the predictions *are no better than taking the mean of the observations*.

2 Comparison against different O_3 metrics

2.1 Boundary conditions — analysis of Mace Head data

For predictions of O_3 concentrations, the boundary conditions used in each model will likely be an important component affecting the overall performance of the models. This section considers predictions at the Mace Head site, which will tend to be more strongly influenced by model boundary conditions than at other sites in the mainland UK. It will also be important to consider the specific assumptions used in each model to generate the boundary conditions. The models each adopt different methods for setting the boundary conditions for O_3, which are summarised in Table 1.2. Given the large influence that 'background' O_3 concentrations have on O_3 predictions in general it is useful to compare how the models differ in this respect.

Simply comparing the annual means shows that there are comparatively large differences in the predictions of O_3 at Mace Head. The annual means are considered more in Section 2.2.1, but are considered briefly in this section for Mace Head. The annual means vary from 32.7 ppb (NAME) to 42.0 ppb (KCL-CMAQ); compared with a measured value of 37 ppb. Other model predictions were EMEP-4-UK 36.7, Hertfordshire-CMAQ 38.2, OSRM 35.5 and AQUM 35.1 ppb.

A better understanding of these differences can be gained if similar air masses are clustered and then analysed. Of particular interest are 'baseline' air masses that will dominate the western fringe of the UK. Two methods have been used to identify these air masses: a cluster analysis of back trajectories and a more refined approach based on careful filtering of air masses arriving at Mace Head as described by Ebinghaus et al. (2011).[3]

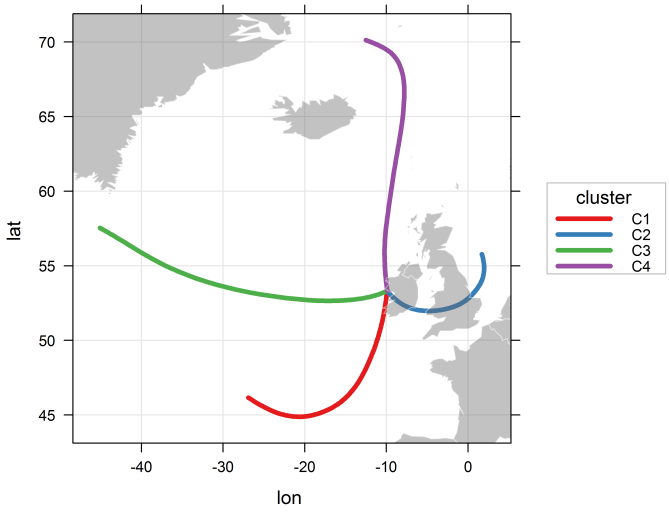

FIGURE 2.1: Mean back trajectories for the 4-cluster solution at Mace Head.

Clustering air masses can help group air mass trajectories that have similar geographic origins and therefore more likely to have similar chemical histories. k-means clustering has been undertaken at the Mace Head site to illustrate these issues. The 96-hour back trajectories were calculated using the HYSPLIT model and were run at 3-hour intervals. Figure 2.1 shows the mean trajectory path for the four clusters.[4] These clusters can broadly be described as Atlantic (cluster 3), Polar (cluster 4), south Atlantic (cluster 1) and European continental (cluster 2).

The mean concentrations of O_3 by model and cluster are summarised in Table 2.1, where the measured values have been subtracted by cluster. Overall, out of the main models evaluated, the EMEP-4-UK provides the most representative predictions. It is now also clear the extent to which the models disagree e.g. for cluster 4 KCL-CMAQ overestimates O_3 concentrations by 6.3 ppb whereas the NAME model underestimates by 5.1 ppb. These differences between the models will likely contribute to differences in their performance for a range of O_3 metrics. However, it is also clear from Table 2.1 that the ensemble model provides the best overall agreement with the measurements.

Whether these different clusters represent different air-masses with different chemical histories can be determined by considering some of the temporal components. For example, it would be expected the seasonal variation in O_3 concentration would be different for baseline air compared with air-masses from a European Origin. Figure 2.2 shows the monthly variation in O_3 concentration both for measured values and for each model. The more refined air-mass allocation method has also been considered to

[3]The air-mass allocations were kindly provided by Professor Dick Derwent for this study.

[4]There are various methods for determining how many clusters should be used, but four was found to nicely differentiate air masses with different O_3 characteristics.

TABLE 2.1: Summary mean O_3 concentrations by model and cluster at Mace Head with measured values subracted (ppb).

model	C1	C2	C3	C4
MEASURED	0.0	0.0	0.0	0.0
AEA-CMAQ	6.3	2.0	2.0	2.6
AQUM	-1.8	-1.4	-3.6	0.9
EMEP4UK	-0.5	1.9	-1.5	-1.3
EMEP4UK-CRI	-0.5	1.2	-1.6	-1.4
ENSEMBLE	1.6	-0.0	-1.5	0.2
Hert-CMAQ	3.3	3.3	-1.2	1.6
KCL-CMAQ	5.6	6.3	3.4	6.3
NAME	-3.7	-1.2	-6.1	-5.1
OSRM	1.8	-11.3	-3.3	-3.9

understand whether different conclusions would be reached compared with the analysis based on back trajectory clustering. The results by month are shown in the right panel of Figure 2.2. Overall, for baseline air the results are very similar.

Considering the monthly variations, the clearest difference between the clusters for measured concentrations is for cluster 2 for European continental air masses. These trajectories tend to show a peak in O_3 concentration during the summer months i.e. coinciding with regional O_3 production. It can be shown that clusters 3 and 4 (Atlantic/Polar origin) show a clear peak in springtime. The most dominant cluster in is cluster 3 (Atlantic), which accounts for around double the number of occurances compared with either cluster 1 or 2. The monthly variation in O_3 concentrations for cluster 3 is captured well by EMEP-4-UK, KCL-CMAQ, Hertfordshire-CMAQ and NAME. The AEA-CMAQ model tends to predict higher concentrations during the summertime and the NAME model tends to underestimate O_3 concentrations in general. The OSRM model on the other hand peaks much earlier in the year (February) compared with the measurements.

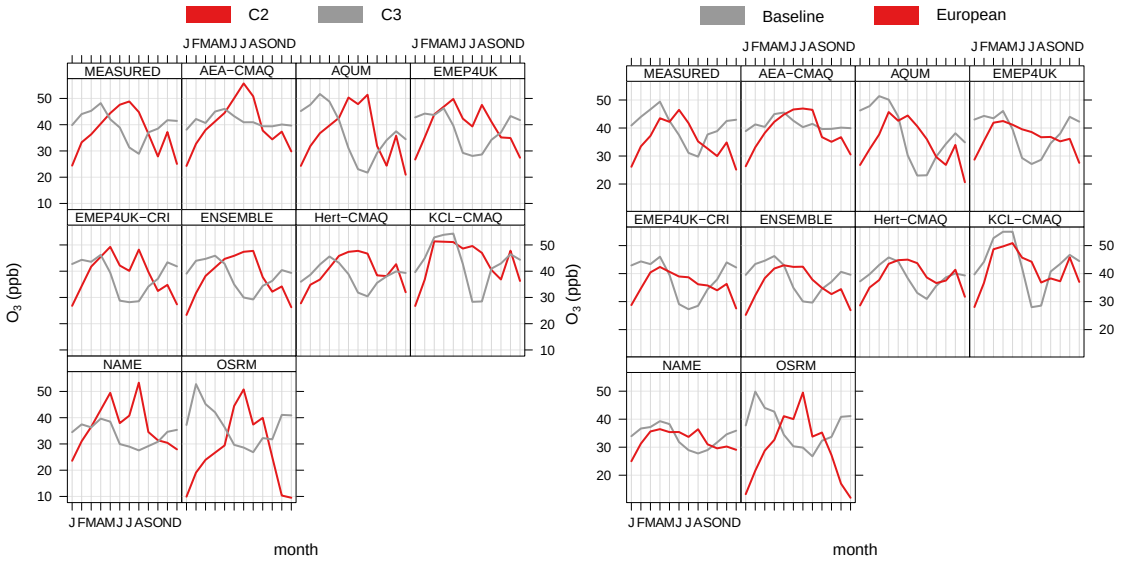

FIGURE 2.2: Comparsion of O_3 concentration seasonal components at Mace Head by trajectory clustering and by allocation for baseline air masses.

The diurnal variations (Figure 2.3) show that most models tend to capture the absolute magnitude of O_3 concentrations reasonably well as well as the lack of diurnal variation in concentration for this marine site. Overall, the NAME model tends to underestimate O_3 concentrations and the OSRM model significantly underestimates O_3 concentrations for European origin air-masses.

Again, for both the seasonal and diurnal variations in concentration for baseline and European air masses, the ensemble model has the best overall agreement with measurements.

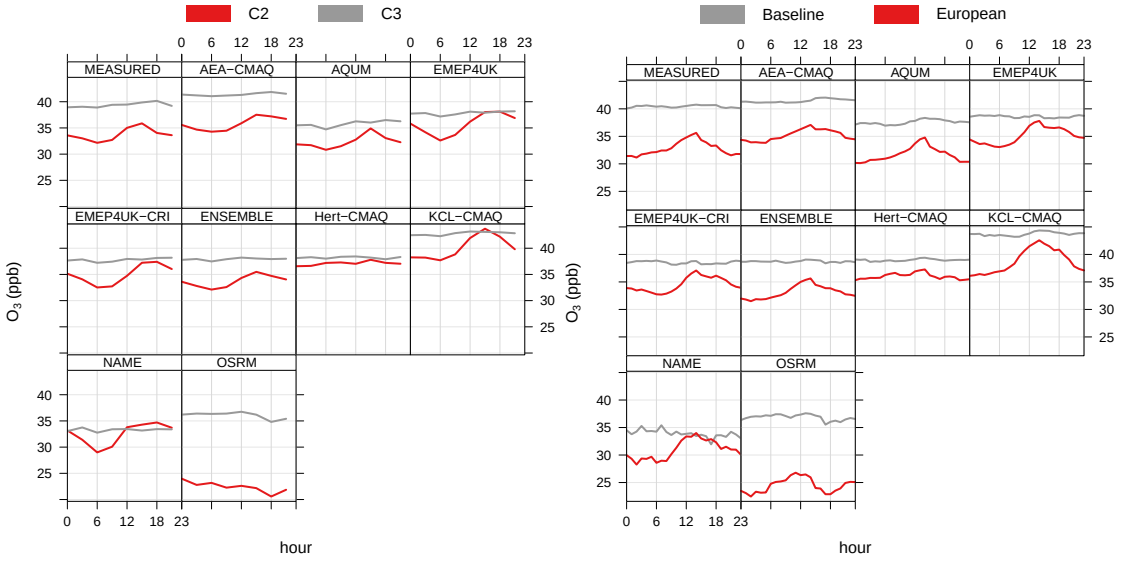

FIGURE 2.3: Comparsion of O_3 concentration diurnal components at Mace Head by trajectory clustering and air mass allocation.

The model performance for specific trajectory clusters can also be calculated. For example, for cluster 3 (Atlantic air masses), the model performance results shown below show that the ensemble, EMEP-4-UK and the University of Herfordshire CMAQ models perform best. Some potentially important observations can be drawn from these results, which could affect O_3 predictions more generally across other sites. The KCL-CMAQ and AEA-CMAQ models tend to show a positive bias compared with other models. The NAME model shows a relatively strong negative bias. With respect to the correlation with measurements, EMEP-4-UK, KCL-CMAQ, Hertfordshire-CMAQ have high correlations whereas AEA-CMAQ, OSRM and NAME have much lower correlations.

There is also consistency in the model agreement between the trajectory clustering approach and the air-mass allocation approach, as seen by comparing Table 2.2 and Table 2.3.

Figure 2.4 shows the conditional quantile plot for cluster 3 results (Atlantic air masses), which provides a more comprehensive analysis of model performance for Atlantic air mass conditions. Figure 2.4 shows more clearly that the NAME model and the OSRM model performance tends to be worse than the other models; the former being due to the under prediction of O_3. It is also clear that the AEA-CMAQ results cover a narrower range than is seen in the observations. Again, the analysis reveals the good performance of the ensemble model, shown by the narrow quantile intervals and the close correspondence of the median (red) line with the perfect model (blue) line.

TABLE 2.2: Summary statistics for Cluster 3 model performance.

data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
ENSEMBLE	1256	1.00	-1.46	3.53	-0.04	0.09	4.45	0.83	0.41
EMEP4UK	1248	1.00	-1.52	4.42	-0.04	0.11	5.71	0.74	0.27
Hert-CMAQ	1256	1.00	-1.16	4.49	-0.03	0.11	5.66	0.69	0.25
EMEP4UK-CRI	1250	1.00	-1.57	4.54	-0.04	0.12	5.88	0.73	0.25
KCL-CMAQ	1256	1.00	3.43	5.24	0.09	0.13	7.08	0.76	0.13
AEA-CMAQ	1252	0.99	2.03	5.74	0.05	0.15	7.29	0.35	0.05
AQUM	1256	1.00	-3.58	6.12	-0.09	0.16	7.09	0.82	-0.02
OSRM	1256	0.98	-3.29	7.15	-0.08	0.18	9.30	0.48	-0.19
NAME	1256	0.97	-6.07	8.42	-0.15	0.21	10.34	0.44	-0.40

TABLE 2.3: Summary statistics for 'baseline air' model performance using the more refined air-mass allocation method.

data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
ENSEMBLE	3646	1.00	-1.72	3.11	-0.04	0.08	3.85	0.87	0.43
EMEP4UK	3613	1.00	-1.81	4.02	-0.04	0.10	5.12	0.78	0.27
EMEP4UK-CRI	3617	1.00	-1.88	4.09	-0.05	0.10	5.21	0.78	0.26
Hert-CMAQ	3645	1.00	-1.44	4.11	-0.04	0.10	5.19	0.71	0.25
KCL-CMAQ	3631	1.00	3.28	5.06	0.08	0.13	6.89	0.78	0.08
AEA-CMAQ	3624	1.00	1.04	5.33	0.03	0.13	6.77	0.29	0.03
AQUM	3646	1.00	-2.83	5.82	-0.07	0.14	6.76	0.83	-0.06
OSRM	3646	0.99	-3.58	6.55	-0.09	0.16	8.49	0.56	-0.19
NAME	3646	0.98	-6.55	8.55	-0.16	0.21	10.42	0.44	-0.56

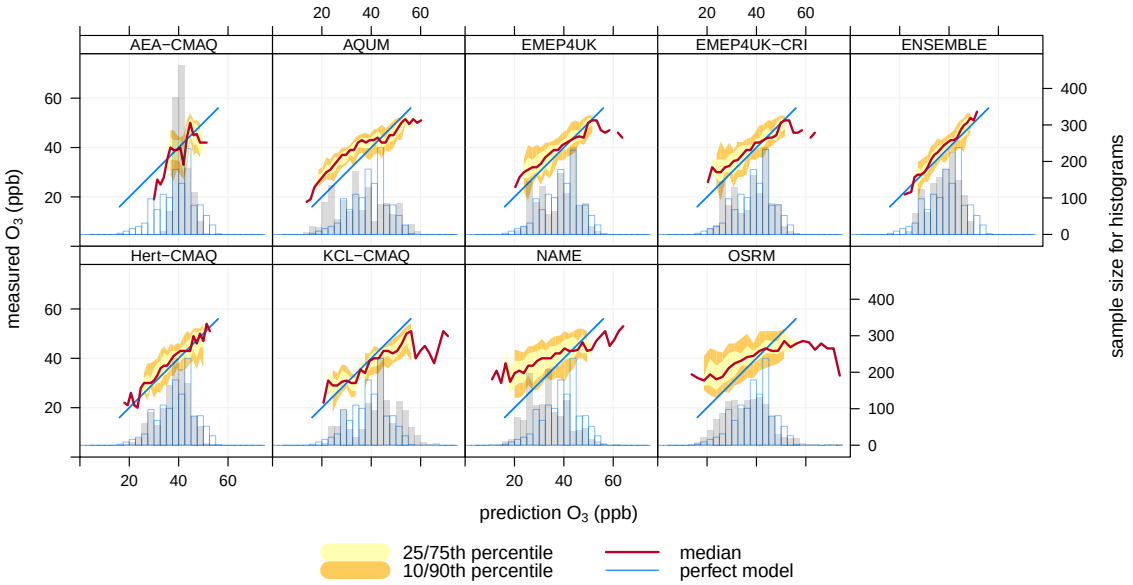

FIGURE 2.4: Conditional quantile plot of O_3 performance at the Mace Head site for cluster 3 back trajectories (Atlantic air). The blue line shows the results for a perfect model. The red line shows the median value of the predictions. The shading shows the predicted quantile intervals i.e. the 25/75th and the 10/90th. A perfect model would lie on the blue line and have a very narrow spread. There is still some spread because even for a perfect model a specific quantile interval will contain a range of values. However, for the number of bins used in this plot the spread will be very narrow. Finally, the shaded histogram shows the counts of predicted values and the blue histogram the counts of the observed values.

Throughout all these comparisons, the ensemble model tended to perform consistently better than the individual models, suggesting that pooling the model results gives a notable improvement in model performance.

2.2 Model performance against commonly used metrics

In this section we compare the model predictions against some commonly used O_3 metrics. The performance against these metrics is important because many relate to UK or European limits which have a direct policy relevance. It is also important to determine the spread in predictions through using different models, as this information provides a good indication of the range in predictions that might reasonably be expected through using different models. This information can also be used to determine whether the models are sufficiently accurate to reliably predict the different O_3 metrics. There are numerous O_3 metrics that have been used in previous studies. For example in the AQEG (2009) report *eight* metrics are considered in Table 1-1. Many of the metrics are highly correlated with one another as can be seen in Figure 2.5. The greatest contrast between metrics is for the annual mean compared with the other metrics.

A summary of the key statistics by site for measured values is shown in Table 2.4.

Table 2.4: Summary measured O_3 statistics by site for 2006. Concentration values are given in $\mu g\ m^{-3}$.

site	data.capture	mean	minimum	maximum	median	max.daily	max.rolling.8	max.rolling.24	percentile.95	percentile.99	roll.8.O3.gt.100	roll.8.O3.gt.120	AOT40
AH	92	72	2	220	72	176	212	182	104	154	43	21	12647
BIR1	99	41	0	186	40	132	176	138	86	132	24	15	6438
BOT	99	52	0	196	52	126	179	133	94	146	31	18	9433
BUSH	98	58	0	158	60	112	143	113	88	102	9	3	3261
CLL2	96	29	0	178	24	105	149	112	70	110	14	5	2797
ESK	99	58	0	176	60	119	162	122	92	117	24	9	6266
GLAZ	74	48	0	188	50	126	172	126	92	146	23	14	6946
HAR	94	54	0	212	56	124	184	127	94	134	31	14	7790
HM	90	58	2	174	56	141	160	142	104	139	34	17	9067
KC1	95	40	0	204	38	117	178	143	92	144	34	17	8338
LB	95	50	0	168	48	120	156	120	90	120	22	11	5271
LH	89	61	0	238	60	169	212	173	110	154	45	28	12790
LN	99	47	0	170	48	114	158	129	80	94	6	5	2150
LON6	97	43	0	194	42	111	182	123	90	136	32	21	7778
MAN3	91	29	0	144	24	109	132	110	72	108	13	2	2254
MAN4	98	33	0	164	30	101	132	108	70	108	14	5	2740
MH	99	74	20	168	76	130	162	145	100	110	37	5	7854
ROCH	99	49	0	200	50	101	150	106	92	124	30	12	6983
SIB	92	57	0	184	56	116	172	121	100	138	38	21	10500
SV	84	72	6	182	72	147	161	148	102	132	47	10	9828
TED	99	51	0	206	52	129	193	140	100	154	42	25	12455
WFEN	86	66	4	278	64	175	260	186	120	174	99	36	23023
YW	96	63	0	234	62	169	213	179	104	136	48	11	10559

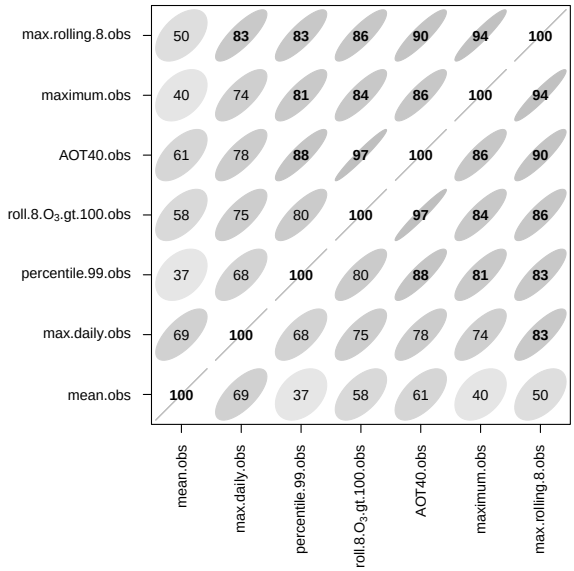

FIGURE 2.5: Correlation matrix showing how different O_3 metrics are correlated. The statistics are based on the all the sites considered in the current study.

It can be shown the analysis of *hourly data over a full year* by site (23 in total), that the ensemble model produces the best results for 15 of the sites and the second best for 4. However, the results for specific metrics differ as outlined below.

2.2.1 Annual mean O_3 predictions

A useful way of summarising the performance of the models is to plot the bias (*NMB*) against the error (*NMGE*), which is shown in Figure 2.6. Sometimes plots such as these are used to show various model evaluation criteria e.g. *NMB* within ± 0.5 and *NMGE*<0.5 and are referred to as 'soccer plots' — so called because the lines that bound the criteria resemble a goal mouth. However, in this study no specific criteria have been used against which the models are judged.

In rural areas the model predictions tend to have a slight positive bias, with the NAME and AQUM models performing best with low bias and error. While it might not be expected that large grid models would predict urban O_3 concentrations well, the KCL-CMAQ model predictions are almost as good as the rural predictions. As shown in Figure 2.6 there is a strong tendency for the models to show a positive bias in urban O_3 predictions. This behaviour is consistent with emissions of NO being mixed into volumes that are too large, diluting them more than they should be and hence resulting in a positive bias in O_3 concentration due to less depletion through the NO + O_3 reaction. Indeed, the positive bias in O_3 concentration predictions in urban areas is reflected by a negative bias in NO_x predictions, shown later in Figure 3.1. Comparing Figure 2.6 with Figure 3.1 reveals for example that the model with the most positive bias for O_3 (Hertfordshire-CMAQ) is also the model with most negative bias for urban NO_x predictions – and the largest grid square of 18 km. The situation for rural areas is more complex and there is no obvious relationship between the O_3 and NO_x predictions. Nevertheless, urban O_3 concentrations are important with respect to exposure and concentrations can be expected to increase in urban areas as NO_x emissions decrease.

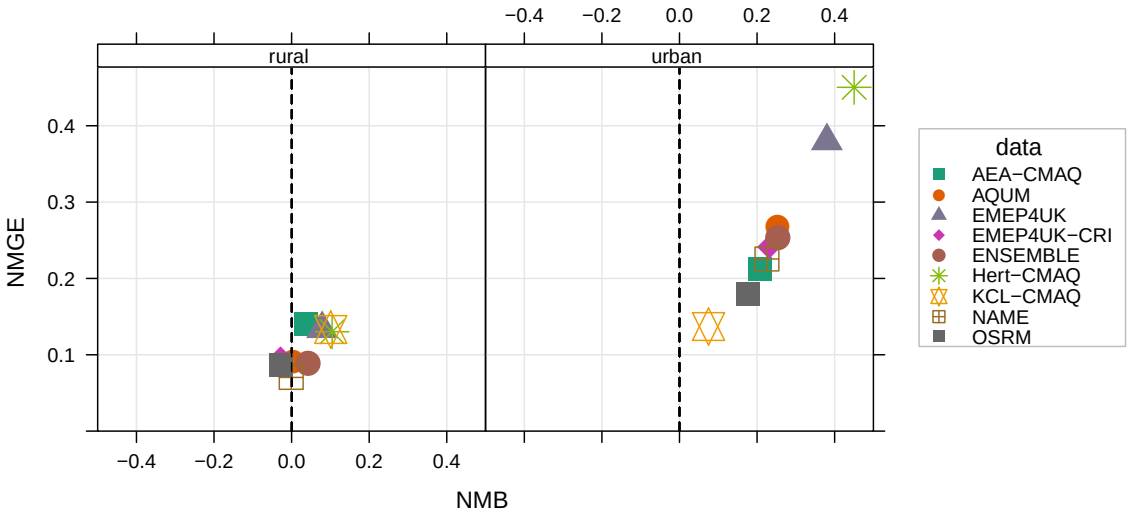

FIGURE 2.6: Comparison of annual mean predicted and measured O_3 concentrations showing the *NMB* against *NMGE* split by rural and urban sites.

The good performance of KCL-CMAQ in urban areas is worthy of note. As discussed above, the performance for annual mean O_3 and annual mean NO_x is better than most other models. There are many potential reasons for the differences between the models in urban areas. However, the KCL-CMAQ model uses a similar grid resolution to other models for the UK (9 km × 9 km) and vertical layers. Note, however, Hertfordshire-CMAQ uses a grid size of 18 km (see Table 1.2), which may contribute to the higher positive bias seen for this model. One potentially important difference with the KCL-CMAQ model compared with other models is the use of updated NO_x emission factors for road vehicles based on vehicle emission remote sensing (Carslaw et al. 2011; Beevers et al. 2012). These updated emissions will likely result in higher NO_x emissions in urban areas compared with currently used emission factors used by other models.

Almost all models predict the annual mean O_3 concentration within a factor of two at all sites, as shown in Table 2.5. The performance by site type is shown in Table 2.5. These results show that the models have very different performances in urban and rural locations. It is now clear for example that there is a considerable positive bias for most models (except KCL-CMAQ) in urban areas and the AEA-CMAQ model performs relatively better in urban areas than rural areas.

TABLE 2.5: Summary statistics for annual mean O_3 performance split by site type.

site.type	data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
rural	NAME	15	1.00	-0.06	4.16	-0.00	0.07	5.60	0.78	0.42
rural	OSRM	16	1.00	-1.67	5.10	-0.03	0.09	7.40	0.54	0.28
rural	ENSEMBLE	16	1.00	2.49	5.22	0.04	0.09	7.14	0.65	0.26
rural	AQUM	16	1.00	0.18	5.37	0.00	0.09	7.05	0.64	0.24
rural	EMEP4UK-CRI	16	1.00	-1.68	5.60	-0.03	0.10	8.22	0.46	0.21
rural	Hert-CMAQ	16	1.00	6.14	7.64	0.10	0.13	9.21	0.60	-0.08
rural	KCL-CMAQ	16	1.00	5.93	7.82	0.10	0.13	9.99	0.64	-0.11
rural	EMEP4UK	16	1.00	4.62	7.87	0.08	0.13	9.35	0.38	-0.12
rural	AEA-CMAQ	16	0.94	2.16	8.23	0.04	0.14	11.65	0.61	-0.17
urban	KCL-CMAQ	7	1.00	2.85	5.20	0.07	0.14	6.12	0.69	0.21
urban	OSRM	7	1.00	6.72	6.80	0.18	0.18	9.00	0.60	-0.03
urban	AEA-CMAQ	7	1.00	7.83	8.06	0.21	0.21	10.66	0.40	-0.23
urban	NAME	2	1.00	7.85	7.85	0.23	0.23	7.86	1.00	-0.35
urban	EMEP4UK-CRI	7	1.00	8.75	9.14	0.23	0.24	10.23	0.81	-0.39
urban	ENSEMBLE	7	1.00	9.62	9.62	0.25	0.25	11.44	0.57	-0.46
urban	AQUM	7	1.00	9.58	10.18	0.25	0.27	11.92	0.33	-0.55
urban	EMEP4UK	7	1.00	14.41	14.41	0.38	0.38	15.71	0.56	-1.20
urban	Hert-CMAQ	7	0.86	17.10	17.10	0.45	0.45	18.44	0.42	-1.60

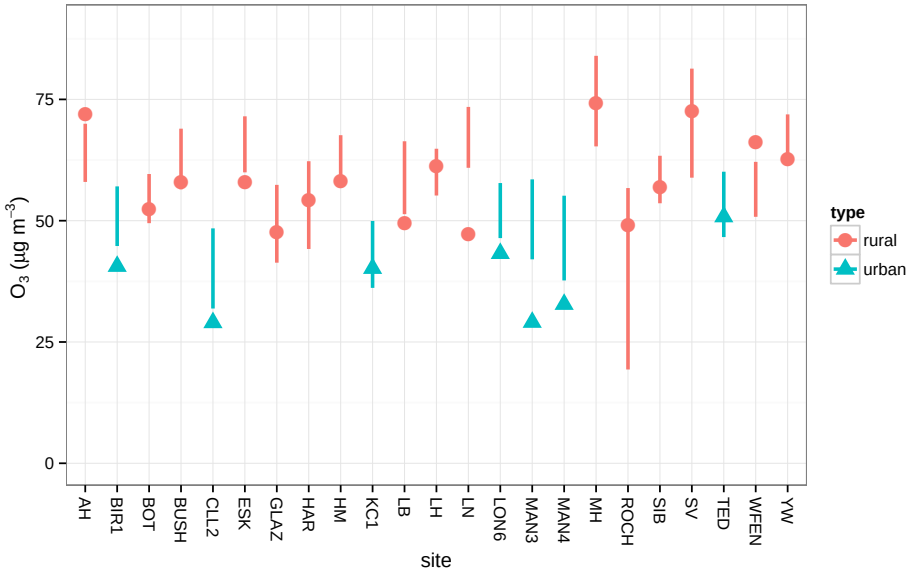

FIGURE 2.7: Comparison of annual mean predicted and measured O_3 concentrations. The solid shape shows the measured value and the lines the range in model predictions. The results are also shown by site type (urban/rural).

The annual mean results can be compared with the results by cluster that were summarised in Section 2.1 to consider whether there is any relationship between the two i.e. whether predictions at a site representative of boundary conditions more generally affects annual mean concentrations. In Table 2.1 it was shown for example the NAME model tended to underestimate O_3 concentrations whereas the KCL-CMAQ model tended to overestimate. Considering Table 2.5 it can be seen that the NAME model does tend to have a much lower mean bias than most models; but certainly not the lowest. However, a model such as Hertfordshire-CMAQ that predicted concentrations of O_3 at Mace Head well does tend to have a relative large positive bias for the annual mean O_3 concentration at other sites. It is not clear therefore that the variation in predictions for boundary-type conditions at Mace Head carries through to other sites or whether other factors become more important e.g. deposition and treatment of atmospheric chemistry. Indeed, the inclusion of the CRI version of the EMEP-4-UK model (which has very

similar performance for Mace Head boundary conditions), differ markedly in their mean bias at rural sites, highlighting the importance of the chemistry scheme used.

2.2.2 AOT40 predictions

The model performance statistics for the AOT40 (the accumulated dose of O_3 over a threshold of 40 ppb, 80 µg m^{-3}) are shown in Table 2.6. The AOT40 was calculated by extracting daylight hours in the growing season, defined as April to September. The model performance for AOT40 predictions is rather mixed and the models predictions cover a wide range of values. Nevertheless, all models predict most sites within a factor of two.

Table 2.6: Summary statistics for AOT40 O_3 performance.

site.type	data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
rural	AQUM	16	0.94	-66.81	2616.00	-0.01	0.29	3457.61	0.68	0.19
rural	EMEP4UK	16	0.81	-1060.09	3095.26	-0.12	0.34	4246.95	0.46	0.04
rural	NAME	15	0.73	-2818.70	3325.31	-0.31	0.37	4862.31	0.57	0.03
rural	Hert-CMAQ	16	0.75	-3316.59	3856.55	-0.37	0.43	5631.79	0.19	-0.19
rural	OSRM	16	0.62	-3271.46	3921.77	-0.36	0.43	5546.70	0.29	-0.21
rural	EMEP4UK-CRI	16	0.81	-3550.85	3983.20	-0.39	0.44	5413.61	0.58	-0.23
rural	KCL-CMAQ	16	0.75	1038.96	4164.59	0.11	0.46	5424.30	-0.00	-0.29
rural	AEA-CMAQ	16	0.75	-2465.82	4294.17	-0.27	0.47	5606.53	0.09	-0.33
rural	ENSEMBLE	16	0.50	-4528.13	4815.99	-0.50	0.53	6177.88	0.50	-0.49
urban	EMEP4UK-CRI	7	1.00	-597.46	1923.37	-0.10	0.31	2415.97	0.85	0.36
urban	KCL-CMAQ	7	0.86	425.88	2159.27	0.07	0.35	2510.30	0.78	0.28
urban	Hert-CMAQ	7	0.71	295.16	2234.39	0.05	0.37	2658.47	0.76	0.26
urban	AQUM	7	0.57	177.25	2512.71	0.03	0.41	2847.06	0.90	0.17
urban	NAME	2	1.00	-627.59	1567.04	-0.14	0.34	1688.04	1.00	0.14
urban	EMEP4UK	7	0.57	2177.17	2627.69	0.36	0.43	2995.11	0.86	0.13
urban	ENSEMBLE	7	0.71	-2695.45	2873.26	-0.44	0.47	3828.40	0.91	0.05
urban	AEA-CMAQ	7	0.71	-1425.44	3025.99	-0.23	0.49	3553.70	0.40	-0.00
urban	OSRM	7	0.43	-3118.20	3232.93	-0.51	0.53	4127.00	0.77	-0.07

The results shown in Figure 2.8 show that the AQUM model tends to show very little bias overall and has the lowest *RMSE*. There is a tendency for other models to show a negative bias, with the exception of KCL-CMAQ.

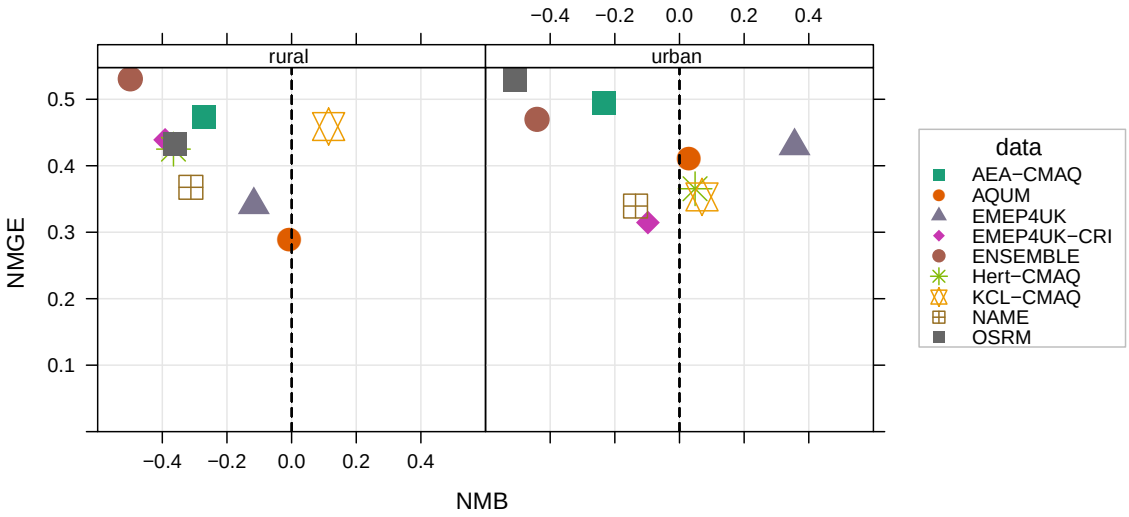

Figure 2.8: Comparison of AOT40 predicted and measured values showing the *NMB* against *NMGE* split by rural and urban sites.

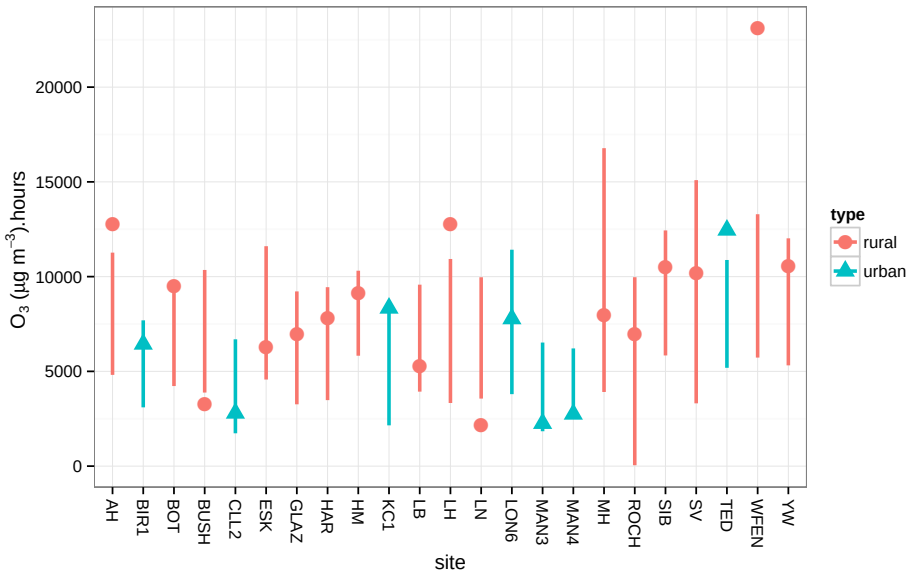

FIGURE 2.9: Comparison of AOT40 predicted and measured O_3 concentrations. The solid circle shows the measured value and the lines the range in model predictions.

2.2.3 Higher concentration metrics

In this section consideration is given to model performance for metrics that consider the higher concentrations of O_3. These metrics tend to be more strongly influenced by summertime regional photochemical episodes.

The first metric considered is the number of days with a daily maximum of running 8-hour means >100 µg m^{-3} (50 ppb). It can be shown from the observed O_3 concentrations across *all* sites that most of these days occur in July (292 site-days) and June (191 site-days). However, there are an appreciable number of days in April and May also (100 and 110 respectively, which will be affected by high baseline O_3 concentrations). The model performance summary is shown in Table 2.7.

TABLE 2.7: Summary statistics for number of days with a daily maximum of running 8-hour means >100 µg m^{-3}.

site.type	data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
rural	NAME	15	0.60	-8.33	15.13	-0.23	0.42	22.66	0.33	-0.06
rural	EMEP4UK	16	0.75	0.25	14.62	0.01	0.41	18.71	0.40	-0.07
rural	EMEP4UK-CRI	16	0.81	-12.31	14.69	-0.35	0.41	21.23	0.57	-0.08
rural	OSRM	16	0.75	3.06	16.19	0.09	0.46	21.73	0.22	-0.19
rural	AQUM	16	0.81	11.75	17.50	0.33	0.49	22.43	0.43	-0.29
rural	AEA-CMAQ	16	0.69	-15.62	18.25	-0.44	0.51	25.35	0.23	-0.34
rural	Hert-CMAQ	16	0.69	-18.44	19.19	-0.52	0.54	26.61	0.36	-0.41
rural	ENSEMBLE	16	0.38	-21.12	21.38	-0.60	0.60	27.34	0.65	-0.57
rural	KCL-CMAQ	16	0.56	24.00	29.75	0.68	0.84	36.67	0.04	-1.18
urban	EMEP4UK-CRI	7	1.00	-0.71	5.29	-0.03	0.21	6.15	0.83	0.45
urban	OSRM	7	0.86	-5.14	6.57	-0.21	0.27	8.65	0.78	0.32
urban	AQUM	7	1.00	3.71	6.86	0.15	0.28	8.02	0.94	0.29
urban	Hert-CMAQ	7	1.00	-0.86	6.86	-0.03	0.28	8.16	0.69	0.29
urban	ENSEMBLE	7	1.00	-9.14	9.14	-0.37	0.37	11.43	0.97	0.05
urban	KCL-CMAQ	7	0.86	7.71	9.43	0.31	0.38	10.93	0.72	0.03
urban	EMEP4UK	7	0.57	10.29	10.29	0.42	0.42	12.05	0.84	-0.06
urban	AEA-CMAQ	7	0.57	-8.43	11.29	-0.34	0.46	13.57	0.16	-0.17
urban	NAME	2	1.00	-6.00	6.00	-0.32	0.32	7.21	1.00	-0.20

The model evaluation statistics show that the models do less well in predicting the number of days where the maximum rolling 8-hour mean is >100 µg m^{-3} compared

with the annual mean. There is no consistent positive or negative bias in the models results: some models tend to over predict while other under predict, as shown in Table 2.7 and Figure 2.10. Predicting this particular metric is difficult for two reasons. First, all models are expected to find it more difficult to predict a short-term mean concentration compared with a longer-term mean, and second, the statistic is a threshold statistic. The latter point is important because if concentrations are close to the threshold, then only a small error in the prediction can have a large effect on the estimate of the number of days where the concentration is greater than the threshold. For this reason, the number of days where the maximum rolling 8-hour mean is >100 µg m^{-3} is a challenging statistic for the models to predict — but nevertheless and important one.

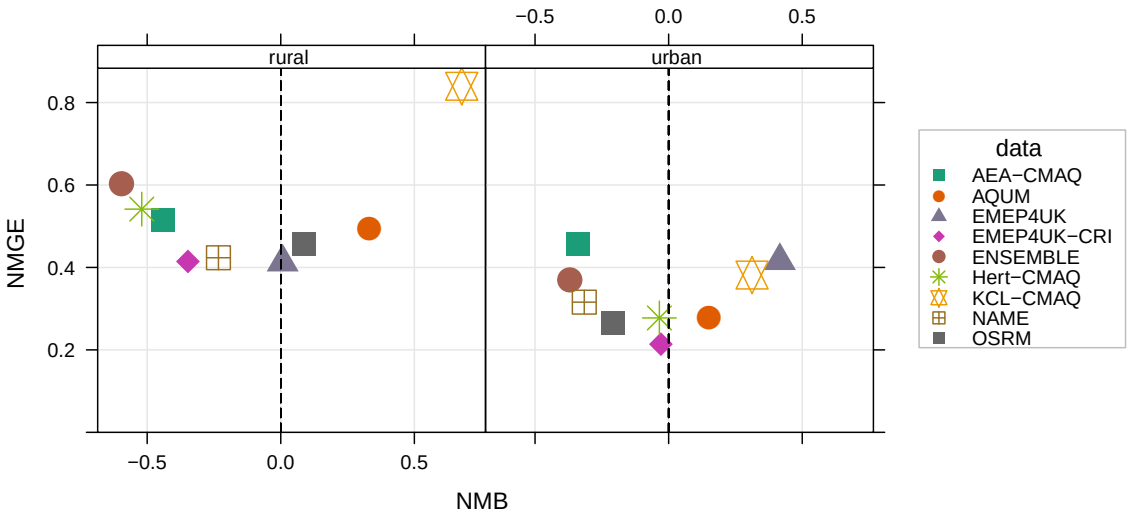

FIGURE 2.10: Comparison of number of days with a daily maximum of running 8-hour means >100 µg m^{-3} predicted and measured values showing the *NMB* against *NMGE* split by rural and urban sites.

The range in predictions across the different models and sites is large at about ±25 days — compared with a typical observed value of about 30 days. It is also interesting to note that unlike many of the other results in this report, the ensemble model does poorly for the number of days where the maximum rolling 8-hour mean is >100 µg m^{-3}. It appears that the reason is that the models predict the timing of the peaks differently and developing an ensemble model of the mean concentration at any one time tends to lower the overall mean values. Indeed, the same is also true of other high concentration metrics including the maximum daily means and the AOT40.

The models produce a wide range of estimates of the number of days metric, which is best shown in Figure 2.11. In general, most models tend to underestimate the number of days where the metric is exceeded. Furthermore, the range in estimates is large (typically around 40 days, which is similar to the magnitude of the exceedances). The Taylor Diagram in Figure A.2 captures the model performance well in that the correlation with measurements is low (maximum r≈0.4) and the RMS error is high.

Given the importance of the exceedance day metric, these findings are important and suggest that all models struggle to produce reliable estimates of the number of exceedance days.

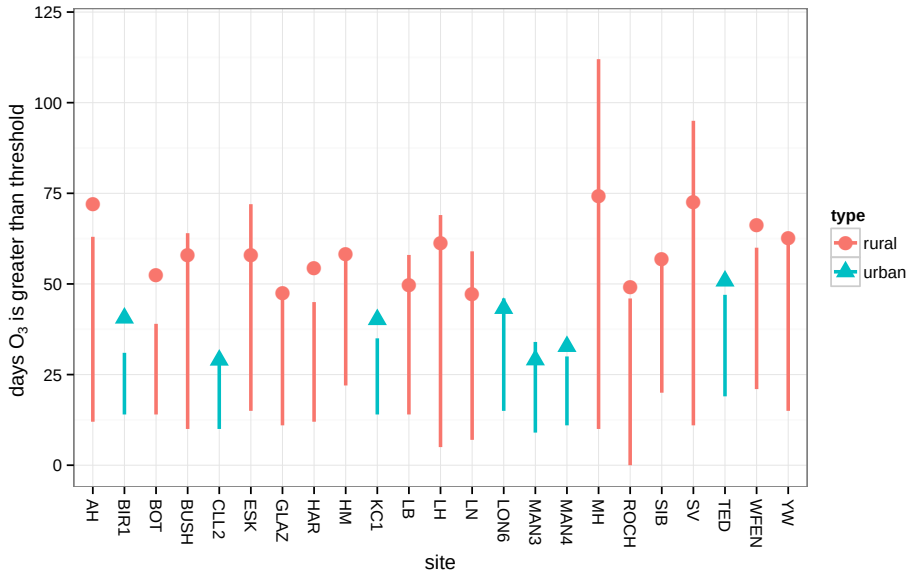

FIGURE 2.11: Comparison of the number of days with a daily maximum of running 8-hour means O_3 concentration >100 µg m^{-3}. The solid shape shows the measured value and the lines the range in model predictions.

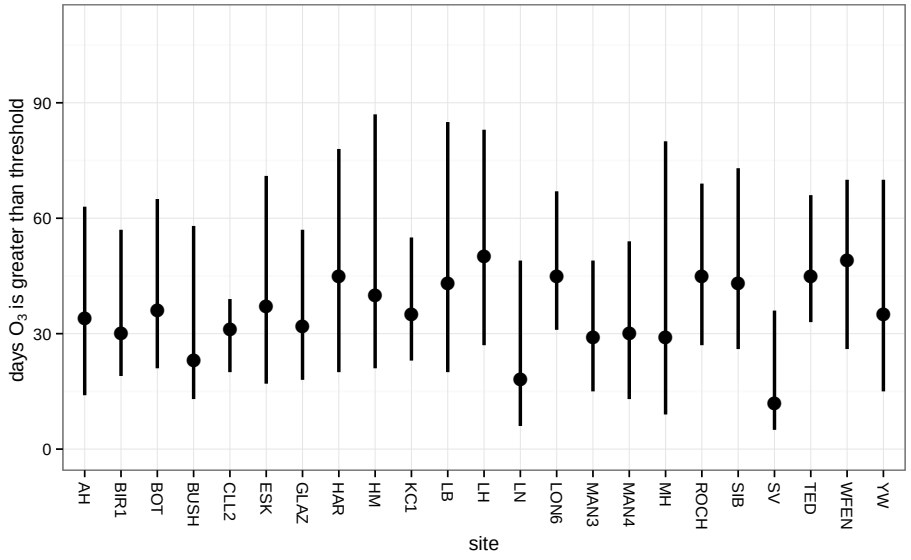

FIGURE 2.12: Comparison of the number of days with a daily maximum of running 8-hour means O_3 concentration >100 µg m^{-3} for the EMEP-4-UK model. The solid circle shows the predicted value and the lines the range in model predictions from a $\pm$10% uncertainty in ozone predictions.

Threshold statistics can be very challenging for models to predict well; particularly when values are close to the threshold. The apparently large differences between the models can caused by relatively small changes in concentration. It is important therefore to understand how sensitive the models are in this respect. As an example, the results from the EMEP-4-UK model have been analysed to show how much of a difference relatively small changes in concentration makes to the number of days where the maximum rolling 8-hour mean is >100 µg m^{-3}.

The results are shown in Figure 2.12 and highlight the extent to which the number of days where the maximum rolling 8-hour mean is >100 µg m^{-3} is sensitive to a $\pm$10%

change in base case O_3 prediction.[5] Figure 2.12 shows very clearly that for a relatively small change in O_3 concentration, the effect on the statistic is large. For example, at Harwell, the estimate varies from 20 to 78 days. Indeed, the results in the $\pm$10% sensitivity test produce similar ranges to the spread across different models as shown in Figure 2.11. Care should be exercised when comparing model predictions for threshold-type statistics, even if they are important health-based standards or limits.

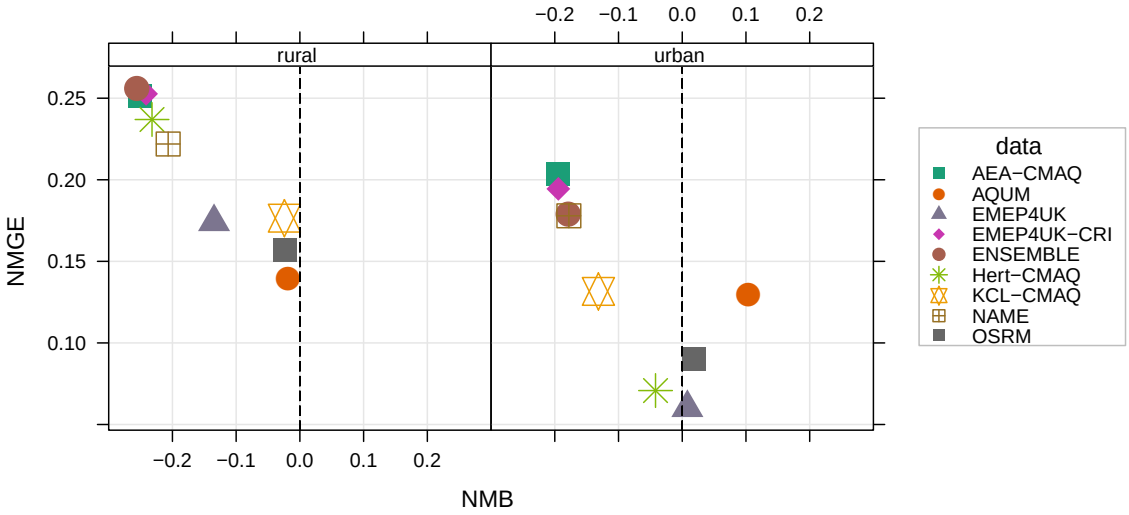

FIGURE 2.13: Comparison of daily maximum O_3 predicted and measured values showing the *NMB* against *NMGE* split by rural and urban sites.

Also considered is a metric that captures *only* summertime episode conditions. While the peak hour metric would correspond to these conditions, there is more uncertainty introduced by considering only one hour e.g. due to measurement uncertainty. Therefore the maximum of the daily means has been considered. All models predict the maximum daily mean within a factor of two at all sites as shown in Figure 2.13, but tend to underestimate the concentration (Table 2.8). Overall, the models are able to predict the maximum daily mean better than the exceedance days metric discussed previously — at least in terms of estimating the absolute value. However, the correlation with measured values is low (r<0.3) and the *RMSE* error is high, which indicates the models are not able to account very well for the variation between sites. Typically the range in predictions of the maximum daily mean is about $\pm$25 µg m^{-3} for a mean values of 129 µg m^{-3}.

It is also interesting to note that there is not a consistent order of model performance for the exceedance days and maximum daily metrics: models that tend to do better with one do not necessarily do better with the other. This observation makes it difficult to identify particular models that are better at predicting the higher concentrations of O_3.

[5] Note that the simple assumption has been made of increasing the O_3 concentration for all hours and sites by $\pm$10%.

TABLE 2.8: Summary statistics for the maximum daily mean O_3 concentration.

site.type	data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
rural	AQUM	16	1.00	-2.58	18.88	-0.02	0.14	22.46	0.35	0.08
rural	OSRM	16	1.00	-3.10	21.27	-0.02	0.16	26.08	-0.07	-0.04
rural	EMEP4UK	16	1.00	-18.23	23.57	-0.13	0.17	31.42	-0.11	-0.15
rural	KCL-CMAQ	16	1.00	-3.29	23.88	-0.02	0.18	29.35	-0.08	-0.16
rural	NAME	15	1.00	-28.10	30.18	-0.21	0.22	37.42	0.13	-0.41
rural	Hert-CMAQ	16	1.00	-31.40	32.07	-0.23	0.24	39.55	0.04	-0.56
rural	AEA-CMAQ	16	1.00	-34.04	34.04	-0.25	0.25	41.89	0.23	-0.66
rural	EMEP4UK-CRI	16	1.00	-32.66	34.19	-0.24	0.25	42.01	-0.04	-0.67
rural	ENSEMBLE	16	1.00	-34.64	34.64	-0.26	0.26	41.45	0.27	-0.69
urban	EMEP4UK	7	1.00	0.93	6.92	0.01	0.06	8.12	0.75	0.28
urban	Hert-CMAQ	7	1.00	-4.82	8.14	-0.04	0.07	12.18	0.27	0.15
urban	OSRM	7	1.00	2.10	10.34	0.02	0.09	12.07	0.32	-0.08
urban	AQUM	7	1.00	11.88	14.92	0.10	0.13	20.01	-0.23	-0.56
urban	NAME	2	1.00	-21.13	21.13	-0.18	0.18	22.04	1.00	-0.56
urban	KCL-CMAQ	7	1.00	-15.12	15.12	-0.13	0.13	17.15	0.68	-0.58
urban	ENSEMBLE	7	1.00	-20.57	20.57	-0.18	0.18	22.72	0.48	-1.15
urban	EMEP4UK-CRI	7	1.00	-22.34	22.34	-0.19	0.19	23.70	0.75	-1.34
urban	AEA-CMAQ	7	1.00	-22.34	23.42	-0.19	0.20	27.24	-0.45	-1.45

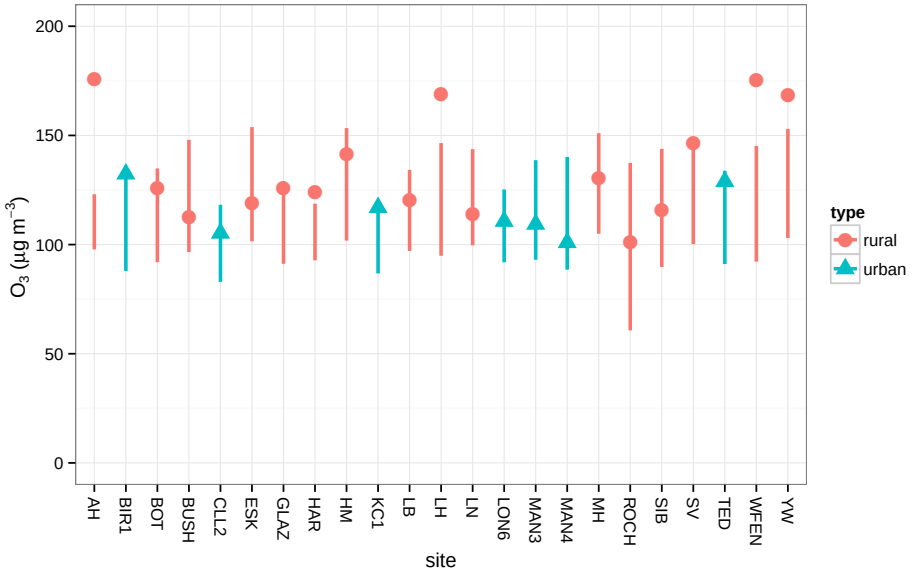

FIGURE 2.14: Comparison of maximum daily O_3 concentration. The solid circle shows the measured value and the lines the range in model predictions.

2.2.4 Comparison of O_3 predictions in relation to other variables

Predictions of other species such as NO_x and NO_2 have also been made by each model together with predictions of meteorological variables, which are considered more in Section 6. The predictions of these other variables can often provide clues as to some of the reasons why model predictions behave as they do. Predictions at the North Kensington site are considered shown in Figure 2.15 as an example. Considering wind speed, most models (except AEA-CMAQ) tend to underestimate wind speed to some extent. There is however no obvious effect of wind speed predictions on O_3 concentrations e.g. the generally worsening performance of the models for high concentrations of O_3 is not associated with a gross over or under-prediction in wind speed. Although the plot is not shown, the model performance for ambient temperature is much better than wind speed and again there is no indication that inadequate temperature predictions are associated

with poor model performance with respect to O_3.

The performance of NO_x predictions is more mixed. There is a tendency for the models to under-predict NO_x across the range of O_3. The exception is KCL-CMAQ which tends to over-predict NO_x for lower O_3. The WRF-Chem results (for June/July) on the other hand tend to over-predict NO_x to a high degree at low O_3 and in this case there is a tendency for O_3 to be underestimated when NO_x is overestimated.

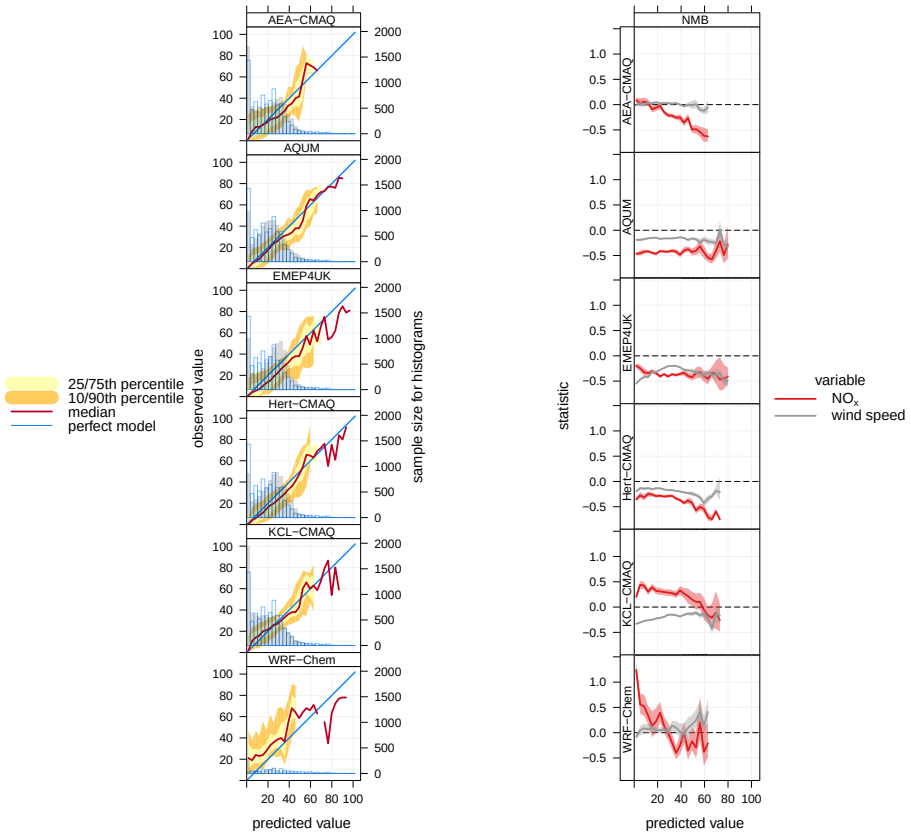

FIGURE 2.15: Conditional quantiles for O_3 concentrations at the North Kensington site. The plot on the right shows how well each model predicts concentrations of NO_x and wind speed (based on predictions at the Heathrow site). The statistic shown in the Normalised Mean Bias.

It is also useful to consider the performance of the models with respect to air mass origin to determine whether there are certain air mass origins that are associated with poor model performance and the extent to which the models agree with one another. As an example, Figure 2.16 shows six back trajectory clusters for the Lullington Heath site calculated using 96-hour HYSPLIT back trajectories. These trajectories provide a wide range of differing source origins from which air masses arrived from.

Figure 2.17 shows several useful results. The plots on the left show conditional quantiles of O_3 concentration by model, together with measured values. The models in general do not cover the full range of observed values i.e. there is an absence of higher O_3 concentrations. The plot on the right shows the proportion of the O_3 concentration intervals split by trajectory cluster calculated in Figure 2.16. There are several interesting features shown in this figure. First, it is clear from the measured concentrations of O_3 that there are higher O_3 concentrations than predicted by most models as already noted. Second, these higher O_3 concentrations (>50 ppb) are nearly all associated with cluster 1 i.e. from continental Europe and under generally anti-cyclonic conditions.

In general, cluster 1 is associated with low (wintertime) and high (summertime) O_3 concentrations and most models capture this behaviour. However, for CMAQ-AEA the highest predicted O_3 concentrations appear to be due to air masses from the Atlantic (cluster 5), suggesting that it misses some of the important high concentration conditions when air masses are from Europe. For the AQUM, Hertfordshire-CMAQ, KCL-CMAQ and EMEP-4-UK models, high O_3 concentrations do derive from cluster 1 air masses, but most other models show a more mixed contribution. For example, while OSRM does predict high O_3 concentrations most of these come from air masses from the south-west (cluster 3).

Clearly, the principal deficiency of most models is in their ability to capture high concentrations of O_3 when air masses arrive from continental Europe under anticyclonic conditions. These are important conditions with respect to O_3 concentrations and dominate the high-O_3 metric statistics.

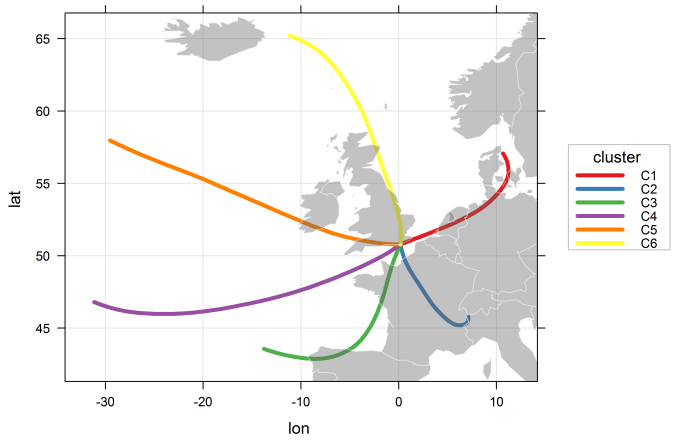

FIGURE 2.16: Back trajectory clusters for the Lullington Heath site for 2006 data.

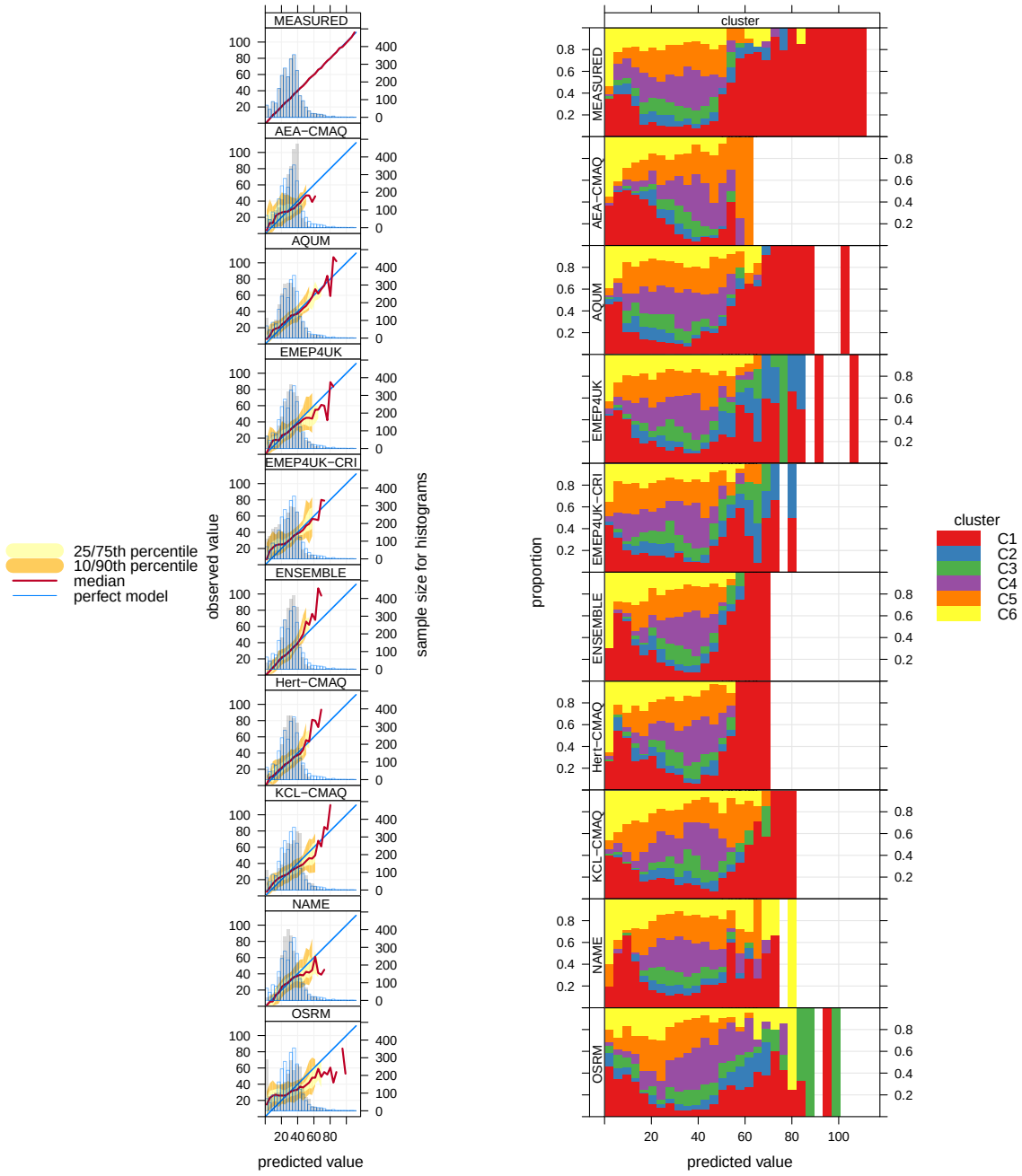

FIGURE 2.17: Conditional quantiles for O_3 concentrations at the Lullington Heath site. The plot on the right shows the origin of the air mass clusters as a proportion.

2.2.5 New air quality index

This section briefly considers the performance of the models against the Defra new daily air quality index. More details concerning the index can be found at `http://uk-air.defra.gov.uk/air-pollution/daqi`. For O_3 the air quality index is shown in Table 2.9, which assumes 8-hour rolling mean concentrations.

Perhaps the most relevant statistic to compare the models against is the mean bias and how that changes through the air quality indexes. Figure 2.18 shows the performance of all the models for June and July 2006. This Figure shows very clearly the general tendency of the models to increasingly underestimate O_3 concentrations as the air quality index increases i.e. as the concentration of O_3 increases itself. The Normalised Mean Bias

TABLE 2.9: The new Defra daily air quality index.

Band	Description	O_3 range (µg m^{-3})
1	Low	0–33
2	Low	34–65
3	Low	66–99
4	Moderate	100–120
5	Moderate	121–140
6	Moderate	141–159
7	High	160–187
8	High	188—213
9	High	214–239
10	Very High	240 or more

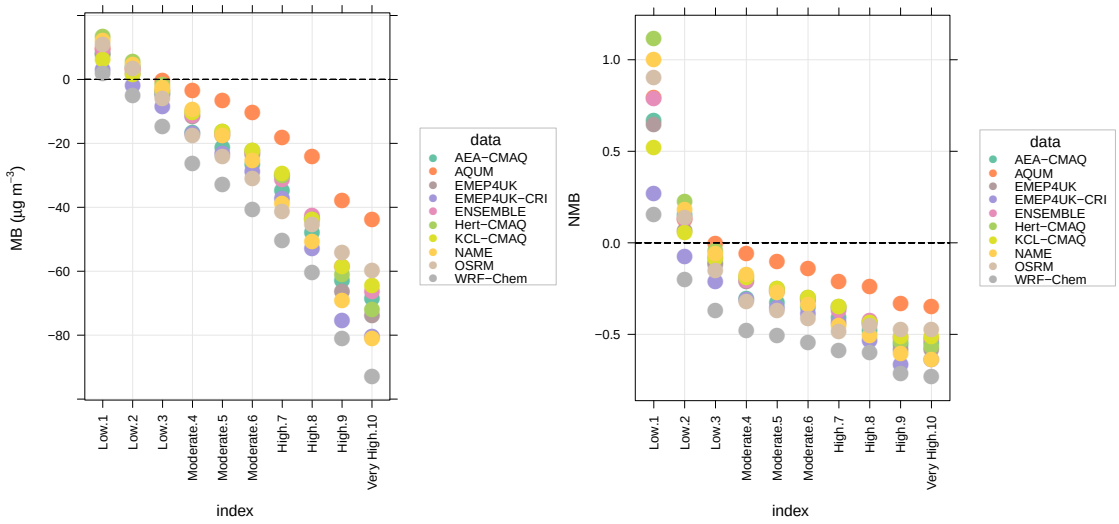

FIGURE 2.18: Comparison of model performance (mean bias, *MB* and normalised mean bias, *NMB*) against the new air quality index for June and July 2006. The index on the x-axis is the index based on observations.

(*NMB*) plot (right) shows that for the highest O_3 index the models typically underestimate O_3 concentrations by a factor of two. However, there is a large range in model performance as the air quality index increases. For the lowest bands (1 and 2) all models tend to overestimate concentrations, but as the bands increase through moderate to high all models increasingly underestimate O_3 concentrations. The best performing model is AQUM, followed by OSRM and KCL-CMAQ. The WRF-Chem model has approximately twice the (negative) bias compared with AQUM at higher air quality indexes and thus tends to underestimate O_3 concentrations by a considerable margin.

The underestimate of O_3 for the high daily indexes can also be expressed in terms of the index itself. For example, when the observed O_3 concentration is at level 7 (High) the corresponding modelled O_3 is typically at level 4, the lowest level on the moderate scale (on average). However, there are large differences between the models. For the same conditions, AQUM for example would estimate the index to be 5 (middle of the moderate range), whereas WRF-Chem would be 3 — clearly considerably less than a value of 7 for observed O_3 concentrations.

It is worth noting that the index is primarily for air quality forecasting purposes but is nevertheless relevant to this report. Indeed, if the same models were used in 'forecast mode' and being driven not be historical meteorological data but forecast data, then in might be expected that model performance would be worse than is shown in this section. However, a worse performance might affect some statistics more than others. It might for example be expected to increase the *RMSE* but not necessarily affect the mean bias.

A common approach to evaluate the quality of forecast predictions is to consider the categorical prediction e.g. was a 'high' prediction forecast by the model? In this case the response is yes or no. There are various methods for evaluating the quality of predictions for these cases and Stephenson (2000) explores approaches based on the *odds ratio*. Specifically, the *odds ratio skill score* is used as a measure of the predictive performance of models for these situations i.e. where there is a yes/no, TRUE/FALSE outcome. As mentioned above, these issues are more relevant to models used for forecasting where the policy maker would like to know whether to issue a warning that a certain level of concentration is likely to be exceeded. Indeed the development of the new daily air quality index by COMEAP used these approaches (COMEAP 2011).

Note that the forecasting performance of some of these models will be considered specifically as a separate exercise during 2013.

3 NO_x and NO_2 comparsions

3.1 Concentration comparisons

In this section the model predictions of NO_x and NO_2 are compared with measurements and with each other. The analysis separately considers urban sites (7 urban background/suburban sites in London, Manchester and Birmingham) and rural background sites (15 sites around the UK). Concentrations of NO_x at urban background sites will be dominated by urban emissions themselves and the comparisons to a large extent provide a test of how well the urban emissions are specified in the models. Given the close coupling of NO-NO_2-O_3 concentrations, good predictions of NO_x ought to help improve O_3 predictions.

For NO_x the model performance statistics are shown below for urban and then rural sites. It is apparent from these statistics that there is a wide range of model performance when it comes to NO_x predictions in urban areas. The most notable difference between the models is the mean bias: several models significantly under-predict urban NO_x concentrations (see the MB and NMB in the table below). In particular, NAME, AQUM, Hertfordshire-CMAQ and EMEP-4-UK all show large under predictions of NO_x. On the other hand, the KCL-CMAQ shows very little bias.

The bias and error for NO_x can be seen more clearly in Figure 3.1. This Figure shows the general negative bias in urban areas and positive bias in rural areas. Not all models show a negative bias in urban areas however. For example, KCL-CMAQ shows very little bias for NO_x; similar to the findings for O_3 considered previously. However, as discussed previously, regional scale models would not necessarily be expected to predict urban NO_x concentrations well and a negative bias would be expected. In rural areas most models tend to show a positive bias, which is relatively large for AQUM and WRF-Chem. The results for NO_2 (not shown) are similar to NO_x with respect to bias.

It should be noted that at rural sites measured concentrations of NO_2 might be expected to overestimate actual NO_2 concentrations due to the use of the chemiluminescence

technique with molybdenum converters (as used in the AURN). This overestimate will depend on the concentrations of interfering species etc. (AQEG 2004; Dunlea et al. 2007). However, as noted above the models tend to show a *positive* bias at rural sites and accounting for these interfering species would tend to make the agreement worse, not better. Throughout a whole year however these effects are likely negligible compared with the modelling uncertainties.

The statistics shown in Table 3.1 and Table 3.2 are based on hourly values, whereas Figure 3.1 is based on annual mean data. The model statistics can vary considerably depending on whether the data are annual or hourly. The choice between the two time averages depends on the aim of the analysis. For example, for annual mean concentrations of NO_2 for which there is a Limit Value it is useful to consider annual mean statistics. Care is needed however when interpreting such results.

Table 3.1: Summary statistics for urban NO_x concentrations based on hourly data.

data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
ENSEMBLE	58060	0.70	-8.24	15.67	-0.26	0.50	32.20	0.62	0.34
KCL-CMAQ	58060	0.65	0.89	17.45	0.03	0.56	31.46	0.61	0.27
EMEP4UK-CRI	57854	0.62	-10.51	17.74	-0.33	0.56	34.38	0.56	0.25
EMEP4UK	57829	0.59	-12.22	18.25	-0.39	0.58	34.89	0.55	0.23
AQUM	58060	0.51	-15.71	18.95	-0.50	0.60	36.53	0.58	0.20
Hert-CMAQ	58053	0.50	-17.31	19.45	-0.55	0.62	37.85	0.55	0.18
AEA-CMAQ	57906	0.58	-6.94	19.72	-0.22	0.63	37.21	0.41	0.17
NAME	15806	0.60	-18.20	24.70	-0.40	0.55	46.50	0.46	0.15
OSRM	58060	0.64	1.84	21.43	0.06	0.68	45.84	0.38	0.10
WRF-Chem	6893	0.50	3.21	15.69	0.16	0.78	22.58	0.31	-0.21

For the rural sites there is no obvious bias towards under prediction of NO_x concentrations. Indeed, two of the models (AQUM and WRF-Chem) show considerable positive bias — see Figure 3.4.

Table 3.2: Summary statistics for rural NO_x concentrations based on hourly data.

data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
Hert-CMAQ	81683	0.55	-1.04	4.23	-0.14	0.59	10.64	0.57	0.30
EMEP4UK	81420	0.52	-0.83	4.54	-0.12	0.63	10.49	0.60	0.25
EMEP4UK-CRI	81454	0.54	-0.09	4.67	-0.01	0.65	10.69	0.60	0.23
ENSEMBLE	81692	0.54	2.50	5.41	0.35	0.76	10.99	0.61	0.11
NAME	81692	0.49	2.46	5.99	0.34	0.84	12.04	0.53	0.01
AQUM	81692	0.49	2.77	6.22	0.39	0.87	13.01	0.54	-0.02
KCL-CMAQ	81692	0.48	3.43	6.46	0.48	0.90	12.58	0.57	-0.06
OSRM	81692	0.47	2.53	7.28	0.35	1.02	16.99	0.39	-0.20
AEA-CMAQ	81514	0.41	8.17	11.34	1.14	1.58	24.61	0.37	-0.87
WRF-Chem	9012	0.34	8.06	9.82	1.57	1.91	16.95	0.42	-1.56

The performance with respect to NO_x concentrations is revealed by considering the temporal components. Figure 3.2 very clearly shows how well the models predict urban (North Kensington) NO_x concentrations — and there is a wide range in model performance. On this basis the KCL-CMAQ and Hertfordshire-CMAQ do well both in terms of the absolute magnitude of NO_x and its diurnal variation. Most models capture the seasonal variation in NO_x concentrations as shown in Figure 3.3, but many models underestimate the magnitude of NO_x during the winter months.

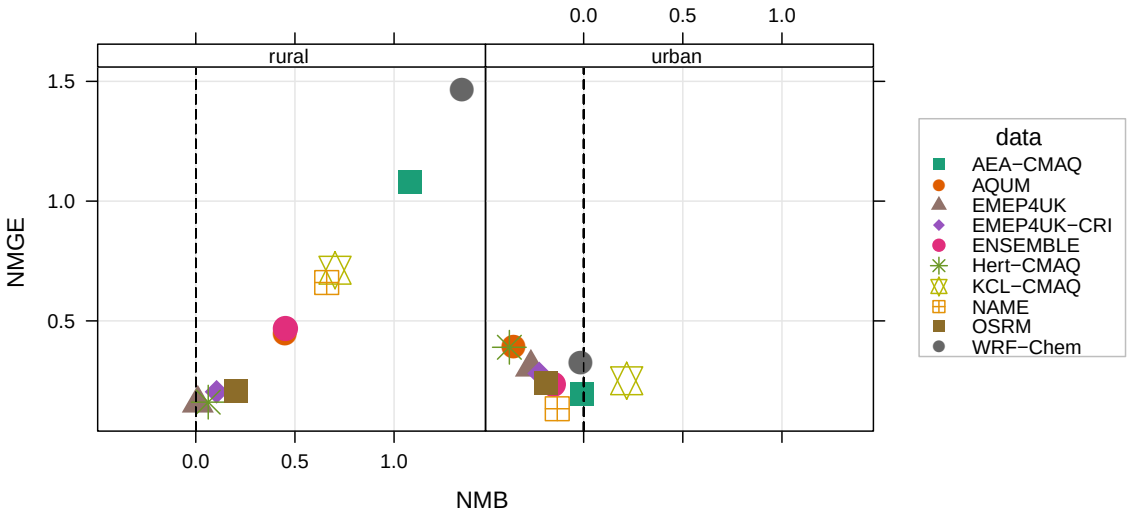

FIGURE 3.1: Comparison of annual mean predicted and measured NO_x concentrations showing the *NMB* against *NMGE* split by rural and urban sites.

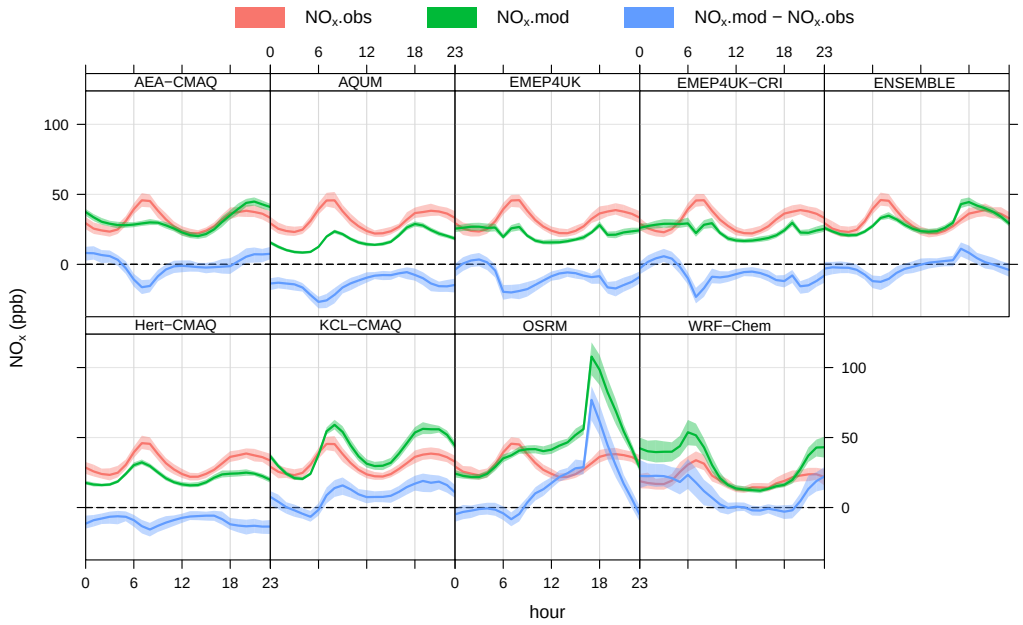

FIGURE 3.2: Diurnal variation in NO_x concentrations predicted at the North Kensington site.

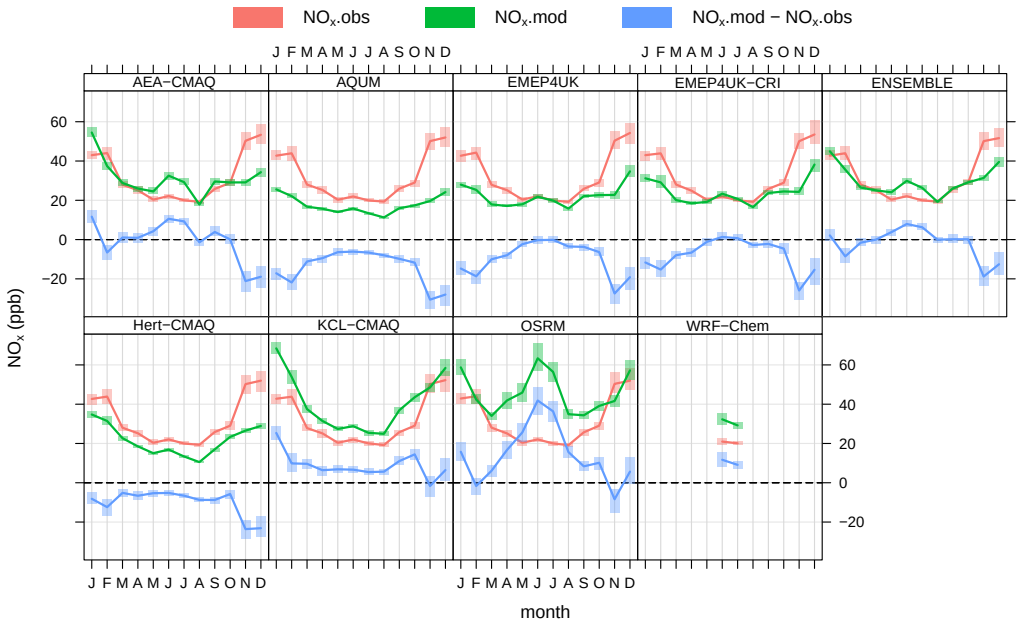

FIGURE 3.3: Seasonal variation in NO_x concentrations predicted at the North Kensington site.

For the rural sites (Harwell in this case) there is no obvious bias towards under prediction of NO_x concentrations. Indeed, two of the models (AEA-CMAQ and WRF-Chem) show considerable positive bias.

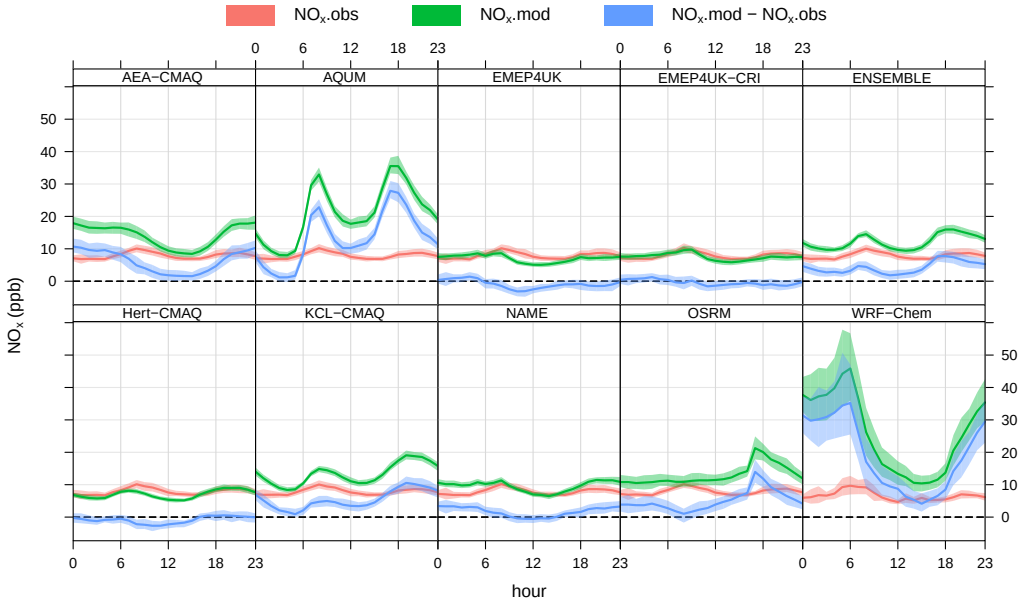

FIGURE 3.4: Diurnal variation in NO_x concentrations predicted at Harwell.

The results for NO_2 are summarised in Table 3.3 (urban) and Table 3.4 (rural).

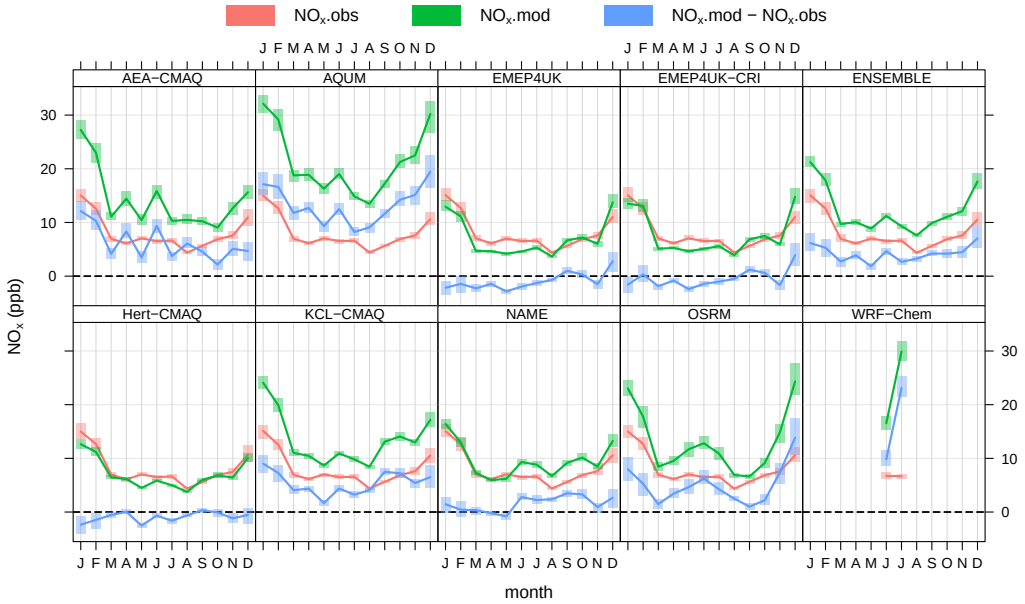

FIGURE 3.5: Seasonal variation in NO_x concentrations predicted at Harwell.

TABLE 3.3: Summary statistics for urban NO_2 concentrations.

data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
ENSEMBLE	58060	0.77	-2.85	6.88	-0.16	0.38	9.58	0.70	0.33
EMEP4UK-CRI	57854	0.68	-4.15	8.19	-0.23	0.45	11.32	0.59	0.21
EMEP4UK	57829	0.64	-4.91	8.58	-0.27	0.47	11.83	0.57	0.17
KCL-CMAQ	58060	0.72	3.97	8.59	0.22	0.47	11.23	0.65	0.17
Hert-CMAQ	58053	0.59	-6.87	8.78	-0.37	0.48	12.17	0.62	0.15
AEA-CMAQ	57906	0.67	-0.23	8.93	-0.01	0.49	12.27	0.54	0.13
AQUM	58060	0.58	-6.49	9.01	-0.35	0.49	12.33	0.58	0.13
OSRM	58060	0.65	-3.60	9.18	-0.20	0.50	13.88	0.44	0.11
WRF-Chem	6893	0.57	-0.17	9.30	-0.01	0.60	12.61	0.38	0.00
NAME	15806	0.69	-3.32	11.27	-0.13	0.45	15.01	0.50	-0.07

TABLE 3.4: Summary statistics for rural NO_2 concentrations.

data	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
Hert-CMAQ	81683	0.56	0.33	2.91	0.06	0.57	4.57	0.70	0.29
EMEP4UK	81420	0.54	0.11	3.16	0.02	0.62	5.25	0.65	0.23
EMEP4UK-CRI	81454	0.55	0.59	3.21	0.12	0.63	5.28	0.66	0.22
ENSEMBLE	81692	0.56	2.37	3.58	0.46	0.70	5.29	0.74	0.13
OSRM	81692	0.49	0.94	4.05	0.18	0.79	6.62	0.47	0.02
AQUM	81692	0.51	2.39	4.12	0.47	0.80	6.47	0.63	-0.00
NAME	81692	0.47	3.33	4.92	0.65	0.96	8.26	0.62	-0.19
KCL-CMAQ	81692	0.47	3.74	4.97	0.73	0.97	7.66	0.69	-0.21
AEA-CMAQ	81514	0.42	5.79	7.10	1.13	1.38	11.73	0.57	-0.72
WRF-Chem	9012	0.34	5.17	6.53	1.28	1.62	9.67	0.43	-1.05

3.2 Effect of model grid size and emission

The models have been shown to have a range of biases when predicting NO_x and O_3. In urban areas in particular there could be important effects due to the model grid size used. Grid size effects also include the locations of the grids themselves e.g. two models may use the same grid size but a different coordinate system which could result in a monitoring site being located in different areas of emission. To investigate this issue in more detail each group was asked to supply the emissions in each grid square corresponding to the location of each monitoring site. Plotting the mean bias in NO_x against model grid size (left plot of Figure 3.6) does show there is a general tendency for models with larger grid cells to have a negative bias for urban NO_x predictions. However, the grid size is not the only factor that will affect the mean bias. Also important is associated NO_x emission in each grid cell. If the mean bias is plotted against the NO_x emission expressed as t km^{-2} (right plot of Figure 3.6) then a much clearer relationship emerges. The right-hand plot shown in Figure 3.6 therefore encapsulates both grid size and emission effects.

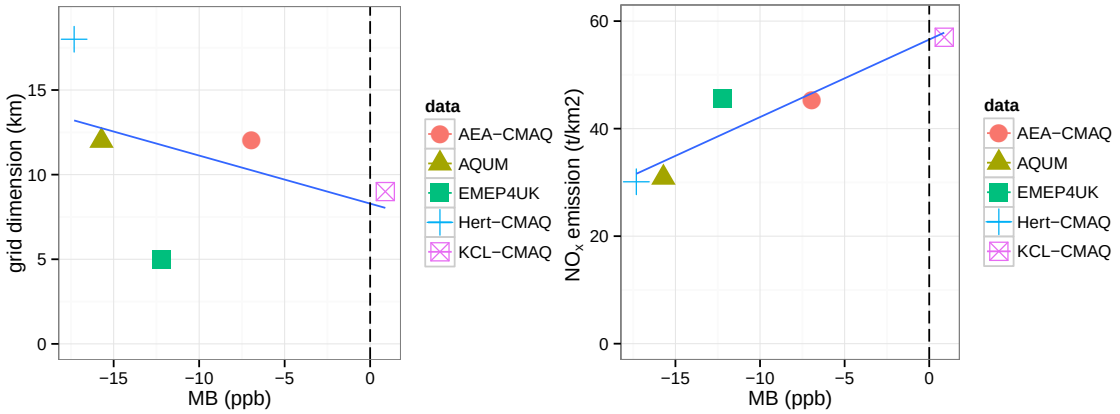

FIGURE 3.6: Left: effect of grid square dimension of NO_x on the mean bias of NO_x concentration for urban sites, right: effect of grid square emission of NO_x on the mean bias of NO_x concentration for urban sites.

It is apparent therefore that the mean bias in urban NO_x predictions is related both to the gird size assumed *and* the NO_x emission assumptions. The best performing model (KCL-CMAQ) has both a relatively small grid size (9 km) and a relatively high emission of NO_x. Comparing the mean urban NO_x emissions in t km^{-2} yields the following: KCL-CMAQ = 57.0, Hertfordshire-CMAQ = 30.1, EMEP-4-UK = 45.6, AQUM = 30.9 and AEA-CMAQ = 45.2.

The higher urban emissions of NO_x assumed in KCL-CMAQ are likely (in part) due to the adoption of more recent vehicle emission data based on recent vehicle emission remote sensing, which assume higher emissions of NO_x compared with previous data particularly for diesel vehicles (Beevers et al. 2012; Carslaw et al. 2011).

4 Effect of precursor emission changes

4.1 Introduction

This section considers how the models respond to changes in precursor emissions. Consideration has been given to the full range of sites and all hours of 2006. While the amount

of processing required is large, this approach has the benefit of being able to understand the impacts the scenarios have on a full year of data and hence various O_3 metrics that require a full year of data. Later in Section 5 there is a focus on considering one month (July) in more depth where concentrations of O_3 are at their highest.

The following scenarios were requested to be run by each model group referred to as S1 to S4:

S1 Reduce total anthropogenic NO_x and VOC by 30% across the UK + Europe

S2 Reduce total anthropogenic NO_x and VOC by 30% across the UK only

S3 Reduce anthropogenic NO_x by 30% across UK + Europe

S4 Reduce anthropogenic VOC by 30% across UK + Europe

Note that UK is defined as sources in the NAEI and therefore will include coastal shipping emissions. These scenarios were considered to understand how the models respond to changes, rather than representing scenarios of direct policy relevance.

Because there are no observations against which the results can be compared, the comparisons focus on the differences between models. However, additional information was requested relating to various indicator species, which will also be considered later in this section. In addition, it is useful to understand the ranges in predicted changes in O_3 and other species across the different models and whether there is broad consistency in the results. In particular it is useful to know for key O_3 metrics the effect of different emissions reduction scenarios.

There are a very large number of model-scenario-receptor combinations and more detail can be provided where necessary. However, only a selection of comparisons is shown here focusing on the annual mean and metrics for higher concentrations of O_3.

4.2 Effect of scenarios on annual mean O_3 concentrations

The effect of the scenarios on the annual mean O_3 concentration is summarised across all receptors in Table 4.1. The broad response of the models is similar across the scenarios in that S1 to S3 (which reduce NO_x *and* VOCs in the UK/Europe or reduce only NO_x; S3) result in increased O_3 concentrations. The reduction of only VOCs (S4) reduces the annual mean O_3 concentration. There is however, a large variation by model shown in Table 4.1. For example for S1 (reducing NO_x and VOCs in Europe) the models vary from increasing O_3 from 0.5 µg m^{-3} (EMEP4UK) to 3.7 µg m^{-3} (AEA-CMAQ), for rural areas. For urban areas there is more of a difference between the models: from 3.8 µg m^{-3} (EMEP-4-UK) to 8.0 µg m^{-3} (KCL-CMAQ). The largest increases in O_3 for scenarios S1 to S3 is in urban locations, which is expected because reductions in NO_x concentration will directly affect the NO + O_3 reaction. The extent to which the models predict an increase in urban areas will also depend on how well the models predict urban NO_x in the first place, which is described in Section 3. For example, the KCL-CMAQ model tends to predict higher urban NO_x concentrations than the other models and consequently tends to predict greater increases in O_3 concentration for scenarios that reduce NO_x compared with other models.

In Figure 4.1 and the following plots, the results are split by site type (rural or urban) and a box and whisker plot is shown to highlight the distribution of changes in O_3 (ΔO_3 with a negative value correponding to a *reduction in O_3 concentration*). The box and whisker

TABLE 4.1: Summary changes in annual mean O_3 concentrations across all receptors by model for 2006 for annual mean concentrations ($\mu g\ m^{-3}$).

site.type	data	S1	S2	S3	S4
rural	AEA-CMAQ	3.7	2.8	4.1	-0.1
rural	AQUM	2.9	2.5	3.4	-0.5
rural	EMEP4UK	0.5	1.2	1.9	-1.5
rural	ENSEMBLE	2.4	2.3	3.1	-0.7
rural	Hert-CMAQ	2.5	2.1	2.6	-0.2
rural	KCL-CMAQ	3.4	3.5	4.0	-0.6
rural	NAME	1.7	2.1	3.2	-1.5
rural	OSRM	1.8	2.1	2.6	-0.7
urban	AEA-CMAQ	6.8	6.1	7.3	-0.2
urban	AQUM	5.5	5.2	6.2	-0.6
urban	EMEP4UK	3.8	4.6	5.3	-1.7
urban	ENSEMBLE	5.7	5.8	6.5	-0.7
urban	Hert-CMAQ	5.4	5.3	5.7	-0.3
urban	KCL-CMAQ	8.0	8.3	8.9	-0.8
urban	NAME	6.4	7.1	8.7	-2.2
urban	OSRM	4.6	5.0	5.4	-0.7

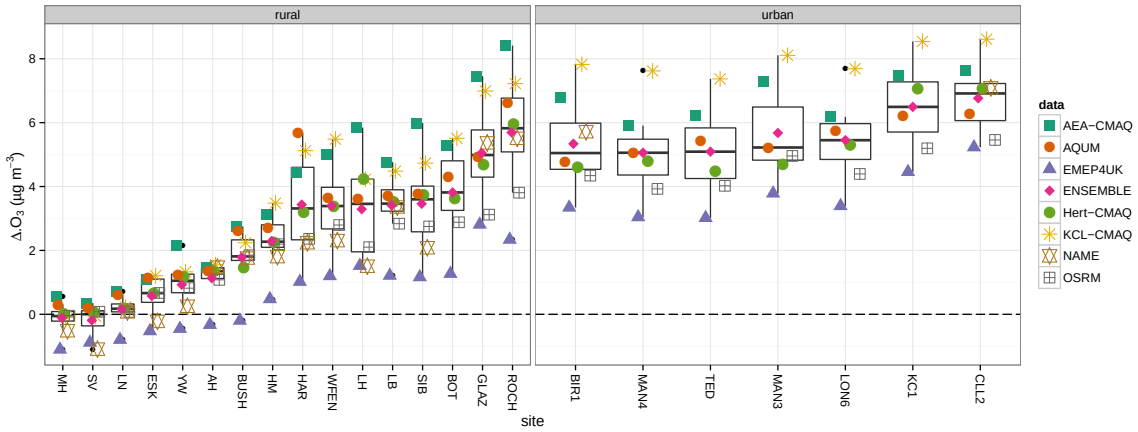

FIGURE 4.1: Model predictions of the change in annual mean O_3 from Scenario 1. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

plot is useful for showing the median response (shown by the horizontal line) and the 25 to 75th percentiles (shown by the box). In addition, the sites have been ordered in terms of their response to a change in O_3 from greatest reduction to least reduction/greatest increase, which can be thought of as 'best' to 'worst' in terms of O_3 concentration change. The benefit of plotting the data in this way is that it is easier to spot patterns in the data e.g. in understanding whether certain types of site have a particular response.

The urban sites tend to give a consistent increase in O_3 concentration, which will be driven by the reduction in urban NO_x. The greatest increase in O_3 is predicted at the central/inner London sites. For the rural sites, those that are in the most remote areas tend to show very little change in O_3 e.g. Mace Head, Strath Vaich and Eskdalemuir. As sites become increasingly influenced by urban areas then there is an tendency for those sites to show increased O_3 concentrations. Overall, however, reducing NO_x and VOCs across the UK and Europe tends to result in increased annual mean O_3 concentrations.

The models give a very similar response to S1 when action is taken to reduce NO_x and

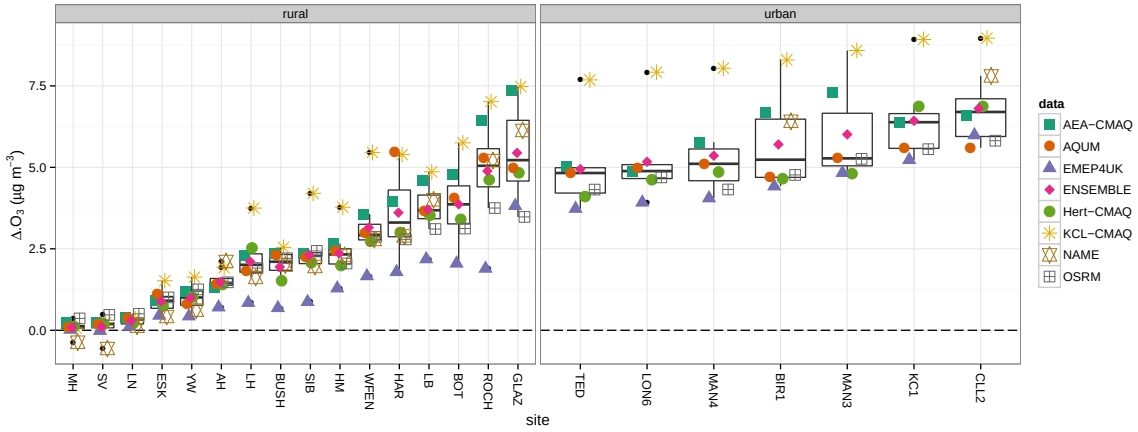

FIGURE 4.2: Model predictions of the change in annual mean O_3 from Scenario 2. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

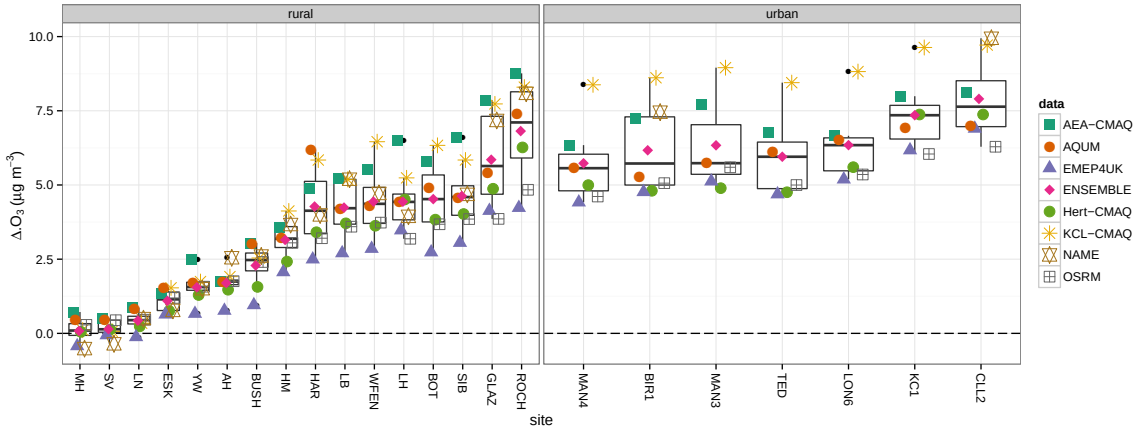

FIGURE 4.3: Model predictions of the change in annual mean O_3 from Scenario 3. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

VOCs in the UK only (S2, Figure 4.2) or when only NO_x is reduced in the UK+Europe (S3, Figure 4.3) i.e. the changes in O_3 are controlled by changes in NO_x emissions.

For almost all scenarios considered, ensemble model produces results that are close to the median of all models and perhaps represents a good 'central estimate'.

Overall for scenario 4 (reduction in UK+Europe VOCs) there is a consistent response of the models in that they all tend to show that O_3 concentrations will decrease in rural and urban locations. The actual change in O_3 concentration is however small at about 1 μg m^{-3}, but is similar for rural and urban sites. Also, rather than the most remote sites showing the most reduction in O_3 concentration, it tends to be sites located in south-east England such as Sibton, Lullington Heath and Wicken Fen.

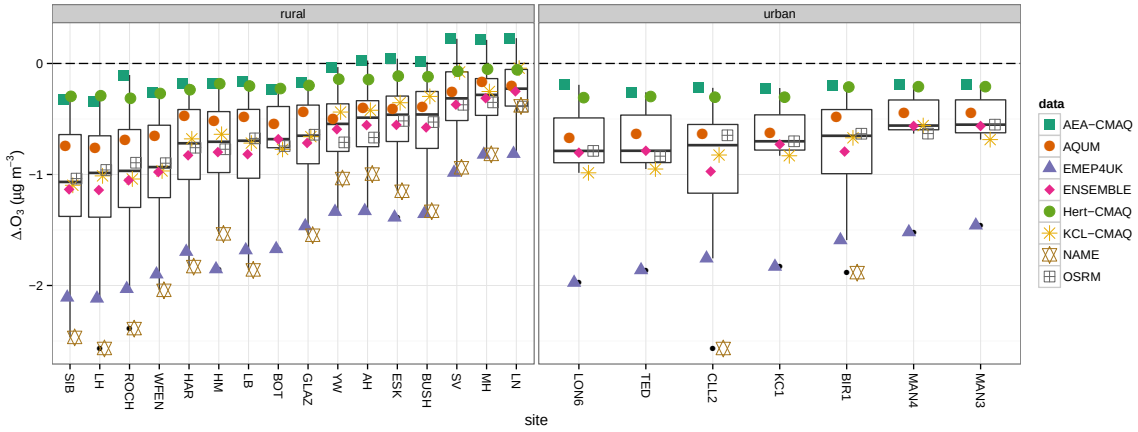

FIGURE 4.4: Model predictions of the change in annual mean O_3 from Scenario S4. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

4.3 Effect of scenarios on higher O_3 concentration metrics

The ability of the models to predict changes in the higher concentration O_3 metrics due to precursor emission changes is an important aspect of their predictive ability. As noted earlier there can be a large range in the accuracy of the models with respect to higher O_3 concentration metrics when compared with measured values. However, it is not certain how consistent a response the models will give to a change in precursor emissions. The use of similar chemistry schemes etc. might for example suggest there is a high degree of consistency. However, as the results in this section show, the response of individual models is mixed.

The predictions for the number of days with a daily maximum of running 8-hour means >100 μg m^{-3} O_3 are shown in Table 4.2 and Figure 4.5 (Scenario S1). These results show a very wide range in behaviour of the models at the different receptors. Considering Scenario S1 in urban areas (Figure 4.5), O_3 concentrations are predicted to increase slightly. In rural areas there is a wider range in responses. Those locations in the most remote areas of the UK and Ireland shows a consistent decrease in the number of days with a daily maximum of running 8-hour means >100 μg m^{-3} O_3. However, in more intermediate sites such as Harwell there is little change in O_3 predicted. The results again highlight the while these general patterns of change are observed, there can be a large difference in the response of individual models.

Reducing NO_x and VOCs in the UK only (Figure 4.6) tends to increase urban O_3 concentrations more than S1 and there is less reduction in O_3 at the more remote sites. These results show the benefit of action taken at a European scale versus UK-only action.

TABLE 4.2: Summary changes in number of days with a daily maximum of running 8-hour means >100 µg m^{-3} O_3 across all receptors by model for 2006. Concentration values are given in µg m^{-3}.

site.type	data	S1	S2	S3	S4
rural	AEA-CMAQ	-3.5	1.4	-1.4	-1.4
rural	AQUM	1.6	1.9	4.2	-2.4
rural	EMEP4UK	-6.1	0.8	1.1	-8.9
rural	ENSEMBLE	-1.7	0.8	2.4	-3.3
rural	Hert-CMAQ	-1.7	0.2	-1.3	-0.6
rural	KCL-CMAQ	-0.2	3.7	3.0	-2.9
rural	NAME	-3.0	1.5	7.3	-7.7
rural	OSRM	-1.5	3.4	2.2	-2.6
urban	AEA-CMAQ	-2.0	3.9	0.1	-1.3
urban	AQUM	5.9	6.3	8.1	-2.4
urban	EMEP4UK	-1.9	4.4	7.3	-8.3
urban	ENSEMBLE	1.1	3.3	5.0	-4.0
urban	Hert-CMAQ	-0.3	3.3	1.4	-1.3
urban	KCL-CMAQ	7.0	10.1	10.7	-4.1
urban	NAME	1.0	7.0	14.0	-6.5
urban	OSRM	2.0	7.1	5.1	-3.3

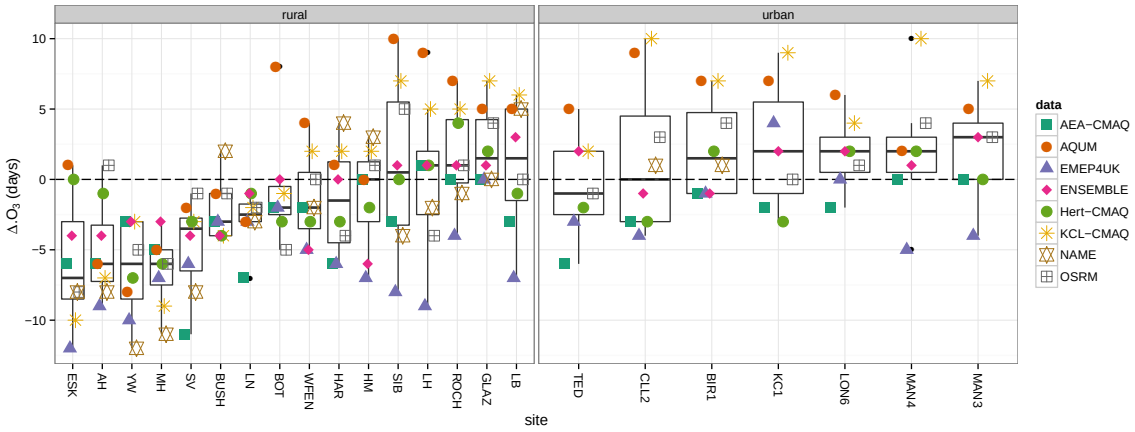

FIGURE 4.5: Model predictions of the change in in number of days with a daily maximum of running 8-hour means >100 µg m^{-3} O_3 for Scenario S1. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

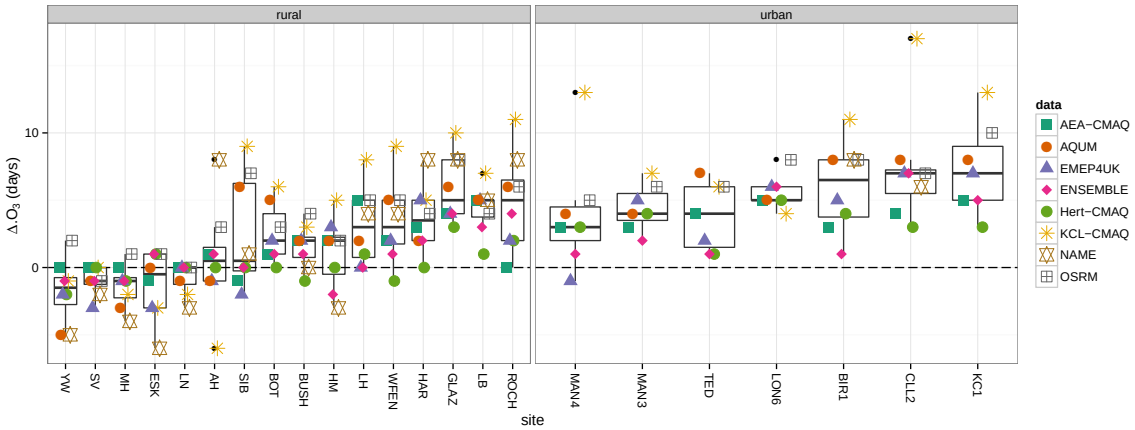

FIGURE 4.6: Model predictions of the change in in number of days with a daily maximum of running 8-hour means >100 µg m^{-3} O_3 for Scenario S2. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

Reducing VOCs in the UK+Europe (S4) again shows that the models tend to agree that the number of days exceeding 100 µg m^{-3} decreases in both rural and urban areas. On average the decrease in exceeding days is modest (<5 days), but some models at some receptors predict over 20 days.

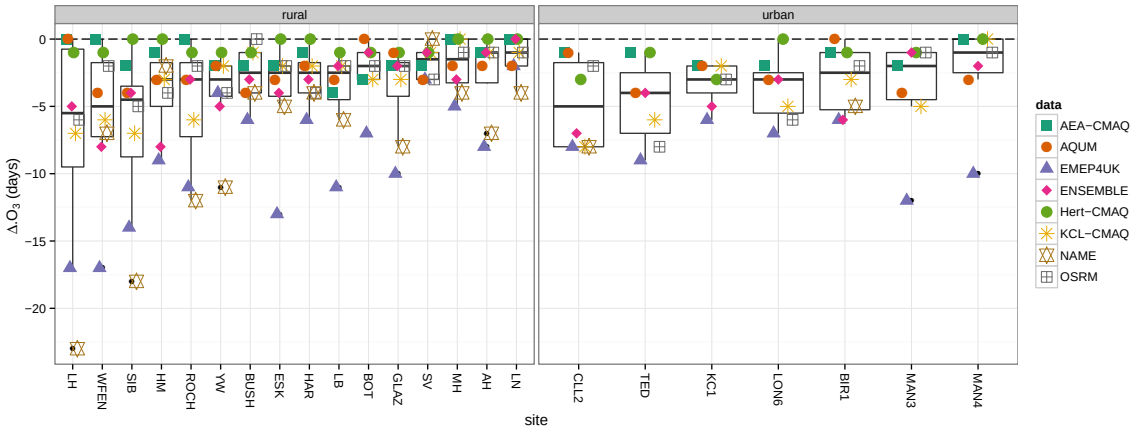

FIGURE 4.8: Model predictions of the change in the number of days with a daily maximum of running 8-hour means >100 µg m^{-3} for Scenario S4. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

The wide variation in the model predictions carries through to the estimates of the maximum daily mean O_3 concentration, shown in Table 4.3. The maximum daily mean statsitic will represent peak sumertime O_3 conditions. The model responses overall are similar to the days with a daily maximum of running 8-hour means >100 µg m^{-3}.

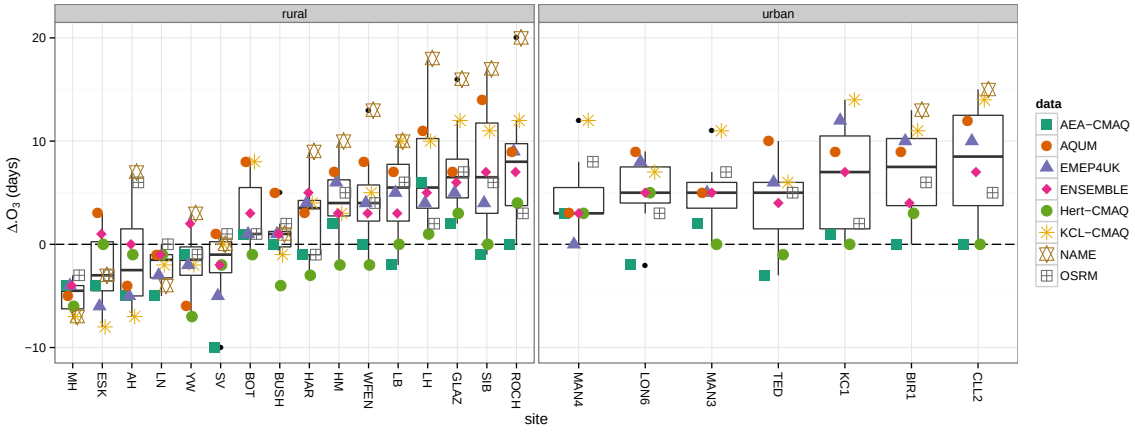

FIGURE 4.7: Model predictions of the change in in number of days with a daily maximum of running 8-hour means >100 µg m^{-3} O_3 for Scenario S3. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

TABLE 4.3: Summary changes in maximum daily O_3 concentration across all receptors by model for 2006. Concentration values are given in µg m^{-3}.

site.type	data	S1	S2	S3	S4
rural	AEA-CMAQ	-2.4	1.8	-1.1	-1.4
rural	AQUM	-3.3	1.2	-0.1	-3.5
rural	EMEP4UK	-6.1	0.6	0.3	-9.1
rural	ENSEMBLE	-3.4	1.2	-0.5	-3.2
rural	Hert-CMAQ	-2.7	0.5	-2.1	-0.8
rural	KCL-CMAQ	2.1	2.4	2.4	-0.4
rural	NAME	-3.0	0.5	3.9	-6.1
rural	OSRM	-1.7	0.7	0.0	-0.5
urban	AEA-CMAQ	1.0	3.6	2.1	-1.0
urban	AQUM	-0.5	4.2	3.3	-2.4
urban	EMEP4UK	-4.0	4.5	2.4	-8.0
urban	ENSEMBLE	-0.2	6.4	2.7	-3.2
urban	Hert-CMAQ	-1.0	5.2	-0.0	-1.2
urban	KCL-CMAQ	9.0	9.5	9.5	-2.0
urban	NAME	3.4	9.6	19.0	-9.4
urban	OSRM	1.1	4.6	1.3	-1.1

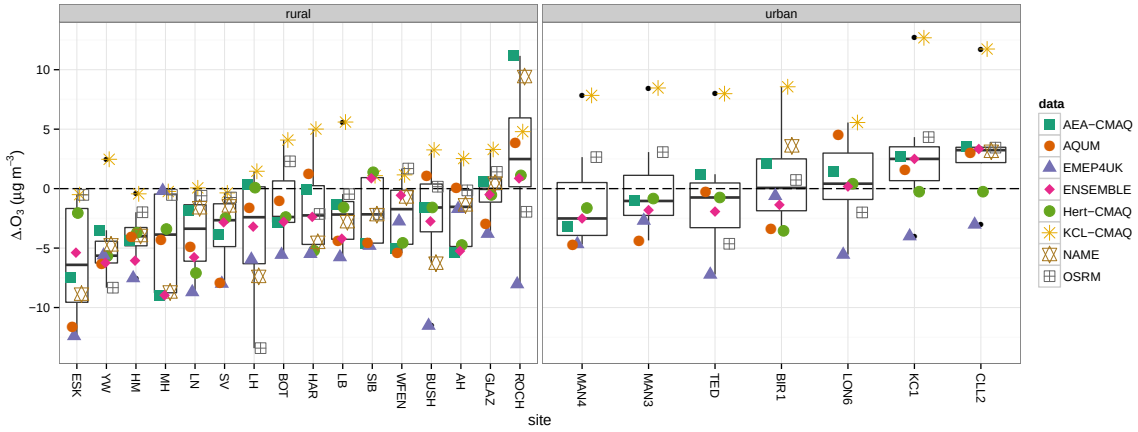

FIGURE 4.9: Model predictions of maximum daily mean O_3 concentration for Scenario S1. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

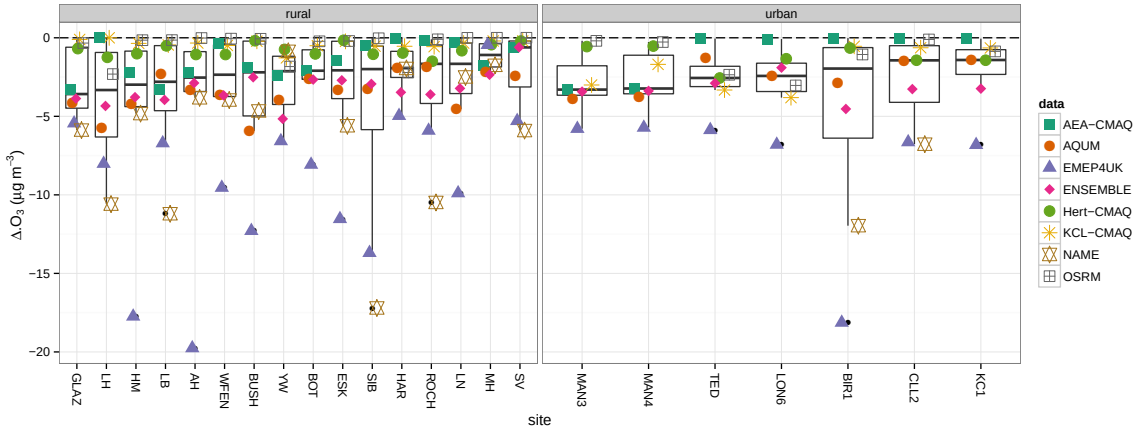

FIGURE 4.10: Model predictions of maximum daily mean O_3 concentration for Scenario S4. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

4.4 Effect on AOT40

The scenario results for the AOT40 metric are shown in Figure 4.11 to Figure 4.14. In general the results follow a similar pattern to those for the maximum daily mean concentration of O_3. Again it is found that S4 is most effective.

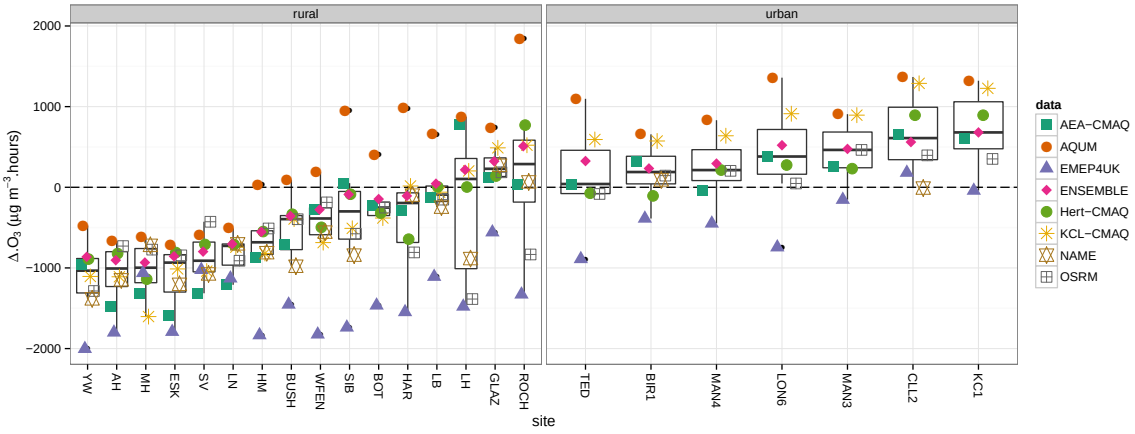

FIGURE 4.11: Model predictions of AOT40 for Scenario S1. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

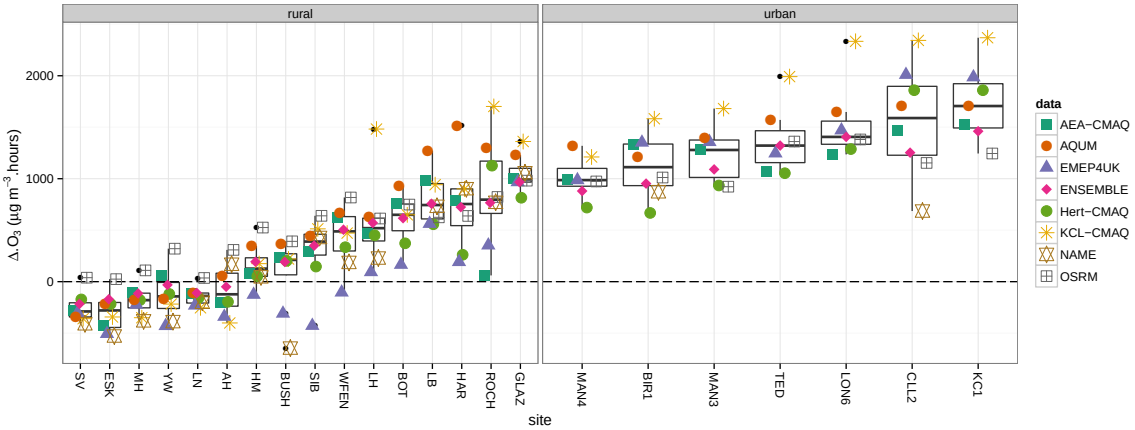

FIGURE 4.12: Model predictions of AOT40 for Scenario S2. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

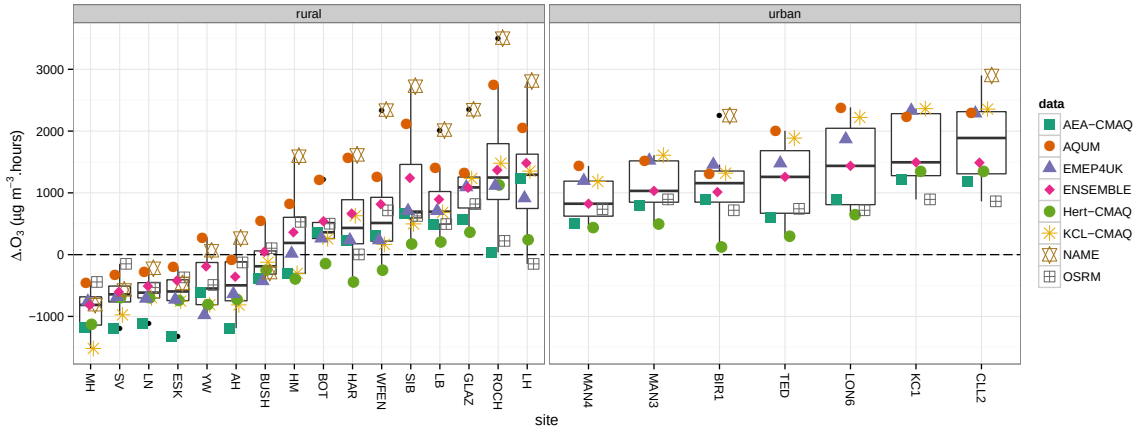

FIGURE 4.13: Model predictions of AOT40 for Scenario S3. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

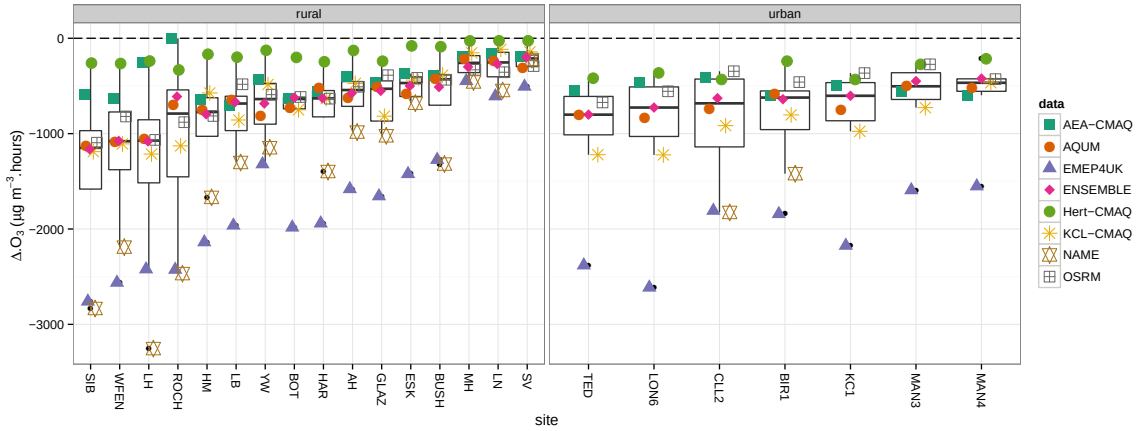

FIGURE 4.14: Model predictions of AOT40 for Scenario S4. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

5 A closer look at episode conditions — July 2006

In this section we consider how well the models perform under important episode conditions when O_3 concentrations are elevated. The initial focus will be on the Harwell site that has often been used for studying O_3 episodes and for which all models have predictions for, including the PTM. The focus will also be on July 2006 when there were several distinct periods when O_3 concentrations were greater than 50 ppb. Figure 5.1 shows the hourly O_3 concentration at the Harwell site for 2006 and highlights several periods where O_3 concentrations were elevated. In particular, the first few days of July and mid-July were periods when the concentration of O_3 >50 ppb.

It is also worth considering the origins of the air masses during July 2006 at Harwell. Figure 5.2 shows the 96-hour back trajectories for July 2006 at Harwell calculated using the HYSPLIT model. The trajectories have been segregated according to O_3 intervals (0–25, 25–50 and >50 ppb) to highlight where the highest O_3 concentration trajectories originate. In fact many of the trajectories for high O_3 (>50 ppb) result from circulatory motion due to a high pressure system with origins in continental Europe.

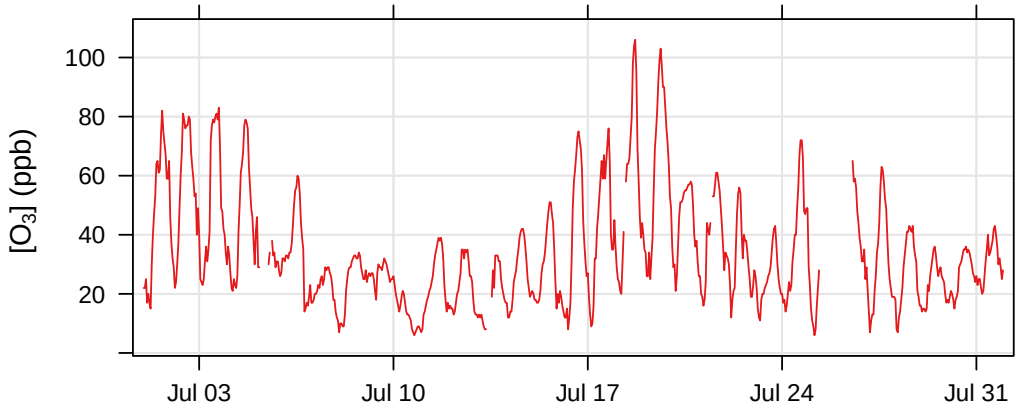

FIGURE 5.1: Hourly O_3 concentrations in July 2006 at the Harwell site.

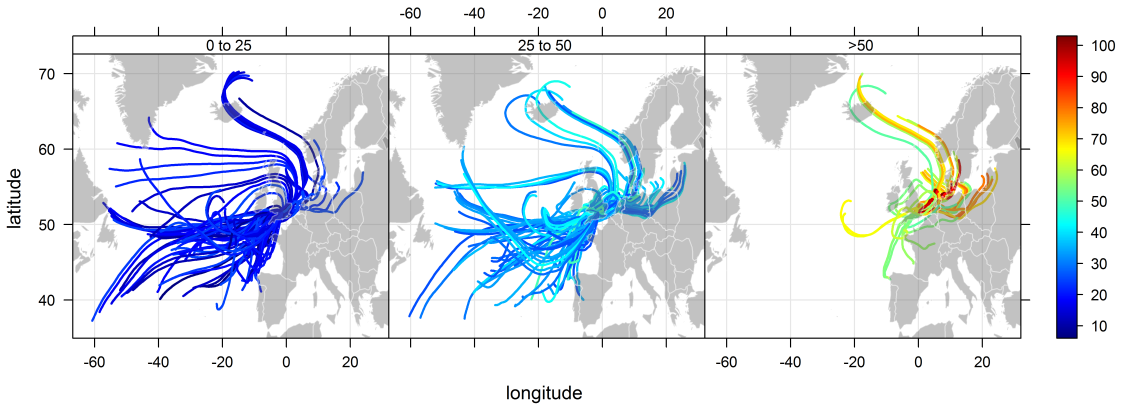

FIGURE 5.2: HYSPLIT 96-hour back trajectories for July 2006 at the Harwell site split by O_3 concentration (0–25, 25–50 and >50 ppb).

5.1 Indicator species

An important concept in the modelling of O_3 is whether concentrations of O_3 are VOC or NO_x-sensitive. A regime is known as VOC-sensitive if a reduction in VOC emissions leads to a greater reduction in O_3 concentrations compared with the same reduction in NO_x emissions. Similarly, a NOx-sensitive regime is one in which a reduction in NO_x emissions would lead to a greater reduction in O_3 than the equivalent reduction in VOCs. Where a location is NO_x or VOC sensitive depends on many factors (notably the relative emissions of NO_x or VOCs). An understanding of these regimes is very important for O_3 modelling because a reduction in VOCs or NO_x may lead to very different i.e. increases or decreases in O_3 concentration.

With respect to evaluation models, many methods have aimed to characterise the responses of models to changes in precursor emissions in this way. Much of this work originates from Sillman (1995), who developed and tested a series of 'indicator' species that could be used to tell whether O_3 concentrations were VOC or NO_x-sensitive. Sillman (1995) considers a range of different indicators to characterise O_3 concentrations as NO_x or VOC-sensitive. A commonly used indicator is to plot the reduction in O_3 concentration in ppb on the y-axis to the ratio of $O_3/(NO_y - NO_x)$. The denominator is often referred to as NO_z (sum of the reaction products).

Sillman (1995) showed that plotting the data in this way showed a transition from VOC to NO_x-sensitive conditions when $O_3/(NO_y - NO_x) > 9$.

Figure 5.3 shows the results of plotting the data according to Sillman (1995) for each model for July across all sites. Most models show a cross-over from VOC to NO_x-sensitive when $O_3/(NO_y - NO_x) > 15$. However, there are some important differences in the way the models respond. The Hertfordshire-CMAQ results are much less sensitive to VOC control than the other models; in other words reducing VOC emissions has very little effect on peak O_3 concentrations at any site. This behaviour is also seen in the analysis in Section 4. Some of the reasons for this behaviour are likely related to the higher biogenic VOC emissions assumed by Hertfordshire-CMAQ (see Table 1.3). Therefore, reducing anthropogenic VOCs would have less of an effect compared with other models that assume no or considerably less BVOC emissions. There may also be effects due to VOC reactivity because BVOCs tend to be more reactive than anthropogenic VOCs.

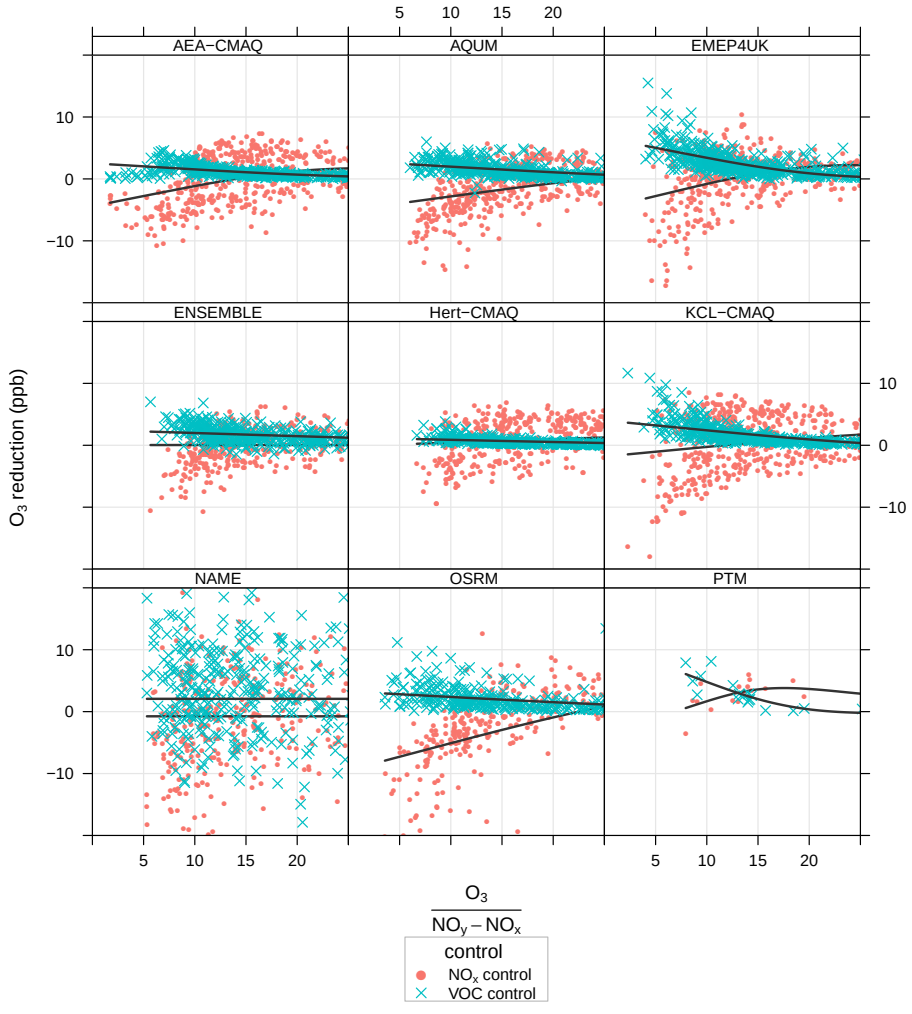

FIGURE 5.3: Plot of O_3 reduction versus $O_3/(NO_y - NO_x)$ for 15:00 GMT, July 2006 for all sites.

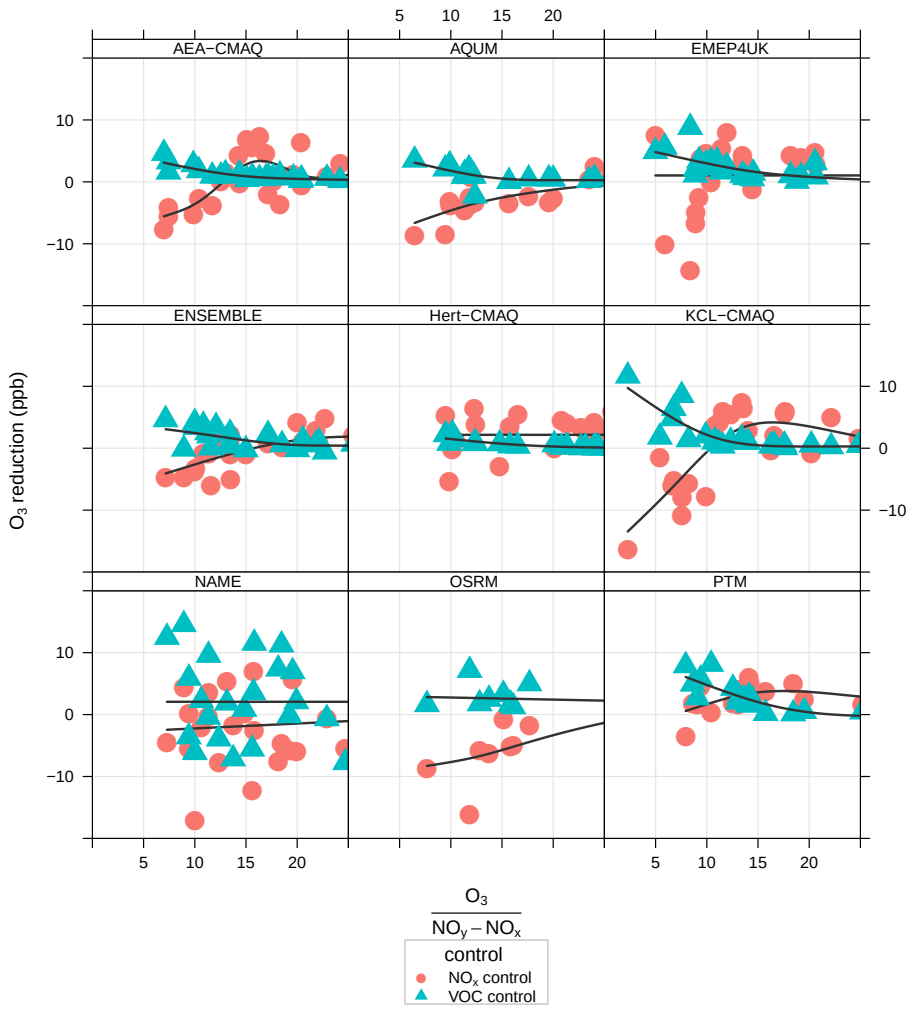

FIGURE 5.4: Plot of O_3 reduction versus $O_3/(NO_y - NO_x)$ at Harwell for 15:00 GMT, July 2006.

5.2 Effect of NO_x vs. VOC control and UK vs. European emissions control

The scenarios allow a more detailed consideration of the efficacy of NO_x vs. VOC control and how the models consider UK vs. UK + European control. These issues are considered in more detail for the Harwell site in Derwent et al. (2013). To make the analysis tractable the maximum daily rolling 8-hour mean O_3 concentrations are presented for July only.

Considering the Harwell site shown in Figure 5.5 it can be seen that all scenarios tend to reduce O_3 concentrations. However, as previously noted there is a large range in responses by model. For example, reducing NO_x + VOC by 30% across the UK and Europe (S1) results in change in O_3 from 1 µg m^{-3} (AQUM) to 17 µg m^{-3} (EMEP-4-UK). A reduction in NO_x only produces a more mixed response from the models; most show a reduction but two models predict an increase in O_3 (AQUM and OSRM). A reduction in only VOC (S4) shows that all models predict a reduction in O_3.

One important issue is that a detailed analysis of a single site might not represent typical model behaviour at other locations. Clearly there will be site specific factors such as the proximity to sources or particular meteorological conditions that will affect the model results. More generally grid based models may not represent the O_3 responses at a particular point in space due to the *incommensurability* between volume-average and point-wise observations (Swall and Foley 2009). For this reason it is worth considering the response in O_3 at other sites. Three other sites are shown in Figure 5.5 as examples of other responses: Lullington Heath, Rochester and Wicken Fen, all the south-east of England.

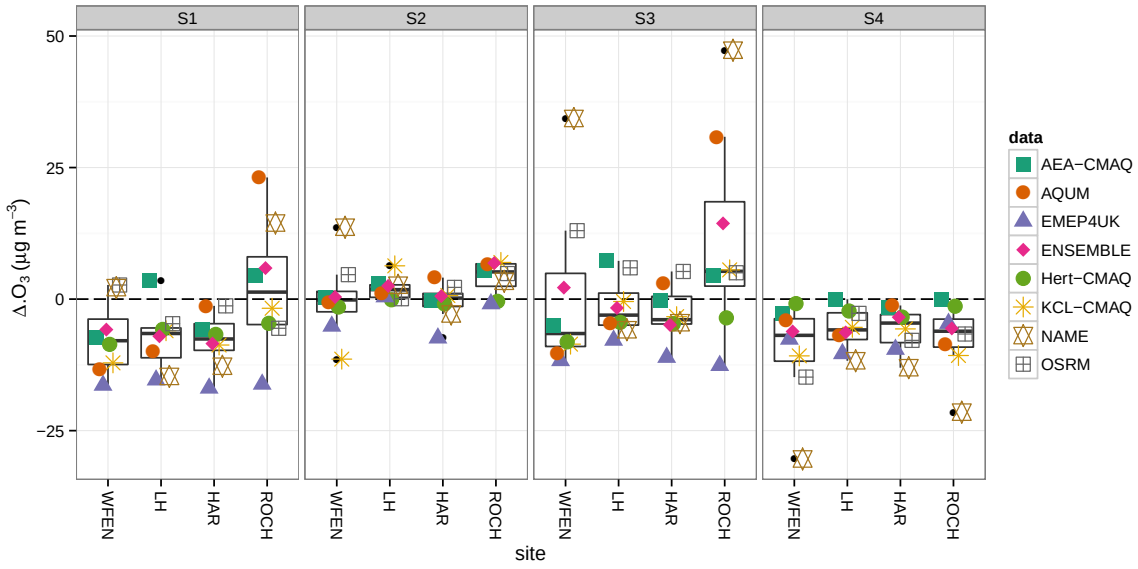

FIGURE 5.5: Effect on maximum rolling 8-hour O_3 concentration for each scenario and four rural sites. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

Considering the four sites together in Figure 5.5 there are two scenarios that the models tend to show consistency with: S1 (NO_x/VOC reduction in the UK and Europe) and S4 (VOC reduction in the UK and Europe). Nevertheless, even here the response for the models covers a wide range of O_3 change. It is clear from Figure 5.5 that the models are most uncertain when predicting the effect of NO_x only reductions, as seen for S3. In this case the response in O_3 can be both large and opposite in sign. In the case of

the Rochester site however, its proximity to London is likely to be important — where reductions in NO_x have a direct titration effect.

A selection of four urban sites has been considered in Figure 5.6. For these sites it would be expected that those scenarios that reduce NO_x emissions would perhaps tend to increase concentrations of O_3. Considering CLL2 (London Bloomsbury, the site embedded in the highest NO_x emissions) it can be seen that S2 most clearly results in the most significant increase in O_3 concentrations (between 1 and 9 µg m^{-3}). This increase is also seen to diminish as the location moves to outer urban areas (KC1 and LON6). However, even in these four urban areas a reduction in NO_x and VOCs across the UK and Europe is still predicted to lead to reductions in O_3 concentration. The other scenario that could lead to increases in urban O_3 is S3 where NO_x alone is reduced in the UK and Europe. However, these increases are only apparent in central and inner London (CLL2 and KC1).

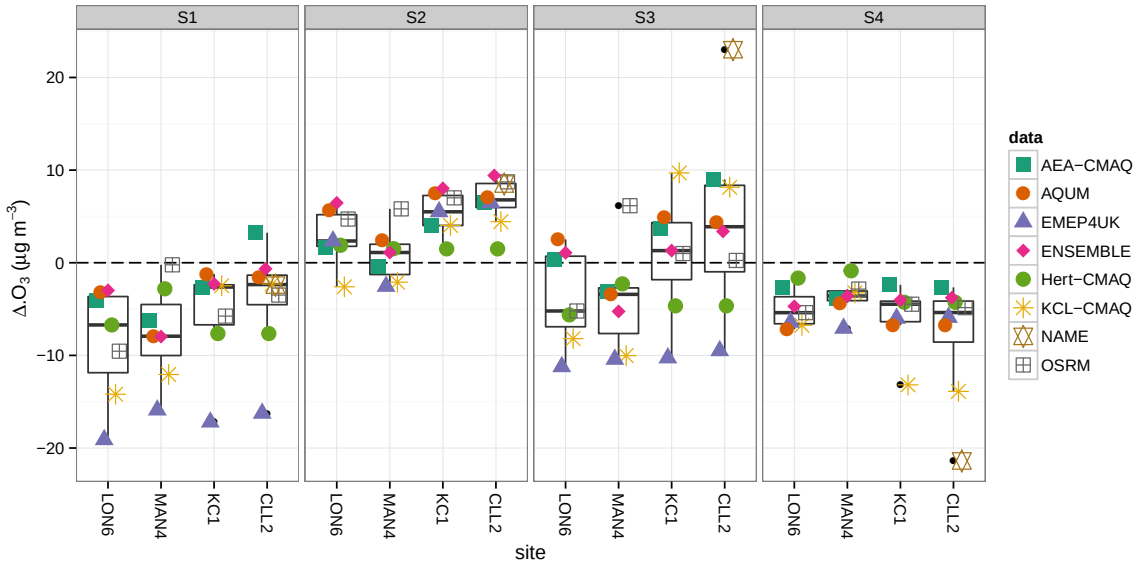

FIGURE 5.6: Effect on maximum rolling 8-hour O_3 concentration for each scenario and four urban sites. Negative values show a *reduction* (improvement) in O_3 concentration. The box and whisker plots help to show where the distribution is centred.

An important issue is the extent to which the models are NO_x or VOC sensitive i.e. whether it is best to reduce NO_x or VOC emissions to reduce O_3 concentrations. Scenarios S3 (30% reduction in NO_x emissions in the UK + Europe) and S4 (30% reduction in VOC emissions in the UK + Europe) can help answer this question. Of particular concern are the higher concentrations of O_3. For this reason consideration has been given to the daily maximum rolling 8-hour mean concentration, although other metrics could also be considered. Similar to Derwent et al. (2013), a day is considered NO_x sensitive if reducing emissions of NO_x gives a greater reduction in O_3 concentration compared with a reduction in VOCs.

Again, to make the analysis tractable July 2006 has been considered. Figure 5.7 shows the results for the models at the Harwell site, indicating whether each day is either NO_x or VOC sensitive. It is clear from Figure 5.7 that the models offer a wide range of results from mostly VOC sensitive (AQUM and NAME) to mostly NO_x sensitive (Hertfordshire-CMAQ). These results indicate that O_3 concentration in July at this site are sensitively balanced between NO_x and VOC sensitivity and that no consensus can be

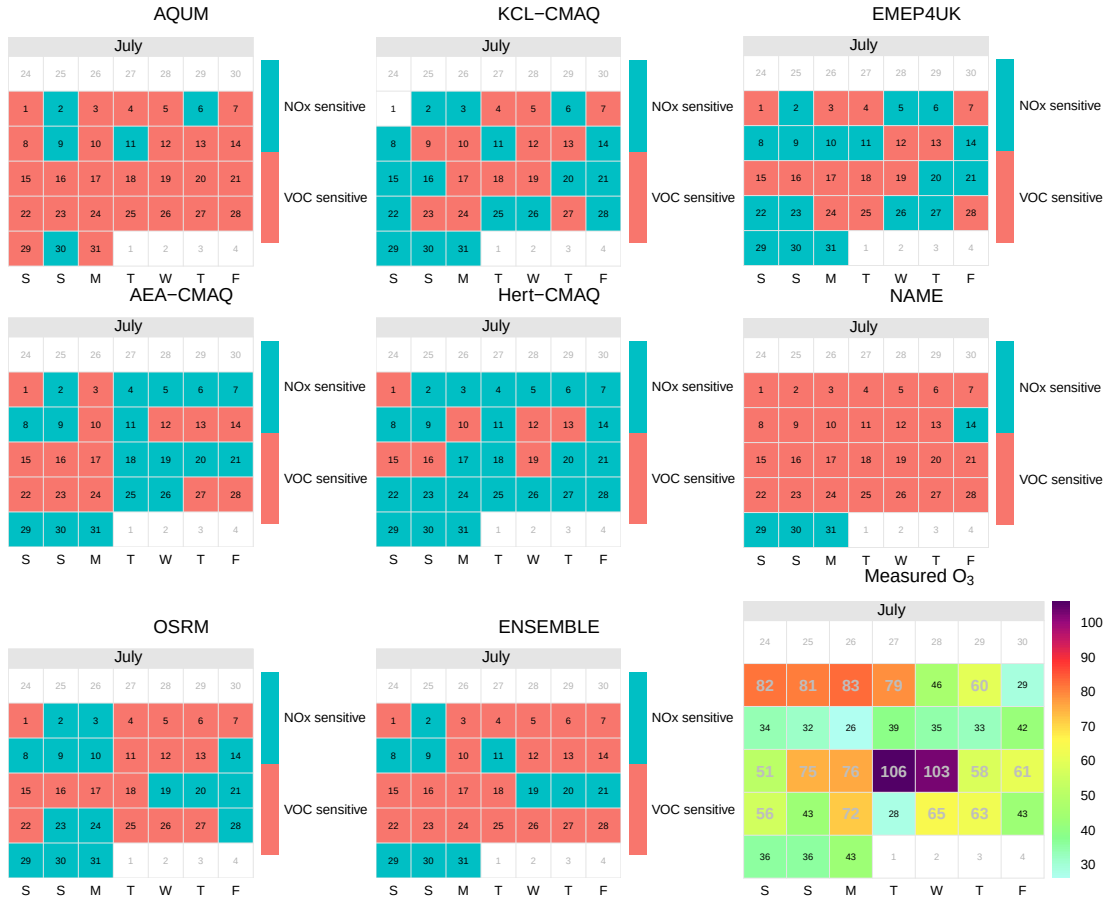

FIGURE 5.7: Plots to show whether the models consider a day in July at Harwell to be sensitive to reductions in NO_x or VOCs. NO_x-sensitive means that a reduction in NO_x emissions results in a greater decrease in O_3 compared with a reduction in VOCs and vice-versa. Also shown are the daily maximum rolling 8-hour mean concentrations of O_3. These results relate to the effect on the daily maximum of the rolling 8-hour mean O_3 concentration.

reached as to which view is more likely to be right. There is only agreement on one of the days (30th July), but 6 out of the 7 models evaluated do indicate the last 3 days of July were NO_x sensitive.

It should be noted however that O_3-NO_x-VOC sensitivity at individual locations and for specific events are often very uncertain and the these results reflect that point (Sillman 1995; Sillman and He 2002). The sensitivity will be very dependent on both emission assumptions (e.g. adequacy of absolute emission estimates and VOC speciation) but also the whole chemistry-transport system. It is therefore perhaps not very surprising that the models do not agree well in this respect. However, from a policy perspective these findings are important because it cannot be said with certainty whether reducing NO_x or VOCs is the most beneficial approach for reducing peak O_3 concentrations.

As discussed above, the results are Harwell seem to be finely balanced, which likely reflects its rural location but which is also influenced by relatively local emissions of NO_x. By taking a wider view of more sites more of a consensus can be reached. Figure 5.8 shows the number of days in July that are considered to be NO_x sensitive or VOC sensitive, split by model and site type. In this plot purple shading shows sites that are more NO_x sensitive and orange shading sites that are more VOC sensitive. This plot highlights several issues. There is general agreement that there are more NO_x-sensitive days at the

rural sites such as Aston Hill, Yarner Wood and Mace Head. For these types of sites the models show that most or all days are NO_x sensitive. There is also agreement at the urban background sites where the situation is 'NO_x saturated' where a reduction in NO_x concentration would increase O_3 concentration and in these cases the days in July are dominated by VOC-sensitive conditions. At intermediate sites (such as Harwell), there tends to be much more spread in the model predictions as noted above.

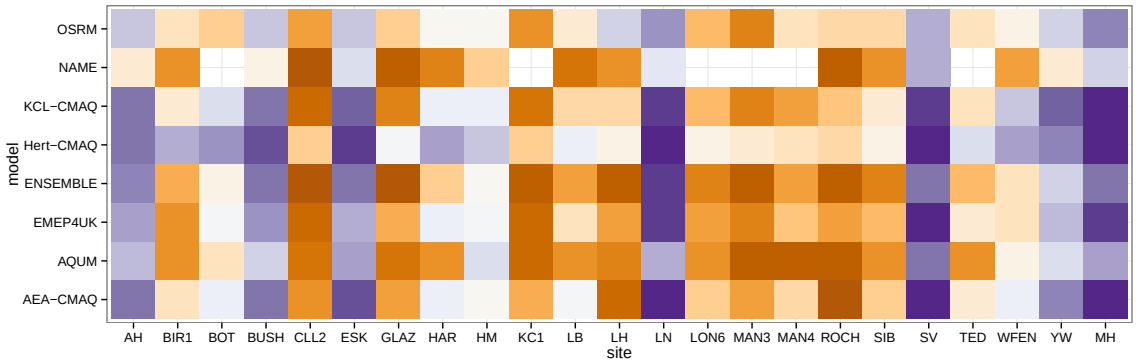

FIGURE 5.8: Plot showing the number of days in July 2006 that the models consider to be NO_x-sensitive or VOC sensitive. Purple colours show days that are more NO_x-sensitive and orange colours show days that are more VOC-sensitive. These results relate to the effect on the daily maximum of the rolling 8-hour mean O_3 concentration.

It is also useful to map the data shown in Figure 5.8 by site and model to consider the spatial distribution of NO_x sensitive days — shown in Figure 5.9 for rural sites. Now it is possible to see that the models tend to show a south-east, north-west gradient, with the south east of England being more VOC-sensitive and the north-west of the UK and Ireland being more NO_x sensitive. It is also clear from Figure 5.9 that AQUM, NAME and EMEP-4-UK predict that the south-east of England is more VOC-sensitive than the other models. The models are therefore reasonably consistent in identifying the spatial variation in NO_x and VOC sensitivity, but as noted previously, in intermediate environments there can be a wide range of model responses.

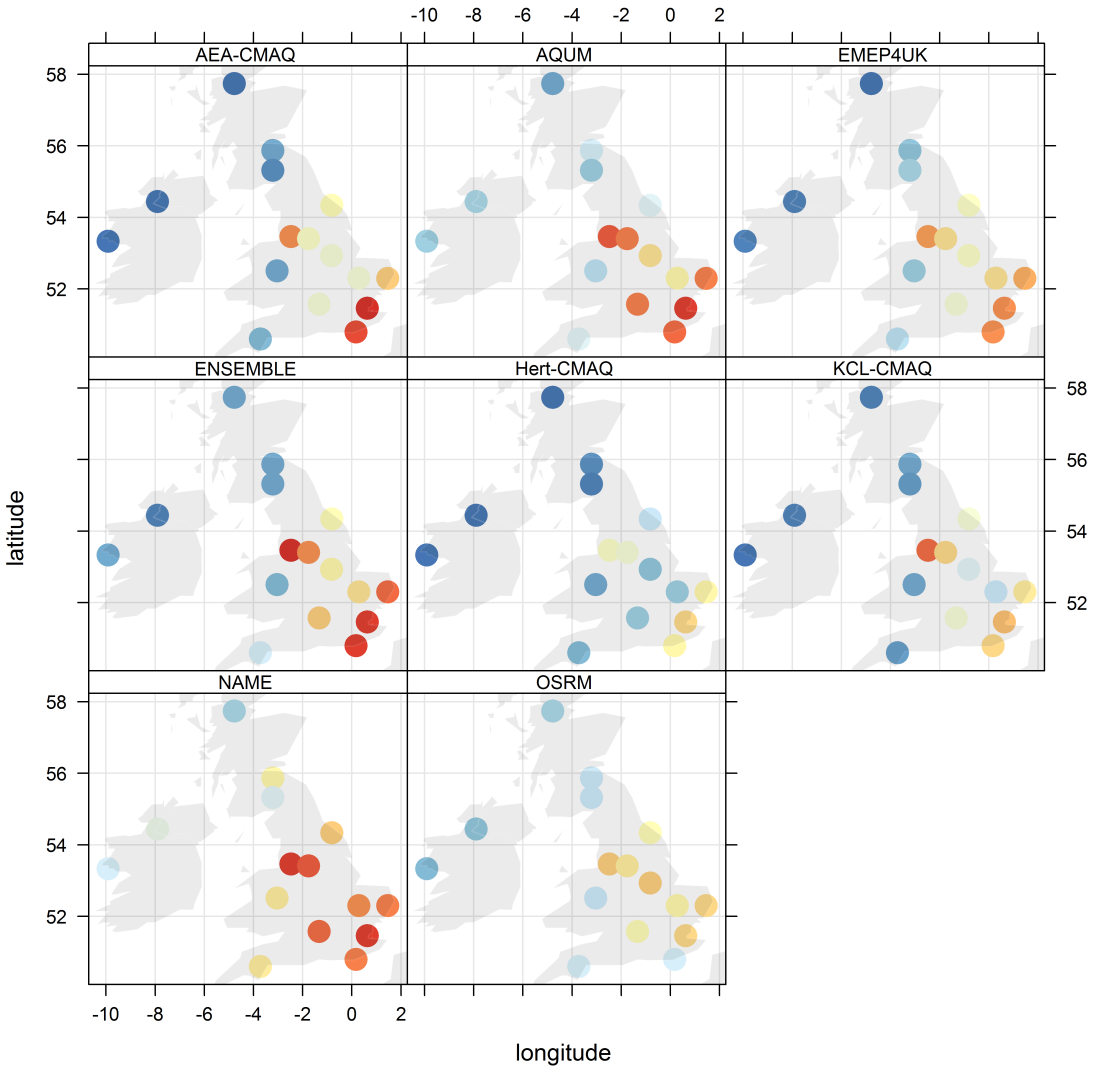

FIGURE 5.9: Maps showing the NO_x-VOC sensitivity in July 2006. Blue shading shows sites that are considered to be more NO_x sensitive, red shows sites that are more VOC-sensitive with yellow being intermediate between NO_x and VOC sensitive. These results relate to the effect on the daily maximum of the rolling 8-hour mean O_3 concentration.

Following the approach of Derwent et al. (2012) consideration has been given to whether UK or UK + European emissions control is most effective in reducing O_3 concentrations. Again, consideration has been given to the daily maximum of the rolling 8-hour means in July. Comparisons between S1 (30% reduction in UK + Europe NO_x and VOC emissions) and S2 (30% reduction in UK only NO_x and VOC emissions) allows an approximate indication of whether UK or European emissions control is most effective at reducing peak O_3 concentrations. A simple approach has been taken where if S1 results in more of s reduction in O_3 compared with S2 it is labelled as a 'European' day i.e. European emissions control is more effective than UK only control.

Similar to the VOC-NO_x sensitive analysis, the results have been presented in a grid showing the number of days in July where European emissions control is more effective than UK control. These results are shown in Figure 5.10. It is clear from Figure 5.10 that the models highlight the clear benefit to UK + European control for peak O_3 concen-

trations compared with UK only control. Most models show that almost every day in July benefits more from UK + European control. However, the NAME model does tend to show that UK only control would be more effective on more days than most other models. There is also an indication in Figure 5.10 that UK only control has more of an effect for sites in the south-east of England (Lullington Heath, Sibton and Rochester).

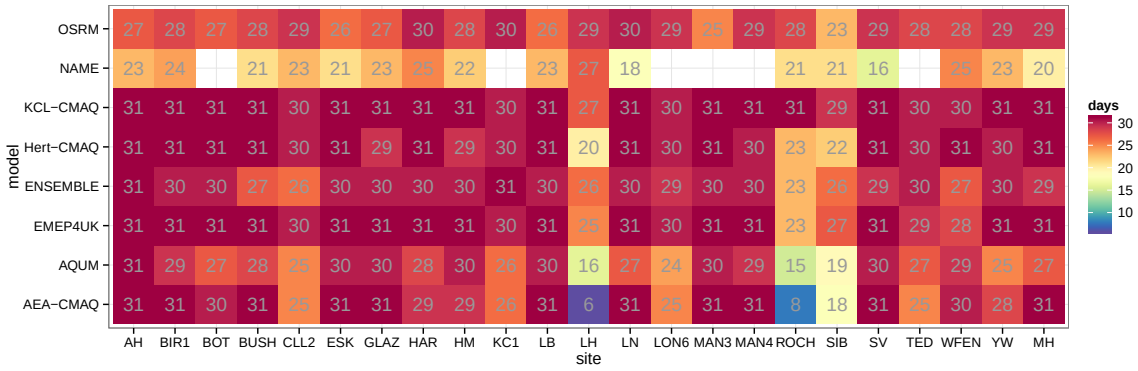

FIGURE 5.10: Plot showing the number of days in July 2006 that the models consider it to be more effective to reduce UK + European emissions of NO_x and VOCs compared with reductions only in the UK. These results relate to the effect on the daily maximum of the rolling 8-hour mean O_3 concentration.

The NO_x-VOC sensitivity issues can be further explored by considering the linkage with the emissions estimates shown in Table 1.3. The area assumed to be Europe in each model differs and for this reason the NO_x/VOC ratio is considered rather than absolute emission estimates. Figure 5.11 shows the NO_x/VOC emission ratio plotted against the number of days in July where each model considers a day to be more NO_x sensitive for Harwell. It is clear from Figure 5.11 that as the NO_x/VOC ratio increases that there is a tendency for the number of NO_x-sensitive days to decrease. Models (or more specifically, the assumptions used in the models) such as Hertfordshire-CMAQ assume much higher biogenic VOC emissions than most other models. At the other extreme is the NAME model that assumes zero biogenic emissions — as shown in Table 1.3.

The results shown in Figure 5.11 follow expectations in that when NO_x emissions are varied in a model with high VOC emissions the conditions are much more NO_x sensitive rather than VOC sensitive. These models have a larger 'buffer' of VOC making them less sensitive to changes in anthropogenic emissions of VOC. It is also clear from Figure 5.11 that these effects can be large; with models showing a wide range of responses from mostly VOC to mostly NO_x sensitive. In other words it seems that much of the variation in the responses of the models to changes in VOC and NO_x emissions is driven by emission assumptions — and in particular the assumptions related to biogenic VOCs at the European scale.

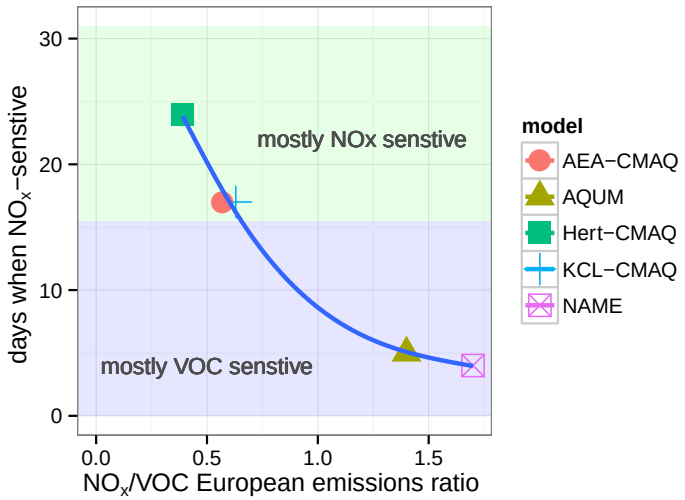

FIGURE 5.11: Plot showing the NO_x/VOC emissions ratio based on European emission against the number of days a model considers the conditions at Harwell to be NO_x-sensitive. These results relate to the effect on the daily maximum of the rolling 8-hour mean O_3 concentration. Only models that were able to provide European anthropogenic and biogenic emissions are included (some groups could not supply biogenic emission estimates.)

6 Meteorological data analysis

The Phase 1 regional model analysis did not give any consideration to the meteorological performance of the models. Details of Met Office surface meteorological stations are given in Table 6.1 and Figure 6.1. Note that most of these surface met stations are coastal, which may have important consequences for comparisons with modelled output e.g. due to sea breezes.

The models produced output for all 11 receptors shown in Table 6.1, but the periods covered by them differed. For example, some models produced hourly output for the whole of 2006 whereas others considered only June and July or more limited periods. For this reason the analysis separately considers individual model performance.

TABLE 6.1: Locations of Met Office surface met stations.

Site	latitude	longitude
Dunkeswell	50.86	-3.24
Weybourne	52.95	1.12
Pembrey Sands	51.71	-4.37
Church Fenton	53.84	-1.20
Edinburgh	55.93	-3.34
Lossiemouth	57.71	-3.32
Benson	51.62	-1.10
Herstmonceux	50.89	0.32
Crosby	53.50	-3.06
Heathrow	51.48	-0.45
Aldergrove	54.66	-6.22

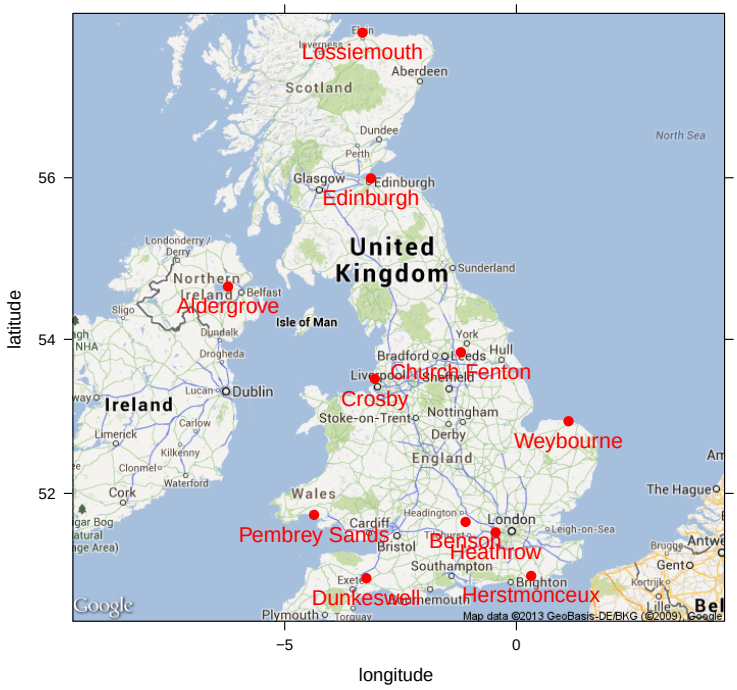

FIGURE 6.1: Map showing the locations of the 11 Met Office surface meteorological stations.

6.1 Wind speed comparsions

In this section brief consideration is given to the comparison of wind speed distributions from each model and how they compare with observed values. There are a very large number of potential comparisons that could be undertaken but these have been limited to providing an overview of key meteorological variables including wind speed, wind direction and temperature. While these sites are in general well-located and should represent conditions over wider areas, there is likely to be some local influences that the models do not capture. For example, at the Herstmonceux site all models tend to over estimate wind speeds.

Figure 6.2 shows the wind speed box and whisker plot for all models at Heathrow, Church Fenton, Lossiemouth and Weybourne, as an example of model performance.

Focusing on July 2006 when all models have predictions available, the following model performance statistics are calculated across all surface met sites. Overall AQUM and NAME produce the best predictions, whereas the performance of EMEP is less good.

TABLE 6.2: Summary statistics for model wind speed performance across all sites for July 2006.

data.mod	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
KCL-CMAQ	8041	0.81	-0.36	1.28	-0.10	0.34	1.72	0.66	0.26
AQUM	8041	0.81	-0.03	1.31	-0.01	0.35	1.69	0.66	0.25
NAME	8041	0.80	-0.04	1.31	-0.01	0.35	1.72	0.65	0.24
Hert-CMAQ	8041	0.78	0.19	1.39	0.05	0.37	1.79	0.64	0.20
AEA-CMAQ	8063	0.76	0.54	1.54	0.14	0.41	1.94	0.60	0.11
EMEP4UK	8041	0.68	-0.96	1.55	-0.26	0.42	2.04	0.61	0.11
WRF-Chem	7986	0.75	0.76	1.68	0.21	0.45	2.15	0.57	0.03

For all sites and months, the overall performance of the models is shown below. Several of the models perform in a very similar way: KCL-CMAQ, AQUM and NAME. The

performance of Hertfordshire-CMAQ, AEA-CMAQ and WRF-Chem is less good for most metrics e.g. they have a higher RMSE and lower correlation coefficient.

TABLE 6.3: Summary statistics for model wind speed performance across all sites for 2006.

data.mod	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
KCL-CMAQ	94465	0.86	-0.15	1.37	-0.03	0.29	1.79	0.80	0.40
AQUM	94465	0.87	-0.00	1.37	-0.00	0.29	1.82	0.79	0.40
NAME	94465	0.86	0.13	1.40	0.03	0.30	1.87	0.78	0.39
EMEP4UK	94103	0.80	-0.41	1.55	-0.09	0.33	2.03	0.76	0.32
Hert-CMAQ	94454	0.80	0.97	1.84	0.21	0.39	2.39	0.75	0.20
AEA-CMAQ	94307	0.77	1.23	2.02	0.26	0.43	2.59	0.72	0.12
WRF-Chem	11612	0.77	0.70	1.62	0.18	0.42	2.07	0.66	0.12

A summary of the wind speed performance over consistent time periods (July 2006) can be shown using a Taylor Diagram (Taylor 2001). In Figure 6.3 the model performance is compared on a site by site basis and the data normalised by the standard deviation of the observed values. The NAME and AQUM models show the best performance with measurements, with correlations between 0.7–0.8 and a variability closer to the measurements than most other models.

A more detailed consideration of wind speed performance can be gained by plotting the conditional quantiles as shown in Figure 6.4. Note that in this Figure the WRF-Chem model only considers part of June and July. However, overall all models show good performance, with good coverage of the wind speed range and with quantile values that are generally close to the observed values. However, Figure 6.4 does highlight that AEA-CMAQ and Hertfordshire-CMAQ tend to overestimate wind speeds, which is seen by the position of the red median line and the histograms.

The diurnal variation in wind speed is also a good indicator of model performance and

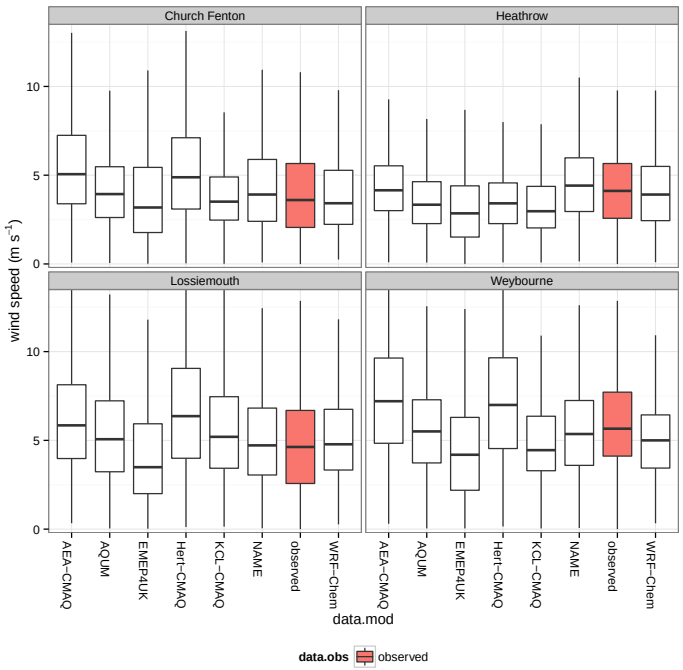

FIGURE 6.2: Box and whisker plot showing the modelled and observed wind speed distributions at Heathrow, Church Fenton, Lossiemouth and Weybourne.

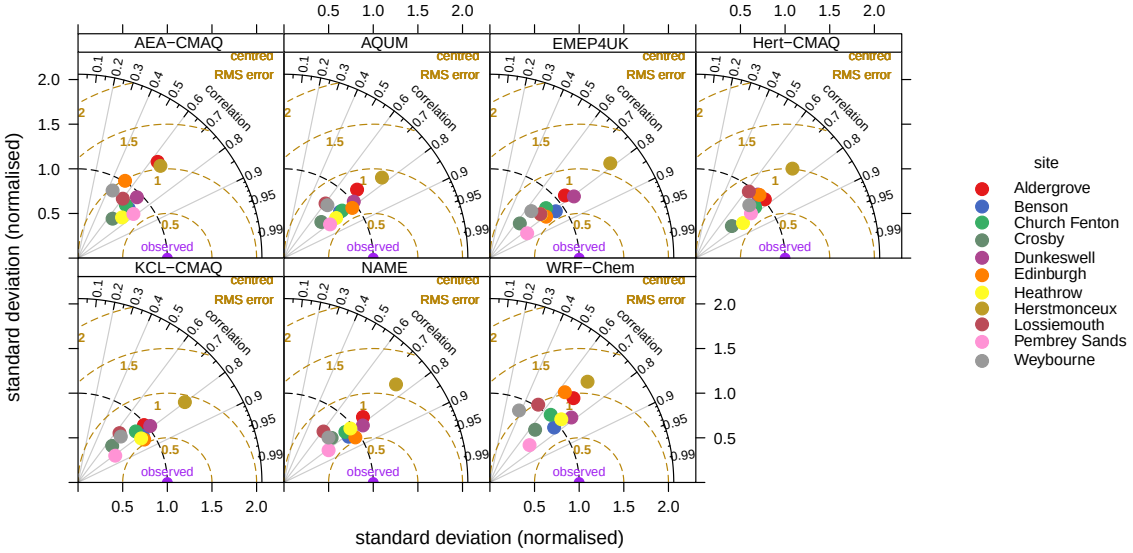

FIGURE 6.3: Taylor Diagram of model wind speed performance in July 2006.

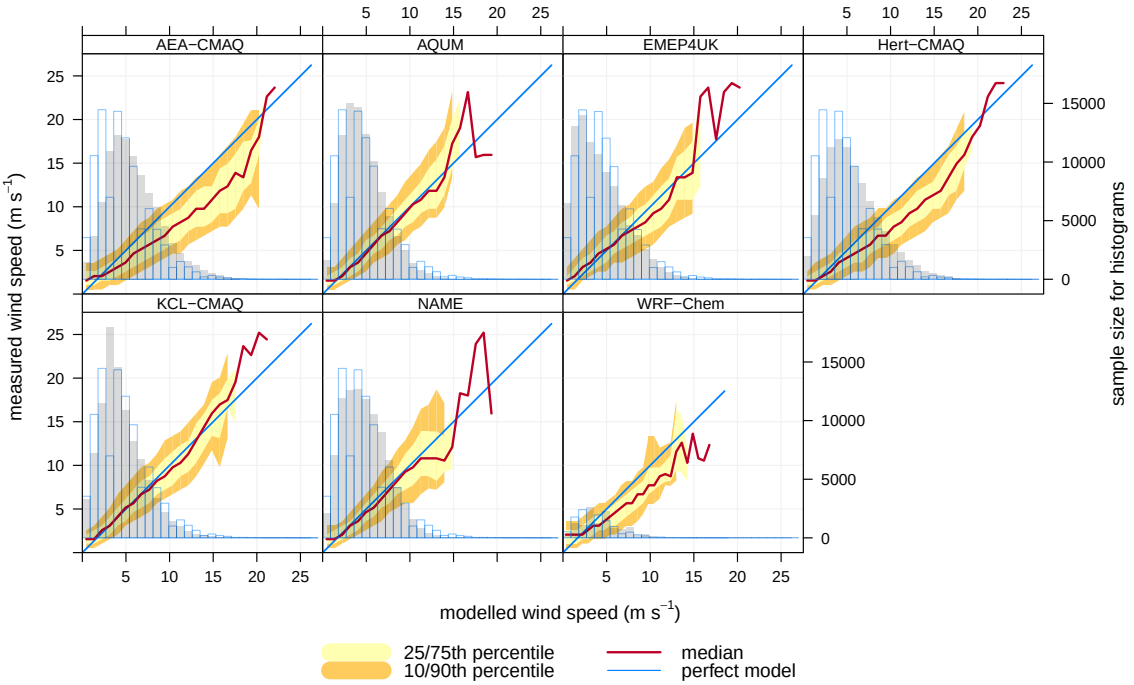

FIGURE 6.4: Conditional quantile plot of wind speed performance. The blue line shows the results for a perfect model. In this case the observations cover a range from 0 to about 25 m s^{-1}. The red line shows the median value of the predictions. The shading shows the predicted quantile intervals i.e. the 25/75th and the 10/90th. A perfect model would lie on the blue line and have a very narrow spread. There is still some spread because even for a perfect model a specific quantile interval will contain a range of values. However, for the number of bins used in this plot the spread will be very narrow. Finally, the shaded histogram shows the counts of predicted values and the histogram shown by the blue line shows the counts for the observed values.

helps to reveal whether models adequately capture the variation in time of day correctly. For example, wind speeds that are too high at night may indicate the model generates an atmosphere that is too unstable and therefore dilutes primary emissions too much and will also affect deposition processes. Figure 6.5 shows the diurnal variations of wind speed and how they compare by model with observed wind speeds. Most models capture the diurnal variation well (particularly the magnitude), although there is a tendency to predict the peak wind speed too early in the day (typically around 12pm). Two of the models do less well in capturing the amplitude of the diurnal wind speed: AEA-CMAQ and Hertfordshire-CMAQ. Both these models clearly overestimate the magnitude of overnight wind speeds.

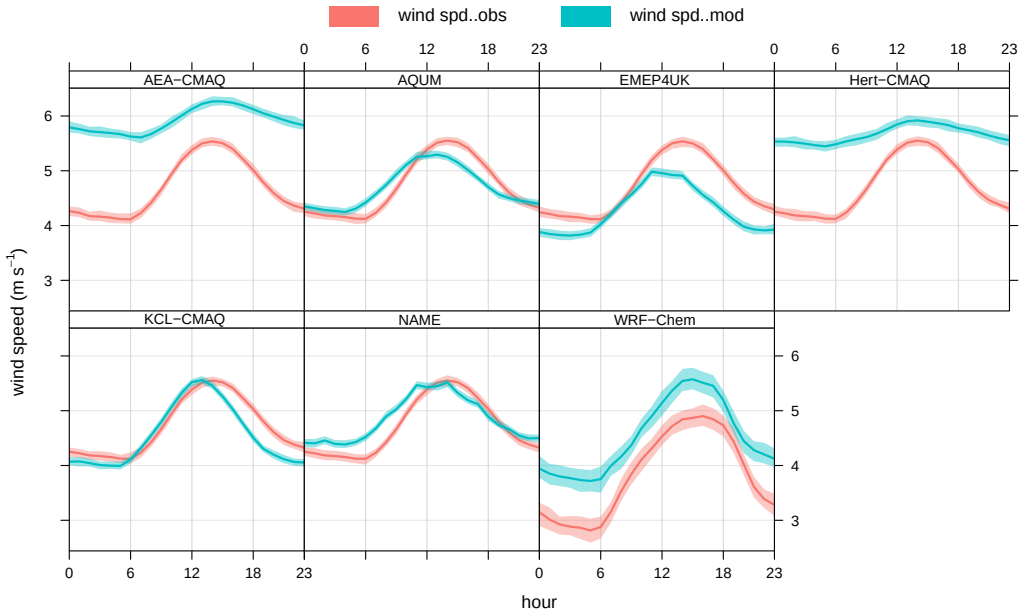

FIGURE 6.5: Diurnal variation in wind speeds predicted across all Met Office surface sites.

6.2 Wind direction comparisons

Wind direction is more tricky to compare than other variables such as wind speed because of the difficulty in presenting angular data. However, it is useful to consider the bias in wind direction, as shown in Figure 6.6. In this plot the modelled − observed wind direction is plotted for Heathrow as an example. There is a tendency for most models to show a positive bias, which also tends to be seen at most other sites. Considering the AEA-CMAQ as an example, there is on average a 20.8 degrees positive bias in wind direction. Also shown in Figure 6.6 is an indication of whether the wind speed tends to be over or underestimated. At this site at least, there is a tendency for the wind speed to be underestimated for AQUM and KCL-CMAQ.

Across all meteorological sites AEA-CMAQ and KCL-CMAQ show a similar positive bias of ≈20 degrees. AQUM and NAME show a slight positive bias of 7 to 12 degrees. Hertfordshire-CMAQ only has a bias of 4 degrees. The analysis also shows that the spread in wind directions for the models tends to be quite narrow and the most important limitation of wind direction predictions in the positive bias shown by several of the models. Note that there appears to be an issue with the EMEP-4-UK wind direction results — likely due to post-processing issues.

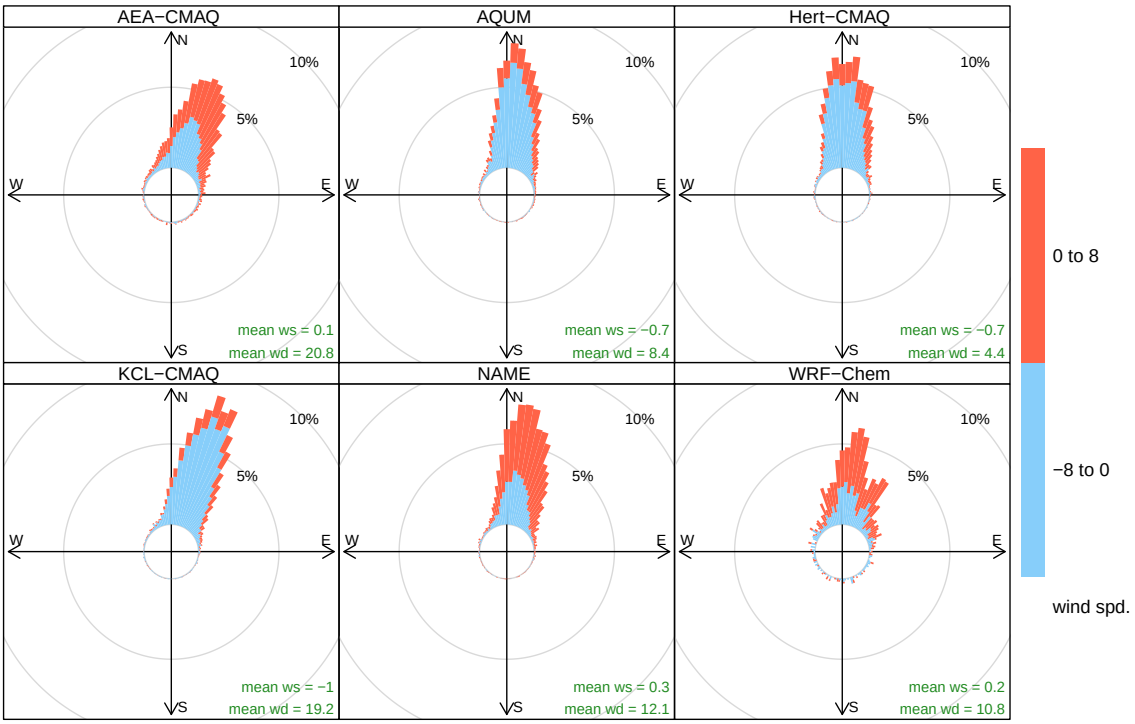

FIGURE 6.6: Wind rose biases comparing observations with model predictions at Heathrow for 2006. The plots show the bias in wind direction (modelled − observed) and wind speed differences.

6.3 Ambient temperature comparsions

The model performance with respect to ambient temperature predictions is generally very good with high correlations, low RMS error and similar variability. However, the models do not tend to capture the temperature variation at the Weybourne site as well as other sites. The results are shown in Figure 6.7 and Figure 6.8.

A summary of the overall model performance for temperature is shown below for July 2006.

TABLE 6.4: Summary statistics for model ambient temperature performance across all sites for July 2006.

data.mod	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
Hert-CMAQ	8131	1.00	-0.07	1.63	-0.00	0.09	2.13	0.88	0.53
AQUM	8131	1.00	0.25	1.65	0.01	0.09	2.21	0.89	0.53
NAME	8131	1.00	-0.80	1.71	-0.04	0.09	2.18	0.89	0.51
KCL-CMAQ	8131	1.00	-0.81	2.11	-0.04	0.11	2.77	0.83	0.40
WRF-Chem	8076	0.99	-1.30	2.14	-0.07	0.11	2.83	0.85	0.39
AEA-CMAQ	8153	1.00	-1.19	2.22	-0.06	0.12	2.95	0.80	0.36
EMEP4UK	8131	1.00	-1.98	2.81	-0.11	0.15	3.52	0.80	0.20

The statistics for all months and sites are:

The diurnal variations in temperature are shown in Figure 6.9, which shows that most models capture the diurnal variation in ambient temperature very well. Nevertheless,

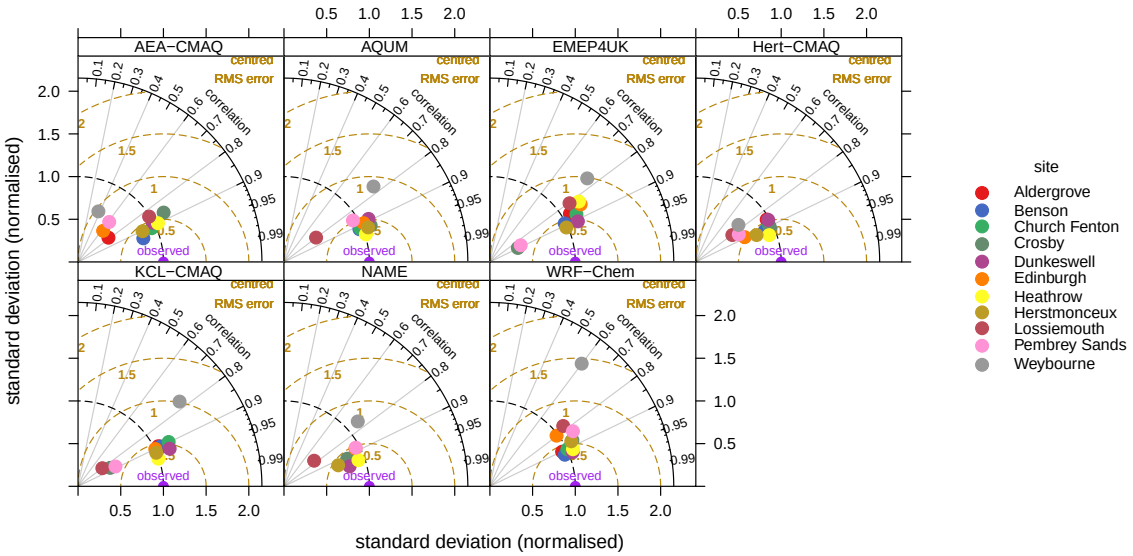

FIGURE 6.7: Taylor Diagram of model ambient temperature performance in July 2006.

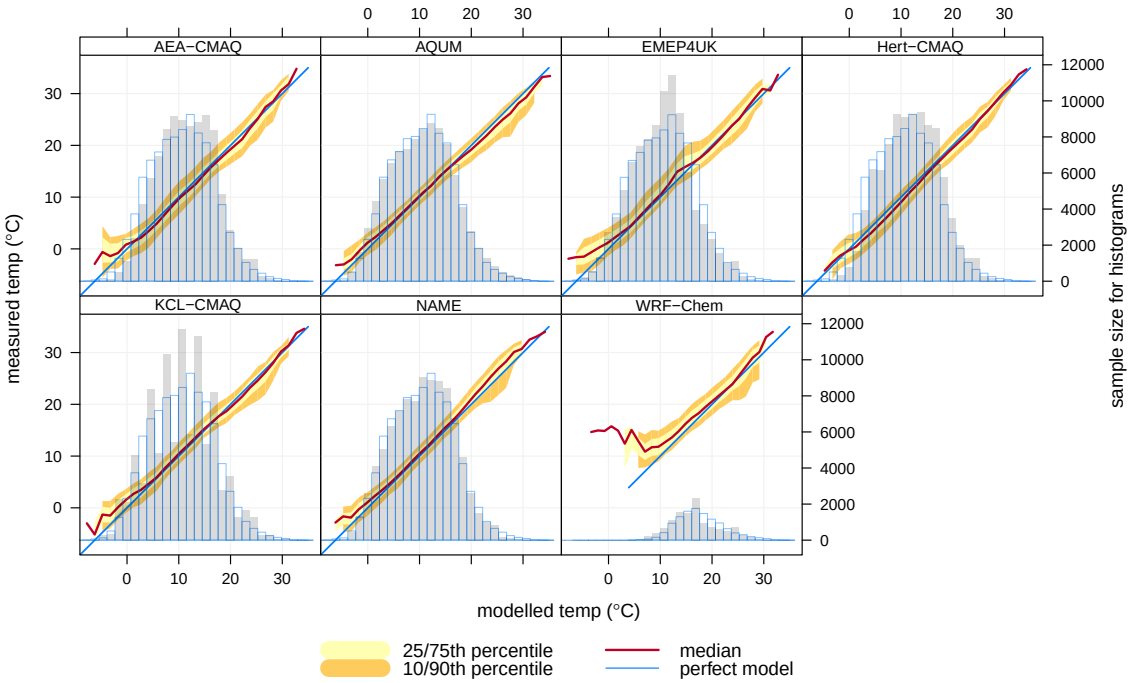

FIGURE 6.8: Conditional quantile plot of ambient temperature performance. The blue line shows the results for a perfect model. In this case the observations cover a range from 0 to about 25 m s^{-1}. The red line shows the median value of the predictions. The shading shows the predicted quantile intervals i.e. the 25/75th and the 10/90th. A perfect model would lie on the blue line and have a very narrow spread. There is still some spread because even for a perfect model a specific quantile interval will contain a range of values. However, for the number of bins used in this plot the spread will be very narrow. Finally, the shaded histogram shows the counts of predicted values and the histogram shown by the blue line shows the counts for the observed values.

TABLE 6.5: Summary statistics for model ambient temperature performance across all sites for 2006.

data.mod	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
AQUM	95178	0.92	-0.07	1.22	-0.01	0.12	1.63	0.97	0.75
NAME	95178	0.92	-0.45	1.31	-0.04	0.12	1.72	0.96	0.74
KCL-CMAQ	95178	0.90	-0.30	1.50	-0.03	0.14	2.00	0.95	0.70
Hert-CMAQ	95167	0.91	0.81	1.59	0.08	0.15	2.12	0.95	0.68
EMEP4UK	94816	0.90	-0.80	1.82	-0.08	0.17	2.39	0.93	0.63
AEA-CMAQ	95020	0.90	0.42	1.87	0.04	0.18	2.48	0.92	0.62
WRF-Chem	11664	0.99	-1.16	2.00	-0.07	0.11	2.63	0.87	0.44

EMEP-4-UK does tend to underestimate nighttime temperatures and Hertfordshire-CMAQ and AEA-CMAQ overestimate nighttime temperatures. Overall, the performance of the models for temperature is better than that for other important variables such as wind speed.

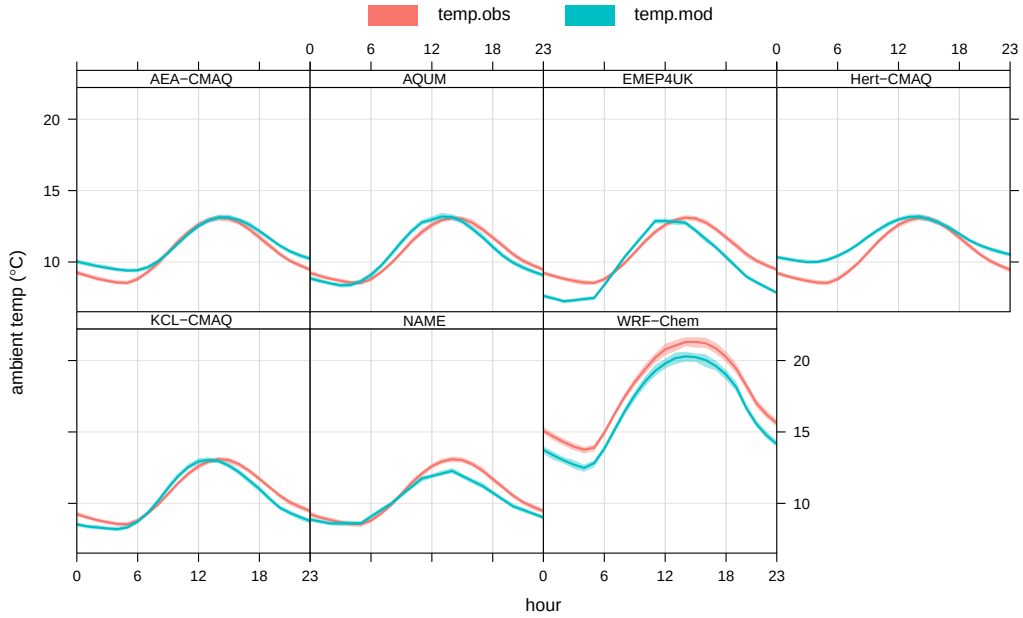

FIGURE 6.9: Diurnal variation in ambient temperature predicted across all Met Office surface sites.

6.4 Relative humidity comparisons

The overall statistics for all months and sites are shown below.

TABLE 6.6: Summary statistics for model relative humidity performance across all sites for July 2006.

data.mod	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
NAME	8128	1.00	-0.19	8.30	-0.00	0.11	10.79	0.81	0.44
Hert-CMAQ	8128	1.00	-1.04	8.97	-0.01	0.12	11.56	0.77	0.39
KCL-CMAQ	8128	1.00	2.03	8.97	0.03	0.12	11.69	0.77	0.39
AEA-CMAQ	8150	1.00	2.87	9.31	0.04	0.13	12.22	0.77	0.37
AQUM	8128	0.99	-4.65	10.20	-0.06	0.14	13.54	0.77	0.31
WRF-Chem	8073	0.99	9.87	12.33	0.13	0.17	15.64	0.73	0.17
EMEP4UK	8128	0.97	15.29	16.81	0.21	0.23	21.25	0.60	-0.14

TABLE 6.7: Summary statistics for model relative humidity performance across all sites in 2006.

data.mod	n	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE
NAME	94658	1.00	-0.20	6.69	-0.00	0.08	8.76	0.80	0.40
AQUM	94658	1.00	-0.78	7.23	-0.01	0.09	9.72	0.79	0.36
KCL-CMAQ	94658	1.00	2.22	7.41	0.03	0.09	9.64	0.76	0.34
AEA-CMAQ	94500	1.00	0.85	8.23	0.01	0.10	10.81	0.69	0.27
Hert-CMAQ	94647	1.00	-4.38	8.90	-0.05	0.11	11.19	0.70	0.21
WRF-Chem	11661	0.99	9.92	12.36	0.13	0.17	15.68	0.71	0.12
EMEP4UK	94296	0.99	7.38	10.02	0.09	0.12	13.57	0.61	0.11

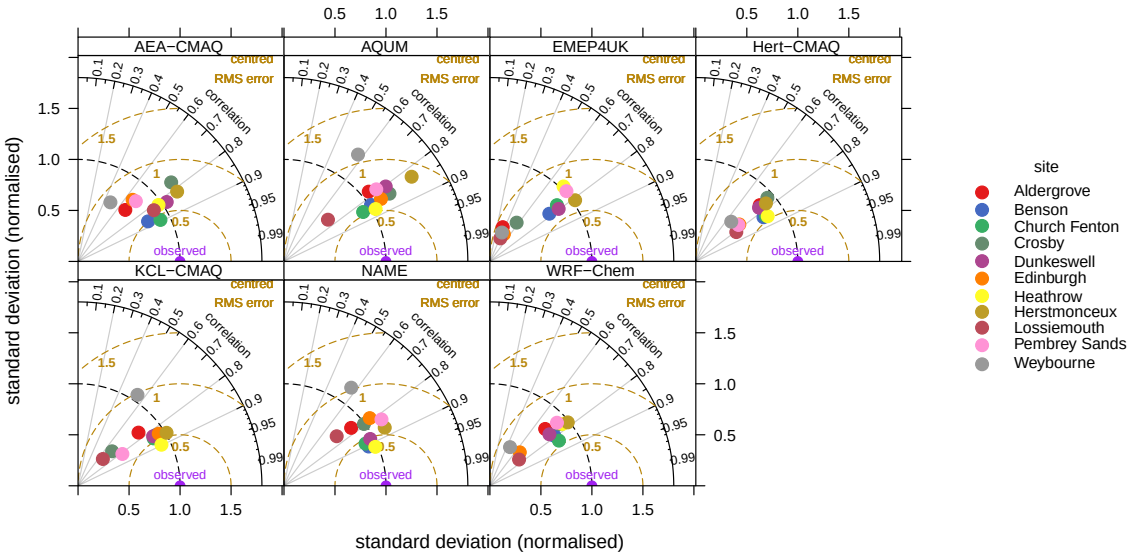

FIGURE 6.10: Taylor Diagram of model relative humidity performance in July 2006.

The conditional quantile plot shown in Figure 6.11 shows that NAME, AQUM and KCL-CMAQ agree best with measured values. The WRF-Chem results indicate that the model tends to predict higher values of RH than the measurements, while the EMEP results look somewhat unusual (some conversion problem?).

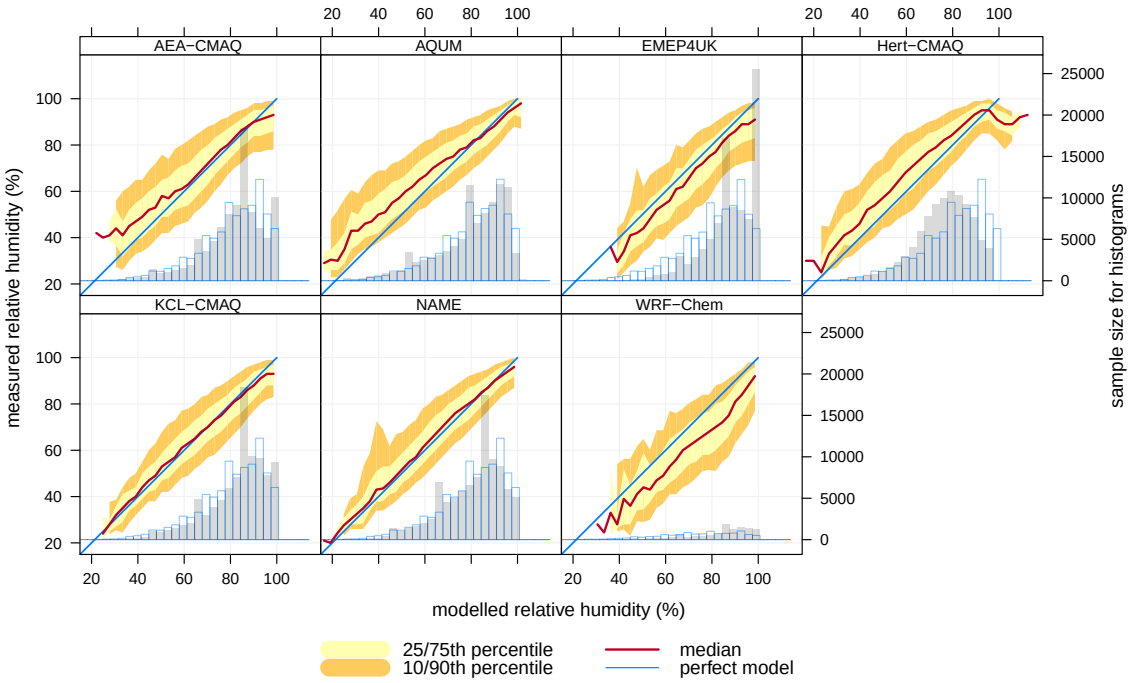

FIGURE 6.11: Conditional quantile plot of relative humidity performance for all sites and month. The blue line shows the results for a perfect model. The red line shows the median value of the predictions. The shading shows the predicted quantile intervals i.e. the 25/75th and the 10/90th. A perfect model would lie on the blue line and have a very narrow spread. There is still some spread because even for a perfect model a specific quantile interval will contain a range of values. However, for the number of bins used in this plot the spread will be very narrow. Finally, the shaded histogram shows the counts of predicted values and the histogram shown by the blue line shows the counts for the observed values.

6.5 Boundary layer height comparisons

This section compares the boundary layer height (BLH) estimates from the models. It was not possible to obtain observed values of BLH and hence models are compared with one another. However, it is possible to consider whether O_3 predictions are affected by BLH predictions and this is also considered. The mean predictions of BLH across all receptors were as follows: AEA-CMAQ = 634, AQUM = 763, EMEP-4-UK = 866, KCL-CMAQ = 270, Hertfordshire-CMAQ = 686 and NAME = 572 m. There are therefore considerable differences in the BLH estimates i.e. a factor of 3.2 between the lowest to highest (KCL-CMAQ to EMEP-4-UK). Such large differences in the models do not appear to translate through to clear differences in predicted concentrations.

Figure 6.12 shows the diurnal variation in BLH by model and by season at the Harwell site. There is a relatively wide variation in BLH between the models and especially during the winter months. In particular, the KCL-CMAQ model has a much lower BLH than the other models.

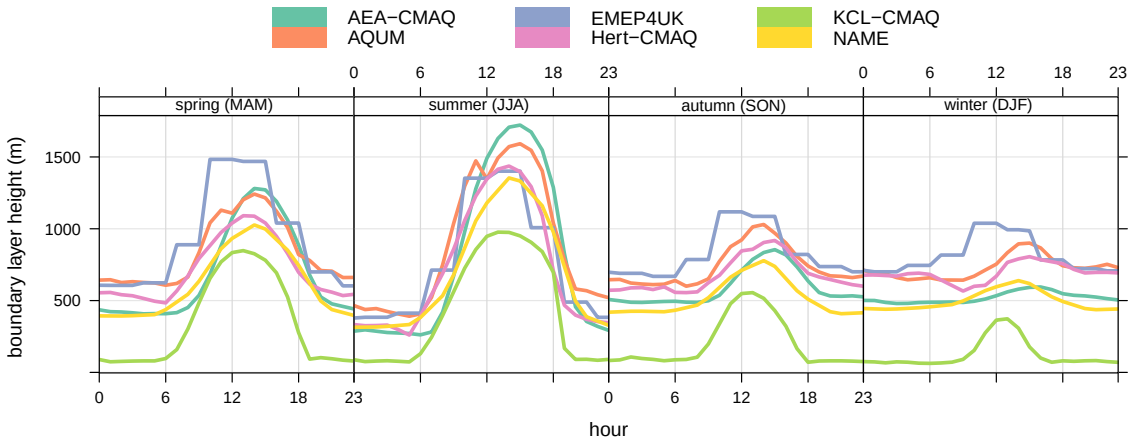

FIGURE 6.12: Diurnal variation in boundary layer height predicted at the Harwell site by season.

The BLH values seem to have very little influence on the predicted concentrations of all pollutants. For example, in Figure 6.13 the conditional quantile plot for NO_x is shown. AQUM and EMEP-4-UK are both shown to overestimate concentrations of NO_x— in particular AQUM. An overestimate of surface NO_x concentrations could be related to an underestimate of BLH, but this is not the case. For example, when the predicted value of NO_x is 50 ppb, the associated boundary layer height varies considerably.

It is difficult to reconcile this behaviour with the overall meteorological predictions. For example, AQUM and KCL-CMAQ predict the diurnal variation in wind speed and temperature rather well (see Figure 6.5 and Figure 6.9).

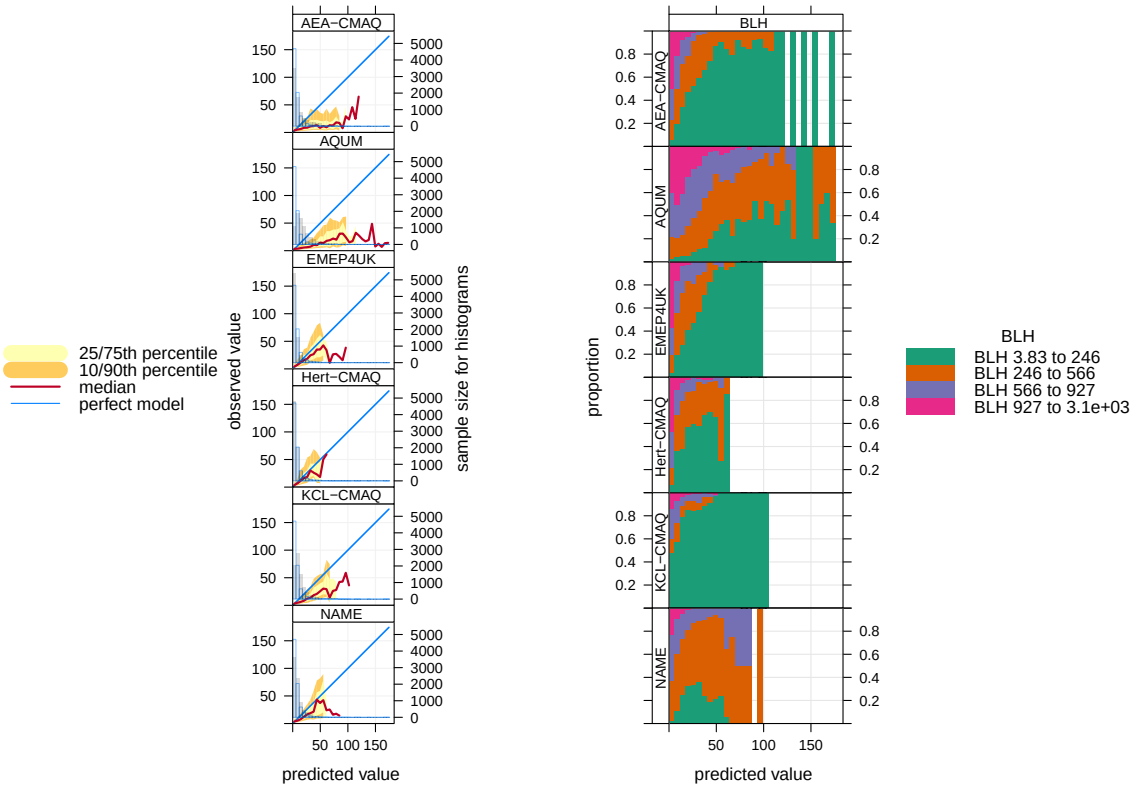

FIGURE 6.13: Conditional quantiles for NO_x concentrations at the Harwell site. The plot on the right shows the corresponding mean values of BLH.

References

Appel, K. W., C. Chemel, S. J. Roselle, X. V. Francis, R.-M. Hu, R. S. Sokhi, S. T. Rao and S. Galmarini (2012). "Examination of the Community Multiscale Air Quality (CMAQ) model performance over the North American and European domains". In: *Atmospheric Environment* 53.SI, 142–155. doi: {10.1016/j.atmosenv.2011.11.016} (cit. on p. 10).

AQEG (2004). *Nitrogen Dioxide in the United Kingdom*. Tech. rep. Report prepared by the Air Quality Expert Group for the Department for Environment, Food, Rural Affairs; Scottish Executive; Welsh Assembly Government; and Department of the Environment in Northern Ireland, March 2004. (cit. on p. 38).

– (2009). *Ozone in the United Kingdom*. Air Quality Expert Group. Report prepared by the Air Quality Expert Group for the Department for Environment, Food, Rural Affairs; Scottish Executive; Welsh Assembly Government; and Department of the Environment in Northern Ireland. (cit. on p. 23).

Beevers, S. D., E. Westmoreland, M. C. de Jong, M. L. Williams and D. C. Carslaw (2012). "Trends in NO_x and NO_2 emissions from road traffic in Great Britain". In: *Atmospheric Environment* 54, 107–116. doi: {10.1016/j.atmosenv.2012.02.028} (cit. on pp. 25, 42).

Carslaw, D. C., S. D. Beevers, J. E. Tate, E. Westmoreland and M. L. Williams (2011). "Recent evidence concerning higher NO_x emissions from passenger cars and light duty vehicles". In: *Atmospheric Environment* 45, pp. 7053–7063 (cit. on pp. 25, 42).

Carslaw, D. C. and K. Ropkins (2012). "openair — An R package for air quality data analysis". In: *Environmental Modelling & Software* 27–28, pp. 52–61. doi: 10.1016/j.envsoft.2011.09.008 (cit. on p. 17).

COMEAP (2011). *Review of the UK Air Quality Index: A report by the Committee on the Medical Effects of Air Pollutants* (cit. on p. 37).

Derwent, D., A. Fraser, J. Abbott, M. Jenkin, P. Willis and T. Murrells (2010). *Evaluating the Performance of Air Quality Models*. Issue 3/June 2010 (cit. on pp. 18, 76).

Dunlea, E. J., S. C. Herndon, D. D. Nelson, R. M. Volkamer, F. San Martini, P. M. Sheehy, M. S. Zahniser, J. H. Shorter, J. C. Wormhoudt, B. K. Lamb, E. J. Allwine, J. S. Gaffney, N. A. Marley, M. Grutter, C. Marquez, S. Blanco, B. Cardenas, A. Retama, C. R. Ramos Villegas, C. E. Kolb, L. T. Molina and M. J. Molina (2007). "Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment". In: *Atmospheric Chemistry and Physics* 7.10, pp. 2691–2704. doi: 10.5194/acp-7-2691-2007 (cit. on p. 38).

Ebinghaus, R., S. G. Jennings, H. H. Kock, R. G. Derwent, A. J. Manning and T. G. Spain (2011). "Decreasing trends in total gaseous mercury observations in baseline air at Mace Head, Ireland from 1996 to 2009". In: *Atmospheric Environment* 45.20, pp. 3475–3480. doi: 10.1016/j.atmosenv.2011.01.033 (cit. on p. 19).

Galmarini, S. and S. T. Rao (2011). "The AQMEII two-continent Regional Air Quality Model evaluation study: Fueling ideas with unprecedented data". In: *Atmospheric Environment* 45.14, pp. 2464–. doi: 10.1016/j.atmosenv.2011.03.025 (cit. on p. 10).

Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock and B. Eder (2005). "Fully coupled "online" chemistry within the WRF model". In:

Atmospheric Environment 39.37, 6957–6975. DOI: {10.1016/j.atmosenv.2005.04.027} (cit. on p. 15).

Hayman, G. D., J. Abbott, T. J. Davies, C. L. Thomson, M. E. Jenkin, R. Thetford and P. Fitzgerald (2010). "The ozone source-receptor model - A tool for UK ozone policy". In: *Atmospheric Environment* 44.34, 4283–4297. DOI: {10.1016/j.atmosenv.2010.06.013} (cit. on p. 14).

Legates, D. R. and G. J. McCabe Jr (1999). "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation". In: *Water Resources Research* 35.1, pp. 233–241 (cit. on pp. 18, 77).

Legates, D. R. and G. J. McCabe (2012). "A refined index of model performance: a rejoinder". In: *International Journal of Climatology* (cit. on pp. 18, 77).

Morrison, N. and H. Webster (2005). "An assessment of turbulence profiles in rural and urban environments using local measurements and numerical weather prediction results". In: *Boundary-Layer Meteorology* 115.2, 223–239. DOI: {10.1007/s10546-004-4422-8} (cit. on p. 14).

Pleim, J., A. Xiu, P. Finkelstein and T. Otte (2001). "A Coupled Land-Surface and Dry Deposition Model and Comparison to Field Measurements of Surface Heat, Moisture, and Ozone Fluxes". English. In: *Water, Air and Soil Pollution: Focus* 1.5-6, pp. 243–252. DOI: 10.1023/A:1013123725860 (cit. on p. 13).

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria (cit. on p. 17).

Rao, S. T., S. Galmarini and K. Puckett (2011). "Air Quality Model Evaluation International Initiative (AQMEII): Advancing the State of the Science in Regional Photochemical Modeling and Its Applications". In: *Bulletin of the American Meteorological Society* 92.1, pp. 23–30 (cit. on p. 10).

Saarikoski, S., M. Sillanpaa, M. Sofiev, H. Timonen, K. Saarnio, K. Teinela, A. Karppinen, J. Kukkonen and R. Hillamo (2007). "Chemical composition of aerosols during a major biomass burning episode over northern Europe in spring 2006: Experimental and modelling assessments". In: *Atmospheric Environment* 41.17, 3577–3589. DOI: {10.1016/j.atmonsenv.2006.12.053} (cit. on p. 11).

Savage, N. H., P. Agnew, L. S. Davis, C. Ordóñez, R. Thorpe, C. E. Johnson, F. M. O'Connor and M. Dalvi (2013). "Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation". In: *Geoscientific Model Development* 6.2, pp. 353–372. DOI: 10.5194/gmd-6-353-2013 (cit. on p. 12).

Sillman, S. and D. He (2002). "Some theoretical results concerning O_3-NO_x-VOC chemistry and NO_x-VOC indicators". In: *J. Geophys. Res* 107.10.1029 (cit. on p. 57).

Sillman, S. (1995). "The use of NO_y, H_2O_2, and HNO_3 as indicators for ozone-NO_x-hydrocarbon sensitivity in urban locations". In: *Journal of Geophysical Research* 100.14, pp. 175–188 (cit. on pp. 5, 52, 53, 57).

Simpson, D., A. Benedictow, H. Berge, R. Bergström, L. D. Emberson, H. Fagerli, C. R. Flechard, G. D. Hayman, M. Gauss, J. E. Jonson, M. E. Jenkin, A. Nyíri, C. Richter, V. S. Semeena, S. Tsyro, J.-P. Tuovinen, Á. Valdebenito and P. Wind (2012). "The EMEP MSC-W chemical transport model ndash; technical description". In: *Atmospheric Chemistry and Physics* 12.16, pp. 7825–7865. DOI: 10.5194/acp-12-7825-2012 (cit. on p. 13).

Solazzo, E., R. Bianconi, R. Vautard, K. W. Appel, M. D. Moran, C. Hogrefe, B. Bessagnet, J. Brandt, J. H. Christensen, C. Chemel, I. Coll, H. Denier Van Der

Gon, J. Ferreira, R. Forkel, X. V. Francis, G. Grell, P. Grossi, A. B. Hansen, A. Jeričević, L. Kraljević, A. I. Miranda, U. Nopmongcol, G. Pirovano, M. Prank, A. Riccio, K. N. Sartelet, M. Schaap, J. D. Silver, R. S. Sokhi, J. Vira, J. Werhahn, R. Wolke, G. Yarwood, J. Zhang, S. T. Rao and S. Galmarini (2012). "Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII". In: *Atmospheric Environment* 53, pp. 60–74. doi: 10.1016/j.atmosenv.2012.01.003 (cit. on p. 15).

Stephenson, D. B. (2000). "Use of the "odds ratio" for diagnosing forecast skill". In: *Weather and Forecasting* 15.2, 221–232 (cit. on p. 37).

Swall, J. L. and K. M. Foley (2009). "The impact of spatial correlation and incommensurability on model evaluation". In: *Atmospheric Environment* 43.6, pp. 1204–1217. doi: DOI:10.1016/j.atmosenv.2008.10.057 (cit. on p. 55).

Taylor, K. E. (2001). "Summarizing multiple aspects of model performance in a single diagram". In: *Journal of Geophysical Research* 106.D7, pp. 7183–7192 (cit. on p. 63).

Vieno, M., A. J. Dore, D. S. Stevenson, R. Doherty, M. R. Heal, S. Reis, S. Hallsworth, L. Tarrason, P. Wind, D. Fowler, D. Simpson and M. A. Sutton (2010). "Modelling surface ozone during the 2003 heat-wave in the UK". In: *Atmospheric Chemistry and Physics* 10.16, pp. 7963–7978. doi: 10.5194/acp-10-7963-2010 (cit. on p. 13).

Witham, C. and A. Manning (2007). "Impacts of Russian biomass burning on UK air quality". In: *Atmospheric Environment* 41.37, 8075–8090. doi: {10.1016/j.atinosenv.2007.06.058} (cit. on p. 11).

Xie, Y. (2013a). *Dynamic Documents with R and knitr*. ISBN 978-1482203530. Chapman and Hall/CRC (cit. on p. 17).

– (2013b). *knitr: A general-purpose package for dynamic report generation in R*. R package version 1.1 (cit. on p. 17).

A Additional plots

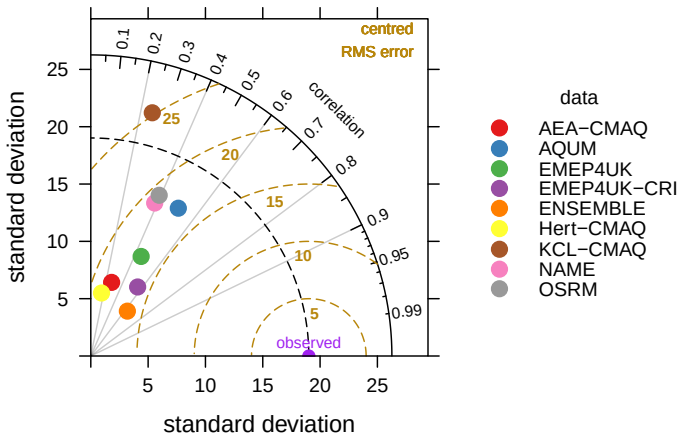

FIGURE A.2: Taylor Diagram of the number of days with a daily maximum of running 8-hour means O_3 >100 µg m^{-3}.

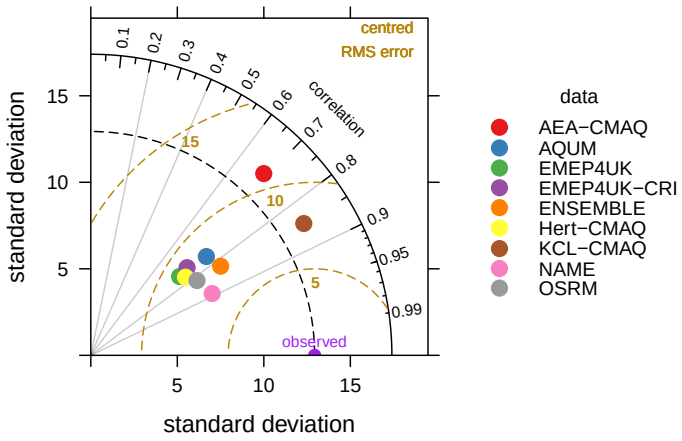

FIGURE A.1: Taylor Diagram of annual mean predicted and measured O_3 concentrations.

B Model performance evaluation statistics

There are a very wide range of evaluation statistics that can be used to assess model performance. There is, however, no single statistic that encapsulates all aspects of interest. For this reason it is useful to consider several performance statistics and also to understand the sort of information or insight they might provide. The performance statistics used here have mostly been guided by Derwent et al. 2010.

In the following definitions, O_i represents the ith observed value and M_i represents the ith modelled value for a total of n observations.

Fraction of predictions within a factor or two, *FAC2*

The fraction of modelled values within a factor of two of the observed values are the fraction of model predictions that satisfy:

$$0.5 \leq \frac{M_i}{O_i} \leq 2.0 \tag{1}$$

Mean bias, *MB*

The mean bias provides a good indication of the mean over or under estimate of predictions. Mean bias in the same units as the quantities being considered.

$$MB = \frac{1}{n} \sum_{i=1}^{N} M_i - O_i \tag{2}$$

Mean Gross Error, *MGE*

The mean gross error provides a good indication of the mean error regardless of whether it is an over or underestimate. Mean gross error is in the same units as the quantities being considered.

$$MGE = \frac{1}{n} \sum_{i=1}^{N} |M_i - O_i| \tag{3}$$

Normalised mean bias, *NMB*

The normalised mean bias is useful for comparing pollutants that cover different concentration scales and the mean bias is normalised by dividing by the observed concentration.

$$NMB = \frac{\sum_{i=1}^{n} M_i - O_i}{\sum_{i=1}^{n} O_i} \tag{4}$$

Normalised mean gross error, *NMGE*

The normalised mean gross error further ignores whether a prediction is an over or underestimate.

$$NMGE = \frac{\sum_{i=1}^{n} |M_i - O_i|}{\sum_{i=1}^{n} O_i} \tag{5}$$

Root mean squared error, *RMSE*

The *RMSE* is a commonly used statistic that provides a good overall measure of how close modelled values are to predicted values.

$$RMSE = \left(\frac{\sum_{i=1}^{n} (M_i - O_i)^2}{n} \right)^{1/2} \tag{6}$$

Correlation coefficient, *r*

The (Pearson) correlation coefficient is a measure of the strength of the linear relationship between two variables. If there is perfect linear relationship with positive slope between the two variables, $r = 1$. If there is a perfect linear relationship with negative slope between the two variables $r = -1$. A correlation coefficient of 0 means that there is no linear relationship between the variables.

$$r = \frac{1}{(n-1)} \sum_{i=1}^{n} \left(\frac{M_i - \overline{M}}{\sigma_M} \right) \left(\frac{O_i - \overline{O}}{\sigma_O} \right) \tag{7}$$

Coefficient of Efficiency, *COE*

The *Coefficient of Efficiency* based on Legates and McCabe (2012) and Legates and McCabe Jr (1999). There have been many suggestions for measuring model performance over the years, but the *COE* is a simple formulation which is easy to interpret.

A perfect model has a $COE = 1$. As noted by Legates and McCabe although the *COE* has no lower bound, a value of $COE = 0.0$ has a fundamental meaning. It implies that the model is no more able to predict the observed values than does the observed mean. Therefore, since the model can explain no more of the variation in the observed values than can the observed mean, such a model can have no predictive advantage.

For negative values of *COE*, the model is less effective than the observed mean in predicting the variation in the observations.

$$COE = 1.0 - \frac{\sum_{i=1}^{n} |M_i - O_i|}{\sum_{i=1}^{n} |O_i - \overline{O}|} \tag{8}$$

C Code used to produce outputs

This section contains all the code that produces the Figures and Tables. The data used in this report are stored in a dropbox folder owned by David Carslaw. Users of the data can contact David Carslaw at mailto:david.carslaw@kcl.ac.uk to request access. The data itself can then be downloaded to a suitable local location. Note that in order for the code to run, users would need to change the path of some of the file locations. There is also a ReadMe.txt file that should be read.

```
## ----loadPackages----------------------------------------------------------
## load all packages we need
library(openair)
library(latticeExtra)
library(plyr)
library(ggplot2)
library(xtable)
library(lattice)
library(reshape2)
## set ggplot plots to B&W
theme_set(theme_bw())
trellis.par.set(list(strip.background = list(col = "white")))
## make sure wd is correctly set
#setwd("~/Projects/modelEvaluation/phase2/regional/")
## ----loadO3----------------------------------------------------------------
load("./data/o3Meas.RData")
load("./data/KCLMod.RData")
load("./data/OSRMMod.RData")
load("./data/HertMod.RData")
load("./data/EMEP4UKMod.RData")
load("./data/EMEP4UK.EXPMod.RData")
load("./data/NAMEMod.RData")
load("./data/AEAMod.RData")
load("./data/AQUMMod.RData")
load("./data/WRFChemMod.RData")
## met data
load("./data/metDataProc.RData")
allMet <- rbind.fill(emepMet, NAMEmet, AQUMmet, WRFChem,  KCLmet, AEAmet,
                     Hertmet)
allMet <- merge(metMeas, allMet, by = c("date", "site"), suffixes = c(".obs", ".mod"))
metLong <- rbind.fill(metMeas, emepMet, NAMEmet, AQUMmet, WRFChem, KCLmet, AEAmet, Hertmet)
## ----loadBLH---------------------------------------------------------------
load("./data/metDataBLH.RData")
BLH <- rbind.fill(AEA.BLH, KCL.BLH, Hert.BLH, AQUM.BLH, NAME.BLH, EMEP.BLH)
## ----makeEnsemble----------------------------------------------------------
## simple averging of everything to represent an ensemble
ENSEMBLE.base <- rbind.fill(EMEP.base, Hert.base, KCL.base, NAME.base, OSRM.base,
                  AEA.base, AQUM.base)
ENSEMBLE.base <- aggregate(subset(ENSEMBLE.base, select = -c(scenario, site, date, data)),
                     subset(ENSEMBLE.base, select = c(site, date)), mean, na.rm = TRUE)
ENSEMBLE.base$data <- "ENSEMBLE"
ENSEMBLE.S1 <- rbind.fill(EMEP.S1, Hert.S1, KCL.S1, NAME.S1, OSRM.S1,
                  AEA.S1, AQUM.S1)
ENSEMBLE.S1 <- aggregate(subset(ENSEMBLE.S1, select = -c(scenario, site, date, data)),
                     subset(ENSEMBLE.S1, select = c(site, date)), mean, na.rm = TRUE)
ENSEMBLE.S1$data <- "ENSEMBLE"
ENSEMBLE.S2 <- rbind.fill(EMEP.S2, Hert.S2, KCL.S2, NAME.S2, OSRM.S2,
                  AEA.S2, AQUM.S2)
ENSEMBLE.S2 <- aggregate(subset(ENSEMBLE.S2, select = -c(scenario, site, date, data)),
                     subset(ENSEMBLE.S2, select = c(site, date)), mean, na.rm = TRUE)
ENSEMBLE.S2$data <- "ENSEMBLE"
ENSEMBLE.S3 <- rbind.fill(EMEP.S3, Hert.S3, KCL.S3, NAME.S3, OSRM.S3,
                  AEA.S3, AQUM.S3)
ENSEMBLE.S3 <- aggregate(subset(ENSEMBLE.S3, select = -c(scenario, site, date, data)),
                     subset(ENSEMBLE.S3, select = c(site, date)), mean, na.rm = TRUE)
ENSEMBLE.S3$data <- "ENSEMBLE"
ENSEMBLE.S4 <- rbind.fill(EMEP.S4, Hert.S4, KCL.S4, NAME.S4, OSRM.S4,
                  AEA.S4, AQUM.S4)
ENSEMBLE.S4 <- aggregate(subset(ENSEMBLE.S4, select = -c(scenario, site, date, data)),
                     subset(ENSEMBLE.S4, select = c(site, date)), mean, na.rm = TRUE)
```

```
ENSEMBLE.S4$data <- "ENSEMBLE"
## ----loadAlloc-----------------------------------------------------------
alloc <- read.table("./data/MH_E_rawbasefit_o3_2006.txt", header =TRUE, skip=9)
alloc <- transform(alloc, date = ISOdatetime(YY, MM, DD, HH, min=0, sec=0, tz="GMT"))
alloc <- subset(alloc, select = c(date, B))
alloc$B <- as.character(alloc$B)
## ----clusterO3MH, w=6,h=5,results='hide',dev='png',dpi=600, out.width='0.6\\textwidth'----
## import back trajectories for Mace Head in 2006
traj <- importTraj("mh", 2006)
## perform cluuster analysis with an angle-based distance matrix
trajMH <- trajCluster(traj, method = "angle", n.cluster= 4)
## ----findClust,fig.keep='none'-------------------------------------------
allMH <- rbind.fill(subset(EMEP.base, site == "MH"),
                    subset(EMEP.EXP.base, site == "MH"),
                    subset(Hert.base, site == "MH"),
                    subset(KCL.base, site == "MH"),
                    subset(NAME.base, site == "MH"),
                    subset(OSRM.base, site == "MH"),
                    subset(AEA.base, site == "MH"),
                    subset(AQUM.base, site == "MH"),
                    subset(ENSEMBLE.base, site == "MH"))
## only need O3
allMH <- subset(allMH, select = c(date, data, O3))
measMH <- subset(o3Meas, site == "MH", select = c(date, O3))
## add Dick's allocation
allMH <- merge(allMH, alloc, by = "date")
measMH <- merge(measMH, alloc, by = "date")
## wide version for model statistics
allWide <- merge(allMH, measMH, by = c("date", "B"), suffix = c(".mod", ".obs"))
measMH[, "data"] <- "MEASURED"
allMH <- rbind.fill(allMH, measMH)
## don't need all trajectory points
allMHTraj <- merge(allMH, subset(trajMH, hour.inc == 0), by = "date")
allWideTraj <- merge(allWide, subset(trajMH, hour.inc == 0), by = "date")
## put measured values first
allMHTraj$data <- relevel(factor(allMHTraj$data), ref = "MEASURED")
## calculate temporal components
clustTv <- timeVariation(subset(allMHTraj, cluster %in% c("C2", "C3")),
                         pollu = "O3",
                         group = "cluster", type = "data",
                         ci = FALSE, col = "Set1", lwd = c(2, 2, 4, 2),
                         ylab = "ozone (ppb)")
## ----clusterMeans, results='hide',echo=FALSE, message=FALSE--------------
tmp <- ddply(subset(allMHTraj, select = c(data, cluster, O3)),
             .(cluster, data), numcolwise(mean), na.rm=TRUE)
tmp <- dcast(tmp, cluster ~ data)
## subract measured values first
id <- which(names(tmp) == "MEASURED")
tmp[, 2:11] <- tmp[2:11] - tmp[, id]
tmp <- melt(tmp)
tmp <- dcast(tmp, variable ~ cluster)
names(tmp)[1] <- "model"
## ----clusterMeansTable,results='asis',echo=FALSE-------------------------
print(xtable(tmp,
             caption = "Summary mean \\ozone concentrations by model and
cluster at Mace Head with measured values subracted (ppb).",
             label = "tab:modMeanClust", digits= 1),
      size= "small",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----Alloc, fig.keep='none'----------------------------------------------
## calculate temporal components
## rename allocation numbers
allMH$B[allMH$B == "1"] <- "Baseline"
allMH$B[allMH$B == "4"] <- "European"
allocDat <- subset(allMH, B %in% c("European", "Baseline"))
## put measured first
allocDat$data <- relevel(factor(allocDat$data), ref = "MEASURED")
allocTv <- timeVariation(allocDat, pollu = "O3",
                         group = "B", type = "data",
                         ci = FALSE,
```

```
                    col = rev(openColours("Set1", 2)),
                    ylab = "ozone (ppb)")
## ----plotClustMonth, w=6, h=6, fig.show='hold', out.width='0.49\\textwidth'----
plot(clustTv, subset = "month")
plot(allocTv, subset = "month")
## ----plotClustDiurnal, w=6, h=6,fig.show='hold', out.width='0.49\\textwidth'----
plot(clustTv, subset= "hour")
plot(allocTv, subset = "hour")
## ----modStatsClust,results='asis',echo=FALSE-----------------------------
tmp <- modStats(subset(allWideTraj, cluster == "C3"), mod = "O3.mod", obs = "O3.obs",
        type =c("cluster", "data"), rank.name = "data")
print(xtable(subset(tmp, select = -cluster),
             caption = "Summary statistics for Cluster 3 model performance.",
             label = "tab:modStatsClust"),
      size = "footnotesize",
      booktabs= TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----modStatsClustAlloc,results='asis',echo=FALSE------------------------
tmp <- modStats(subset(allWide, B == "1"), mod = "O3.mod", obs = "O3.obs",
        type =c("B", "data"), rank.name = "data")
print(xtable(subset(tmp, select = -B),
             caption = "Summary statistics for `baseline air' model performance using the more refined air-mass allocation metho
             label = "tab:modStatsClustAlloc"),
      size = "footnotesize",
      booktabs= TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----condQuantClust,w=10,h=6,out.width='1\\textwidth', results='hide'----
conditionalQuantile(subset(allWideTraj, cluster == "C3"), mod = "O3.mod",
                    obs = "O3.obs",
                    type = "data", xlab = "prediction o3 (ppb)",
                    ylab = "measured o3 (ppb)")
## ----siteTypes-----------------------------------------------------------
## useful to sub-select later
urban.sites <- c("CLL2", "BIR1", "KC1", "TED", "LON6", "MAN3", "MAN4")
rural.sites <- c("AH", "BOT", "BUSH", "ESK", "GLAZ", "HAR", "HM", "LB", "LH",
                 "LN", "MH", "ROCH", "SIB", "SV", "WFEN", "YW")
## ----combineBase---------------------------------------------------------
all.data <- rbind.fill(AEA.base, AQUM.base, EMEP.base, EMEP.EXP.base,
                       Hert.base, KCL.base, NAME.base,
                       OSRM.base, WRFChem.base, ENSEMBLE.base)
all.data <- merge(all.data, o3Meas, by = c("site", "date"),
                  suffixes=c(".mod", ".obs"))
all.data <- transform(all.data, NOx.obs = NO.obs + NO2.obs,
                      NOx.mod = NO.mod + NO2.mod,
                      OX.obs = NO2.obs + O3.obs, OX.mod = NO2.mod + O3.mod)
## index
all.data <- ddply(all.data, .(data), rollingMean, pollutant="O3.obs",
                  new.name = "O3.roll.obs")
all.data <- ddply(all.data, .(data), rollingMean, pollutant="O3.mod",
                  new.name = "O3.roll.mod")
all.data <- transform(all.data, index = cut(2 * O3.roll.obs,
                                breaks = c(0, 34, 66, 100, 121, 141, 160, 188,
                                                    214, 240, 500),
                                labels=c("Low.1", "Low.2", "Low.3",
                                "Moderate.4", "Moderate.5", "Moderate.6",
                                "High.7", "High.8",
                                "High.9", "Very High.10")))
## ----statsMeas-----------------------------------------------------------
statsMeas <- aqStats(transform(o3Meas, O3 = O3 * 2), pollutant = "O3")
## ----printStats,results='asis', cache=FALSE, fig.align='default', echo=FALSE----
print(xtable(subset(statsMeas,select = -c(pollutant, year)), digits = 0,
             caption = "Summary measured \\ozone statistics by site for 2006.
Concentration values are given in \\ug.",
             label = "tab:o3StatsMeas"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top",
      rotate.colnames = TRUE)
## ----summaryStat---------------------------------------------------------
```

```
## statistics summary
## for each model,  merge with measurements and calcuate summary stats
meta <- importMeta() ## AURN site details
statSite <- function (x) {
    ## calc AQ stats by site to allow latitude to be used in AOT40
    res <- aqStats(x, poll = "o3", latitude = meta$latitude[meta$code == x$site[1]])
    res
}
makeSumStat <- function(thedata) {
    ## first merge model data with observations
    group <- thedata$data[1]
    thedata <- merge(subset(thedata, select = c(date, site, O3)),
                     subset(o3Meas, select = c(date, site, O3)),
                     by = c("date", "site"), suffixes = c(".mod", ".obs"))
    ## calculate aq stats for measurements/model
    ## convert to ug/m3 for calculations
    obs <-  ddply(transform(thedata, o3 = 2 * O3.obs), .(site), statSite)
    mod <-  ddply(transform(thedata, o3 = 2 * O3.mod), .(site), statSite)
    thedata <- merge(obs, mod, by = "site", suffixes = c(".obs", ".mod"))
    thedata$data <- group
    thedata
}
## summary of key model stats, matching measurements on a per site basis
stats <- ldply(list(EMEP.base, EMEP.EXP.base, AQUM.base, KCL.base, AEA.base,
                    Hert.base, OSRM.base,
                    NAME.base, ENSEMBLE.base), makeSumStat)
## add site.type
stats$site.type[stats$site %in% rural.sites] <- "rural"
stats$site.type[stats$site %in% urban.sites] <- "urban"
## ----corPlot,w=6,h=6,out.width='0.5\\textwidth'-------------------------
corPlot(stats, poll=c("mean.obs", "maximum.obs", "max.daily.obs", "max.rolling.8.obs",
                "percentile.99.obs", "roll.8.O3.gt.100.obs","AOT40.obs"),
        col = c("white", "white", "grey"))
## ----annualOzone,w=8,h=4,out.width='1\\textwidth'------------------------
annStat <- modStats(stats, obs = "mean.obs", mod = "mean.mod",
                    type = c("site.type", "data"), rank = "data")
scatterPlot(annStat, x="NMB", y="NMGE", group = "data", type = "site.type",
            cex = 2, pch = c(15:19, 8, 11, 12), col = "Dark2", ref.x = 0,
            xlim = c(-0.5, 0.5), ylim = c(0, NA))
## ----modPerMeanSiteType, results='asis'---------------------------------
print(xtable(annStat,
             caption = "Summary statistics for annual mean \\ozone performance split by site type.",
             label = "tab:modStatsMeanO3SiteType"),
      size = "scriptsize",
      booktabs= TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----meanPointrange,w=8,h=5,out.width='0.8\\textwidth'-------------------
tmp <- with(stats, aggregate(mean.mod, list(site = site, type = site.type), range))
tmp <- data.frame(tmp[, c(1, 2)], as.data.frame(tmp[, 3]))
tmp[, "obs"] <- with(stats, aggregate(mean.obs, list(site = site, type = site.type), mean))[, 3]
names(tmp)[3:4] <- c("ymin", "ymax")
ggplot(tmp, aes(site, obs, ymin= ymin, ymax = ymax, col = type, shape = type)) +
    geom_pointrange(size = 1) +
    ylim(0, 90) + ylab(quickText("o3 (ug/m3)")) +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----modPerAOT40, results='asis'----------------------------------------
print(xtable(modStats(stats, obs = "AOT40.obs", mod = "AOT40.mod",
                      type = c("site.type", "data"), rank ="data"),
             caption = "Summary statistics for AOT40 \\ozone performance.",
             label = "tab:modStatsAOT40O3"),
      size = "scriptsize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----AOT40Ozone,w=8,h=4,out.width='1\\textwidth'------------------------
AOT40Stat <- modStats(stats, obs = "AOT40.obs", mod = "AOT40.mod",
                      type = c("site.type", "data"), rank = "data")
scatterPlot(AOT40Stat, x="NMB", y="NMGE", group = "data", type = "site.type",
            cex = 2, pch = c(15:19, 8, 11, 12), col = "Dark2", ref.x = 0,
            xlim = c(-0.6,  0.6), ylim = c(0, NA))
## ----AOT40Pointrange,w=8,h=5,out.width='0.8\\textwidth'-----------------
```

```
tmp <- with(stats, aggregate(AOT40.mod, list(site = site, type = site.type), range))
tmp <- data.frame(tmp[, c(1, 2)], as.data.frame(tmp[, 3]))
tmp[, "obs"] <- with(stats, aggregate(AOT40.obs, list(site = site, type = site.type), mean))[, 3]
names(tmp)[3:4] <- c("ymin", "ymax")
ggplot(tmp, aes(site, obs, ymin = ymin, ymax = ymax, col =type, shape = type)) +
    geom_pointrange(size = 1) +
    ylab(quickText("o3 (ug/m3).hours")) +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----modPerRoll, results='asis'----------------------------------------
print(xtable(modStats(stats, obs = "roll.8.O3.gt.100.obs", mod = "roll.8.O3.gt.100.mod",
                      type = c("site.type", "data"), rank ="data"),
                caption = "Summary statistics for number of days with a daily maximum
of running 8-hour means $>$100~\\ug.",
                label = "tab:modStatsDays100O3"),
          size = "scriptsize",
          booktabs = TRUE,
          include.rownames = FALSE,
          caption.placement = "top")
## ----roll8Ozone,w=8,h=4,out.width='1\\textwidth'------------------------
roll8Stat <- modStats(stats, obs = "roll.8.O3.gt.100.obs", mod = "roll.8.O3.gt.100.mod",
                      type = c("site.type", "data"), rank = "data")
scatterPlot(roll8Stat, x="NMB", y="NMGE", group = "data", type = "site.type",
            cex = 2, pch = c(15:19, 8, 11, 12), col = "Dark2", ref.x = 0,
            ylim = c(0, NA))
## ----roll8Pointrange,w=8,h=5,out.width='0.8\\textwidth'-----------------
tmp <- with(stats, aggregate(roll.8.O3.gt.100.mod, list(site = site, type = site.type), range))
tmp <- data.frame(tmp[, c(1, 2)], as.data.frame(tmp[, 3]))
tmp[, "obs"] <- with(stats, aggregate(mean.obs, list(site = site, type = site.type), mean))[, 3]
names(tmp)[3:4] <- c("ymin", "ymax")
ggplot(tmp, aes(site, obs, ymin =ymin, ymax = ymax, col = type, shape = type)) +
    geom_pointrange(size = 1) +
    ylim(0, 120) + ylab(quickText("days o3 is greater than threshold")) +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----statSenstivity----------------------------------------------------
## consider a +/- 10% change in base case EMEP concentrations
EMEP.minus <- transform(EMEP.base, O3 = O3 * 0.9)
EMEP.plus <- transform(EMEP.base, O3 = O3 * 1.1)
EMEP.minus$data <- "MINUS"
EMEP.plus$data <- "PLUS"
sens <- ldply(list(EMEP.base,EMEP.minus, EMEP.plus), makeSumStat)
## just need the one statistic
emep.sens <- subset(sens, select = c(site, roll.8.O3.gt.100.mod, data))
emep.sens <- melt(emep.sens)
emep.sens <- dcast(emep.sens, ... ~ data)
## ----emepSensFig,w=8,h=5,out.width='0.8\\textwidth'---------------------
ggplot(emep.sens, aes(site, EMEP4UK, ymin = MINUS,  ymax = PLUS)) +
    geom_pointrange(size = 1) + ylim(0, 110) +
    ylab(quickText("days o3 is greater than threshold")) +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----maxDailyOzone,w=8,h=4,out.width='1\\textwidth'---------------------
roll8Stat <- modStats(stats, obs = "max.daily.obs", mod = "max.daily.mod",
                      type = c("site.type", "data"), rank = "data")
scatterPlot(roll8Stat, x="NMB", y="NMGE", group = "data", type = "site.type",
            cex = 2, pch = c(15:19, 8, 11, 12), col = "Dark2", ref.x = 0,
            xlim = c(-0.3, 0.3), ylom = c(0, NA))
## ----modMax, results='asis'--------------------------------------------
print(xtable(modStats(stats, obs = "max.daily.obs", mod = "max.daily.mod",
                      type = c("site.type", "data"), rank= "data"),
             caption = "Summary statistics for the maximum daily mean \\ozone concentration.",
             label = "tab:modStatsMaxDailyO3"),
      size = "scriptsize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----maxDailyPointrange,w=8,h=5,out.width='0.8\\textwidth'--------------
tmp <- with(stats, aggregate(max.daily.mod, list(site = site, type = site.type), range))
tmp <- data.frame(tmp[, c(1, 2)], as.data.frame(tmp[, 3]))
tmp[, "obs"] <- with(stats, aggregate(max.daily.obs, list(site = site, type = site.type), mean))[, 3]
names(tmp)[3:4] <- c("ymin", "ymax")
ggplot(tmp, aes(site, obs, ymin = ymin, ymax = ymax, col = type, shape = type)) +
    geom_pointrange(size =1) +
    ylim(0, 200) + ylab(quickText("o3 (ug/m3)")) +
```

```
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----funDateMod----------------------------------------------------------
## function to make it easy to extract a site/months, combine with o3Meas
combineAll <- function(site = "HAR", month = 6:7){
    sub <- rbind.fill(AEA.base, AQUM.base, EMEP.base, Hert.base, KCL.base, NAME.base,
                      OSRM.base, WRFChem.base, ENSEMBLE.base)
    sub <- sub[sub$site %in% site, ]
    sub <- selectByDate(sub, month = month)
    sub <- merge(sub, o3Meas, by = c("site", "date"), suffixes = c(".mod", ".obs"))
    ## make total nox available
    sub <- transform(sub, NOx.obs = NOx, NOx.mod = NO2.mod + NO.mod,
                     OX.obs = O3.obs + NO2.obs, OX.mod = O3.mod + NO2.mod)
    sub
}
## ----KC1CondEval,w=12,h=10,fig.keep='last', out.width='0.8\\textwidth'----
tmp <-  merge(subset(all.data, site == "KC1"), subset(allMet, site == "Heathrow"),
              by.x = c("date", "data"), by.y = c("date", "data.mod"))
conditionalEval(tmp, obs= "O3.obs", mod = "O3.mod", var.obs=c("NOx.obs", "ws.obs"),
                var.mod =c("NOx.mod", "ws.mod"), statistic = "NMB", type = "data",
                var.names=c("nox", "wind speed"))
## ----LHclusterTraj,w=7,h=5,dev='png',dpi=600, out.width='0.6\\textwidth'----
traj <- importTraj("lh", 2006)
clusters <- trajCluster(traj, method ="Angle", n.cluster=6, col = "Set1")
## merge data but not non-full years
tmp <- merge(subset(all.data, !data %in% c("WRF-Chem") & site == "LH"), clusters, by = "date")
## add measured values for reference
tmp1 <- subset(tmp, data == "KCL-CMAQ")
tmp1 <- transform(tmp1, O3.mod = O3.obs, data ="MEASURED")
o3Clust <- rbind(tmp, tmp1)
## ----LHcluster,w=12,h=13,fig.keep='last', out.width='1\\textwidth'-------
## put measured at top
o3Clust$data <- relevel(factor(o3Clust$data), ref = "MEASURED")
conditionalEval(o3Clust, obs = "O3.obs", mod = "O3.mod", statistic = "cluster",
                type = "data", col.var = "Set1")
## ----AQindexMB,w=6,h=6,fig.show='hold', out.width='0.49\\textwidth'------
## celculate 8-hour running means,  by model/site...
res <- selectByDate(all.data, month = 6:7)
res <- modStats(res, obs = "O3.roll.obs", mod = "O3.roll.mod",
                type = c("index", "data"), rank.name="data")
scatterPlot(res, x = "index", y = "MB", group = "data", pch = 16, cex=2, col ="Set2", ref.y=0,
            ylab = "MB (ug/m3)")
scatterPlot(res, x = "index", y = "NMB", group = "data", pch = 16, cex=2, col ="Set2", ref.y=0)
## ----modStatsurb, results='asis'-----------------------------------------
nox.stat <- modStats(subset(all.data, site %in% urban.sites), obs = "NOx.obs", mod = "NOx.mod",
          type = "data", rank.name="data")
print(xtable(nox.stat,
             caption = "Summary statistics for urban \\nox concentrations based on hourly data.",
             label = "tab:modStatsUrbNOx"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----modStatrural, results='asis'----------------------------------------
print(xtable(modStats(subset(all.data, site %in% rural.sites), obs = "NOx.obs", mod = "NOx.mod",
          type = "data", rank.name="data"),
             caption = "Summary statistics for rural \\nox concentrations based on hourly data.",
             label = "tab:modStatsRuralNOx"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----annualNOxNO2--------------------------------------------------------
annual <- aggregate(subset(all.data, select = c(NOx.obs, NOx.mod, NO2.obs, NO2.mod)),
                    subset(all.data, select = c(site, data)), mean, na.rm = TRUE)
annual$site.type <- "rural"
id <- which(annual$site %in% urban.sites)
annual$site.type[id] <- "urban"
## ----noxAnnual,w=8,h=4,out.width='1\\textwidth'--------------------------
annStat <- modStats(annual, obs = "NO2.obs", mod = "NO2.mod",
                    type = c("site.type", "data"), rank = "data")
scatterPlot(annStat, x = "NMB", y = "NMGE", group = "data", type = "site.type",
            cex = 2, pch = c(15:19, 8, 11, 12), col = "Dark2", ref.x = 0)
```

```
## ----timeVarNOx, fig.keep='none'----------------------------------------
timeVarNOxUrb <- timeVariation(subset(all.data, site == "KC1"),
                               pollu = c("NOx.obs", "NOx.mod"),
                     type = "data", ylab = "nox (ppb)", difference = TRUE)
timeVarNOxRural <- timeVariation(subset(all.data, site == "HAR"),
                                 pollu = c("NOx.obs", "NOx.mod"),
                     type = "data", ylab = "nox (ppb)", difference = TRUE)
## ----timeVarNOxDiurnal,w=10,h=6,out.width='0.9\\textwidth'---------------
plot(timeVarNOxUrb, subset = "hour")
## ----timeVarNOxDiurnalSeas,w=10,h=6,out.width='0.9\\textwidth'-----------
plot(timeVarNOxUrb, subset = "month")
## ----timeVarNOxDiurnalRural,w=10,h=6,out.width='0.9\\textwidth'----------
plot(timeVarNOxRural, subset = "hour")
## ----timeVarNOxDiurnalSeasRural,w=10,h=6,out.width='0.9\\textwidth'------
plot(timeVarNOxRural, subset = "month")
## ----modStatsurbNO2, results='asis'--------------------------------------
print(xtable(modStats(subset(all.data, site %in% urban.sites), obs = "NO2.obs", mod = "NO2.mod",
          type = "data", rank.name ="data"),
             caption = "Summary statistics for urban \\notwo concentrations.",
             label = "tab:modStatsUrbNO2"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----modStatruralNO2, results='asis'-------------------------------------
print(xtable(modStats(subset(all.data, site %in% rural.sites), obs = "NO2.obs", mod = "NO2.mod",
          type = "data", rank.name ="data"),
             caption = "Summary statistics for rural \\notwo concentrations.",
             label = "tab:modStatsRuralNO2"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----gridSize------------------------------------------------------------
## read grid data
gridDat <- read.csv("./data/grid Emissions.csv", header = TRUE)
gridDat$site.type[gridDat$site %in% urban.sites] <- "urban"
gridDat$site.type[gridDat$site %in% rural.sites] <- "rural"
## average emissions by location
gridRes <- ddply(gridDat, .(site.type, data), numcolwise(mean))
gridRes <- subset(gridRes, select = c(site.type, data, t.km2, x))
gridRes <- merge(gridRes, nox.stat, by = "data")
gridRes <- subset(gridRes, site.type == "urban")
## ----gridEffect,w=5,h=3.6,fig.show='hold', out.width='0.49\\textwidth'----
ggplot(gridRes, aes(x = MB, y = x, color = data, shape = data)) +
    geom_point(size = 5) +
    geom_smooth(aes(group = 1), method = "lm", se = FALSE) +
    ylab("grid dimension (km)") +
    xlab("MB (ppb)") +
    ylim(0, 19) +
    geom_vline(lty = 5)
ggplot(gridRes, aes(x = MB, y = t.km2, color = data, shape = data)) +
    geom_point(size = 5) +
    ylim(0, 60) +
    geom_smooth(aes(group = 1), method = "lm", se = FALSE) +
    ylab(quickText("nox emission (t/km2)")) +
    xlab("MB (ppb)") +
    geom_vline(lty = 5)
## ----calcO3ScenStats-----------------------------------------------------
prepScenStat <- function(mod = "KCL", name = "KCL-CMAQ", month = 1:12) {
    ## KCL
    stats.base <- aqStats(transform(selectByDate(get(paste(mod, ".base", sep = "")),
                                                 month = month), O3 = O3 * 2),
                          pollutant = "O3")
    stats.S1 <- aqStats(transform(selectByDate(get(paste(mod, ".S1", sep = "")),  month = month)
                                  , O3 = O3 * 2),
                        pollutant = "O3")
    names(stats.S1) <- paste(names(stats.S1), ".S1", sep = "")
    stats.S1[, 3:16] <- stats.S1[, 3:16] - stats.base[, 3:16]
    stats.S2 <- aqStats(transform(selectByDate(get(paste(mod, ".S2", sep = "")), month = month)
                                  , O3 = O3 * 2),
                        pollutant = "O3")
```

```
    names(stats.S2) <- paste(names(stats.S2), ".S2", sep = "")
    stats.S2[, 3:16] <- stats.S2[, 3:16] - stats.base[, 3:16]
    stats.S3 <- aqStats(transform(selectByDate(get(paste(mod, ".S3", sep = "")), month = month)
                                  , O3 = O3 * 2),
                        pollutant = "O3")
    names(stats.S3) <- paste(names(stats.S3), ".S3", sep = "")
    stats.S3[, 3:16] <- stats.S3[, 3:16] - stats.base[, 3:16]
    stats.S4 <- aqStats(transform(selectByDate(get(paste(mod, ".S4", sep = "")), month = month),
                                  O3 = O3 * 2),
                        pollutant = "O3")
    names(stats.S4) <- paste(names(stats.S4), ".S4", sep = "")
    stats.S4[, 3:16] <- stats.S4[, 3:16] - stats.base[, 3:16]
    res <- data.frame(stats.base, stats.S1, stats.S2, stats.S3, stats.S4)
    res[, "data"] <-name
    res
}
stats.KCL <- prepScenStat(mod = "KCL", name = "KCL-CMAQ")
stats.AEA <- prepScenStat(mod = "AEA", name = "AEA-CMAQ")
stats.Hert <- prepScenStat(mod = "Hert", name = "Hert-CMAQ")
stats.EMEP <- prepScenStat(mod = "EMEP", name = "EMEP4UK")
stats.OSRM <- prepScenStat(mod = "OSRM", name = "OSRM")
stats.NAME <- prepScenStat(mod = "NAME", name = "NAME")
stats.AQUM <- prepScenStat(mod = "AQUM", name = "AQUM")
stats.ENSEMBLE <- prepScenStat(mod = "ENSEMBLE", name = "ENSEMBLE")
stats.scen <- rbind.fill(stats.KCL, stats.AEA, stats.Hert, stats.EMEP, stats.OSRM,
                    stats.NAME, stats.AQUM, stats.ENSEMBLE)
stats.scen$site.type <- "rural"
stats.scen$site.type[which(stats.scen$site %in% urban.sites)] <- "urban"
## ----tableScenMean,results='asis', cache=FALSE, fig.align='default', echo=FALSE----
res <- ddply(subset(stats.scen, select = c(data,site.type, mean.S1, mean.S2, mean.S3, mean.S4)),
             .(site.type, data), numcolwise(mean))
names(res)[3:6] <- paste("S", 1:4, sep = "")
print(xtable(res, digits =1,
             caption = "Summary changes in annual mean \\ozone
concentrations across all receptors by model for 2006 for annual mean concentrations (\\ug).",
             label = "tab:ScenMean"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames =  FALSE,
      caption.placement = "top")
## ----meanO3S1,w=13,h=5,out.width='1\\textwidth'--------------------------
## re-order by median
ggplot(stats.scen, aes(x = reorder(site, mean.S1, FUN = "median"), y = mean.S1)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y=mean.S1, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----meanO3S2,w=13,h=5,out.width='1\\textwidth'--------------------------
## re-order by median
ggplot(stats.scen, aes(x = reorder(site, mean.S2, FUN = "median"), y = mean.S2)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = mean.S2, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values = c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----meanO3S3,w=13,h=5,out.width='1\\textwidth'--------------------------
## re-order by median
ggplot(stats.scen, aes(x = reorder(site, mean.S3, FUN = "median"), y = mean.S3)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = mean.S3, color = data, shape = data), size = 4,
```

```
              position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----meanO3S4,w=13,h=5,out.width='1\\textwidth'-------------------------
ggplot(stats.scen, aes(x = reorder(site, mean.S4, FUN = "median"), y = mean.S4)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = mean.S4, color = data, shape = data), size = 4,
              position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----tableScenRoll100,results='asis', cache=FALSE, fig.align='default', echo=FALSE----
res <- ddply(subset(stats.scen, select = c(data, site.type, roll.8.O3.gt.100.S1,
                                roll.8.O3.gt.100.S2, roll.8.O3.gt.100.S3,
                                roll.8.O3.gt.100.S4)), .(site.type, data),
                  numcolwise(mean))
names(res)[3:6] <- paste("S", 1:4, sep = "")
print(xtable(res, digits = 1,
             caption = "Summary changes in number of days with a daily maximum of running 8-hour means $>$100~\\ug \\ozone across
             label = "tab:ScenMaxRoll100"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----roll100O3S1,w=13,h=5,out.width='1\\textwidth'-----------------------
ggplot(stats.scen, aes(x = reorder(site, roll.8.O3.gt.100.S1, FUN = "median"),
                       y = roll.8.O3.gt.100.S1)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = roll.8.O3.gt.100.S1, color = data, shape = data),
              size = 4,
              position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (days)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----roll100O3S2,w=13,h=5,out.width='1\\textwidth'-----------------------
ggplot(stats.scen, aes(x = reorder(site, roll.8.O3.gt.100.S2, FUN = "median"),
                       y = roll.8.O3.gt.100.S2)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = roll.8.O3.gt.100.S2, color = data, shape = data),
              size = 4,
              position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (days)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----roll100O3S3,w=13,h=5,out.width='1\\textwidth'-----------------------
ggplot(stats.scen, aes(x = reorder(site, roll.8.O3.gt.100.S3, FUN = "median"),
                       y = roll.8.O3.gt.100.S3)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = roll.8.O3.gt.100.S3, color = data, shape = data),
              size = 4,
              position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (days)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
```

```
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----roll100O3S4,w=13,h=5,out.width='1\\textwidth'-----------------------
ggplot(stats.scen, aes(x = reorder(site, roll.8.O3.gt.100.S4, FUN = "median"),
                       y = roll.8.O3.gt.100.S4)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = roll.8.O3.gt.100.S4, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (days)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----tableScenMaxdaily,results='asis', cache=FALSE, fig.align='default', echo=FALSE----
res <- ddply(subset(stats.scen, select = c(data, site.type, max.daily.S1, max.daily.S2,
                                           max.daily.S3, max.daily.S4)), .(site.type, data),
                    numcolwise(mean))
names(res)[3:6] <- paste("S", 1:4, sep = "")
print(xtable(res, digits = 1,
             caption = "Summary changes in maximum daily \\ozone concentration across all receptors by model for 2006. Concentrat
             label = "tab:ScenMaxDaily"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----MaxDailyS1,w=13,h=5,out.width='1\\textwidth'------------------------
ggplot(stats.scen, aes(x = reorder(site, max.daily.S1, FUN = "median"),
                       y = max.daily.S1)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = max.daily.S1, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----MaxDailyS4,w=13,h=5,out.width='1\\textwidth'------------------------
ggplot(stats.scen, aes(x = reorder(site, max.daily.S4, FUN = "median"),
                       y = max.daily.S4)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = max.daily.S4, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales = "free_x")
## ----AOT40S1,w=13,h=5,out.width='1\\textwidth'---------------------------
ggplot(stats.scen, aes(x = reorder(site, AOT40.S1, FUN = "median"), y = AOT40.S1)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = AOT40.S1, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3.hours)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales ="free_x")
## ----AOT40S2,w=13,h=5,out.width='1\\textwidth'---------------------------
ggplot(stats.scen, aes(x = reorder(site, AOT40.S2, FUN = "median"), y = AOT40.S2)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = AOT40.S2, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
```

```
    ylab(quickText("delta.o3 (ug/m3.hours)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales ="free_x")
## ----AOT40S3,w=13,h=5,out.width='1\\textwidth'--------------------------
ggplot(stats.scen, aes(x = reorder(site, AOT40.S3, FUN = "median"), y = AOT40.S3)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = AOT40.S3, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3.hours)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales ="free_x")
## ----AOT40S4,w=13,h=5,out.width='1\\textwidth'--------------------------
ggplot(stats.scen, aes(x =  reorder(site, AOT40.S4, FUN = "median"), y = AOT40.S4)) +
    geom_boxplot(fill = "white") +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y = AOT40.S4, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3.hours)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5)) +
    facet_grid(~ site.type, scales ="free_x")
## ----JulyO3, w=6, h=3, out.width='0.9\\textwidth'-----------------------
timePlot(selectByDate(subset(o3Meas, site == "HAR"), month = 7), poll="O3",
         ylab = "[O3] (ppb)", key = FALSE)
## ----JulyTraj-----------------------------------------------------------
traj <- importTraj("har", 2006)
## merge with measurements
traj <- merge(traj, subset(o3Meas, site == "HAR"), by = "date")
## intervals of O3
traj$interval <- cut(traj$O3, breaks = c(0, 25, 50, 105), labels = c("0 to 25", "25 to 50",  ">50"))
traj$day <- as.Date(traj$date)
## ----JulyTrajPlot,w=10,h=4, dev='png', dpi=600, out.width='1\\textwidth'----
trajPlot(selectByDate(traj, month =7), plot.type = "1", poll = "O3", lwd = 2,
         type = "interval", col = "jet", layout = c(3, 1), main = "")
## ----indicatorFun, message=FALSE, results='hide'------------------------
## prepare indicator; takes NOx and VOC control data
indPrep <- function(base, noxCont, vocCont, month = 7, hour = 15){
    model <- unique(base$data)
    base <- selectByDate(base, month = month, hour = hour)
    noxCont <- selectByDate(noxCont, month = month, hour = hour)
    vocCont <- selectByDate(vocCont, month = month, hour = hour)
    tmp <- merge(base, noxCont, by = c("date", "site"))
    tmp <- subset(tmp, select = c(date, site, O3.x, O3.y, NO.x, NO2.x, NOy.x))
    tmp <- transform(tmp, delta.O3 = O3.x - O3.y, NOx = NO.x + NO2.x, O3 = O3.x, control = "NOx control")
    names(tmp)[7] <- "NOy"
    tmp <- subset(tmp, select = c(date, site, O3, delta.O3, NOx, NOy, control))
    tmp1 <- merge(base, vocCont, by = c("date", "site"))
    tmp1 <- subset(tmp1, select = c(date, site, O3.x, O3.y, NO.x, NO2.x, NOy.x))
    tmp1 <- transform(tmp1, delta.O3 = O3.x - O3.y, NOx = NO.x + NO2.x, O3 = O3.x, control = "VOC control")
    names(tmp1)[7] <- "NOy"
    tmp1 <- subset(tmp1, select = c(date, O3, site, delta.O3, NOx, NOy, control))
    tmp <- rbind.fill(tmp, tmp1)
    tmp <- transform(tmp, ind = O3 / (NOy - NOx), NOz = NOy - NOx)
    tmp$data <- model
    tmp
}
AEA.ind <- indPrep(AEA.base, AEA.S3, AEA.S4)
AQUM.ind <- indPrep(AQUM.base, AQUM.S3, AQUM.S4)
EMEP.ind <- indPrep(EMEP.base, EMEP.S3, EMEP.S4)
Hert.ind <- indPrep(Hert.base, Hert.S3, Hert.S4)
KCL.ind <- indPrep(KCL.base, KCL.S3, KCL.S4)
NAME.ind <- indPrep(NAME.base, NAME.S3, NAME.S4)
OSRM.ind <- indPrep(OSRM.base, OSRM.S3, OSRM.S4)
ENSEMBLE.ind <- indPrep(ENSEMBLE.base, ENSEMBLE.S3, ENSEMBLE.S4)
```

```
## import PTM results
ptm.ind <- import("./data/PTM.csv")
ptm.ind$data <- as.character(ptm.ind$data)
ind.res <- rbind.fill(AEA.ind, AQUM.ind, EMEP.ind, Hert.ind, KCL.ind, NAME.ind, OSRM.ind,
                      ptm.ind, ENSEMBLE.ind)
## ----SillmanInd,w=8,h=9,out.width='0.8\\textwidth'----------------------
scatterPlot(subset(ind.res, ind >= 0), x = "ind", y = "delta.O3", group="control",
            pch = c(16, 4), cex = c(0.5, 1),
            type = "data", xlim = c(0, 25), ylim = c(-20, 20), ylab = "o3 reduction (ppb)",
            xlab = expression(frac(O [3], NO[y] - NO[x])), smooth =T, ci = F,
            key.position="bottom", layout = c(3, 3))
## ----SillmanIndHar,w=8,h=9,out.width='0.8\\textwidth'--------------------
scatterPlot(subset(ind.res, site == "HAR" & ind >0), x = "ind", y = "delta.O3", group="control",
            pch = c(16, 17),
            type = "data", xlim = c(0, 25), ylim = c(-20, 20), ylab = "o3 reduction (ppb)",
            xlab = expression(frac(O [3], NO[y] - NO[x])), smooth =T, ci = F, cex = 2,
            key.position="bottom", layout = c(3, 3))
## ----julyStats----------------------------------------------------------
jul.KCL <- prepScenStat(mod = "KCL", name = "KCL-CMAQ", month = 7)
jul.AEA <- prepScenStat(mod = "AEA", name = "AEA-CMAQ", month = 7)
jul.Hert <- prepScenStat(mod = "Hert", name = "Hert-CMAQ", month = 7)
jul.EMEP <- prepScenStat(mod = "EMEP", name = "EMEP4UK", month = 7)
jul.OSRM <- prepScenStat(mod = "OSRM", name = "OSRM", month = 7)
jul.NAME <- prepScenStat(mod = "NAME", name = "NAME", month = 7)
jul.AQUM <- prepScenStat(mod = "AQUM", name = "AQUM", month = 7)
jul.ENSEMBLE <- prepScenStat(mod = "ENSEMBLE", name = "ENSEMBLE", month = 7)
##jul.WRFChem <- prepScenStat(mod = "WRF-Chem", name = "WRF-Chem", month = 7)
jul.scen <- rbind.fill(jul.KCL, jul.AEA, jul.Hert, jul.EMEP, jul.OSRM,
                       jul.NAME, jul.AQUM, jul.ENSEMBLE)
## select what we need
few.scen <- subset(jul.scen, site %in% c("HAR", "LH", "ROCH", "WFEN"))
few.scen <- few.scen[, c(1, grep("max.rolling.8", names(few.scen)), 81)]
names(few.scen)[3:6] <- c("S1", "S2", "S3", "S4")
few.scen <- melt(few.scen)
few.scen <- subset(few.scen, variable != "max.rolling.8")
urb.scen <- subset(jul.scen, site %in% c("CLL2", "KC1", "LON6", "MAN4"))
urb.scen <- urb.scen[, c(1, grep("max.rolling.8", names(urb.scen)), 81)]
names(urb.scen)[3:6] <- c("S1", "S2", "S3", "S4")
urb.scen <- melt(urb.scen)
urb.scen <- subset(urb.scen, variable != "max.rolling.8")
## ----Julyroll8,w=10,h=5,out.width='1\\textwidth'-------------------------
few.scen$site <- reorder(few.scen$site, few.scen$value,
                         FUN = median)
ggplot(few.scen, aes(x = site, y=value)) +
    geom_boxplot() +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y=value, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values =c(15:19, 8, 11, 12)) +
    xlab("site") +
    facet_grid(~ variable) +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----Julyroll8Urb,w=10,h=5,out.width='1\\textwidth'----------------------
urb.scen$site <- reorder(urb.scen$site, urb.scen$value,
                         FUN = median)
ggplot(urb.scen, aes(x = site, y=value)) +
    geom_boxplot() +
    geom_hline(yintercept = 0, lty = 5) +
    geom_point(aes(x = site, y=value, color = data, shape = data), size = 4,
               position = position_dodge(.5)) +
    scale_color_brewer(palette = "Dark2") +
    ylab(quickText("delta.o3 (ug/m3)")) +
    scale_shape_manual(values= c(15:19, 8, 11, 12)) +
    xlab("site") +
    facet_grid(~ variable) +
    theme(axis.text.x = element_text(angle= -90, hjust = 0, vjust = 0.5))
## ----dailySensitivity---------------------------------------------------
## function to work out if max rolling 8 is VOC or NOx senstive
prepRollStat <- function(mod = "KCL", name = "KCL-CMAQ", month = 1:12) {
    stats.base <- selectByDate(get(paste(mod, ".base", sep = "")),  month = month)
```

```
    stats.base <- subset(stats.base, select = c(site, date, O3))
    stats.base <- rollingMean(stats.base, pollutant = "O3", alight = "right")
    stats.base <- timeAverage(stats.base, "day", data.thresh = 75, statistic = "max")
    stats.base <- stats.base[, -3] ## don't need hourly O3
    names(stats.base)[3] <- "base"
    stats.S3 <- selectByDate(get(paste(mod, ".S3", sep = "")),  month = month)
    stats.S3 <- subset(stats.S3, select = c(site, date, O3))
    stats.S3 <- rollingMean(stats.S3, pollutant = "O3", alight = "right")
    stats.S3 <- timeAverage(stats.S3, "day", data.thresh = 75, statistic = "max")
    stats.S3 <- stats.S3[, -3] ## don't need hourly O3
    names(stats.S3)[3] <- "S3"
    stats.S4 <- selectByDate(get(paste(mod, ".S4", sep = "")),  month = month)
    stats.S4 <- subset(stats.S4, select = c(site, date, O3))
    stats.S4 <- rollingMean(stats.S4, pollutant = "O3", alight = "right")
    stats.S4 <- timeAverage(stats.S4, "day", data.thresh = 75, statistic = "max")
    stats.S4 <- stats.S4[, -3] ## don't need hourly O3
    names(stats.S4)[3] <- "S4"
    res <-merge(stats.base, stats.S3, by = c("site", "date"))
    res <-merge(res, stats.S4, by = c("site", "date"))
    res[, "data"] <- name
    res
}
roll.KCL <- prepRollStat(mod = "KCL", name = "KCL-CMAQ", month = 7)
roll.AEA <- prepRollStat(mod = "AEA", name = "AEA-CMAQ", month = 7)
roll.Hert <- prepRollStat(mod = "Hert", name = "Hert-CMAQ", month = 7)
roll.EMEP <- prepRollStat(mod = "EMEP", name = "EMEP4UK", month = 7)
roll.OSRM <- prepRollStat(mod = "OSRM", name = "OSRM", month = 7)
roll.NAME <- prepRollStat(mod = "NAME", name = "NAME", month = 7)
roll.AQUM <- prepRollStat(mod = "AQUM", name = "AQUM", month = 7)
roll.ENSEMBLE <- prepRollStat(mod = "ENSEMBLE", name = "ENSEMBLE", month = 7)
roll.res <- rbind.fill(roll.KCL, roll.AEA, roll.Hert, roll.EMEP, roll.OSRM,
                       roll.NAME, roll.AQUM, roll.ENSEMBLE)
## VOC or NOx dominated
roll.res$sens <- ifelse((roll.res$S3 - roll.res$base) < (roll.res$S4 - roll.res$base),
                        "NOx", "NOx")
## add a numeric value calendarPlot can work with
roll.res$bin <- ifelse((roll.res$S3 - roll.res$base) < (roll.res$S4 - roll.res$base),
                        1, 0)
## ----JulySens,w=4,h=4,fig.show='hold',out.width='0.32\\textwidth'--------
plotCal <- function(model) {
    calendarPlot(subset(roll.res, site == "HAR" & data == model),
             pollutant = "bin", breaks = c(-1, 0.1, 1),
             labels = c("VOC sensitive", "NOx sensitive"),
             year = 2006, cols = "hue", main = model)
}
l_ply(c("AQUM", "KCL-CMAQ", "EMEP4UK", "AEA-CMAQ", "Hert-CMAQ", "NAME", "OSRM", "ENSEMBLE"),
      plotCal)
## also show measured values
o3HAR <- subset(o3Meas, site.name == "Harwell")
o3HAR <- selectByDate(o3HAR, month = 7)
o3HAR <- rollingMean(o3HAR, pollutant = "O3", alight = "right")
o3HAR <- timeAverage(o3HAR, "day", statistic = "max")
calendarPlot(o3HAR, pollutant = "O3", year = 2006, annotate = "value", lim = 50,
             col.lim = c("black","grey"),
             main = "Measured ozone", col = "increment")
## ----sensPlot,w=12,h=4,out.width='1\\textwidth'--------------------------
nox.sens <- ddply(roll.res, .(site, data, sens), numcolwise(sum), na.rm = TRUE)
nox.sens <- subset(nox.sens, sens == "NOx", select = c(data, site, bin))
names(nox.sens) <- c("model", "site", "days")
nox.sens$days <- nox.sens$days - 15.5 ## for more effective plotting
 ggplot(nox.sens, aes(x = site, y= model, fill = days)) +
    geom_tile() +
    scale_fill_gradientn(colours = openColours("PuOr", 31)) +
    theme(legend.position="none")
## ----sensMap,w=8,h=8,dev='png', dpi=600, out.width='1\\textwidth'--------
nox.sens <- merge(nox.sens, meta, by.x = "site", by.y = "code")
scatterPlot(subset(nox.sens, site.type == "Rural Background"),
            x = "longitude",y = "latitude", z = "days",
            map = TRUE, cex =2, map.alpha = 0.1, col = "RdYlBu",
            type = "model", main = NULL, key = FALSE)
## ----UKEuropeSens--------------------------------------------------------
## function to work out if max rolling 8 is VOC or NOx senstive
```

```
prepUKEurStat <- function(mod = "KCL", name = "KCL-CMAQ", month = 1:12) {
    stats.base <- selectByDate(get(paste(mod, ".base", sep = "")),  month = month)
    stats.base <- subset(stats.base, select = c(site, date, O3))
    stats.base <- rollingMean(stats.base, pollutant = "O3", alight = "right")
    stats.base <- timeAverage(stats.base, "day", data.thresh = 75, statistic = "max")
    stats.base <- stats.base[, -3] ## don't need hourly O3
    names(stats.base)[3] <- "base"
    ## S1 is UK + Europe
    stats.S1 <- selectByDate(get(paste(mod, ".S1", sep = "")),  month = month)
    stats.S1 <- subset(stats.S1, select = c(site, date, O3))
    stats.S1 <- rollingMean(stats.S1, pollutant = "O3", alight = "right")
    stats.S1 <- timeAverage(stats.S1, "day", data.thresh = 75, statistic = "max")
    stats.S1 <- stats.S1[, -3] ## don't need hourly O3
    names(stats.S1)[3] <- "S1"
    ## S2 is UK only
    stats.S2 <- selectByDate(get(paste(mod, ".S2", sep = "")),  month = month)
    stats.S2 <- subset(stats.S2, select = c(site, date, O3))
    stats.S2 <- rollingMean(stats.S2, pollutant = "O3", alight = "right")
    stats.S2 <- timeAverage(stats.S2, "day", data.thresh = 75, statistic = "max")
    stats.S2 <- stats.S2[, -3] ## don't need hourly O3
    names(stats.S2)[3] <- "S2"
    res <-merge(stats.base, stats.S1, by = c("site", "date"))
    res <-merge(res, stats.S2, by = c("site", "date"))
    res[, "data"] <- name
    res
}
S1S2.KCL <- prepUKEurStat(mod = "KCL", name = "KCL-CMAQ", month = 7)
S1S2.AEA <- prepUKEurStat(mod = "AEA", name = "AEA-CMAQ", month = 7)
S1S2.Hert <- prepUKEurStat(mod = "Hert", name = "Hert-CMAQ", month = 7)
S1S2.EMEP <- prepUKEurStat(mod = "EMEP", name = "EMEP4UK", month = 7)
S1S2.OSRM <- prepUKEurStat(mod = "OSRM", name = "OSRM", month = 7)
S1S2.NAME <- prepUKEurStat(mod = "NAME", name = "NAME", month = 7)
S1S2.AQUM <- prepUKEurStat(mod = "AQUM", name = "AQUM", month = 7)
S1S2.ENSEMBLE <- prepUKEurStat(mod = "ENSEMBLE", name = "ENSEMBLE", month = 7)
S1S2.res <- rbind.fill(S1S2.KCL, S1S2.AEA, S1S2.Hert, S1S2.EMEP, S1S2.OSRM,
                       S1S2.NAME, S1S2.AQUM, S1S2.ENSEMBLE)
## VOC or NOx dominated
S1S2.res$sens <- ifelse((S1S2.res$S1 - S1S2.res$base) < (S1S2.res$S2 - S1S2.res$base),
                        "Europe", "UK")
## add a numeric value calendarPlot can work with
S1S2.res$bin <- ifelse((S1S2.res$S1 - S1S2.res$base) < (S1S2.res$S2 - S1S2.res$base),
                       1, 0)
## ----UKsensPlot,w=12,h=4,out.width='1\\textwidth'------------------------
UK.sens <- ddply(S1S2.res, .(site, data, sens), numcolwise(sum), na.rm = TRUE)
UK.sens <- subset(UK.sens, sens == "Europe", select = c(data, site, bin))
names(UK.sens) <- c("model", "site", "days")
ggplot(UK.sens, aes(x = site, y = model, fill = days)) +
    geom_tile() +
    scale_fill_gradientn(colours = openColours("default", 31)) +
    geom_text(aes(label = days), color = "grey60")
## ----sensEmi,w=5,h=3.6,out.width='0.6\\textwidth'------------------------
## simple NOx-VOC senstivity vs. emissions
## make data frame
sensEmi <- data.frame(model = c("KCL-CMAQ", "NAME", "AQUM", "Hert-CMAQ", "AEA-CMAQ"),
                      ratio = c(0.63,  1.70,  1.40,  0.39, 0.57),
                      days = c(17, 4, 5, 24, 17))
library(MASS)
library(splines)
ggplot(sensEmi, aes(ratio, days, color = model, shape = model)) +
    geom_rect(aes(NULL, NULL), alpha = 0.02, xmin = -10, xmax = 2, ymin = 0,
              ymax = 15.5, fill="blue", color = "transparent") +
    geom_rect(aes(NULL, NULL), alpha = 0.02, xmin = -10, xmax = 2, ymin = 15.5,
              ymax = 31, fill = "green", color = "transparent") +
    geom_text(x = 0.5, y = 4, label = "mostly VOC senstive", size = 4, color = "grey30") +
    geom_text(x = 1.2, y = 22, label = "mostly NOx senstive", size = 4, color = "grey30") +
    geom_point(size = 5) + ylim(0, 31) +
    xlim(0,  1.7) +
    xlab(quickText("NOx/VOC European emissions ratio")) +
    ylab(quickText("days when nox-senstive")) +
    stat_smooth(aes(group=1), method = "lm",formula = y ~ ns(x, 2), size = 1, se = FALSE)
## ----metSites,w=7,h=7,out.width='.65\\textwidth',message=FALSE-----------
## codes to prepare regional data
```

```
## map of met sites
metLoc <- read.csv("./data/metSites.csv", header = TRUE)
GoogleMapsPlot(metLoc, latitude = "latitude", longitude = "longitude",  col = "red",
               labels = list(labels = "Site", col = "red", cex=1.2, pos = 1, font = 2),
               cex = 1.5, maptype = "roadmap", verbose = 0)
## ----modStatsWS, results='asis'----------------------------------------
print(xtable(modStats(selectByDate(allMet, month = 7),  obs = "ws.obs",
                      mod = "ws.mod", type = "data.mod", rank.name = "data.mod"),
             caption = "Summary statistics for model wind speed performance across all sites for July 2006.",
             label = "tab:modStatsWSJuly"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----modStatsWSall, results='asis'-------------------------------------
print(xtable(modStats(allMet, obs = "ws.obs",  mod = "ws.mod",
                      type = "data.mod", rank.name = "data.mod"),
             caption = "Summary statistics for model wind speed performance across all sites for 2006.",
             label = "tab:modStatsWSAll"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----wsEMEP,w=8,h=8,out.width='0.6\\textwidth',cache=TRUE---------------
ggplot(subset(allMet, site %in% c("Heathrow", "Church Fenton", "Weybourne", "Lossiemouth")),
       aes(data.mod, ws.mod)) +
    geom_boxplot(outlier.shape = NA) +
     geom_boxplot(aes(data.obs, ws.obs, fill = data.obs), outlier.shape = NA) +
    coord_cartesian(ylim = c(-0.2, 13.5)) +
    opts(legend.position = "bottom",  axis.text.x = theme_text(angle= -90, hjust = 0)) +
    facet_wrap( ~ site) +
    ylab(quickText("wind speed (m/s)"))
## ----taylorWS,w=10,h=8,out.width='1\\textwidth',cache=TRUE---------------
TaylorDiagram(selectByDate(allMet,  month =7),  obs = "ws.obs",  mod = "ws.mod",
              group = "site",  norm = TRUE,  type = "data.mod")
## ----wsCondQuant,w=10,h=8,out.width='1\\textwidth'----------------------
conditionalQuantile(allMet,  obs = "ws.obs",  mod = "ws.mod",
                    type = "data.mod", xlab = "modelled wind speed (m/s)",
                    ylab = "measured wind speed (m/s)")
## ----diurnalws, fig.keep='none'----------------------------------------
tmp <- timeVariation(allMet, pollutant=c("ws.obs", "ws.mod"),
                     type = "data.mod", ylab = "wind speed (m/s)")
## ----wsDiurnal,w=10,h=6,out.width='0.9\\textwidth'----------------------
plot(tmp, subset = "hour")
## ----metSub,w=9,h=6,out.width='1\\textwidth'----------------------------
pollutionRose(subset(allMet, site == "Heathrow" & data.mod != "EMEP4UK"),
              ws = "ws.obs", wd = "wd.obs",
              ws2 = "ws.mod", wd2 = "wd.mod", type = "data.mod", grid = 5,
              angle = 3)
## ----modStatstemp, results='asis'--------------------------------------
print(xtable(modStats(selectByDate(allMet,  month = 7),  obs = "temp.obs",
                      mod = "temp.mod",  type = "data.mod", rank.name = "data.mod"),
             caption = "Summary statistics for model ambient temperature performance across all sites for July 2006.",
             label = "tab:modStatsTempJuly"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----modStatstempAll, results='asis'-----------------------------------
print(xtable(modStats(allMet,  obs = "temp.obs", mod = "temp.mod",
                      type = "data.mod", rank.name = "data.mod"),
             caption = "Summary statistics for model ambient temperature performance across all sites for 2006.",
             label = "tab:modStatsTempAll"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----taylorTemp,w=10,h=8,out.width='1\\textwidth'-----------------------
TaylorDiagram(selectByDate(allMet, month =7), obs = "temp.obs", mod = "temp.mod",
              group = "site",  norm = TRUE, type = "data.mod")
## ----tempCondQuant,w=10,h=8,out.width='1\\textwidth'--------------------
conditionalQuantile(allMet,  obs = "temp.obs",  mod = "temp.mod",
```

```
                    type = "data.mod", xlab = "modelled temp (degreesC)", ylab = "measured temp (degreesC)")
## ----diurnalTemp, fig.keep='none'-------------------------------------
tmp <- timeVariation(allMet, pollutant=c("temp.obs", "temp.mod"),
                     type = "data.mod", ylab = "ambient temp (degreesC)")
## ----tempDiurnal,w=10,h=6,out.width='0.9\\textwidth'-------------------
plot(tmp, subset = "hour")
## ----modStatsRH, results='asis'---------------------------------------
print(xtable(modStats(selectByDate(allMet,  month = 7), obs = "rh.obs",
                      mod = "rh.mod",  type = "data.mod", rank.name = "data.mod"),
             caption = "Summary statistics for model relative humidity performance across all sites for July 2006.",
             label = "tab:modStatsRHJuly"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----modStatsRHall, results='asis'------------------------------------
print(xtable(modStats(allMet, obs = "rh.obs",  mod = "rh.mod",
                      type = "data.mod", rank.name = "data.mod"),
             caption = "Summary statistics for model relative humidity performance across all sites in 2006.",
             label = "tab:modStatsRHAll"),
      size = "footnotesize",
      booktabs = TRUE,
      include.rownames = FALSE,
      caption.placement = "top")
## ----taylorRH,w=10,h=8,out.width='1\\textwidth'------------------------
TaylorDiagram(selectByDate(allMet,  month =7),  obs = "rh.obs",  mod = "rh.mod",
              group = "site", norm = TRUE, type = "data.mod")
## ----rhCondQuant,w=10,h=8,out.width='1\\textwidth'---------------------
conditionalQuantile(allMet,  obs = "rh.obs",  mod = "rh.mod",
                    type = "data.mod", xlab = "modelled relative humidity (%)",
                    ylab = "measured relative humidity (%)")
## ----BLHadd-----------------------------------------------------------
allHar <- combineAll("HAR", month= 1:12)
## merge with BLH
allHar <- merge(allHar, subset(BLH, site == "HAR"), by= c("date", "data"))
## ----timeVarBLH,fig.keep='none'---------------------------------------
plt <- timeVariation(allHar, pollutant = "BLH", group = "data", type = "season",
                     ci = FALSE, lwd= 3, ylab = "boundary layer height (m)",
                     cols = "Set2", key.columns = 3)
## ----timeVarBLHMonth,w=10,h= 4,out.width='1\\textwidth'----------------
plot(plt, subset= "hour")
## ----BLHcondNOx,w=12,h=8,fig.keep='last', out.width='1\\textwidth'-----
conditionalEval(allHar, obs = "NOx.obs", mod = "NOx.mod", type = "data",
                statistic = "BLH", col.var ="Dark2")
## ----annualTaylor,w=5,h=5,out.width='0.6\\textwidth'-------------------
TaylorDiagram(stats, obs = "mean.obs", mod = "mean.mod",group = "data")
## ----roll8Taylor,w=5,h=5,out.width='0.6\\textwidth'--------------------
TaylorDiagram(stats, obs = "roll.8.O3.gt.100.obs", mod = "roll.8.O3.gt.100.mod",group = "data")
```