Geosci. Model Dev. Discuss., 5, 3131–3182, 2012
www.geosci-model-dev-discuss.net/5/3131/2012/
doi:10.5194/gmdd-5-3131-2012

Geoscientific Model Development Discussions

Air quality modelling using the Met Office Unified Model: model description and initial evaluation

N. H. Savage, P. Agnew, L. S. Davis, C. Ordóñez, R. Thorpe, C. E. Johnson, F. M. O'Connor, and M. Dalvi

Met Office, Fitzroy Road, Exeter, EX1 3PB, UK

Received: 24 August 2012 – Accepted: 26 September 2012 – Published: 10 October 2012

Correspondence to: N. H. Savage (nicholas.savage@metoffice.gov.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The on-line air quality model AQUM (Air Quality in the Unified Model) is a limited-area forecast configuration of the Met Office Unified Model which uses the UKCA (UK Chemistry and Aerosols) sub-model. AQUM has been developed with two aims: as an operational system to deliver regional air quality forecasts and as a modelling system to enable air quality studies to be conducted to inform policy decisions relating to emissions controls. This paper presents a description of the model and the methods used to evaluate the performance of the forecast system. Results are presented of evaluation studies conducted for a year-long period of operational forecast trials and several past cases of poor air quality episodes. To place the model performance in context we compare AQUM ozone forecasts with those of another forecasting system, the MACC ensemble, for a 5-month period. The results demonstrate that AQUM has a large dynamic range of modelled ozone levels and has a good level of responsiveness to elevated ozone episode conditions – a characteristic which is essential for forecasting poor air quality episodes. An analysis of the variation of model skill with forecast lead-time is presented and the insights this provides to the relative sources of error in air quality modelling are discussed.

1 Introduction

Regional air quality models have evolved rapidly in sophistication over the last ten years. Off-line chemical transport models (CTM), configured with constant or climatological chemical lateral boundary conditions (LBCs) have been superseded by regional models coupled to global models, with the latter providing spatially and temporally evolving boundary fluxes of key chemical species. The GEMS project (Global and regional Earth-system atmosphere Monitoring using Satellite and in-situ data – see Hollingsworth et al., 2008) and its successor MACC (http://www.gmes-atmosphere.eu/) have played a major role in these developments with the creation of a system of global

GMDD
5, 3131–3182, 2012

models for reactive gases and aerosols (operated by the European Centre for Medium-Range Weather Forecasting – ECMWF) providing boundary fluxes to European regional air quality models. The ECMWF global models and some of the more advanced regional models incorporate data assimilation of key chemical and aerosol species, thus adding a further degree of sophistication. Another example of increasing model sophistication concerns the transition from off-line to on-line modelling. In the latter the meteorological and chemical evolution of the atmosphere are modelled within the same system, with the potential to include feedbacks of composition on meteorology. Examples of possible feedbacks include direct aerosol effects due to radiation scattering and indirect effects such as the nucleation of cloud droplets by particulates. A new collaborative project – COST-ES1004 (http://www.eumetchem.info/) – has been initiated with the objective of clarifying and quantifying the improvements to meteorological forecasts by including on-line composition modelling.

The UK Met Office Unified Model (MetUM) is a weather and climate modelling system which is used across a very wide range of spatial and temporal scales, from short range weather forecasting at 1.5 km resolution (Price et al., 2011) to multi-decadal simulations in an Earth system model configuration (Collins et al., 2011). We have developed a configuration of the MetUM for use as an on-line regional air quality model – AQUM (Air Quality in the Unified Model). This model builds on the work of the United Kingdom Chemistry and Aerosols (UKCA) project (Morgenstern et al., 2009; O'Connor et al., 2012), which has constructed a new framework for atmospheric composition modelling within the MetUM. The type of parameterisations used in UKCA and the level of complexity in representing the earth system can be selected as appropriate to the problem under investigation. AQUM has been developed to fulfil two purposes: (i) the operational delivery of daily air quality forecasts and (ii) to enable atmospheric modelling studies to address scientific and air quality policy-related questions.

In Sect. 2 of this paper we present an overview of the AQUM modelling system. We then describe the verification methodology we have used to evaluate the model in Sect. 3. In Sect. 4 we present a summary of the evaluation studies we have carried out

GMDD
5, 3131–3182, 2012

on both operational forecasts and particular past air quality episodes. A summary of the results is presented in Sect. 5 and a short description of the model developments planned for AQUM is given.

2 Model description

2.1 Physical model overview

AQUM has a horizontal resolution of around 12 km and 38 vertical levels up to a model top height of 39 km. The model domain can be viewed in Fig. 9. The model physics configuration is based on the Met Office's North Atlantic and European Model (NAE). A description of this configuration is given in Bush et al. (2006), although some further minor developments were made to the model prior to it forming the basis of the AQUM model described here.

The MetUM dynamical core is non-hydrostatic and fully compressible, and no shallow atmosphere approximations are made. Semi-implicit, semi-Lagrangian time-integration methods are used and a positive definite semi-Lagrangian tracer advection scheme is used to advect aerosols and gases (Davies et al., 2005). Boundary layer mixing (including that of aerosols and gases) is parameterised with a non-local, first order closure, multi-regime scheme (Lock et al., 2000). Convection is represented with a mass flux scheme with downdraughts, momentum transport and CAPE (convective available potential energy) closure (Gregory and Rowntree, 1990). The land surface scheme, MOSES II, is a nine tile, flux blended surface exchange approach and includes an urban tile (Essery et al., 2003). The model uses the Edwards-Slingo flexible multi-band two stream code for long and short wave radiation with 6 SW and 9 LW bands (Edwards and Slingo, 1996). Wilson and Ballard (1999) microphysics is employed, and extended to include prognostic ice and snow, rain and graupel. The model uses the diagnostic cloud scheme described by Smith (1990).

GMDD
5, 3131–3182, 2012

2.2 Gas phase chemistry scheme

The AQUM gas phase chemistry is a further development of the UKCA tropospheric chemistry scheme (O'Connor et al., 2012) but with a new chemical mechanism added specifically for regional air quality (RAQ) modelling. This RAQ mechanism includes 40 transported species (16 of them emitted), 18 non-advected species, 116 gas-phase reactions (see Supplement, Table S1) and 23 photolysis reactions (Table S2). Removal by wet and dry deposition is considered for 19 and 16 species, respectively. Unlike the standard tropospheric chemistry described in O'Connor et al. (2012), this scheme includes the oxidation of both C2-C3 alkenes (ethene and propene), isoprene and aromatic compounds such as toluene and o-xylene, as well as the formation of organic nitrate. It is adapted from the mechanism presented in Collins et al. (1997) with the additional reactions described in Collins et al. (1999) and some further modifications (in particular to the isoprene and aromatic chemistry mechanisms). All reaction rates have been updated to the IUPAC recommendation values (Atkinson et al., 2004, 2006). Note that sulphur chemistry is not currently included in the RAQ mechanism but is treated in the aerosol scheme (see below). The concentrations are updated using a backward Euler solver with a time-step of 75 s in the studies described in this paper.

Dry deposition of gases and aerosols is based on a multiple resistance approach (Wesley, 1989). Wet deposition is parameterised as first order loss rate, calculated as a function of the model's three-dimensional convective and large-scale precipitation in a manner adapted from Giannakopoulos (1998) and Giannakopoulos et al. (1999). Photolysis rates are calculated with the on-line photolysis scheme Fast-J (Wild et al., 2000) which is coupled to the modelled liquid water and ice content, and sulphate aerosols on a time-step basis.

GMDD
5, 3131–3182, 2012

2.3 Aerosol scheme

The current AQUM configuration uses the Coupled Large-scale Aerosol Simulator for Studies in Climate (CLASSIC) aerosol module. A short description will be given here – for further details see Appendix A of Bellouin et al. (2011) and references therein.

The scheme contains six prognostic tropospheric aerosol types: ammonium sulphate, mineral dust, fossil fuel black carbon (FFBC), fossil fuel organic carbon (FFOC), biomass burning aerosols and ammonium nitrate. In addition there is a diagnostic aerosol scheme for sea salt and a fixed climatology of secondary organic aerosols (SOA).

The model has two-way coupling of oxidants between the aerosol and gas phase chemistry schemes. Thus emissions of sulphur dioxide (SO_2) and dimethyl sulphide (DMS) are oxidised into sulphate aerosol (SO_4^{2-}) by oxidants whose concentrations are calculated in the RAQ chemistry scheme, and the depleted oxidant fields are then passed back to the RAQ scheme to ensure consistency. Sulphate aerosol is represented by Aitken and accumulation modes and an additional tracer for sulphate dissolved in cloud droplets. Sulphate mass is assumed to all be in the form of ammonium sulphate [$(NH_4)_2SO_4$]. Emissions of DMS from oceans are parameterised as function of wind speed based on the approach of Wanninkhof (1992), with sea water concentrations from a $1^\circ \times 1^\circ$ climatology (Kettle et al., 1999).

Ammonia (NH_3) is a transported tracer in the model and it reacts initially with any sulphuric acid (H_2SO_4) present to form ammonium sulphate; after all the sulphuric acid has been depleted, any excess ammonia can react with nitric acid (HNO_3) to form ammonium nitrate aerosol (NH_4NO_3). Thermal decomposition of ammonium nitrate to nitric acid and ammonia is permitted according to the equilibrium model described by Ackermann et al. (1995). Nitric acid concentrations are derived from the RAQ scheme and depleted by nitrate aerosol formation. The mass of nitrate aerosol formed goes into an accumulation mode and cloud formation transforms some of the accumulation mode into the dissolved mode, as with sulphate.

GMDD
5, 3131–3182, 2012

The mineral dust scheme has six size bins covering radii from 0.0316 µm to 3.16 µm. The emissions fluxes depend on vegetation fraction, soil roughness length and moisture, and near-surface wind speeds. The dust in these six bins is transported and deposited by gravitational settling, turbulence and below cloud scavenging.

FFBC, FFOC and biomass burning aerosols have three modes – fresh, aged and in-cloud. Ageing is represented as an exponential decay. Sea salt is represented in a diagnostic manner with number concentrations over the open ocean calculated as a function of the wind speed at a height of 10 m (O'Dowd et al., 1999). Biogenic secondary organic aerosols are included as a three dimensional fixed climatology (Derwent et al., 2003).

The direct radiative effects of all aerosols are included in the model by use of wavelength dependent scattering and absorption coefficients calculated off-line according to Mie theory. The off-line calculations include the effects of hygroscopic growth for the sulphate, sea-salt, nitrate, biomass burning, FFOC and biogenic aerosols. All aerosol species except mineral dust and FFBC are considered to act as cloud condensation nuclei. The parameterisations of the first and second indirect effects are as described by Jones et al. (2001).

2.4 Lateral boundary conditions

Lateral boundary conditions (LBCs) use the method of Davies (1976) which involves relaxing the interior flow near the boundaries towards the externally prescribed flow. Relaxation involves blending of the LBCs and the limited area model (LAM) over several grid-points. For further details see Davies (2012).

Model LBCs are a combination of chemistry and aerosol data from the GEMS or MACC global models (Flemming et al., 2009) and meteorological data from Met Office weather forecast models. The use of meteorological data from the Met Office models improves consistency with the dynamics of the AQUM model. Meteorological LBCs (and initial conditions) come from the MetUM global model forecasts (for case studies) or from the NAE (for operational forecasts). For pragmatic reasons the reanalyses

GMDD
5, 3131–3182, 2012

and forecasts produced by the GEMS and MACC projects have been used to provide boundary conditions for the composition fields. This avoids the computational and maintenance costs of running an operational global model chemistry and aerosol configuration, and benefits from the data assimilation carried out by the GEMS and MACC projects. A re-analysis prepared as part of the GEMS project was used to provide chemical LBCs for the case studies we have conducted, whilst real-time MACC global model fields are used for the AQUM operational forecasts.

2.5 Model resolution, domain and initialisation

AQUM is currently operated with a 12 km horizontal resolution grid covering much of Western Europe (see Fig. 9). The resolution and domain were selected to enable regional scale ozone and particulate matter (PM) events impacting the UK to be modelled. For the period of the model evaluations described in Sect. 4 the operational forecasts were made out to two days ahead (although the system has since been upgraded and now provides forecasts out to 5 days ahead). No independent data assimilation cycle was used in the forecasts, however the initial conditions for the meteorology used Met Office analyses and so inherit assimilated fields produced using 4-D-variational assimilation. There is no data assimilation of chemical species either directly or indirectly. However the chemical LBCs used from the GEMS/MACC global model benefit from the data assimilation within those models.

2.6 Emissions

The anthropogenic pollutant emissions used in AQUM are derived from three datasets. The highest resolution emissions dataset is from the UK National Atmospheric Emissions Inventory (NAEI, MacCarthy et al., 2011) which has a 1 km resolution and covers the UK only. Outside of the UK, emissions are taken from the European Monitoring and Evaluation Programme (EMEP) emissions datasets, which cover Europe at 50 km resolution (http://www.ceip.at/webdab-emission-database/

GMDD
5, 3131–3182, 2012

emissions-as-used-in-emep-models/). Finally, a 5 km resolution gridded shipping emissions dataset produced by Entec UK Ltd on behalf of Defra (Whall et al., 2010), is used to represent emissions for waters around the UK. Where they overlap, shipping data from Entec replaces data from the NAEI and EMEP SNAP sector 8 ("Other mobile sources and machinery") in the NOS ("North Sea") and ATL ("Remaining North-East Atlantic Region") regions. This process ensures there is no duplication of shipping emissions. The Entec dataset was only compiled for 2007, so for other years these data are scaled according to published totals from EMEP. Data from all three sources are interpolated to the AQUM 12 km grid prior to merging.

Six key families of pollutants are provided in the emissions datasets described above: carbon monoxide (CO), sulphur oxide gases (SO_x), volatile organic compounds (VOC), nitrogen oxides ($NO_x = NO_2 + NO$), fine particulate matter with a diameter of 2.5 µm or less ($PM_{2.5}$), PM coarse with a diameter from 2.5 to 10 µm (defined as PM_{10}–$PM_{2.5}$) and NH_3. For use in AQUM the non-methane VOC component of emissions is partitioned into the species required by the RAQ chemical mechanism: formaldehyde, ethene, propene, isoprene, o-xylene, toluene, acetaldehyde, ethane, propane, butane, acetone and methanol. The inventory total VOC emitted mass is apportioned amongst these species according to the tabulated data for 2006 given by Dore et al. (2008), in a manner which ensures the total VOC mass is accounted for. This same report provides further information that we have used to provide a separate traffic-specific speciation of emitted VOC over the UK.

For gas phase emissions, AQUM currently has a simple treatment of the vertical emission profile: all emissions are spread equally over the first four model levels (20, 80, 180 and 320 m). This profile was selected on the basis of sensitivity tests under ozone episode conditions. Clearly this is a significant over-simplification, but a more sophisticated treatment is currently being developed (see Sect. 5). However in practice the representation of a physically realistic profile is limited in an Eulerian model by the spacing of model grid levels and numerical diffusion.

GMDD
5, 3131–3182, 2012

The CLASSIC aerosol scheme used by AQUM requires emissions of specific aerosol and gas phase species: FFOC, FFBC, biomass burning, DMS and S in SO_2 for sulphate production, NO_x and ammonia for nitrate aerosol production, mineral dust and sea salt. The last two species have on-line source terms which depend on meteorology and surface properties, rather than being generated by fixed emission fields. Nitrate aerosol can be regarded as being formed entirely from gas phase precursors (secondary aerosol). Sulphate aerosol is largely secondary but also contains a small primary component. We currently model the latter by emitting an equivalent amount of gas phase SO_2 which is then oxidised to sulphate within CLASSIC. The UK anthropogenic emissions of SO_x required by CLASSIC are split into high- and low-level components, representing emissions from chimneys and surface sources respectively. Volcanic SO_x emissions are derived from the 3-dimensional climatology of Andres and Kasgnoc (1998). Other gas phase emissions for secondary aerosol production are accounted for by the emissions derived from the three inventories described above. However emissions are required for primary particulate matter. The emissions datasets generated by EMEP, NAEI and ENTEC provide only total PM; thus in order to use these high resolution emissions datasets we must apportion the total PM amongst the different primary species required by CLASSIC. In order to achieve this we have used a dataset compiled by TNO for the GEMS project (Visschedijk et al., 2007). This dataset provides an estimate of the percentage contribution of key aerosol species to the total PM in each SNAP sector. The largest contribution in all sectors of the TNO speciation is "Other primary emissions", i.e. non specific PM. Some definite choice must be made about how to apportion this mass amongst the CLASSIC species. We have apportioned both "fine" and "coarse other" primary PM_{10} to FFBC in the CLASSIC scheme. This choice is somewhat arbitrary and is simply a device to enable all emitted PM to be accounted for. The vertical distribution of aerosol sources are split into high (320 m) and surface sources of sulphate, black carbon and organic carbon fossil fuel, according to data provided by NAEI.

GMDD
5, 3131–3182, 2012

Several other emissions datasets are used in AQUM. For aircraft emissions, a 2002 dataset taken from the AERO2K project as described in Eyers et al. (2004) is used. Biomass burning emissions of aerosols are taken from year 2000 values from the Global Fire Emissions Database (GFED) version 1 (Randerson et al., 2005). Biogenic emissions of isoprene are from Poupkou et al. (2010) at 0.125° resolution and scaled according to the totals provided by the lower resolution GEMS dataset.

2.7 Model configuration for forecasts and hindcasts studies

Beginning in April 2010 AQUM was run in the Met Office's operational forecast suite, carrying out a two-day forecast once a day. The model forecast was initialised with meteorological fields from the 0Z analysis of the NAE model and run for 48 h. The total time taken to run the suite was approximately forty five minutes including the time to prepare the lateral boundary conditions, which combined meteorology from the NAE with chemistry from the GEMS or MACC global forecasts. The actual forecast component took approximately 30 min to run using 2 nodes (64 processors) on an IBM Power 6. This system was used until January 2012, at which point it was upgraded to generate a five day forecast. In this paper we evaluate the results from the first year of operational forecasts from this system.

To supplement the operational forecasts evaluated here, we have also conducted a case study to examine an additional pollution episode (July 2006). This used a similar set up to the forecasts, although due to data availability issues, the initial meteorological analyses were from the global model rather than the NAE. We also examine some episode periods within the year of operational model output in more detail.

GMDD
5, 3131–3182, 2012

3 Model evaluation methodology

3.1 Bias and error metrics

A wide range of methods and metrics for comparing meteorological forecasts with observed quantities have been developed (see for example Wilks, 2006). Mean error (bias) and root mean square error remain important metrics for estimating forecast errors. However when verifying chemical species concentration values some important differences arise compared to verifying standard meteorological fields such as temperature or wind speed. For example, spatial or temporal variations can be much greater and the differences between model and observed values ("model errors") are frequently much larger in magnitude. Under these circumstances it becomes more convenient to work in terms of metrics which can be related to a multiplicative rather than additive error between forecast and observation. Another problem arises when we wish to compare forecast errors for different pollutants: since typical concentrations can vary quite widely between different pollutant types, a given bias or error value can have a quite different significance. It is useful therefore to consider bias and error metrics which are normalised with respect to observed concentrations and hence which can provide a consistent scale regardless of pollutant type. We employ a bias metric termed the "modified normalised mean bias" (MNMB):

$$\mathrm{MNMB} = \frac{2}{N} \sum_{i} \left(\frac{f_i - o_i}{f_i + o_i} \right) \quad (1)$$

In this equation f_i and o_i represent the model (forecast) and observed values respectively at site i or at times i for a given site. The use of a normalisation factor of the mean of the observed and forecast value gives a measure of forecast bias which performs symmetrically with respect to under and over-prediction and is bounded by the values −2 to +2. This approach is adopted by Seigneur et al. (2000), and Cox and Tikvart (1990). It is also useful to understand how the modified normalised mean bias

GMDD
5, 3131–3182, 2012

relates to the multiplicative model error. If we define α_i as the ratio of forecast to observed value

$$f_i = \alpha_i o_i \tag{2}$$

then the mean value of α is given, to a good approximation, by

$$\overline{\alpha} \approx \frac{2 + \mathrm{MNMB}}{2 - \mathrm{MNMB}} \tag{3}$$

Therefore if the model has a MNMB of +1 for example then on average the model predictions are three times the observations, while a MNMB of −0.5 indicates that the forecasts are on average 0.6 times the observations.

Similarly, we use the fractional gross error, FGE, as the indicator of overall forecast error

$$\mathrm{FGE} = \frac{2}{N} \sum_i \left| \frac{f_i - o_i}{f_i + o_i} \right| \tag{4}$$

This is essentially a version of the commonly used "mean absolute error", normalised in a manner which performs symmetrically with respect to under and over-prediction and is bounded by the values 0 to +2. MNMB indicates the extent to which the model systematically under or over-predicts the set of observations, whilst FGE gives a measure of the overall forecast error.

The MNMB and FGE can be combined in a "soccer" plot, which gives a convenient visual representation of the model error characteristics. In these plots (see Fig. 1 for an example) the MNMB is plotted on the x-axis, and FGE on the y-axis. Results for each station are plotted as a point. A perfect forecast would appear as a point at the origin, with the magnitude of any discrepancy increasing with distance from this point. Three boxes mark out maximum bias/error combinations of 15 %/35 %, 30 %/50 %, and 60 %/75 %. A systematic bias appears as a linear grouping of points. If other random sources of error dominate, the resulting pattern will be a scatter of points. This

GMDD
5, 3131–3182, 2012

representation is a convenient way of presenting the statistics across a range of sites, with the quality of the overall forecast and any strong common characteristics or contrasts between the statistics at rural and urban sites being immediately apparent.

An additional metric for comparing forecast and observation fields is the Pearson correlation coefficient (R). This indicates the extent to which temporal patterns in the forecast match those in the observations at a single site or for an ensemble of sites. Another simple metric we use, which is convenient for giving a broad-scale impression of overall forecast skill, is termed "FAC2". This is the fraction of model predictions where the forecast value is within a factor of 2 (either greater or smaller) of the observed value.

3.2 Threshold exceedance skill scores

The verification measures described above provide information about the forecast errors under all conditions, regardless of the magnitude of pollutant concentration. However it is desirable to have metrics which provide information regarding forecast skill specifically at those times when pollutant levels are elevated and pose a greater risk to human health. It is important to assess the skill that models possess in predicting exceedance of given thresholds. The odds ratio is constructed from a standard 2×2 contingency table (Stephenson, 2000) and is defined as:

$$\theta = \frac{ad}{bc} \tag{5}$$

Where is a is the number of correct forecasts of an event, b is number of false alarms, c is the number of missed forecasts and d is the number of correct rejections.

The Odds Ratio Skill Score (ORSS) can be constructed from the odds ratio via a simple transformation

$$\mathrm{ORSS} = \frac{\theta - 1}{\theta + 1} \tag{6}$$

This score ranges from −1 to +1. Forecasts having a strong negative (positive) association with observations have ORSS values tending to −1 (+1), whilst random forecasts

GMDD
5, 3131–3182, 2012

have ORSS tending to zero. In addition the hit rate H (the proportion of events occurring which were correctly forecast) and false alarm rate F (the proportion of forecasts of events occurring which were incorrect forecasts), defined as follows, are valuable metrics for assessing forecast performance:

$$H = \frac{a}{a+c} \tag{7}$$

$$F = \frac{b}{b+d} \tag{8}$$

We have used an hourly average ozone concentration of $100\,\mu g\,m^{-3}$ as the threshold for defining an event in the categorical analyses conducted in Sect. 4. According to the current UK "Daily Air Quality Index" an 8-h rolling mean value of this magnitude is the threshold for the designation of "Moderate" levels of air pollution due to ozone.

Many of the above methods for characterising performance of air quality models were adopted by the GEMS project, based on a report by Agnew et al. (2007), where further discussion of verification issues is given.

4 Results of model evaluation

We have evaluated AQUM against hourly observations of O_3, NO_2, NO, PM_{10}, and $PM_{2.5}$ from the UK Automatic Urban and Rural (AURN) observing network. Observations from around 70 rural, remote, urban background and suburban sites were used, although not all species are measured at every site. More information about the AURN, including the location of all sites, is available at http://uk-air.defra.gov.uk/networks/network-info?view=aurn. The periods analysed are the 12 month period 1 May 2010–30 April 2011 and then the poor air quality episodes in July 2006, June 2010 and April 2011. In addition we have analysed the period June to October 2011 to conduct a comparison of the operational AQUM forecasts with those of the MACC Regional Air Quality Ensemble.

GMDD
5, 3131–3182, 2012

4.1 Evaluation of one year of operational forecasts

4.1.1 Meteorology of May 2010–April 2011

The 12 month period was characterised by a climatologically average start, followed by a relatively unsettled summer 2010. Autumn started warm and settled, but ended cold, leading into an unusually cold December, followed by an average January, and warm February and spring. The period ended with an exceptional spell of warm, settled weather for the time of year, with April 2011 being the warmest on record in the UK and also one of the sunniest and driest. More details are available from the monthly weather summaries provided at http://www.metoffice.gov.uk/climate/uk.

4.1.2 Ozone

Ozone production and build up is favoured by strong sunshine, light winds and elevated temperatures, and thus episodes of high ozone concentrations tend to be more frequent and severe in the summer. However in the Northern Hemisphere the background concentrations are generally highest in the spring (Monks, 2000); this together with enough insolation (necessary for regional ozone production) during those months means that spring-time ozone episodes are also frequently observed. The meteorology was not generally favourable for ozone production in summer 2010, with cool, unsettled and over-cast conditions. In fact the highest levels of ozone during the period occurred during the exceptionally warm April 2011, which saw around twice the frequency of elevated ozone compared to any other month (Table 1).

Model performance metrics for the forecasts during this 12 month period are shown in Table 2 for ozone (first column). The model has a correlation coefficient with observations of 0.68, a modest positive bias of 8.38 $\mu g\,m^{-3}$ and 77 % of model predictions are within a factor of 2 of the observations. The false alarm rate for a threshold of 100 $\mu g\,m^{-3}$ is very low at only 3 % and the hit rate is 57 %. The soccer plot for ozone at urban background (orange) and rural (green) stations is shown in Fig. 1. At urban

GMDD
5, 3131–3182, 2012

stations the model has a positive bias but there is no clear systematic bias for rural stations. Both bias and other sources of error contribute to the fractional gross error, and both are higher for the urban than the rural stations. These results will be interpreted in the following section in the context of the model's performance for NO_x.

4.1.3 NO_2

The model performance metrics for NO_2 are also shown in Table 2. The correlation coefficient of 0.57 is lower than for ozone and there is a negative bias of $-6.10\,\mu g\,m^{-3}$ which is of a similar magnitude (but opposite in sign) to that of ozone. However as NO_2 concentrations are lower than ozone, the magnitude of the MNMB is much greater, with values of 0.12 for ozone and −0.26 for NO_2 (corresponding to α = 1.13 and 0.77 respectively).

Figure 2 shows the soccer plot for NO_2. There is a large negative bias at urban sites which dominates the overall error at these sites. At rural sites there is generally a positive bias, but the error displays a more random characteristic rather than the systematic trend for urban sites. It should be borne in mind that NO_2 measurements made using the chemiluminescence technique with molybdenum converters (as used in the AURN) may over-estimate the true concentration by up to 30 % (Steinbacher et al., 2007).

As a regional air quality model at ∼ 12 km resolution, AQUM does not adequately resolve the sources of primary NO and NO_2 emission (typically dominated by road transport and combustion at point sources). In view of this we have not presented a systematic evaluation of model NO predictions. However it is worthwhile noting that there is a strong negative bias for NO predictions which dominates the error characteristics. This pattern of over-estimation of NO_x at rural sites and under-estimation at urban ones is consistent with the model resolution being too coarse to properly resolve sources of NO_x. In an Eulerian model primary emissions are instantaneously spread over an entire grid box, thus giving apparently lower concentrations close to source regions than occur in reality. Corresponding with this there is a spurious increase in

GMDD
5, 3131–3182, 2012

concentrations further away from the source regions. These effects combine to give the pattern of biases observed for primary pollutants at rural and urban sites. This aspect of model performance is likely to improve as model resolution increases. The under-prediction of NO could cause the under-estimation of the ozone loss by titration in the model, which would be consistent with the positive bias found for this species at urban sites.

4.1.4 Particulate matter

The model performance statistics for PM_{10} in Table 2 show that overall it is the most challenging pollutant to model accurately. It has the lowest correlation coefficient (0.52) and the greatest negative bias (MNMB = −0.67), which implies that on average the model predictions for PM_{10} are only half of the observed concentrations. FGE (0.83) is also the highest of the three pollutants.

Figure 3 shows the soccer plot for PM_{10}. Urban sites consistently have a negative bias and although there are few data available from rural sites, a negative bias is generally exhibited. The under-forecasting of concentrations of particulate matter is a widespread problem in most present-day forecasts systems. Inspection of the MACC regional ensemble models (http://macc-raq.gmes-atmosphere.eu/) shows that all models exhibit a negative bias to some extent. Although there are currently too few observational data available for $PM_{2.5}$ to make a systematic comparison, the available evidence suggests that this component of PM has a smaller negative bias in AQUM. Thus most of the under-represented PM_{10} is in the coarse component from $PM_{2.5}$ to PM_{10}. Emissions in this size regime are typically due to sea salt, wind-blown dust and matter re-suspended by road transport. These last two components of PM emission dependent sensitively on the assumptions made regarding surface properties and are difficult to model accurately.

GMDD
5, 3131–3182, 2012

4.2 Model evaluation during pollution episodes

A key requirement of modelling and forecast systems is the ability to represent the rapid rise and fall of pollutant concentrations which occur around episodes of poor air quality. Whilst most models can be tuned to give reasonable monthly or annual averages, a more discriminating test is whether models can respond in episode conditions and demonstrate a wide dynamic range, predicting the onset and termination of elevated pollutant concentrations. In the next two sections we assess the performance of the model during periods of moderate and high ozone by comparing the year of operational data, May 2010–April 2011 as a whole, to individual months where significant episodes occurred, in June 2010 and April 2011 as well as the additional month of July 2006 which was modelled in hindcast mode. The geographical locations of particular sites for which we show results are depicted in Fig. 4.

4.2.1 July 2006

This month was exceptionally warm and sunny and therefore produced some of the most significant ozone episodes that the UK has experienced since 2003. Consequently this month is a demanding test of the model because ozone levels were particularly high at their peak and because the ozone episodes came in separate phases. There were 3 significant ozone episodes, separated by days when ozone levels were low. This period is therefore a good test of the model's dynamic range in modelling the rapid build-up of ozone, the maintenance of high levels during the episode and the reduction at the end.

Figure 5 shows modelled and observed hourly ozone for Aston Hill, a rural site on the border between England and Wales. At this site there are 3 ozone episodes; an initial one from the 1st to the 5th July, a second one from the 17th to the 20th and third more modest episode covering 24–25 July. The model generally agrees well with the observations at this site throughout the month, both in terms of predicting actual ozone levels and episode duration. However it did not predict the highest concentration occurring on

GMDD
5, 3131–3182, 2012

19 July. The low values of ozone between the episodes are well reproduced, showing that the model is able to capture abrupt changes in ozone concentration as episodes conditions arise and then dissipate. Similar results were found for most rural and urban sites. Figure 6 summarises the good model performance for ozone at most sites during July 2006; the summary performance statistics are given in Table 3. The bias of 1.99 $\mu g\,m^{-3}$ is particularly low compared to the 12 month period.

A key requirement for a forecast system is to be able to predict ozone concentration levels greater than a given threshold. Using a threshold of 100 $\mu g\,m^{-3}$, Table 3 also shows the categorical metrics (hit rate, false alarm rate, ORSS) for July 2006. Compared to the 12 month period metrics (Table 2), the hit rate is much higher for July 2006 (0.71 versus 0.57 respectively). In view of the fact that there is a lower positive bias, this demonstrates that the model predicts this episode well.

4.2.2 June 2010

The weather over the UK in June 2010 was mainly dry and sunny, particularly in the second half when it became very warm, reaching a maximum of 30.9 °C in Gravesend (South-East England) on 27 June. Although sunshine levels were below average in Scotland, they were around 50 % above average in South Wales and South-West England, where it was the third sunniest June since 1929. The majority of the rainfall occurred during the second week; it became unsettled at the very end of the month (Met Office, UK weather summary for June 2010, http://www.metoffice.gov.uk/climate/uk/2010/june.html).

There were two main poor air quality episodes this month, with high levels of both ozone and PM_{10}. The first period was from the 3rd to the 6th June, during which elevated levels of all the key air quality pollutants were observed across Southern England, with ozone reaching a maximum of 172 $\mu g\,m^{-3}$ at Weybourne on the 6th. PM_{10} peaked at 96 $\mu g\,m^{-3}$ in Thurrock on the 5th (see Fig. 7). The model captures this first episode well, both in timing and in magnitude. There were other short-lived peaks of PM throughout the month and in general the model exhibited a negative bias. However

GMDD
5, 3131–3182, 2012

from 22nd–28th a longer duration episode occurred. During this episode PM_{10} reached a maximum of 89 $\mu g\,m^{-3}$ in Leamington Spa on the 28th, while a peak value of ozone was recorded in Weybourne on the 27th, at a concentration of 194 $\mu g\,m^{-3}$; 40 other sites observed peak ozone concentrations of above 100 $\mu g\,m^{-3}$ and of these, 6 sites measured ozone concentrations higher than 150 $\mu g\,m^{-3}$. Figure 8 shows the time series of ozone concentrations for the Harwell site (a rural location around 30 miles west of London) and illustrates the extent of the episode. Here the model captures the general characteristics and higher peak concentrations of the episode well. This is reflected in the high ORSS score of 0.95. The ozone prediction performance statistics for the whole month are shown in Table 3.

A contour plot showing the daily maximum values of ozone across the model domain is shown for 27 June in Fig. 9. In this figure the observed daily maxima are overplotted as colour-coded squares. It can be seen that the model predicts ozone levels higher than 150 $\mu g\,m^{-3}$ across a large swathe of South-Eastern England, compared to observed concentrations, where only two sites – Weybourne and Sibton – actually reached these levels. This trend to over-forecast ozone levels is continued across the entire month, as indicated by the relatively large positive model bias of 20.37 $\mu g\,m^{-3}$ (see Table 3) for the thirty day period. This also results in a high hit rate of 0.86 and a false alarm rate somewhat higher than for other episodes (0.14).

4.2.3 April 2011

The meteorology of May 2010–April 2011 was not generally conducive to the build-up of ozone, with the summer lacking extended periods of clear skies and high temperatures. Instead, a period of elevated ozone occurred during April 2011, which was unusually warm and sunny. The combination of these conditions with the elevated background concentrations noted above resulted in some of the poorest air quality episodes over the UK for the whole of 2011. The elevated ozone levels occurred together with a major PM episode. Meteorologically, high pressure dominated the UK weather throughout the month, resulting in mainly fine, warm weather. Daily

GMDD
5, 3131–3182, 2012

maximum temperatures were well above normal – by as much as 6 °C in South-East England, with a maximum of 27.8 °C recorded in Wisley, Surrey, on 23 April. It was also one of the driest and sunniest months of April on record, although Scotland had near to above normal rainfall (Met Office, UK weather summary for April 2011, http://www.metoffice.gov.uk/climate/uk/2011/april.html).

There were widespread elevated ozone levels which peaked during the period 20th–23rd. Due to the contribution of background ozone levels the onset of the episode is not especially pronounced, as demonstrated by the time series for Harwell shown in Fig. 10, where ozone levels were generally high throughout the month. Figure 11 shows the model error characteristics for April 2011 for ozone, and other statistics are given in Table 3. While the bias and RMSE are generally comparable to the other episodes, the hit rate is significantly lower. This is likely to be because the ozone concentrations were close to the threshold value for much of April, so that small errors in the model forecast concentration values could often result in incorrect classification as a hit or false alarm.

The most notable feature for April 2011 was the major PM episode which occurred approximately 18–23 April and affected the whole of the UK. A maximum PM_{10} concentration of 142 $\mu g\,m^{-3}$ was observed in Thurrock on 21 April. A time series at this site of modelled PM_{10} concentrations compared to observed PM_{10} is shown in Fig. 12. AQUM predicts the overall evolution of the episode well, but under-predicts the observed concentrations on the three days which saw the maximum PM_{10} levels.

Provisional speciated PM observations are available for this episode at the rural Harwell site (S. Telling, personal communication, 2011). A time series plot of measured $PM_{2.5}$ and its components at this site is shown in Fig. 13. On 22 April PM_{10} reached a maximum concentration of 105 $\mu g\,m^{-3}$; most of this was in the $PM_{2.5}$ component with a concentration of 98 $\mu g\,m^{-3}$. The largest component of $PM_{2.5}$ was nitrate aerosol, with a peak concentration of 56 $\mu g\,m^{-3}$. The modelled speciated $PM_{2.5}$ concentrations are shown in Fig. 14 for comparison with the observations. AQUM correctly predicts the overall magnitude of $PM_{2.5}$ and the relative contributions of nitrate, ammonium and sulphate aerosol components. However the model does not predict the worsening of

GMDD
5, 3131–3182, 2012

$PM_{2.5}$ values from the 20th to the 22nd and significantly over-predicts values on 23 April and other days.

4.3 Comparison with the MACC regional air quality ensemble

MACC (Monitoring Atmospheric Composition and Climate) was a project funded under the European Union Seventh Framework Programme FP7 to develop and implement trial elements of an atmospheric composition and climate service. One element of MACC is a European air quality forecast service. Seven different models forecasting for a European domain contribute to an ensemble median forecast out to three days ahead. In order to place the performance of AQUM in the context of other similar air quality forecast models we have conducted a comparison between AQUM and the MACC ensemble forecast over the period June to October 2011.

Summer 2011 was generally cooler and slightly wetter than average, in particular it was wetter than 2010. However there were some periods of fine weather, and the last few days of September and the first week in October were very warm and sunny for this time of year. There were several periods of elevated ozone during June–October: in both early and late July as well as at the end of September and early October. Both AQUM and the MACC ensemble captured these events fairly well (see, for example, a time series of ozone at Harwell in Fig. 15). The times series plot in Fig. 16 compares the ozone concentration bias for the MACC ensemble and AQUM. In general AQUM has a higher bias than the MACC ensemble, with mean values over the whole period of $13.28\,\mu g\,m^{-3}$ for AQUM compared to $4.10\,\mu g\,m^{-3}$ for the MACC ensemble. This figure illustrates that the positive bias of both model systems rises during the episode periods. Table 4 shows a summary of performance metrics. The performance of the MACC ensemble is somewhat better than that of AQUM for most metrics (hit rate is a notable exception). However the difference between the two model systems is not large. One would expect the ensemble performance of a collection of well configured models to be better than that of any single member, hence the greater skill of the ensemble forecast is not surprising. The hit rate of AQUM (where a hit is defined as the forecast of

GMDD
5, 3131–3182, 2012

exceeding the 100 µg m^{-3} ozone concentration threshold) is 0.64, significantly better than the value of 0.27 achieved by the MACC ensemble. This is clearly related to the higher bias of AQUM, but for a forecast model designed for issuing health impact warnings the higher hit rate is arguably a more important characteristic than a lower bias, as long as the false alarm rate does not increase unacceptably as a result. The false alarm rate for AQUM is only 4 %, whilst the MACC ensemble has no false alarms.

4.4 Variation of model skill with forecast lead time

A further area where we have evaluated AQUM is to examine the variation of skill with forecast lead time. AQUM operational forecasts extend out to two days, thus it is possible to compare forecasts made two days ahead with those made one day ahead. We have analysed AQUM day 1 and day 2 ozone forecasts over the period May 2010 to April 2011 and the results are presented in Table 5. In contrast to meteorological variables, where one generally finds a significant decrease in forecast skill with lead time, ozone forecasts exhibit a weak dependence on lead-time for all metrics. This is consistent with our general observation that, for air quality forecasting, a 24 h persistence forecast (i.e. assuming the next day has the same air quality as the current day) usually exhibits a substantial level of skill. The contrasting behaviour with meteorological forecasts indicates that the factors controlling errors differ in the two types of forecast, and that the impact of typical errors in meteorology does not dominate other sources of error in ozone forecasts, such as emissions or the representation of atmospheric chemistry.

5 Summary and future developments

We have presented a description of a new on-line air quality model AQUM, which is based on the Met Office Unified Model and uses the UKCA sub-model for describing atmospheric chemistry processes. A variety of metrics for assessing model performance

GMDD
5, 3131–3182, 2012

have been described and the importance of using metrics which assess both mean performance and skill in predicting exceedance of threshold concentration values is emphasised. We have evaluated AQUM against routine, hourly observations from the UK AURN observing network. Averaged over the course of a full year, the model exhibits a positive bias for ozone of around 8 $\mu g\,m^{-3}$. The model exhibits good dynamic range in simulating ozone and case studies of elevated ozone episodes demonstrate that the model reproduces time series of measured ozone concentrations at individual sites well. For NO_2 the model exhibits a negative bias for urban sites and positive bias for rural sites. This is likely to be a consequence of the fact that, at 12 km resolution, AQUM does not adequately resolve the main sources of NO_x (i.e. road traffic and combustion point sources). This results in the dilution of emissions close to source regions (urban areas) and enhanced emissions in regions distant from sources (rural areas). For PM_{10} the model generally exhibits a negative bias, in common with many air quality models. This is likely to be the result of emissions which are not represented in the annual average inventories, such as re-suspension of deposited coarse PM, sea salt or wind blown dust. However evaluation of the model during particular episodes, such as April 2011, demonstrates that it can represent fine PM reasonably well.

AQUM is being actively developed by the Met Office and UKCA academic partners. Priority areas for future developments include: (i) an improved representation of emissions, allowing different vertical and temporal profiles to be applied according to sector and interactive emissions of biogenic VOCs; (ii) implementation of the modal aerosol scheme "UKCA-GLOMAP-mode" (Mann et al., 2010). This more sophisticated scheme will allow the time evolution of aerosol modes, the separate modelling of aerosol mass and number and an improved representation of sea salt and prognostic secondary organic aerosols; (iii) the implementation of a post-processing system to apply a bias correction to forecasts. The observation that a 24 h persistence forecast for air quality generally displays considerable skill suggests that measured values from the previous 24 h period can be used to derive a bias correction to new forecasts. We have begun development work to explore the potential of this and initial results appear promising.

GMDD
5, 3131–3182, 2012

(iv) A final priority for development is to increase the model resolution. The Met Office currently runs a 1.5 km resolution, 70 level meteorological model which resolves smaller scale convection and gives an improved forecast for precipitation. In the near future we plan to develop a version of AQUM at this resolution. In addition to providing improvements in the meteorological parameters in AQUM this will allow an improved representation of emissions.

Supplementary material related to this article is available online at: http://www.geosci-model-dev-discuss.net/5/3131/2012/gmdd-5-3131-2012-supplement.pdf.

Copyright statement

Acknowledgements. The development of AQUM has only been made possible by initial developments of the UKCA project. In particular we wish to acknowledge the contributions made by O. Morgenstern, P. Young, G. Zeng and J. Pyle. The development of the UKCA model was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) and the Natural Environment Research Council, through the NCAS initiative. Funding for AQUM is provided by the Met Office Public Weather Service. We are indebted to M. Sanderson for his assistance in the initial implementation of the RAQ chemistry scheme in AQUM. The speciated particulate matter data used in this paper are produced under contract to Defra and were kindly provided by S. Telling at AEA Technology and M. Twigg and C. Braban at the Centre for Ecology and Hydrology, Edinburgh. Data from the AURN are provided under contract to Defra by AEA Technology. Data from the MACC project were kindly provided by F. Cheroux of Meteo-France. This study contributes to COST Action ES1004 on online integrated modelling.

GMDD
5, 3131–3182, 2012

References

Ackermann, I. H., Hass, H., Memmesheimer, M., Ziegenbein, C., and Ebel, A.: The parametrization of the sulphate-nitrate-ammonia aerosol system in the long-range transport model EURAD, Meteorol. Atmos. Phys., 57, 101–114, 1995.

Agnew, P., Mittermaier, M. P., Honore, C., Elbern, H., Coll, I., Vautard, R., and Peuch., V.-H.: Evaluation of GEMS regional air quality forecasts, GEMS project report, available at: http://gems.ecmwf.int/do/get/PublicDocuments/1533/1402?showfile=true (last access: 2 October 2012), 2007.

Andres R. J. and Kasgnoc A. D.: A time-averaged inventory of sub-aerial volcanic sulfur emissions. J. Geophys. Res.-Atmos, 103, 25251–25261, 1998.

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., and Troe, J.: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume I – gas phase reactions of O_x, HO_x, NO_x and SO_x species, Atmos. Chem. Phys., 4, 1461–1738, doi:10.5194/acp-4-1461-2004, 2004.

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., Troe, J., and IUPAC Subcommittee: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II – gas phase reactions of organic species, Atmos. Chem. Phys., 6, 3625–4055, doi:10.5194/acp-6-3625-2006, 2006.

Bellouin, N., Rae, J., Jones, A., Johnson, C., Haywood, J., and Boucher, O.: Aerosol forcing in the Climate Model Intercomparison Project (CMIP5) simulations by HadGEM2-ES and the role of ammonium nitrate, J. Geophys Res.-Atmos., 116, D20206, doi:10.1029/2011JD016074, 2011.

Bush, M., Bell, S., Christidis, N., Renshaw, R., MacPherson, B., and Wilson, B.: Development of the North Atlantic European model (NAE) into an operational model, Forecasting Research Technical Report No. 47, Met Office, UK, 2006.

Collins, W. J., Stevenson, D. S., Johnson, C. E., and Derwent, R. G.: Tropospheric ozone in a global-scale three-dimensional Lagrangian model and its response to NO_x emission controls, J. Atmos. Chem., 26, 223–274, 1997.

Collins, W. J., Stevenson, D. S., Johnson, C. E., and Derwent, R. G.: Role of convection in determining the budget of odd hydrogen in the upper troposphere, J. Geophys. Res.-Atmos., 104, 26927–26941, 1999.

GMDD
5, 3131–3182, 2012

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4, 1051–1075, doi:10.5194/gmd-4-1051-2011, 2011.

Cox, W. M. and Tikvart, J. A.: A statistical procedure for determining the best performing air quality simulation model, Atmos. Environ., 24, 2387–2395, 1990.

Davies, H. C.: A lateral boundary formulation for multi-level prediction models, Q. J. Roy. Meteor. Soc., 102, 405–418, 1976.

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere, Q. J. Roy. Meteor. Soc., 131, 1759–1782, 2005.

Davies T.: Lateral boundary conditions for limited area models, Q. J. Roy. Meteor. Soc., in preparation, 2012.

Derwent, R. G., Jenkin, M. E., Collins, W. J., Johnson, C. E., and Stevenson, D. S.: The global distribution of secondary particulate matter in a 3-D Lagrangian chemistry transport model, J. Atmos. Chem., 44, 57–95, 2003.

Dore, C. J., Murrells, T. P., Passant, N. R., Hobson, M. M., Thistlethwaite, G., Wagner, A., Li, Y., Bush, T., King, K. R., Norris, J., Coleman, P. J., Walker, C., Stewart, R. A., Tsagatakis, I., Conolly, C., Brophy, N. C. J., and Hann, M. R.: UK emissions of air pollutants 1970 to 2006, available at: http://uk-air.defra.gov.uk/reports/cat07/0810291043_NAEI_2006_Report_Final_Version(3).pdf (last access: 2 October 2012), 2008.

Edwards J. M. and Slingo, A.: Studies with a flexible new radiation code, 1, choosing a configuration for a large-scale model, Q. J. Roy. Meteor. Soc., 122, 689–719, 1996.

Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M., and Taylor, C. M.: Explicit representation of subgrid heterogeneity in a GCM land surface scheme, J. Hydrometeorol., 4, 530–543, 2003.

Eyers, C. J., Norman, P., Middel. J., Plohr, M., Atkinson, K., and Christou, R. A.: Aero2k global aviation emission inventories for 2002 and 2025, available at: http://www.aero-net.org/pdf-docs/AERO2K_Global_Aviation_Emissions_Inventories_for_2002_and_2025.pdf (last access: 16 January 2006), 2004.

Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G., and Stein, O.: Coupling global chemistry transport models to ECMWF's integrated forecast system, Geosci. Model Dev., 2, 253–265, doi:10.5194/gmd-2-253-2009, 2009.

GMDD
5, 3131–3182, 2012

Giannakopoulos, C.: Three dimensional modelling of the concentration and deposition of tropospheric trace gases, Ph.D. thesis, University of Cambridge, Cambridge, UK, 1998.
Giannakopoulos, C., Chipperfield, M. P., Law, K. S., and Pyle, J. A.: Validation and intercomparison of wet and dry deposition schemes using ^{210}Pb in a global three-dimensional off-line chemical transport model, J. Geophys. Res.-Atmos., 104, 23761–23784, 1999.
Gregory, D. and Rowntree, P. R.: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure, Mon. Weather Rev., 118, 1483–1506, 1990.
Jones, A., Roberts, D. L., Woodage, M. J., and Johnson, C. E.: Indirect sulphate aerosol forcing in a climate model with an interactive sulphur cycle, J. Geophys. Res.-Atmos., 106, 20293–20310, 2001.
Hollingsworth, A., Engelen, R. J., Benedetti, A., Dethof, A., Flemming, J., Kaiser, J. W., Morcrette, J.-J., Simmons, A. J., Textor, C., Boucher, O., Chevallier, F., Rayner, P., Elbern, H., Eskes, H., Granier, C., Peuch, V.-H., Rouil, L., and Schultz, M. G.: Toward a monitoring and forecasting system for atmospheric composition: the GEMS project, B. Am. Meteorol. Soc., 89, 1147–1164, 2008.
Kettle, A. J., Andreae, M. O., Amouroux, D., Andreae, T. W., Bates, T. S., Berresheim, H., Bingemer, H., Boniforti, R., Curran, M. A. J., DiTullio, G. R., Helas, G., Jones, G. B., Keller, M. D., Kiene, R. P., Leck, C., Levasseur, M., Malin, G., Maspero, M., Matrai, P., McTaggart, A. R., Mihalopoulos, N., Nguyen, B. C., Novo, A., Putaud, J. P., Rapsomanikis, S., Roberts, G., Schebeske, G., Sharma, S., Simo, R., Staubes, R., Turner, S., and Uher, G.: A global database of sea surface dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month, Glob. Biogeochem. Cy., 13, 399–444, 1999.
Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., and Smith, R. N. B.: A new boundary layer mixing scheme, Part I: Scheme description and single-column model tests, Mon. Weather Rev., 128, 3187–3199, 2000.
MacCarthy, J., Pang, Y., Murrells, T. P., Passant, N., Martinez, C., Thomas, J., Thistlethawaite, G., and Misselbrook, T.: Air quality pollutant inventories for England, Scotland, Wales and Northern Ireland: 1990–2009, AEA report number AEAT/ENV/R/3225, 2011.
Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., and Johnson, C. E.: Description and evaluation of GLOMAP-mode: a modal

GMDD
5, 3131–3182, 2012

global aerosol microphysics model for the UKCA composition-climate model, Geosci. Model Dev., 3, 519–551, doi:10.5194/gmd-3-519-2010, 2010.
Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 1: The stratosphere, Geosci. Model Dev., 2, 43–57, doi:10.5194/gmd-2-43-2009, 2009.
Monks, P. S.: A review of the observations and origins of the spring ozone maximum, Atmos. Environ., 34, 3545–3561, 2000.
O'Connor, F. M., Johnson, C. E., Morgenstern, O., Sanderson, M. G., Young, P. J., Zeng, G., Collins, W. J., and Pyle, J. A., Evaluation of the new UKCA climate-composition model, Part 2, The Troposphere, in preparation, 2012.
O'Dowd, C. D., Lowe, J. A., and Smith, M. H.: Coupling sea-salt and sulphate interactions and its impact on cloud droplet concentration predictions, Geophys. Res. Lett., 26, 9, doi:10.1029/1999GL900231, 1999.
Poupkou, A., Giannaros, T., Markakis, K. Kioutsioukis, I., Curci, G., Melas, D., and Zerefos, C.: A model for European biogenic volatile organic compound emissions: Software development and first validation, Environ. Modell. Softw., 25, 1845–1856, 2010.
Price, J. D., Vosper, S., Brown, A., Ross, A., Clark, P., Davies, F., Horlacher, V., Claxton, B., McGregor, J. R., Hoare, J. S., Jemmett-Smith, B., and Sheridan, P.: COLPEX: field and numerical studies over a region of small hills, B. Am. Meteorol. Soc., 92, 1636–1650, 2011.
Randerson, J. T., Kasibhatla, P. S., Kasischke, E. S., Hyer, E. J., Giglio, L., Collatz, G. J., and van der Werf, G. R.: Global fire emissions database (GFED), version 1, available at: http://daac.ornl.gov (last access: 2 October 2012), 2005.
Seigneur, C., Pun, B., Pai, P., Louis, J.-F., Solomom, P., Emery, C., Morris, R., Zahniser, M., Worsnop, D., Koutrakis, P., White, W., and Tombach, I.: Guidance for the performance evaluation of three dimensional air quality modelling systems for particulate matter and visibility, J. Air Waste Ma., 50, 588–599, 2000.
Smith, R. N. B.: A scheme for predicting layer clouds and their water-content in a general-circulation model, Q. J. Roy. Meteor. Soc., 116, 435–460, 1990.
Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevot, A. S. H., and Hueglin, C.: Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, J. Geophys. Res.-Atmos., 112, D11307, doi:10.1029/2006JD007971, 2007.

GMDD
5, 3131–3182, 2012

Stephenson, D.: Use of the "Odds Ratio" for diagnosing forecast skill, Weather Forecast., 15, 221–232, 2000.
Visschedijk, A., Zanveld, P., and van der Gon, H.: A high resolution gridded European emission database for the EU integrated project GEMS, TNO report 2007-A-R0233/B., 2007.
Wanninkhof, R.: Relationship between wind-speed and gas-exchange over the ocean, J. Geophys. Res.-Oceans, 97, 7373–7382, 1992.
Wesley M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical-models, Atmos. Environ., 23, 1293–1304, 1989.
Whall, C., Scarborough, T., Stavrakaki, A., Green, C., Squire, J., and Noden, R.: UK Ship Emissions Inventory, Entec UK Ltd, London, UK, available at: http://uk-air.defra.gov.uk/reports/cat15/1012131459_21897_Final_Report_291110.pdf (last access: 2 October 2012), 2010.
Wild, O., Zhu, X., and Prather, M. J.: Fast-J: accurate simulation of in- and below-cloud photolysis in tropospheric chemical models, J. Atmos. Chem., 37, 245–282, 2000.
Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, 2nd Edn., Academic Press, London, 2006.
Wilson, D. R. and Ballard, S. P.: A micro-physically based precipitation scheme for the UK Meteorological Office Unified Model, Q. J. Roy. Meteor. Soc., 125, 1607–1636, 1999.

GMDD
5, 3131–3182, 2012

Table 1. Prevalence of ozone episode conditions by month between 1 May 2010 to 30 April 2011. For each month, the percentage of hourly average observations (across all sites) where the ozone concentration exceeds the given threshold is shown.

Month	Percentage of obs $\geq 80\,\mu g\,m^{-3}$	Percentage of obs $\geq 100\,\mu g\,m^{-3}$	Percentage of obs $\geq 120\,\mu g\,m^{-3}$	Percentage of obs $\geq 150\,\mu g\,m^{-3}$
May 2010	17.767	1.825	0.184	0.008
Jun	15.762	5.190	1.621	0.176
Jul	3.640	0.856	0.285	0.046
Aug	2.414	0.396	0.061	0.027
Sep	4.247	0.337	0.024	0.0
Oct	1.905	0.044	0.0	0.0
Nov	1.59	0.0	0.0	0.0
Dec	0.873	0.003	0.0	0.0
Jan 2011	3.509	0.038	0.013	0.005
Feb	7.707	0.054	0.0	0.0
Mar	11.068	0.275	0.003	0.0
Apr	34.785	9.248	2.790	0.356

GMDD
5, 3131–3182, 2012

Table 2. Model performance metrics for the period 1 May 2010 and 30 April 2011. These statistics are based on all hourly values in each day. An ozone threshold of 100 $\mu g m^{-3}$ was used in the calculation of the ORSS, hit rate and false alarm rate metrics. The categorical metrics are presented only for ozone; these metrics add little value for interpreting the results for NO_2 and PM_{10} due to the large negative bias in model predictions for these species.

Metric	O_3	NO_2	PM_{10}
Correlation	0.68	0.57	0.52
Bias ($\mu g m^{-3}$)	8.38	−6.10	−9.17
RMSE ($\mu g m^{-3}$)	22.83	18.66	17.28
MNMB	0.12	−0.26	−0.67
FGE	0.49	0.69	0.83
FAC2	0.77	0.57	0.43
ORSS	0.95	–	–
Hit rate	0.57	–	–
False alarm rate	0.03	–	–

GMDD
5, 3131–3182, 2012

Table 3. Model performance metrics for ozone for the three case study periods, July 2006, June 2010 and April 2011. An ozone threshold of 100 $\mu g\,m^{-3}$ was used in the calculation of the ORSS, hit rate and false alarm rate metrics.

Metric	Jul 2006	Jun 2010	Apr 2011
Correlation	0.75	0.66	0.59
Bias ($\mu g\,m^{-3}$)	1.99	20.37	7.09
RMSE ($\mu g\,m^{-3}$)	25.94	28.26	24.89
MNMB	0.07	0.33	0.14
FGE	0.31	0.38	0.35
FAC2	0.89	0.86	0.86
ORSS	0.92	0.95	0.80
Hit rate	0.71	0.86	0.43
False alarm rate	0.09	0.14	0.08

GMDD
5, 3131–3182, 2012

Table 4. Performance metrics for AQUM and MACC ensemble ozone forecasts for the period 1 June to 31 October 2011.

Metric	AQUM	MACC Ensemble
Correlation	0.62	0.71
Bias ($\mu g\,m^{-3}$)	13.28	4.10
RMSE ($\mu g\,m^{-3}$)	22.68	15.71
FAC2	0.85	0.89
ORSS	0.96	0.99
Hit rate	0.64	0.27
False alarm rate	0.04	0.0

GMDD
5, 3131–3182, 2012

Table 5. AQUM performance metrics for ozone forecasts over the period 1 May 2010 to 30 April 2011, illustrating variation with forecast lead time.

Metric	Day 1	Day 2
Correlation	0.68	0.66
Bias ($\mu g m^{-3}$)	8.38	9.04
RMSE ($\mu g m^{-3}$)	22.83	23.79
FAC2	0.77	0.76
ORSS	0.95	0.95
Hit rate	0.57	0.56
False alarm rate	0.03	0.04

GMDD
5, 3131–3182, 2012

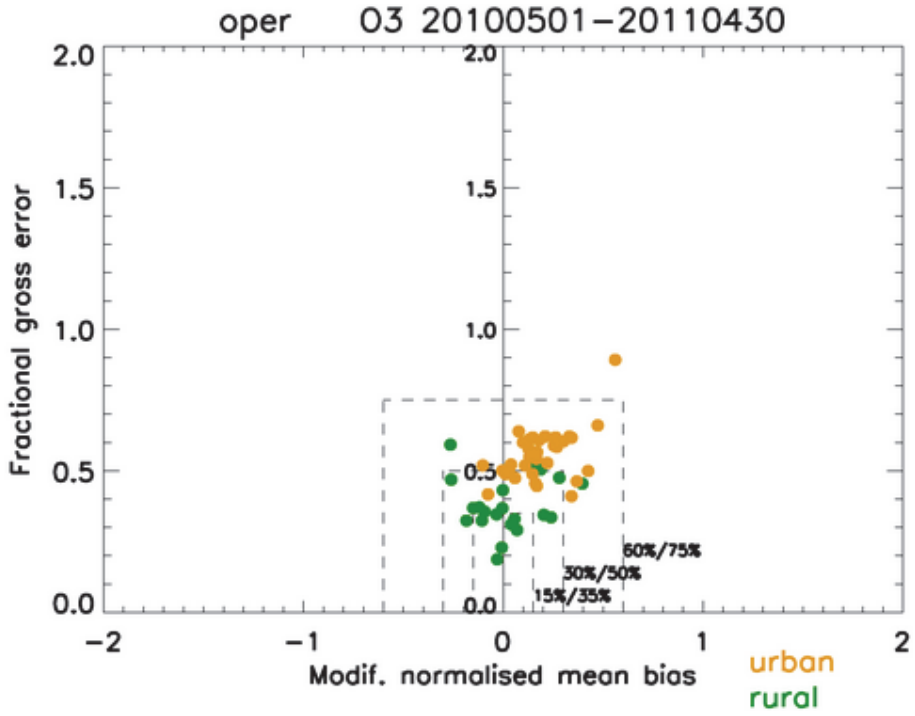

Fig. 1. Soccer plot showing fractional gross error as a function of the modified normalised mean bias relative to hourly observations of ozone for the period 1 May 2010 to 30 April 2011.

GMDD
5, 3131–3182, 2012

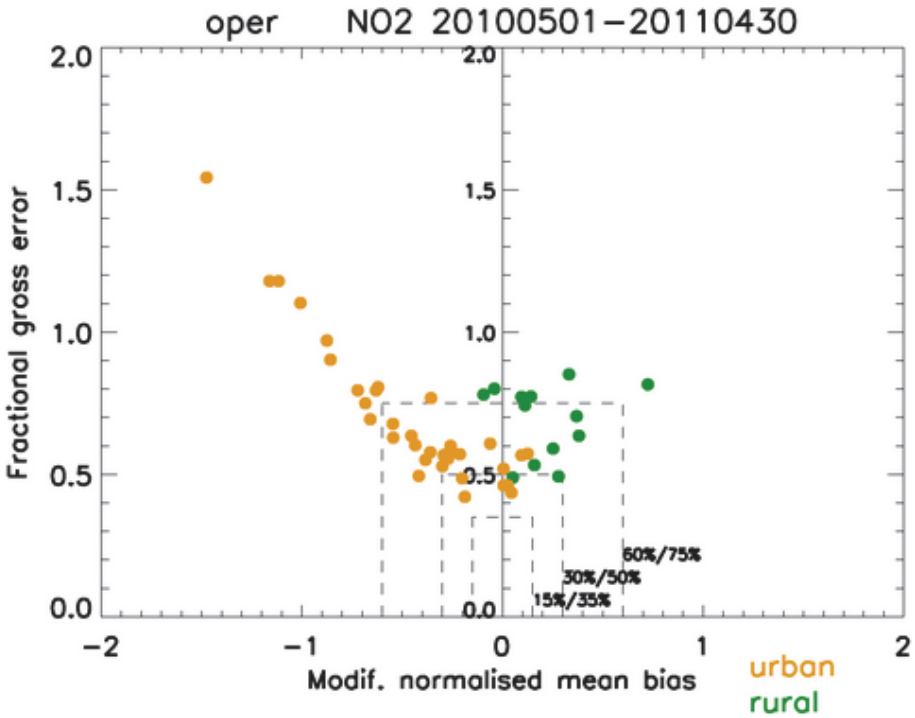

Fig. 2. Soccer plot for NO_2 for the period 1 May 2010 to 30 April 2011.

GMDD
5, 3131–3182, 2012

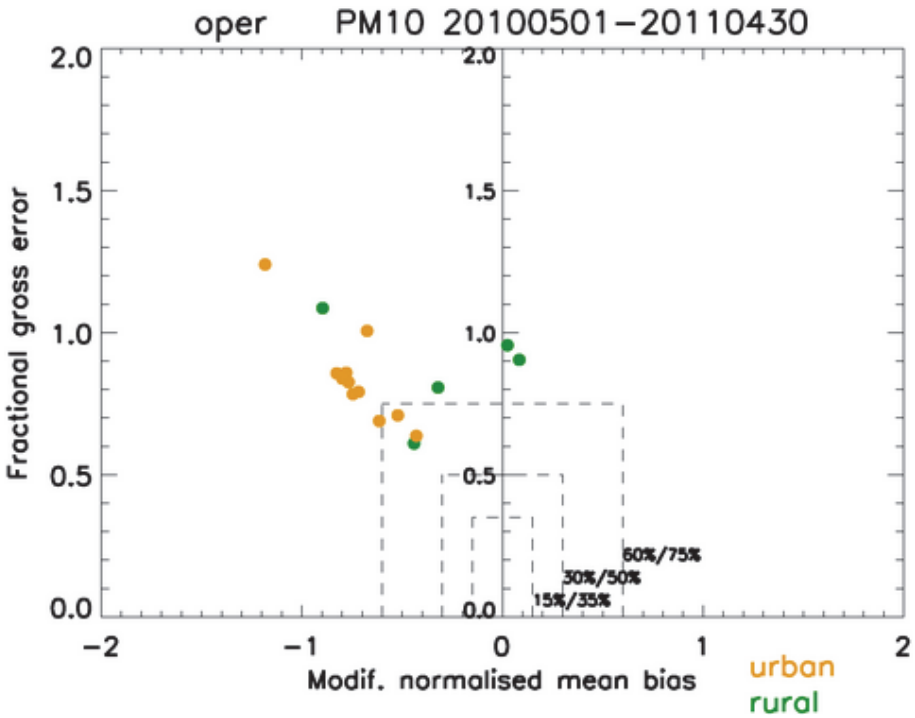

Fig. 3. Soccer plot for PM_{10} for the period 1 May 2010 to 30 April 2011.

GMDD
5, 3131–3182, 2012

AQUM model description

N. H. Savage et al.

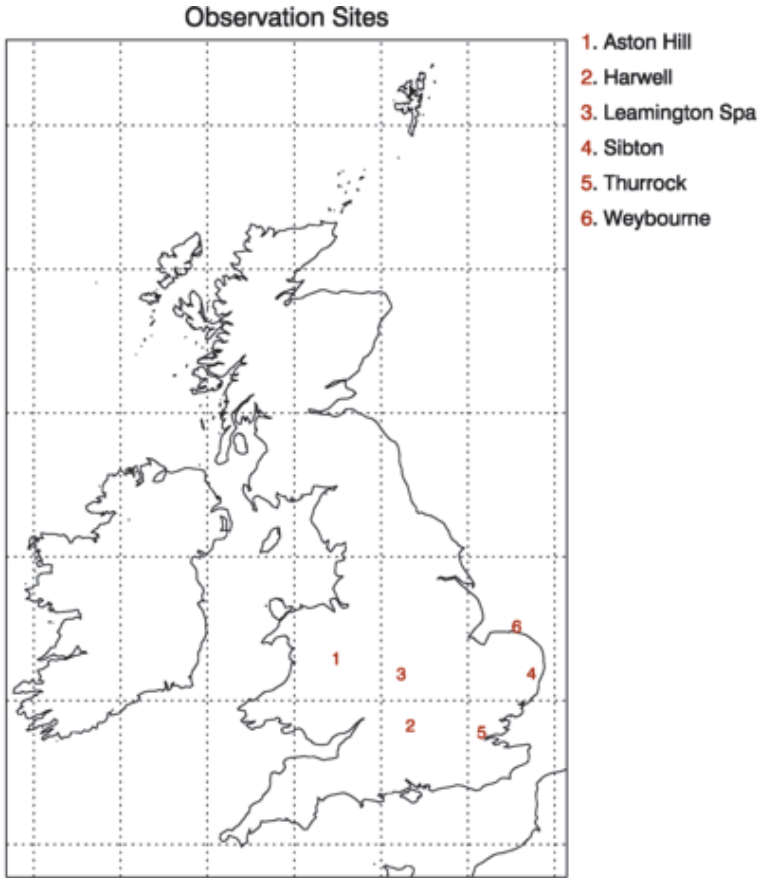

Fig. 4. Location of specific air quality observing sites in the UK referred to in the text and other Figures.

GMDD
5, 3131–3182, 2012

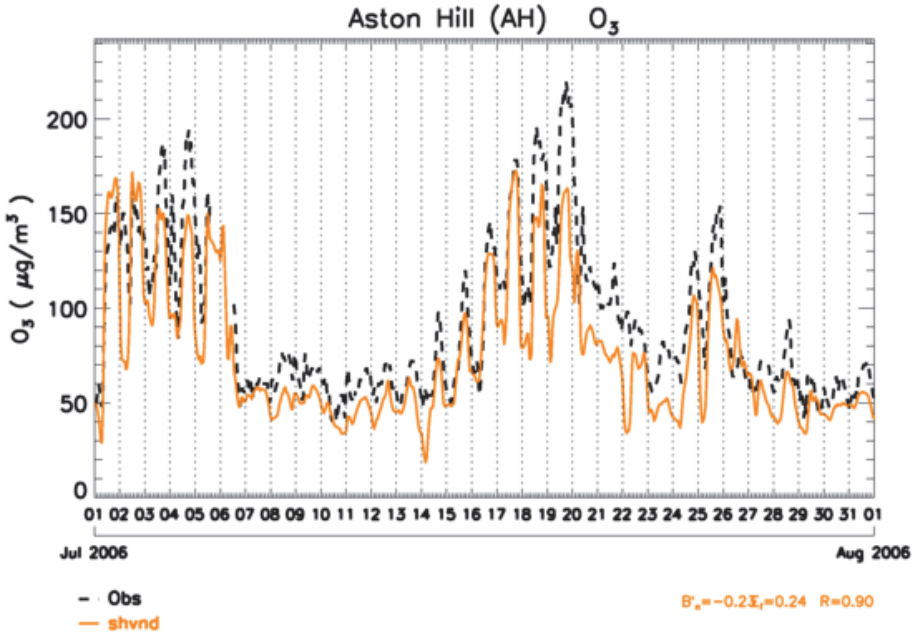

Fig. 5. Time series of ozone concentrations ($\mu g\, m^{-3}$) for the rural site at Aston Hill, for July 2006 (rural site) for July 2006. Observed concentrations are shown as the black dashed line and the model output as the solid orange line.

GMDD
5, 3131–3182, 2012

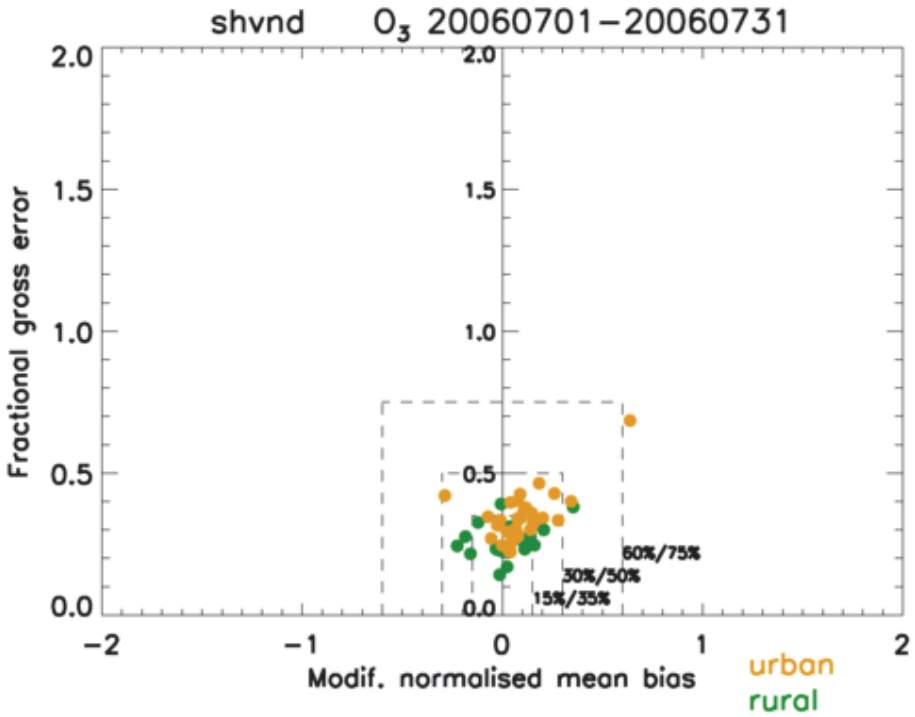

Fig. 6. Soccer plot for ozone for July 2006.

GMDD
5, 3131–3182, 2012

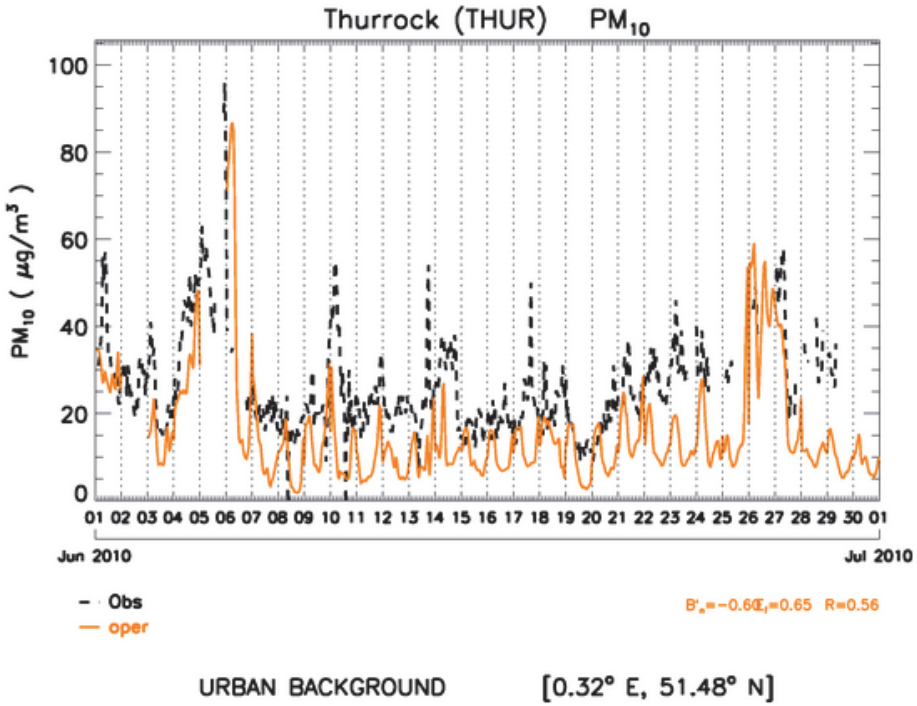

Fig. 7. Time series of PM_{10} concentrations ($\mu g\,m^{-3}$) for the urban background site at Thurrock, east of London, for June 2010. Observed concentrations are shown as the black dashed line and the model output as the solid orange line.

GMDD
5, 3131–3182, 2012

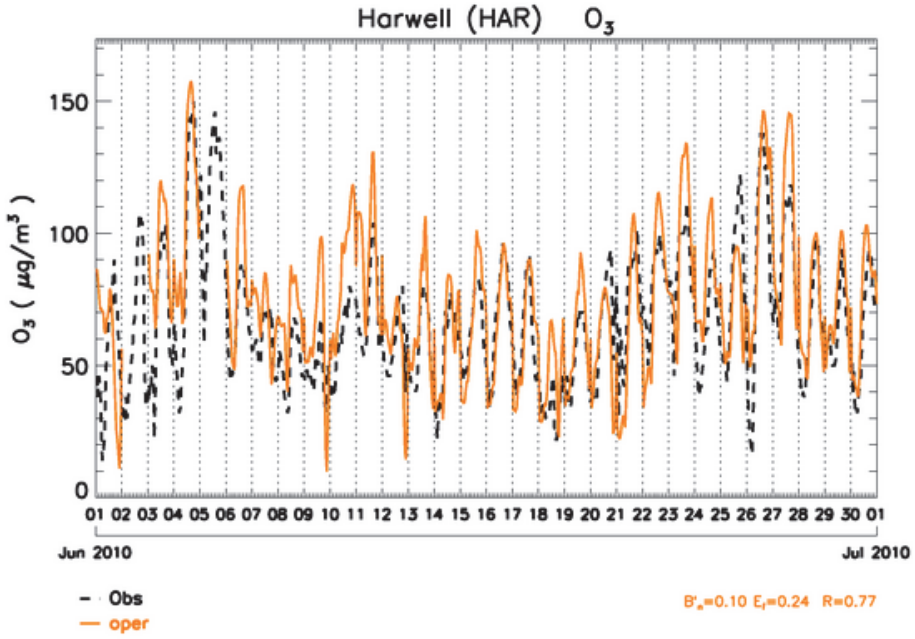

Fig. 8. Time series of ozone concentrations ($\mu g\,m^{-3}$) for the rural site Harwell for June 2010. Observed concentrations are shown as the black dashed line and the model output as the solid orange line.

GMDD
5, 3131–3182, 2012

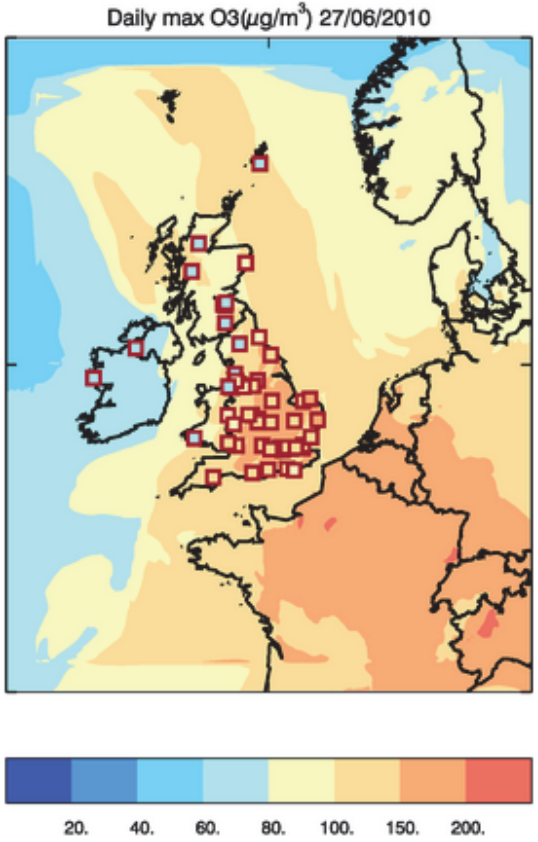

Fig. 9. Maximum modelled hourly ozone concentrations ($\mu g m^{-3}$) for 27 June 2010, with observed concentrations over-plotted within squares.

GMDD
5, 3131–3182, 2012

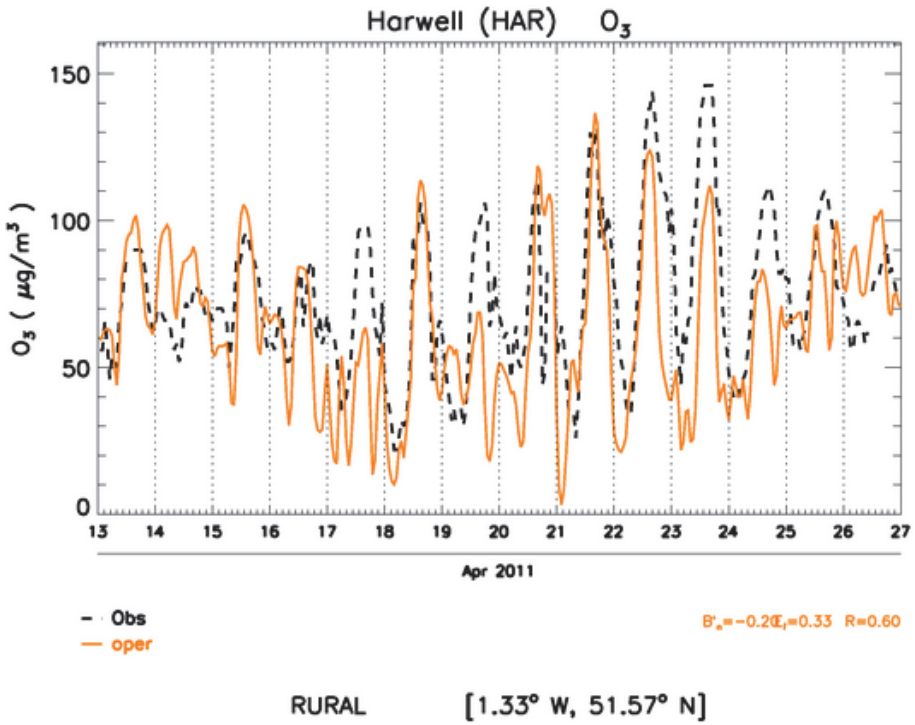

Fig. 10. Time series of ozone concentrations ($\mu g\,m^{-3}$) for the rural site at Harwell, for April 2011. Observed concentrations are shown as the black dashed line and the model output as the solid orange line.

GMDD
5, 3131–3182, 2012

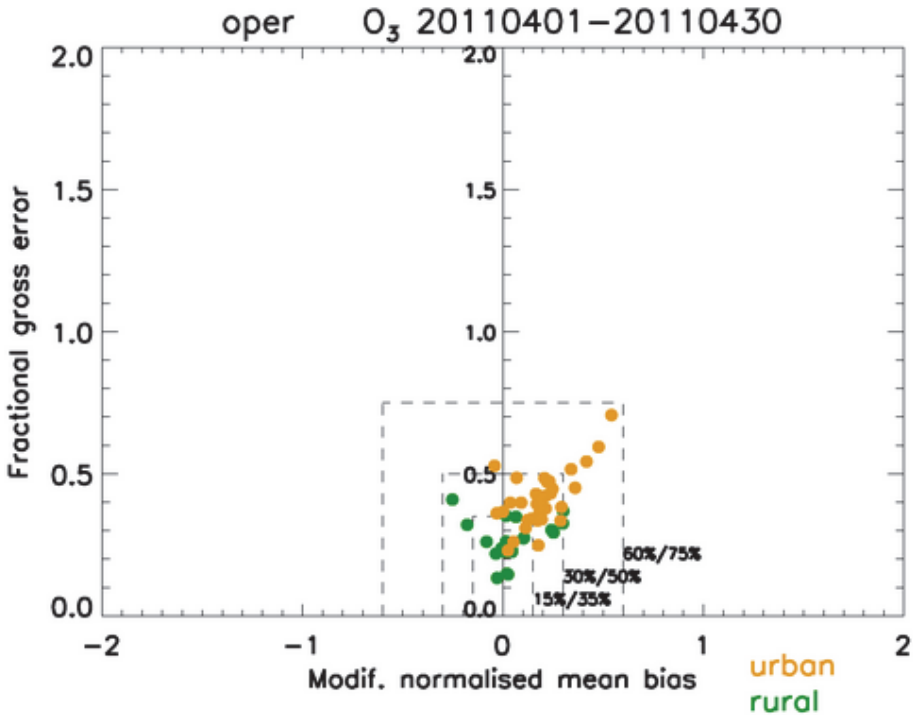

Fig. 11. Soccer plot for ozone for April 2011.

GMDD
5, 3131–3182, 2012

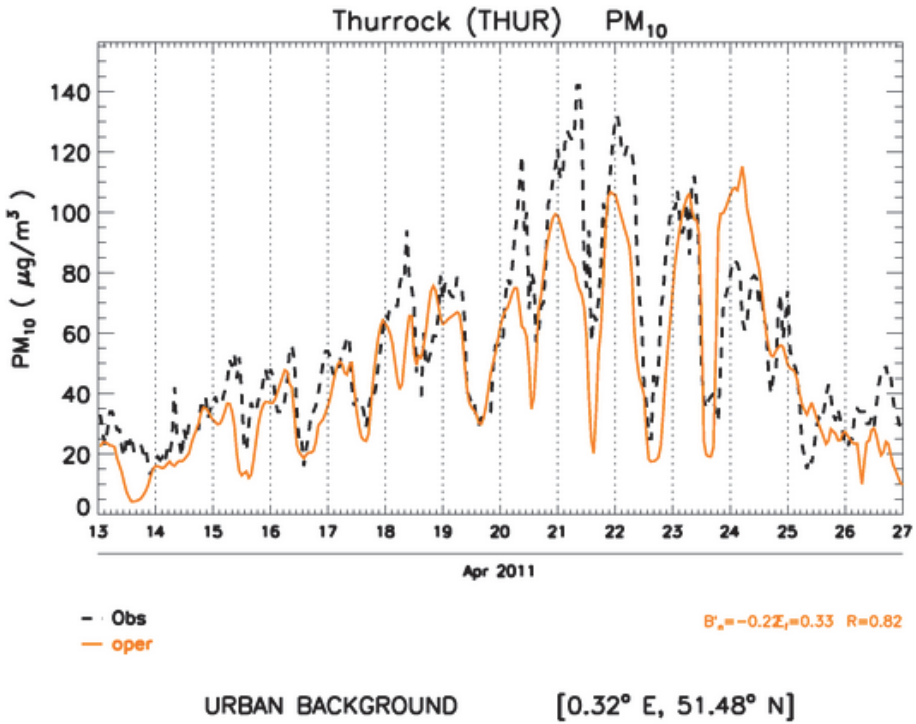

Fig. 12. Time series of PM_{10} concentrations ($\mu g\,m^{-3}$) for the urban background site of Thurrock, east of London at the peak of the April 2011 episode. Observed concentrations are shown as the black dashed line and the model output as the solid orange line.

GMDD
5, 3131–3182, 2012

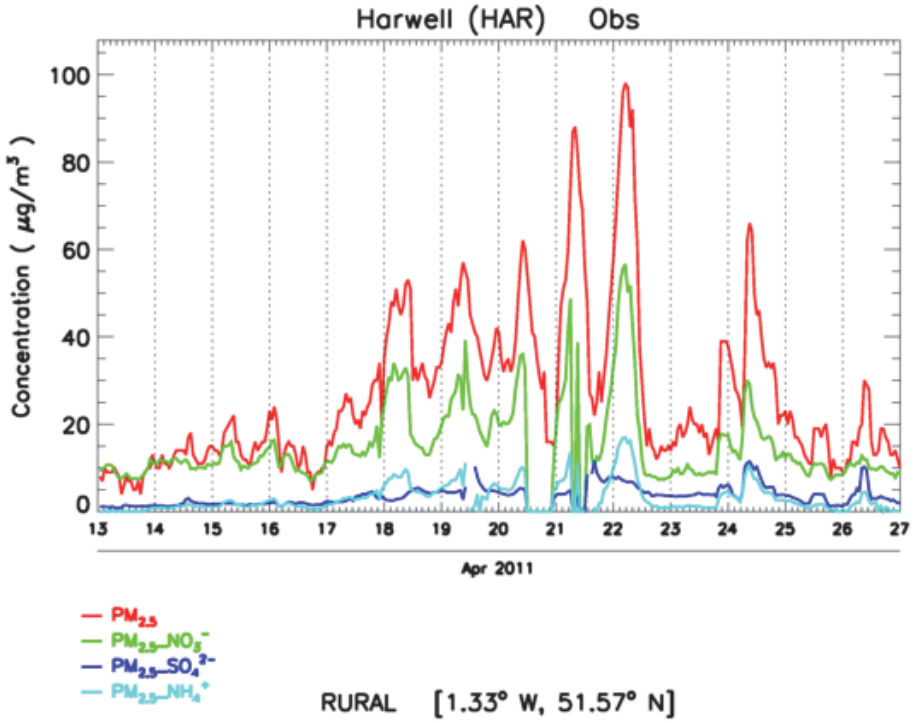

Fig. 13. Provisional speciated $PM_{2.5}$ measurements at Harwell (rural site) for the peak of the April 2011 episode.

GMDD
5, 3131–3182, 2012

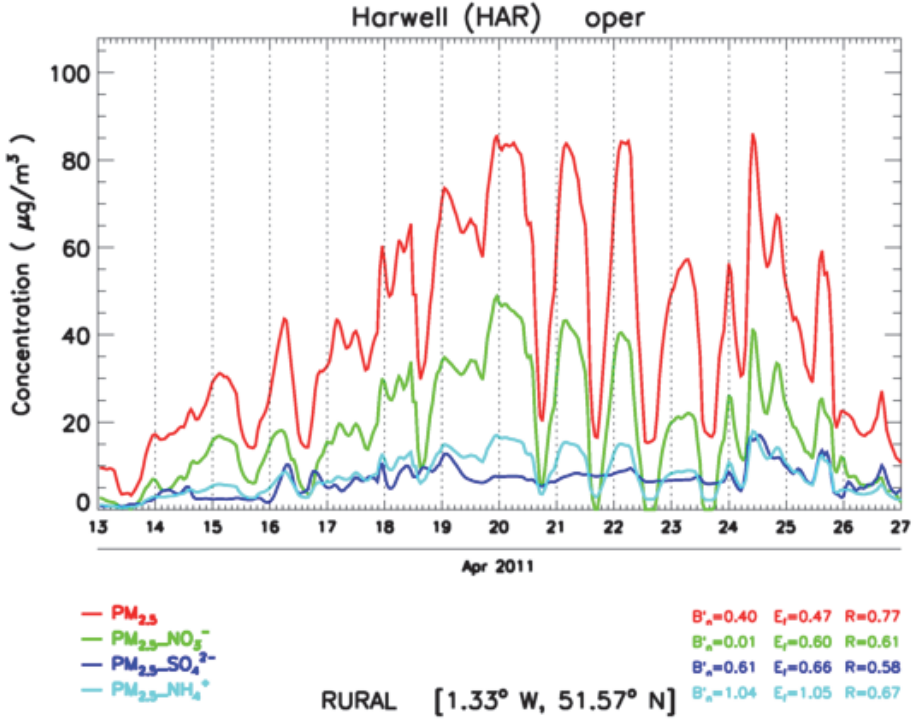

Fig. 14. Speciated $PM_{2.5}$ model forecasts for Harwell (rural site) for the peak of the April 2011 episode.

GMDD
5, 3131–3182, 2012

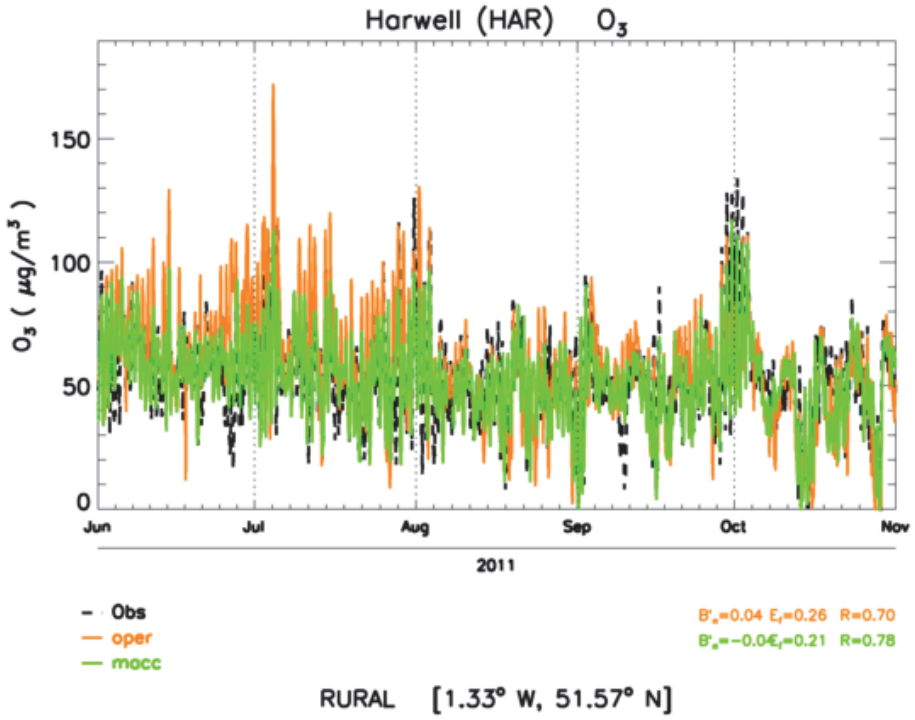

Fig. 15. AQUM and MACC ensemble predictions of ozone compared to hourly observations at Harwell (rural site) for 1 June–31 October 2011. The observations are the black dashed line, AQUM output in orange and MACC ensemble predictions in green.

GMDD
5, 3131–3182, 2012

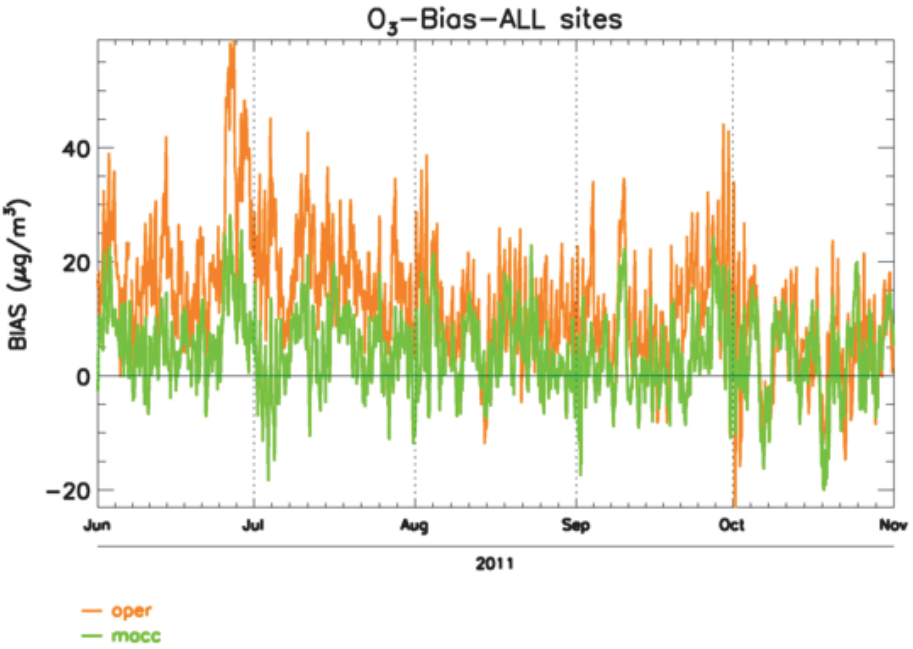

Fig. 16. Time series of ozone bias comparing for AQUM operational output in orange and MACC ensemble in green.

GMDD
5, 3131–3182, 2012