Accepted Manuscript

Authors: C. Chemel, R.S. Sokhi, Y. Yu, G.D. Hayman, K.J. Vincent, A.J. Dore, Y.S. Tang, H.D. Prain, B.E.A. Fisher

Please cite this article as: Chemel, C., Sokhi, R.S., Yu, Y., Hayman, G.D., Vincent, K.J., Dore, A.J., Tang, Y.S., Prain, H.D., Fisher, B.E.A. Evaluation of a CMAQ simulation at high resolution over the UK for the calendar year 2003, Atmospheric Environment (2010), doi: 10.1016/j.atmosenv.2010.03.029

Evaluation of a CMAQ simulation at high resolution over the UK for the calendar year 2003

C. Chemel [a,*], R. S. Sokhi [a], Y. Yu [a,b], G. D. Hayman [c], K. J. Vincent [d],

A. J. Dore [e], Y. S. Tang [e], H. D. Prain [a], B. E. A. Fisher [f]

[a] *Centre for Atmospheric & Instrumentation Research, University of Hertfordshire, College Lane Campus, Hatfield, Herts AL10 9AB, UK*

[b] *Institute of Plateau Atmospheric Physics, Cold and Arid Regions Environmental & Engineering Research Institute, Chinese Academy of Sciences, Donggang Xilu 322, Lanzhou, Gansu 730000, P. R. China*

[c] *Hayman Atmospheric Consulting, 22 The Old Pound, Abingdon, Oxon OX13 6BQ, UK*

[d] *AEA Energy & Environment, The Gemini Building, Fermi Avenue, Harwell International Business Centre, Didcot, Oxon OX11 0QR, UK*

[e] *Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian EH26 0QB, UK*

[f] *Risk and Forecasting Science, Environment Agency, Kings Meadow House, Kings Meadow Road, Reading, Berks RG1 8DQ, UK*

Abstract

A comprehensive 'operational' evaluation of the performance of the Community Multiscale Air Quality (CMAQ) modelling system version 4.6 was conducted in support of pollution assessment in the UK for the calendar year 2003. The model was run on multiple grids using one-way nests down to a horizontal resolution as fine as 5 km over the whole of the UK. The model performance was evaluated for pollutants with standards and limit values (*e.g.* O_3, PM_{10}) and acid deposition species (*e.g.* NH_3, SO_4^{2-}, NO_3^-, NH_4^+) against data from operational national monitoring networks. The key performance characteristics of the modelling system were found to be variable according to acceptance criteria and to depend on the type (*e.g.* urban, rural) and location of the sites, as well as on the time of the year. As regards the techniques that were used for 'operational' evaluation, performance generally complied with expected levels and ranged from good (*e.g.* O_3, SO_4^{2-}) to moderate (*e.g.* PM_{10}, NO_3^-). At a few sites low correlations and large standard deviations for some species (*e.g.* SO_2) suggest that these sites are subject to local factors (*e.g.* topography, sources) that are not well described in the model. Overall, the model tends to over predict O_3 and under predict aerosol species (except SO_4^{2-}). Discrepancies between predicted and observed concentrations may be due to a variety of intertwined factors, which include inaccuracies in meteorological predictions, chemical boundary conditions, temporal variability in emissions, and uncertainties in the treatment of gas and aerosol chemistry. Further work is thus required to investigate the respective contributions of such factors on the predicted concentrations.

Keywords: Air quality; Numerical simulation; Model performance; Evaluation

2

1. Introduction

In Europe, pollutants released into the environment are regulated under the European Community (EC) Directive 96/61/EC, which covers integrated pollution prevention and control. Air pollutants such as sulphur dioxide (SO_2), nitrogen oxides (NO_x), volatile organic compounds (VOCs) and particulate matter (PM) smaller than 10 μm in aerodynamic diameter (PM_{10}), which are emitted particularly from industrial sources, fall under these control regimes. Regulation of these pollutants is necessary to minimize their adverse impact on air quality, and the environment as a whole, requiring accurate and realistic assessment. As an example, NO_x and VOCs under the action of sunlight can lead to the creation of ozone (O_3). Nitric oxide (NO) can be oxidized into harmful nitrogen dioxide (NO_2) by reacting with O_3. Pollutants such as NO_2, O_3, VOCs (*e.g.* Benzene) and PM_{10} are all harmful to human health and thus are subjected to limit values specified by the EC Directive 2008/50/EC on ambient air quality and cleaner air for Europe.

Emissions of SO_2, NO_x and PM_{10} from sources, such as power stations, petroleum refineries and steelworks, are controlled by the EC Directive 2001/80/EC on the limitation of emissions of certain pollutants into the air from large combustion plants. As a result of such controls it is hoped that the harm to people and damage to the environment will be reduced. Specifically, the reduction in emissions should lead to reduced environmental impact including ground-level O_3 and deposition of pollutants. Also, specific measures are often taken at national levels to comply with EC obligations and potentially further reduce pollution levels, as is the case in the UK under the National Air Quality Strategy (UK Department for Environment, Food and Rural Affairs (Defra), 2007). On a broader scale, as part of the Convention on Long-Range Transboundary Air Pollution, the main pollutants associated with industrial sources (namely, SO_2, NO_x, VOCs, and ammonia (NH_3)) are subjected to emission ceilings set for 2010 in the 1999 Gothenburg Protocol to abate acidification, eutrophication and ground-level O_3. Specific sources, such as combustion plants and electricity power stations, are controlled by the protocol through strict emission limit values.

Numerical models play a key role in assessing the contribution of regulated sources to regional air quality. Examples of recent applications in the UK include the works by Abbott *et al.* (2006) and Yu *et al.* (2007). Some of the most challenging air quality problems involve complex multi-pollutant and multi-scale interactions and coupling between atmospheric chemistry and dynamics. This is reflected through the complex non-linear relationships between emissions, chemical transformations and transport mechanisms with the added dimension of contributions from surrounding and long-range transport sources. In his review of plume chemistry, Hewitt (2001) concluded that comprehensive air quality models are eventually more ap-

* Corresponding author. Tel.: 441707286143; fax: 441707284208.

E-mail address: c.chemel@herts.ac.uk (C. Chemel).

⁵⁹ propriate than simpler modelling approaches because they can account for non-linear interactions involving

⁶⁰ multiple pollutants and multiple scales.

⁶¹ A number of simple air quality models, including the Fine Resolution Atmospheric Multi-pollutant Ex-

⁶² change (FRAME) model (Singles *et al.*, 1998), the Hull Acid Rain Model (HARM, Metcalfe *et al.*, 2005),

⁶³ and the Trajectory model with Atmospheric Chemical Kinetics (TRACK, Lee *et al.*, 2000), have been ap-

⁶⁴ plied to the UK to estimate sulphur and nitrogen deposition. These models were found to give a reasonable

⁶⁵ representation of annual average measured values for gas and aerosol concentrations in air as well as wet

⁶⁶ deposition (Dore *et al.*, 2007). They have also been successfully applied to estimate future changes in sul-

⁶⁷ phur and nitrogen deposition and exceedance of critical loads to support policy on abatement of pollutant

⁶⁸ emissions (Metcalfe *et al.*, 2001; Matejko *et al.*, 2009), and deposition from regulated emissions sources

⁶⁹ (Abbott *et al.*, 2006; Vieno *et al.*, 2009a). Simple models with a fast simulation speed also offer the opportu-

⁷⁰ nity for multiple simulations for use in integrated assessment modelling (Oxley *et al.*, 2003) and uncertainty

⁷¹ studies (Page *et al.*, 2004).

⁷² A major disadvantage with simple models however is their simple representation of meteorology and use

⁷³ of straight line trajectories. More comprehensive models, such as the Community Multiscale Air Quality

⁷⁴ (CMAQ) modelling system (US Environment Protection Agency, 1999; Byun and Schere, 2006), allow

⁷⁵ an integrated approach to representation of meteorological, chemical and physical processes. The year to

⁷⁶ year variation in meteorology and its impact on sulphur and nitrogen deposition can be assessed with such

⁷⁷ complex models. Furthermore, they can simultaneously represent processes influencing a number of envi-

⁷⁸ ronmental issues including surface O_3, PM, and acidic and nitrogen deposition. The detailed parameteri-

⁷⁹ zation of photo-oxidation is important not just to calculate ground-level O_3 but also to drive the oxidation

⁸⁰ processes influencing the chemical conversion of emitted gases which contribute to acidification.

⁸¹ Although advanced air quality models, such as the CMAQ modelling system, have been applied inter-

⁸² nationally for research and real regulatory applications (*e.g.* Gilliland *et al.*, 2008), they have not been used

⁸³ by regulators in the UK as operational tools. Published works by Sokhi *et al.* (2006), Yu *et al.* (2007, 2008)

⁸⁴ have demonstrated the potential of the CMAQ modelling system to be used for pollution assessment in

⁸⁵ the UK over short-term episodic periods (typically in the order of a week or so). These studies have pro-

⁸⁶ vided a sound foundation for the UK Environment Agency to consider the merits (and disadvantages) of

⁸⁷ using advanced air quality model, such as the CMAQ modelling system, as one of its primary air pollution

⁸⁸ assessment tools. In this context, the present study is a first step in evaluating the practicability and perfor-

⁸⁹ mance of the CMAQ modelling system for a year-long simulation at high resolution (5-km horizontal grid

⁹⁰ resolution) over the whole of the UK.

⁹¹ It is worth noting that only a few published works actually report on performance characteristics of the

CMAQ modelling system for long-term simulations. Several studies discussed its performance in reproducing field campaigns and/or short-term episodic conditions worldwide (*e.g.* Zhang *et al.*, 2006a; 2006b; 2006c, in the US, Brulfert *et al.*, 2007, in Canada, Jiménez *et al.*, 2006, in Spain, Yu *et al.*, 2008, in the UK, Fu *et al.*, 2008, in East Asia). While such studies are invaluable sources of information to detail dynamical and chemical processes involved under given circumstances, they are inevitably limited to some, possibly non-representative, episodic conditions. Evaluations of long-term simulations with the CMAQ modelling system were mainly performed for the US (see for instance Eder and Yu, 2006; Gilliland *et al.*, 2006; Hogrefe *et al.*, 2006; Phillips and Finkelstein, 2006; Tesche *et al.*, 2006; Hogrefe *et al.*, 2007; Appel *et al.*, 2008; Spak and Holloway, 2009). It is unwise to extend or translate results of these studies to other regions without re-appraisal. To our knowledge, the only long-term studies conducted with the CMAQ modelling system for Europe were those by Jiménez-Guerrero *et al.* (2008) and Matthias (2008). Jiménez-Guerrero *et al.* (2008) investigated the performance characteristics of the CMAQ modelling system over the North-Western Mediterranean at a horizontal resolution of 2 km for the entire year 2004. The model performance was found to be effective in both coastal and inland areas but with a tendency to over estimate O_3 levels and under estimate other photochemical pollutants (NO_2, CO, and PM_{10}). Matthias (2008) applied the CMAQ modelling system to simulate PM distribution in Europe with a nest over the North Sea, for the years 2000 and 2001. The horizontal grid resolution was 54 km for the European domain and 18 km for the nested domain, annual anthropogenic emissions being kept the same for both domains. The model performance was not found to be highly sensitive to horizontal grid resolution.

The outline of the paper is as follows. The modelling system and its setup are presented in § 2. The air quality monitoring networks that are used for comparison with model results are also presented. In § 3, a comprehensive 'operational' evaluation of the performance of the modelling system is conducted for the species with limit values, and those contributing to acid deposition. Modelled concentrations are compared with measurements for a range of sites across the UK. Results of this evaluation are discussed in light of the type (*e.g.* urban, rural) and location of the sites, as well as time of the year. Conclusions and suggestions for further work are given in § 4.

2. Modelling system and monitoring networks

The modelling system is based on CMAQ version 4.6, with the Advanced Research core of the Weather Research and Forecasting model version 3.0.1.1 (Skamarock *et al.*, 2008), simply referred to as WRF hereafter, as the meteorological driver, and the Sparse Matrix Operator Kernel Emissions (SMOKE, Houyoux *et al.*, 2000) version 2.4, as the emission preprocessing tool.

The simulation was conducted for the year 2003, which contained several pollution episodes throughout

5

124 the year (*e.g.* calm weather smogs in February and March, and heatwaves in July and August). The model

125 run was started on December 2002 (to handle seasonal variations) with a one-week spin-up time to minimize

126 the impact of initial conditions (see for instance Berge *et al.*, 2001). The following subsections provide

127 details of each of the main components of the system, along with indications of the modifications we made

128 to adapt it for this study.

129 *2. 1.   Setup of CMAQ and WRF*

130 CMAQ is a comprehensive air quality modelling system based on the 'one atmosphere' concept in

131 which complex interactions between atmospheric pollutants on urban, regional and hemispheric scales are

132 treated in a consistent framework. It is designed for assessing the impact of multiple pollutants including

133 tropospheric O_3 and other oxidants, speciated PM, and acid deposition species. It can simulate complex

134 atmospheric processes that transport and transform these pollutants in a dynamic environment over a broad

135 range of time scales from minutes to days and weeks. US Environment Protection Agency (1999) and Byun

136 and Schere (2006) give a thorough description of the CMAQ modelling system including its formulation

137 and applications.

138 The model was run on multiple grids using one-way nests down to a horizontal resolution of 5 km. Three

139 domains using horizontal resolutions of 45 km, 15 km, and 5 km were used. The outer (coarser) domain

140 covers most of Europe while the innermost domain encompasses the whole of the UK and includes the

141 Republic of Ireland (see Fig. 1). The computations were made on 15 vertical levels up to 50 hPa. The grid

142 was stretched along the vertical axis to accommodate a high resolution within the boundary layer (9 layers

143 up to about 2000 m above ground level) and close to the ground surface (first layer approximately 40-m

144 deep). Digital elevation, soil type, landcover data, and the other characteristics of the soil and the ground

145 surface (*e.g.* monthly surface albedo) were derived from the default geographical data that is provided with

146 the WRF preprocessing system (Skamarock *et al.*, 2008).

Fig. 1

147 Chemical interactions for the gas-phase chemistry were treated with the Carbon Bond mechanism CB05

148 (Sarwar *et al.*, 2008) and associated Euler Backward Iterative (EBI) solver (Hertel *et al.*, 1993). This chem-

149 ical mechanism was extended, compared with its predecessor CB-IV (Gery *et al.*, 1989), to better support

150 PM modelling needs such as the formation of secondary organic aerosols (SOAs). Inorganic reactions were

151 also updated to better account for the range of conditions of temperature, pressure, and chemical environ-

152 ment encountered in annual simulations at scales ranging from urban to continental. The tri-modal approach

153 to aerosol size distribution based on that of the Regional Particulate Model (RPM, Binkowski and Shankar,

154 1995), which discriminates PM into coarse PM and speciated $PM_{2.5}$ (*i.e.* PM smaller than 2.5 μm in aero-

155 dynamic diameter), was used in order to model PM (see Binkowski and Roselle, 2003). The subspecies

156 considered are sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), sodium (Na^+), chloride (Cl^-), water

157 (H_2O), and organics from precursors of anthropogenic and biogenic origin. Each mode (namely, Aitken,

158 accumulation, and coarse) is subjected to both wet and dry deposition. The aerosol module that we used

159 (referred to as AERO4 in the chemical-transport model) treats sea-salt aerosols and contains calculations

160 of thermodynamic equilibrium between the accumulation mode and the gas phase treated within the ISOR-

161 ROPIA equilibrium module (Nenes *et al.*, 1999).

162 Chemical initial and boundary conditions for the outer domain were derived from monthly mean concen-

163 trations, modelled by the UK Met Office Lagrangian chemistry-transport model STOCHEM (Collins *et al.*,

164 2000), for the year 2000. The model uses a horizontal resolution of $5°$ and 9 vertical levels up to 150 hPa.

165 STOCHEM is coupled to the Hadley Centre climate model HadCM3 (Gordon *et al.*, 2000), to provide the

166 required meteorological forcing. There is no aerosol module implemented in the model and the chemical

167 scheme incorporates the chemistry of several gas species (*e.g.* NO_x, O_3, methane, isoprene). We used the

168 default profile available in the CMAQ modelling system for PM species. Further work is required to re-

169 fine initial and boundary conditions for PM. The initial and boundary conditions for the gas species were

170 prepared for species required for the RADM2 chemical mechanism (Stockwell *et al.*, 1990) and mapped

171 to those required for the CB05 chemical mechanism using existing programmes in the CMAQ modelling

172 system. For the RADM2 species that were not available in STOCHEM (SULF, PAA, ORA1, ORA2, NO3,

173 HC5, HC8, OLI, ACO3, TPAN, HONO, DCB, ONIT, CSL, TERP, HO, HO2, MACR, MVK, ASO4I, NU-

174 MATKN, NUMACC, ASOIL, NUMCOR, SRFATKN, and SRFACC), the default profiles specified in the

175 CMAQ modelling system were used.

176 The WRF model was used as the meteorological driver for the CMAQ modelling system. The Meteorology-

177 Chemistry Interface Processor (known as MCIP) version 3.4.1 (Otte and Pleim, 2009) was used to translate

178 WRF meteorological data to the format required by CMAQ. The grids for the WRF simulation match those

179 of the CMAQ simulation but with 38 vertical levels and 5 grid cells more in each horizontal direction. The

180 38 vertical levels were collapsed in MCIP to the 15 levels used in the CMAQ calculation. Meteorologi-

181 cal initial and lateral boundary conditions of the outer domain were derived from the European Centre for

182 Medium-range Weather Forecasts (ECMWF) gridded analyses available every 6 h with a horizontal res-

183 olution of $0.5°$ on operational pressure levels up to 50 hPa for vertically distributed data, and surface and

184 soil levels for surface and deep-soil data. A grid nudging technique (Four-Dimensional Data Assimilation,

185 Stauffer and Seaman, 1990) was employed for the outer domain every 6 h in order to constrain the model

186 towards the analyses and to shorten the spin-up time (see also Otte, 2008a,b). The model was reinitialized

187 every calendar month. A relaxation zone covering 5 grid cells around each domain was employed to smooth

188 gradients near the lateral boundaries. These halos were discarded when meteorological data was processed

with MCIP.

We used the YSU non-local boundary-layer parameterization scheme (Hong *et al.*, 2006). The Monin-Obukhov surface layer scheme was used to provide surface forcing in terms of momentum, heat, and moisture fluxes. The land-surface energy budget was calculated by the Noah soil-vegetation model (Ek *et al.*, 2003). Other physics options that we used include the CAM3 radiation package (Collins *et al.*, 2006), the microphysical scheme by Thompson *et al.* (2004, 2006), and the ensemble cumulus scheme introduced by Grell and Dévényi (2002) for the two grids with a horizontal resolution larger than 5 km. For the finer-resolved grid with a horizontal resolution of 5 km, convection was explicitly resolved.

2. 2. *Preparation of emissions*

The CMAQ modelling system requires hourly emissions data of primary pollutants. SMOKE has been developed for this purpose and can be adapted to process annual emissions data (from point, line and area sources) into temporally-resolved, spatially-distributed and speciated emissions files ready for chemical-transport model. We took into account the influence of meteorology and land cover heterogeneities by using spatial surrogates including land use, road network, and population density. SMOKE can also handle the projection of the domains and reactivity controls. Reactivity control packets, by source category or specific source, allow for different VOC profiles from different emissions processes, including substituting a compound of lower reactivity for a compound of higher reactivity. We used annual anthropogenic emissions data from the European Monitoring and Evaluation Programme (EMEP, Vestreng *et al.*, 2005) for area sources using a horizontal resolution of 50 km and from the European Pollutant Emission Register (EPER, Pulles *et al.*, 2007) for point sources for grid cells outside the UK. For the UK, we used the UK National Atmospheric Emissions Inventory (NAEI, Dore *et al.*, 2005), which provides annual emissions from point sources and area sources at a horizontal resolution of 1 km.

The use of SMOKE for European or UK applications is not straightforward since all the input data, required by SMOKE, have to be in a specific format, which was developed for US applications. Currently the formats of the emissions dataset that are used by EMEP for Europe, and those of the NAEI, for the UK, differ significantly from the required format. Furthermore, the original US temporal and speciation profiles released with SMOKE need to be replaced with profiles reflecting European activity patterns and fuel consumption situations. The adaptation that we made to accommodate European and UK emissions is discussed in detail by Yu *et al.* (2007, 2008). As well as these adaptations, we made the following changes: (*i*) temporal profiles for different pollutants in the UK were refined, (*ii*) speciation profiles for VOCs were specifically developed for the CB05 chemical mechanism using source information in Europe and the UK, and (*iii*) biogenic emissions were calculated online with WRF using the methodology proposed by

8

221 Guenther *et al.* (1995) and detailed in Yu *et al.* (2008).

222 *2. 3. Monitoring networks*

223   Modelled concentrations of species with limit values (namely, carbon monoxide (CO), NO_2, O_3, PM_{10},
224 and SO_2) and acid deposition species (namely, NH_3, SO_2, nitric acid (HNO_3), and hydrogen chloride (HCl)
225 for gases, and SO_4^{2-}, NO_3^-, NH_4^+, Cl^-, and Na^+ for aerosols) are compared with measurements from the
226 UK Automatic Urban and Rural Network (AURN) and Acid Deposition Monitoring Network (ADMN),
227 respectively, to evaluate the performance of the modelling system. The spatial coverage of both monitoring
228 networks is displayed in Fig. 2, along with the type (*e.g.* urban, rural) of the sites. Traffic monitoring sites
229 were discarded for this study as being too strongly influenced by local sources. The automatic sites in
230 the AURN provide hourly concentrations. The non-automatic sites in the ADMN measure concentrations
231 averaged over a monthly sampling period. We selected only the sites using denuder-based samplers, which
232 monitor acid gases and aerosol components.

Fig. 2

233 **3.   Model evaluation**

234 *3. 1.   Rationale*

235   To have sufficient confidence in the performance of such a complex modelling system, it is necessary to
236 undertake a more detailed evaluation than just analyzing the final species concentrations. Meteorological
237 data has been evaluated separately and this evaluation is not reported in this paper. We found that the grid
238 nudging technique that we used for the outer domain did constrain the meteorological fields to remain close
239 to observational data (as expected). Given that other simpler models have already been adopted as policy
240 tools in the UK, it is important to assess the performance characteristics of the modelling system according
241 to acceptance criteria which conform to the UK Environment Agency's policy on the use of dispersion
242 models. Basic elements of this policy include that the assessment models should be fit for purpose, be
243 based on established peer-reviewed scientific principles, and be evaluated and documented.

244   No universal consensus has been reached so far on good practices to evaluate model performance. Den-
245 nis *et al.* (2010) provided a comprehensive review of tools and criteria which are widely used to evaluate
246 regional-scale photochemical air quality modelling systems. Most of the techniques commonly used for
247 'operational' evaluation (see Dennis *et al.*, 2010, and references therein, for detailed information) are exam-
248 ined in our work in the next subsections. These techniques make use of time series, scatter plots, statistical
249 metrics, Taylor diagrams, and 'bugle plots'. Appendix A provides the definition of the statistical metrics
250 that are used in our study. Since such 'operational' evaluation can generate a very large number of plots,

251 we decided to focus mainly on O_3 and PM species in the text and to refer to Appendix B for other species.

252 *3. 2.   Time series and scatter plots*

253 Time series of observed and predicted maximum daily running 8-hour mean O_3 mixing ratios at four

254 sites (namely, Ladybower, Harwell, Manchester Piccadilly, and North Kensington (see Fig. 2)) are shown

255 in Fig. 3. Those sites were selected as being representative for rural (Ladybower and Harwell) and urban

256 background (Manchester Piccadilly and North Kensington) sites. The altitudes above sea level of the sites

257 at Ladybower, Harwell, Manchester Piccadilly, and North Kensington are 367, 126, 55, and 25 m, respec-

258 tively. Time series of CO, NO_2, PM_{10}, and SO_2 at these sites are provided in Appendix B. Predicted values

259 of the modelled variables were extracted from the first vertical layer of the innermost model grid. The model

260 captures the temporal variability of O_3 quite well. O_3 concentrations are relatively unbiased at Ladybower

261 and North Kensington, under predicted at Harwell, and over predicted at Manchester Piccadilly (see Ta-

262 bles 1 to 4 of Appendix B). CMAQ tends to over predict the O3 mixing ratios lower than about 30 ppbv at

263 Manchester Piccadilly, while generally reproducing the larger values. This over prediction of low values is

264 also visible at the other three sites. Large discrepancies can be noted on a few days during the spring and

265 summer seasons. Determining accurately the reasons for these differences in terms of the treatment of the

266 key processes within the modelling system may be premature. Yu *et al.* (2008) suggested that uncertainties

267 in the emissions of O_3 precursors (*e.g.* NO_x and VOCs) might be the primary cause for these discrepancies

268 although other factors such as chemical boundary conditions may play an important role as well. Vieno

269 *et al.* (2009b) examined factors that influenced O_3 levels during the August 2003 heatwave in the UK.

270 Ozone imported from outside of the UK was found to be the largest contributor to the high O_3 levels in

271 the south of England. Dry deposition of O_3, when switched off in their model, was found to elevate O_3

272 concentration by up to 50 ppbv at night-time. We performed a similar model calculation by switching off

273 O_3 dry deposition for the summer months (June, July, and August). Results of this calculation (not shown)

274 confirmed that dry deposition did play a major role in increasing ground level O_3 mixing ratios. The scatter

275 plots of the observed and predicted maximum daily running 8-hour mean O_3 mixing ratios at those sites

276 are presented in Fig. 4. Scatter plots for NO_2 and PM_{10} are given in Appendix B. Over prediction of more

277 than a factor of two occur mostly for O_3 mixing ratios less than 20 ppbv. This result is consistent with the

278 findings of Yu *et al.* (2008) during a high O_3 episode in the UK in 2001 and several other studies in the US

279 (*e.g.* Smyth *et al.*, 2006). For O_3 levels higher than 60 ppbv, O_3 mixing ratios are clearly under estimated,

280 especially at Harwell and North Kensington. It is worth noting that none of the observed exceedances of O_3

281 over 60 ppbv at those sites are reproduced by the model. The results from the time series and scatter plots

282 for pollutants with limit values presented in this subsection and Appendix B indicate satisfactory overall

283 performance. Nonetheless, the performance of the modelling system is only qualitatively assessed by using

284 time series and scatter plots. A quantification of the model performance is proposed in the next subsections.

285 *3. 3.   Statistical metrics and Taylor diagrams*

286   Statistics are calculated separately for all species and all sites displayed in Fig. 2 because of their distinct

287 characteristics. Rather than making an average of statistical metrics over the sites (as done for instance in

288 Zhang *et al.*, 2006c), statistical metrics for each site can be plotted on a map to account for their variability

289 from one site to another. The resulting maps for the maximum daily running 8-hour mean O_3 mean bias

290 (MB) and root-mean square error (RMSE) are displayed in Fig. 5. Maps for the daily mean PM_{10} MB and

291 RMSE, along with tables summarizing a range of statistical metrics for CO, NO_2, O_3, PM_{10}, and SO_2 at

292 the four sites discussed in § 3.2 are provided in Appendix B. Fig. 5 indicates that the modelling system

293 tends to under estimate O_3 in the south of the UK and to over estimate O_3 in the north. The source of

294 this difference in performance has not yet been identified although it is likely to be associated with local

295 environmental factors (*e.g.* emissions from the industrial sector). It is worth noting that the largest values

296 of RMSE are concentrated within the Greater London area, where the sub-grid variability in emissions and

297 ground surface properties is enhanced.

Fig. 5

298   The performance of our modelling system is comparable to that of similar modelling systems exercised

299 in Europe (*e.g.* Schmidt *et al.*, 2001; Bessagnet *et al.*, 2004; Vautard *et al.*, 2007). For maximum daily

300 running 8-hour mean O_3, the normalized mean bias (NMB) and normalized mean error (NME) considering

301 all predicted/observed pairs of values from all the AURN sites are 5.34 % and 28.84 %, respectively (see

302 Table 1). These values fulfill the skill criteria $|NMB| \leq 15$ % and NME ≤ 35 % for O_3 suggested by Russell

303 and Dennis (2000). In contrast to O_3, the values of NMB and NME for daily mean PM_{10} (-34.00 % and

304 52.83 %, respectively) do not fulfill those skill criteria suggested for O_3, even though they almost fulfill less

305 stringent criteria that are often used for PM_{10} (*e.g.* $|NMB| \leq 50$ % and NME ≤ 50 %). As for other species

306 with standards and limit values at Ladybower, Harwell, Manchester Piccadilly, and North Kensington (see

307 Tables 1 to 4 of Appendix B), most of the skill scores comply with acceptance criteria. Table 1 gives

308 categorical statistics (see for instance Eder *et al.*, 2006) associated with maximum daily running 8-hour

309 mean O_3 and daily mean PM_{10}, along with the actual exceedance and non exceedance numbers a, b, c, and

310 d (see Fig. 4) used in their calculation. The accuracy (A) exceeds 90 % for both O_3 and PM_{10}. The bias

311 (B) is close to zero for O_3, which indicates that the modelling system greatly under predicted exceedances

312 ($B \ll 1$). As regards PM_{10}, exceedances are slightly under predicted ($B < 1$). The hit rate (H), also known

313 as probability of detection, is close to zero for O_3, which means that the modelling system barely produced

314 any exceedance that actually occurred. The false alarm ratio (FAR) is high for both O_3 and PM_{10}, which

315 indicates that a large proportion of the exceedances that were predicted by the modelling system did not
316 actually occur. These PM_{10} exceedances were predicted although the total number of exceedances were
317 under predicted. Further work is required to understand the conditions whereby PM_{10} peaks.

Table 1

318   Comparisons of predicted and measured O_3 and PM_{10} are further examined using Taylor diagrams (Tay-
319 lor, 2001). These diagrams convey some statistical metrics in a convenient way to evaluate model per-
320 formance. Time correlation between observed and predicted values (*i.e.* correlation coefficient, r) is rep-
321 resented along with the normalized standard deviation of predicted values in a polar plot. The standard
322 deviation of predicted values is normalized by that of observed values in order to mask the differences in
323 absolute values at the different sites. The normalized standard deviation is sometimes referred to as skill
324 variance (SKVAR). Taylor diagrams for maximum daily running 8-hour O_3 and daily mean PM_{10} con-
325 sidering all predicted/observed pairs of values for each AURN site for 2003 are shown in Fig. 6. Low
326 correlations and large SKVAR values for PM_{10} at a few sites indicate that these sites are subject to sources
327 that can be highly variable in composition, space, and time (Monks *et al.*, 2009) and thus could not be well
328 described in the model. As regards O_3, the Taylor diagram shows a more homogeneous pattern across the
329 sites. Predicted standard deviations for O_3 are smaller than their observed counterparts. This means that the
330 modelling system under estimate the variability of the maximum daily running 8-hour mean for O_3 at those
331 sites.

Fig. 6

Fig. 7

332   Fig. 7 gives the NMB and NME for acidifying and eutrophying gases and aerosols. Model perfor-
333 mance is highly variable and depends on the species, months of the year, and sites. Overall, our results
334 are consistent with those of Tesche *et al.* (2006) for inorganic aerosols (SO_4^{2-}, NO_3^-, and NH_4^+). SO_4^{2-}
335 is generally well reproduced by the modelling system. The NMB is slightly negative during the colder
336 months (-17.25 % averaged over the first and last quarters of the year) and slightly positive during the
337 warmer months (22.81 % averaged over the rest of the year). NO_3^- and NH_4^+ are under estimated during
338 the colder months while being better simulated during the warmer months. Model performance for NH_4^+
339 follows rather closely that of NH_3 and NO_3^-. The fact that NH_3 is grossly under estimated during the colder
340 months reduces dramatically NO_3^- and NH_4^+ formation, the level of NH_3 being the limiting factor in the
341 formation of ammonium nitrate (NH_4NO_3) during these months. NH_3 is clearly over estimated at sites
342 2, 8, and 11 (see Fig. 2 for the location of the sites). These sites are located in heterogeneous landscapes
343 (moorland type for site 2 and woodland type for sites 8, and 11), for which the sub-grid spatial variability
344 in emissions is expected to be strong. Model performance for HNO_3 is similar to that of SO_2. Both species
345 are over estimated at sites 4, 5, and 11. Two of these sites (4 and 11) are located in remote places, where one
346 would expect larger discrepancies due to the localized environmental displacement of very low background
347 values. HCl is under estimated by a factor of about 2. Further work is required to identify possible reasons

Fig. 8

348 for the observed discrepancies. In particular, the coarse particle mode in CMAQ version 4.6 is treated as

349 dry and chemically inert with a fixed geometric standard deviation of 2.2, which is clearly a limitation for

350 an accurate description of sea-salt particles. The upgrade of CMAQ to version 4.7 for future work looks

351 promising since it includes a chemically interactive coarse particle mode that enables dynamic transfer of

352 HNO_3, sulphuric acid (H_2SO_4), HCl, and NH_3 between coarse particles and the gas phase (Kelly *et al.*,

353 2009).

354 *3. 4. Bugle plots*

355 'Bugle plots' for maximum daily running 8-hour mean O_3 and daily mean PM_{10} considering all pre-

356 dicted/observed pairs of values for each AURN site during each season for 2003 are shown in Fig. 8 in

357 order to examine how model performance varies as a function of concentration (see Boylan and Russel,

358 2006, for further details on such plots). Model performance complies with expected levels, namely both

359 the mean fractional bias (MFB) and mean fractional error (MFE) fall under the values for the performance

360 criteria set by Boylan and Russel (2006) at most of the sites during each season. For O_3, best performance

361 is obtained during spring and summer, when concentrations are highest. Most of the values for these sea-

362 sons lie within the performance goal. Worse performance is obtained during winter and autumn, when

363 concentrations are lowest. The 'bugle plots' for PM_{10} show as for O_3 that performance improves when

364 concentrations increase. However, in contrast to O_3, PM_{10} does not reveal a clear seasonal trend in terms

365 of performance. This confirms that PM_{10}, as a complex mixture, is more variable in time than is O_3.

366 **4. Concluding remarks**

367 The UK Environment Agency is considering advanced air quality modelling as one possible tool for

368 air pollution assessment. Before the UK Environment Agency can make an informed decision whether to

369 include it as one of its assessment tools, it requires sound scientific information on its performance. With

370 that goal in mind, this study provides the first 'operational' evaluation of a CMAQ simulation for a year-

371 long simulation at high resolution (5-km horizontal resolution) over the whole of the UK. The simulation

372 was conducted for the year 2003 which contained several pollution episodes throughout the year (*e.g.* calm

373 weather smogs in February and March, and heatwaves in July and August). The performance characteristics

374 for pollutants with standards and limit values (namely, CO, NO_2, O_3, PM_{10}, and SO_2) and acid deposition

375 species (namely, NH_3, SO_2, HNO_3, and HCl for gases, and SO_4^{2-}, NO_3^-, NH_4^+, Cl^-, and Na^+ for aerosols)

376 were evaluated in an 'operational' sense. The main findings of this evaluation study are summarized in the

377 following.

- The performance characteristics of the modelling system were found to be variable according to acceptance criteria and to depend on the type (*e.g.* urban, rural) and location of the sites, as well as time of the year (*e.g.* for NH_3).

- As regards the techniques that were used for 'operational' evaluation, performance generally conformed to expected levels and ranged from good (*e.g.* O_3, SO_4^{2-}) to moderate (*e.g.* PM_{10}, NO_3^-). The moderate performance for PM_{10} is reflected by the moderate performance for NO_3^- and NH_4^+. At a few sites low correlations and large standard deviations for some species (*e.g.* SO_2) suggest that these sites are subject to sources that are not well described in the model. Overall, the model tends to over predict O_3 and under predict aerosol species (except SO_4^{2-}). Reasons for these discrepancies have not been clearly identified yet.

One has to be aware of the limitations of the approach to model evaluation that we used in our work. Evaluation techniques that aim at comparing predicted values of the modelled variables with measurements provide only an overall evaluation of model performance (Dennis *et al.*, 2010). Indeed, these comparisons do not examine whether the results of the model are correct for the right reasons nor how sensitive is the model performance to chemical and meteorological processes. Such an evaluation (often referred to as 'diagnostic' evaluation) complements the 'operational' and is being considered for future work. In particular, further work is needed to evaluate the capabilities of the modelling system to (*i*) predict the response of regional ozone concentrations to changes in emissions of NO_x and VOCs, and (*ii*) calculate the contribution of regulated industrial emissions to size speciated PM concentrations and associated chemical species. This 'diagnostic' evaluation will involve comparison with simpler methods that are already adopted as policy tools in the UK such the TRACK-ADMS modelling system, combining TRACK and the Atmospheric Dispersion Modelling System (ADMS, Carruthers *et al.*, 1994), for annual audits, the Photochemical Ozone Creation Potential (POCP) method (Derwent *et al.*, 1998) and Ozone Source-Receptor Model (OSRM, Hayman *et al.*, 2002) for O_3, and FRAME for acid deposition.

14

409 Talent Project' No. O827631002. This model performance evaluation study benefited from the synthesizing

410 work of COST Action 728 on model evaluation.

Appendices

411

A   Statistical metrics

412

413 Evaluation of model performance through statistical metrics focuses on measures that compare a set of

414 N predicted concentrations $\mathcal{P}_i$ with their counterpart observed concentrations $\mathcal{O}_i$, where i refers to a given

415 time and/or location. Standard metrics used for air quality performance evaluation are detailed in numerous

416 papers (*e.g.* Dennis *et al.*, 2010, and references therein) and only the ones that are used in our work (main

417 text and Appendix B) are reported hereafter. The means of N predictions and observations are defined as

418
$$\overline{\mathcal{P}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{P}_i \quad \text{and} \quad \overline{\mathcal{O}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}_i,$$

419 respectively. The standard deviations of N predictions and observations are defined as

420
$$\sigma_{\mathcal{P}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathcal{P}_i - \overline{\mathcal{P}})^2} \quad \text{and} \quad \sigma_{\mathcal{O}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathcal{O}_i - \overline{\mathcal{O}})^2},$$

421 respectively. The variables a, b, c, and d used to calculate the categorical statistics A, B, H, and FAR

422 represent all the exceedances that did not occur, exceedances that did occur, exceedances that were not

423 predicted and not observed, and exceedances that were not predicted but observed, respectively (see Fig. 4).

424 <u>A</u>ccuracy (no unit, in %):

425
$$A = \left(\frac{b+c}{a+b+c+d} \right) \times 100$$

426 <u>B</u>ias (no unit):

427
$$B = \frac{a+b}{b+d}$$

428 Correlation coefficient, r (no unit):

429
$$r = \sum_{i=1}^{N} (\mathcal{P}_i - \overline{\mathcal{P}})(\mathcal{O}_i - \overline{\mathcal{O}}) / (\sigma_{\mathcal{P}} \, \sigma_{\mathcal{O}})$$

430 <u>F</u>actor <u>O</u>f <u>EX</u>ceedance (no unit, range $[-50, 50]$ %):

$$431 \quad \text{FOEX} = \left[\left(\frac{1}{N} \sum_{i=1}^{N} i | (\mathcal{P}_i > \mathcal{O}_i) \right) - 0.5 \right] \times 100$$

432   Fraction of predictions within a <u>F</u>actor <u>O</u>f <u>2</u> of observations (no unit, in %):

$$433 \quad \text{FO2} = \left(\frac{1}{N} \sum_{i=1}^{N} i | \left(0.5 \le \frac{\mathcal{P}_i}{\mathcal{O}_i} \le 2 \right) \right) \times 100$$

434   <u>F</u>alse <u>A</u>larm <u>R</u>atio (no unit, in %):

$$435 \quad \text{FAR} = \left(\frac{a}{a+b} \right) \times 100$$

436   <u>F</u>ractional <u>B</u>ias (no unit, range $[-2, 2]$):

$$437 \quad \text{FB} = \sum_{i=1}^{N} (\mathcal{P}_i - \mathcal{O}_i) / \sum_{i=1}^{N} [(\mathcal{P}_i + \mathcal{O}_i)/2]$$

438   <u>F</u>ractional <u>E</u>rror (no unit, range $[0, 2]$):

$$439 \quad \text{FE} = \sum_{i=1}^{N} |\mathcal{P}_i - \mathcal{O}_i| / \sum_{i=1}^{N} [(\mathcal{P}_i + \mathcal{O}_i)/2]$$

440   <u>H</u>it Rate (no unit, in %):

$$441 \quad \text{H} = \left(\frac{b}{b+d} \right) \times 100$$

442   <u>I</u>ndex of <u>A</u>greement (no unit, range $[0, 1]$):

$$443 \quad \text{IA} = 1 - \frac{\sum_{i=1}^{N} \left[(\mathcal{P}_i - \overline{\mathcal{P}}) - (\mathcal{O}_i - \overline{\mathcal{O}}) \right]^2}{\sum_{i=1}^{N} \left[|\mathcal{P}_i - \overline{\mathcal{P}}| - |\mathcal{O}_i - \overline{\mathcal{O}}| \right]^2},$$

444   <u>M</u>ean <u>B</u>ias (in unit of concentration):

$$445 \quad \text{MB} = \frac{1}{N} \sum_{i=1}^{N} (\mathcal{P}_i - \mathcal{O}_i)$$

446   <u>M</u>ean <u>E</u>rror (in unit of concentration):

$$447 \quad \text{ME} = \frac{1}{N} \sum_{i=1}^{N} |\mathcal{P}_i - \mathcal{O}_i|$$

448 Mean Fractional Bias (no unit, range $[-200, 200]$ %):

449
$$\text{MFB} = \sum_{i=1}^{N} \left\{ (\mathcal{P}_i - \mathcal{O}_i)/[(\mathcal{P}_i + \mathcal{O}_i)/2] \right\} \times 100$$

450 Mean Fractional Error (no unit, range $[0, 200]$ %):

451
$$\text{MFE} = \sum_{i=1}^{N} \left\{ |\mathcal{P}_i - \mathcal{O}_i|/[(\mathcal{P}_i + \mathcal{O}_i)/2] \right\} \times 100$$

452 Normalized Mean Bias (no unit, in %):

453
$$\text{NMB} = \sum_{i=1}^{N} (\mathcal{P}_i - \mathcal{O}_i)/\sum_{i=1}^{N} \mathcal{O}_i \times 100$$

454 Normalized Mean Error (no unit, in %):

455
$$\text{NME} = \sum_{i=1}^{N} |\mathcal{P}_i - \mathcal{O}_i|/\sum_{i=1}^{N} \mathcal{O}_i \times 100$$

456 Root Mean Square Error (in unit of concentration):

457
$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathcal{P}_i - \mathcal{O}_i)^2}$$

458 SKill VARiance (no unit):

459
$$\text{SKVAR} = \sigma_{\mathcal{P}}/\sigma_{\mathcal{O}}$$

460 **B   Supplementary materials**

461 Supplementary data associated with this article can be found in the online version.

462 **References**

463 Abbott, J., Stedman, J. R., Vincent, K. J., 2006. Annual Audits of the Contribution to Pollutant Concentrations from Processes Regu-

464   lated by the Environment Agency: Application of Method. Report for the R&D project No. SC030172/SR3, published by the UK

465   Environment Agency, Reading, UK, ISBN 978-1844326853, 50 pp.

466 Appel, K. W., Bhave, P. V., Gilliland, A. B., Sarwar, G., Roselle, S. J., 2008. Evaluation of the community multiscale air quality

467   (CMAQ) model version 4.5: Sensitivities impacting model performance; Part II – particulate matter. Atm. Environ. 42, 6057–6066.

468 Berge, E., Huang, H.-C., Chang, J., Liu, T.-H., 2001. A study of the importance of initial conditions for photochemical oxdant

469   modeling. J. Geophys. Res. 106, D1, 1347–1363.

Bessagnet, B., Hodzic, A., Vautard, R., Beekmann, M., Cheinet, S., Honoré, C., Liousse, C., Rouil, L., 2004. Aerosol modeling with CHIMERE – preliminary evaluation at the continental scale. Atm. Environ. 38, 2803–2817.

Binkowski, F. S., Roselle, S. J., 2003. Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component. 1. Model description. J. Geophys. Res. 108, D6, 4183.

Binkowski, F. S., Shankar, U., 1995. The Regional Particulate Matter Model. 1. Model description and preliminary results. J. Geophys. Res. 100, D12, 26191–26209.

Boylan, J. W., Russel, A. G., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atm. Environ. 40, 4946–4959.

Brulfert, G., Galvez, O., Yang, F., Sloan, J. J., 2007. A regional modelling study of the high ozone episode of June 2001 in southern Ontario. Atm. Environ. 41, 3777–3788.

Byun, D., Schere, K. L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl. Mech. Rev. 59, 51–77.

Carruthers, D. J., Holroyd, R. J., Hunt, J. C. R., Weng, W.-S., Robins, A. G., Apsley, D. D., Thompson, D. J., Smith, F. B., 1994. UK-ADMS: A new approach to modelling dispersion in the earth's atmospheric boundary layer. J. Wind Eng. Ind. Aerodyn. 52, 139–153.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Briegleb, B. P., Bitz, C. M., Lin, S.-J., Zhang, M., 2006. The formulation and atmospheric simulation of the Community Atmosphere Model Version 3 (CAM3). J. Climate 19, 2144–2161.

Collins, W. J., Stevenson, D. S., Johnson, C. E., Derwent, R. G., 2000. The European regional ozone distribution and its links with the global scale for the years 1992 and 2015. Atm. Environ. 34, 255–267.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S. T., Scheffe, R., Schere, K., Steyn, D., Venkatram, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. Env. Fluid Mech., in press.

Derwent, R. G., Jenkin, M. E., Saunders, S. M., Pilling, M. J., 1998. Photochemical ozone creation potentials for organic compounds in north west europe calculated with a master chemical mechanism. Atm. Environ. 32, 2429–2441.

Dore, A. J., Vieno, M., Tang, Y. S., Dragosits, U., Dosio, A., Weston, K. J., Sutton, M. A., 2007. Modelling the atmospheric transport and deposition of sulphur and nitrogen over the United Kingdom and assessment of the influence of so2 emissions from international shipping. Atm. Environ. 41, 2355–2367.

Dore, C. J., Watterson, J. D., Murrells, T. P., Passant, N. R., Hobson, M. M., Baggott, S. L., Thistlethwaite, G., Goodwin, J. W. L., King, K. R., Adams, M., Walker, C., Downes, M. K., Coleman, P. J., Stewart, R. A., Wagner, A., Sturman, J., Conolly, C., Lawrence, H., Cumine, P. R., 2005. UK emissions of air pollutants 1970 to 2003. 17th annual report from the UK National Atmospheric Emissions Inventory (NAEI), produced by the National Environmental Technology Centre (NETCEN), published by AEA Energy & Environment, Didcot, Oxon, UK, 167 pp.

Eder, B., Kang, D., Mathur, R., Yu, S., Schere, K., 2006. An operational evaluation of the Eta–CMAQ air quality forecast model. Atm. Environ. 40, 4894–4905.

Eder, B., Yu, S., 2006. A performance evaluation of the 2004 release of Models-3 CMAQ. Atm. Environ. 40, 4811–4824.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J. D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res.

508   108, D22/8851.

509   Fu, J. S., Jang, C. J., Streets, D. G., Li, Z., Kwok, R., Park, R., Han, Z., 2008. MICS-Asia II: Modeling gaseous pollutants and
510   evaluating an advanced modeling system over East Asia. Atm. Environ. 42, 3571–3583.

511   Gery, M. W., Whitten, G. Z., Killus, J. P., Dodge, M. C., 1989. A photochemical kinetics mechanism for urban and regional scale
512   computer modeling. J. Geophys. Res. 94, D10, 12925–12956.

513   Gilliland, A. B., Appel, K. W., Pinder, R. W., Dennis, R. L., 2006. Seasonal NH_3 emissions for the continental united states: Inverse
514   model estimation and evaluation. Atm. Environ. 40, 4986–4998.

515   Gilliland, A. B., Hogrefe, C., Pinder, R. W., Godowitch, J. M., Foley, K. L., Rao, S. T., 2008. Dynamic evaluation of regional air
516   quality models: Assessing changes in o_3 stemming from changes in emissions and meteorology. Atm. Environ. 42, 5110–5123.

517   Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., Wood, R. A., 2000. The simulation
518   of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments.
519   Clim. Dyn. 16, 147–168.

520   Grell, G. A., Dévényi, D., 2002. A generalized approach to parameterizing convection combining ensemble and data assimilation
521   techniques. Geophys. Res. Lett. 121, D14/1693.

522   Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W. A., Pierce, T.,
523   Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., Zimmerman, P., 1995. A global model of natural volatile organic compound
524   emissions. J. Geophys. Res. 100, D5, 8873–8892.

525   Hayman, G. D., Jenkin, M. E., Pilling, M. J., Derwent, R. G., 2002. Modelling of tropospheric ozone formation. Report for the UK
526   Department for Environment, Food and Rural Affairs (Defra) under Contract No. EPG 1/3/143, published by AEA Technology,
527   Abingdon, Oxon, UK (Report No. AEAT/ENV/R/1029, Issue 2), 227 pp.

528   Hertel, O., Berkowicz, R., Christensen, J., Hov, Ø., 1993. Test of two numerical schemes for use in atmospheric transport-chemistry
529   models. Atm. Environ. 27, 2591–2611.

530   Hewitt, C. N., 2001. The atmospheric chemistry of sulphur and nitrogen in power station plumes. Atm. Environ. 35, 1155–1170.

531   Hogrefe, C., Hao, W., Civerolo, K., Ku, J.-Y., Sistla, G., Gaza, R. S., Sedefian, L., Schere, K., Gilliland, A., Mathur, R., 2007. Daily
532   simulation of ozone and fine particulates over New York State: findings and challenges. J. Appl. Meteor. Clim. 46, 961–979.

533   Hogrefe, C., Porter, P. S., Gego, E., Gilliland, A., Gilliam, R., Swall, J., Irwin, J., Rao, S. T., 2006. Temporal features in observed and
534   simulated meteorology and air quality over the Eastern United States. Atm. Environ. 40, 5041–5055.

535   Hong, S.-Y., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes.
536   Mon. Weath. Rev. 134, 2318–2341.

537   Houyoux, M. R., Vukovich, J. M., Coats Jr., C. J., Wheeler, N. M., Kasibhatla, P. S., 2000. Emission inventory development and
538   processing for the Seasonal Model for Regional Air Quality (SMRAQ) project. J. Geophys. Res. 105, D7, 9079–9090.

539   Jiménez, P., Jorba, O., Parra, R., Baldasano, J. M., 2006. Evaluation of MM5-EMICAT2000-CMAQ performance and sensitivity in
540   complex terrain: High-resolution application to the northeastern Iberian Peninsula. Atm. Environ. 40, 5056–5072.

541   Jiménez-Guerrero, P., Jorba, O., Baldasano, J. M., Gass'o, S., 2008. The use of a modelling system as a tool for air quality management:
542   Annual high-resolution simulations and evaluation. Sci. Tot. Env. 390, 323–340.

543   Kelly, J. T., Bhave, P. V., Nolte, C. G., Shankar, U., Foley, K. M., 2009. Simulating emission and chemical evolution of coarse sea-salt
544   particles in the Community Multiscale Air Quality (CMAQ) model. Geosci. Model Dev. Discuss. 2, 1335–1374.

545   Lee, D. S., Kingdon, R. D., Jenkin, M. E., Webster, A., 2000. Modelling the contribution of different sources of sulphur to atmospheric

546   deposition in the United Kingdom. Environ. Model. Assess. 5, 105–118.

547   Matejko, M., Dore, A. J., Hall, J., Dore, C. J., Blás, M., Kryza, M., Smith, R., Fowler, D., 2009. The influence of long term trends in pol-

548   lutant emissions on deposition of sulphur and nitrogen and exceedance of critical loads in the United Kingdom. Environ. Sci. Policy

549   12, 882–896.

550   Matthias, V., 2008. The aerosol distribution in Europe derived with the Community Multiscale Air Quality (CMAQ) model: compari-

551   son to near surface in situ and sunphotometer measurements. Atmos. Chem. Phys. 8, 5077–5097.

552   Metcalfe, S. E., Whyatt, J. D., Broughton, R., Derwent, R. G., Finnegan, D., Hall, J., Mineter, M., O'Donoghue, M., Sutton, M. A.,

553   2001. Developing the Hull Acid Rain Model: its validation and implication for policy makers. Environ. Sci. Policy 4, 25–37.

554   Metcalfe, S. E., Whyatt, J. D., Nicholson, J. P. G., Derwent, R. G., Heywood, E., 2005. Issues in model validation: assessing the

555   performance of a regional-scale acid deposition model using measured and modelled data. Atm. Environ. 39, 587–598.

556   Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., and 58 co-authors, 2009. Atmospheric composition change – global

557   and regional air quality. Atm. Environ. 43, 5268–5350.

558   Nenes, A., Pandis, S. N., Pilinis, C., 1999. Continued development and testing of a new thermodynamic aerosol module for urban and

559   regional air quality models. Atm. Environ. 33, 1553–1560.

560   Otte, T. L., 2008a. The impact of nudging in the meteorological model for retrospective air quality simulations. Part I: Evaluation

561   against national observation networks. J. Appl. Meteor. Clim. 47, 1853–1867.

562   Otte, T. L., 2008b. The impact of nudging in the meteorological model for retrospective air quality simulations. Part II: Evaluating

563   collocated meteorological and air quality observations. J. Appl. Meteor. Clim. 47, 1868–1887.

564   Otte, T. L., Pleim, J. E., 2009. The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system. Geosci.

565   Model Dev. Discuss. 2, 1449–1486.

566   Oxley, T., ApSimon, H., Dore, A. J., Sutton, M. A., Hall, J., Heywood, E., Gonzales del Campo, T., Warren, R., 2003. The UK

567   Integrated Assessment Model, UKIAM: A national scale approach to the analysis of strategies for abatement of atmospheric

568   pollutants under the convention on long-range transboundary air pollution. Integrated Assessment 4, 236–249.

569   Page, T., Whyatt, J. D., Beven, K. J., Metcalfe, S. E., 2004. Uncertainty in modelled estimates of acid deposition across Wales: a

570   GLUE approach. Atm. Environ. 38, 2079–2090.

571   Phillips, S. B., Finkelstein, P. L., 2006. Comparison of spatial patterns of pollutant distribution with CMAQ predictions. Atm. Environ.

572   40, 4999–5009.

573   Pulles, T., Kuenen, J., Pesik, J., Cadman, J., Wagner, A., 2007. EPER review report 2004. Report for the European Commission under

574   Contract No. 070402/2006/440841/MAR/C4, published by the European Environment Agency, Copenhagen, Denmark, 121 pp.

575   Russell, A., Dennis, R., 2000. NARSTO critical review of photochemical models and modeling. Atm. Environ. 34, 2283–2324.

576   Sarwar, G., Luecken, D., Yarwood, G., Whitten, G. Z., Carter, W. P. L., 2008. Impact of an updated carbon bond mechanism on

577   predictions from the CMAQ modeling system: preliminary assessment. J. Appl. Meteor. Clim. 47, 3–14.

578   Schmidt, H., Derognat, C., Vautard, R., Beekmann, M., 2001. A comparison of simulated and observed ozone mixing ratios for the

579   summer of 1998 in Western Europe. Atm. Environ. 35, 6277–6297.

580   Singles, R., Sutton, M. A., Weston, K. J., 1998. A multi-layer model to describe the atmospheric transport and deposition of ammonia

581   in Great Britain. Atm. Environ. 32, 393–399.

582   Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., Powers, J. G., 2008. A

583   description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, NCAR, Boulder, CO, USA,

125 pp.

Smyth, S. C., Jiang, W., Yin, D., Roth, H., Giroux, E., 2006. Evaluation of CMAQ O_3 and $PM_{2.5}$ performance using Pacific 2001 measurement data. Atm. Environ. 40, 2735–2749.

Sokhi, R. S., San José, R., Kitwiroon, N., Fragkou, E., Pérez, J. L., Middleton, D. R., 2006. Prediction of ozone levels in London using the MM5-CMAQ modelling system. Env. Modelling & Sofware 21, 566–576.

Spak, S. N., Holloway, T., 2009. Seasonality of speciated aerosol transport over the Great Lakes region. J. Geophys. Res. 114, D08302.

Stauffer, D. R., Seaman, N., 1990. Use of Four-Dimensional Data Assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. Mon. Weath. Rev. 118, 1250–1277.

Stockwell, W. R., Middleton, P., Chang, J. S., Tang, X., 1990. The Second Generation Regional Acid Deposition Model Chemical Mechanism for Regional Air Quality Modeling. J. Geophys. Res. 95, D10, 16343–16367.

Taylor, K. E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106, D7, 7183–7192.

Tesche, T. W., Morris, R., Tonnesen, G., McNally, D., Boylan, J., Brewer, P., 2006. CMAQ/CAMx annual 2002 performance evaluation over the eastern US. Atm. Environ. 40, 4906–4919.

Thompson, G., Field, P. R., Hall, W. D., Rasmussen, R. M., 2006. A new bulk microphysical parameterization for WRF (& MM5). In: Proc. of the 8th WRF Users' Workshop. 11–15 June 2007, Boulder, CO, USA.

Thompson, G., Rasmussen, R. M., Manning, K., 2004. Explicit forecasts of winter precipitation using an improved bulk micro-physics scheme. Part I: Description and sensitivity analysis. Mon. Weath. Rev. 132, 519–542.

UK Department for Environment, Food and Rural Affairs (Defra), 2007. The Air Quality Strategy for England, Scotland, Wales and Northern Ireland (Volume 1). Cm 7169 NIA 61/06-07, published by TSO, Norwich, UK, ISBN 978-0-10-171692-5, 54 pp.

US Environment Protection Agency, 1999. Science algorithms of the EPA Models-3 Community Multiscale Air Quality modeling system. Byun, D. W., Ching, J. K. S., Eds. Technical report EPA/600/R-99/030, US Environment Protection Agency, Research Triangle Park, NC, USA, 727 pp.

Vautard, R., Builtjes, P. H. J., Thunis, P., Cuvelier, C., Bedogni, M., Bessagnet, B., Honoré, C., Moussiopoulos, N., Pirovano, G., Schaap, M., Stern, R., Terrason, L., Wind, P., 2007. Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry tramsport models over four european cities within the CityDelta project. Atm. Environ. 41, 173–188.

Vestreng, V., Breivik, K., Adams, M., Wagner, A., Goodwin, J., Rozovskaya, O., Pacyna, J. M., 2005. Inventory Review 2005, Emission Data reported to LRTAP Convention and NEC Directive, Initial review of HMs and POPs. Technical Report MSC-W 1/2005, published by the Meteorological Synthesizing Centre - West (MSC-W) of EMEP, The Norwegian Meteorological Institute, Blindern, Oslo, Norway, ISSN 0804-2446, 131 pp.

Vieno, M., Dore, A. J., Bealey, W. J., Stevenson, D. S., Sutton, M. A., 2009a. The importance of source configuration in quantifying footprints of regional atmospheric sulphur deposition. Sci. Tot. Env., in press.

Vieno, M., Dore, A. J., Stevenson, D. S., Doherty, R., Heal, M. R., Reis, S., Hallsworth, S., Terrason, L., Wind, P., Fowler, D., Simpson, D., Sutton, M. A., 2009b. Modelling surface ozone during the 2003 heat wave in the uk. Atmos. Chem. Phys. Discussion 9, 19509–19544.

Yu, Y., Sokhi, R. S., Kitwiroon, N., Middleton, D. R., Fisher, B., 2008. Performance characteristics of MM5-SMOKE-CMAQ for a summer photochemical episode in southeast England, United Kingdom. Atm. Environ. 42, 4870–4883.

Yu, Y., Sokhi, R. S., Middleton, D. R., 2007. Estimating contributions of Agency-regulated sources to secondary pollutants using CMAQ and NAME III models. Report for the R&D project No. P4-120/3, published by the UK Environment Agency, Reading,

622   UK, ISBN 978-1844328314, 101 pp.

623   Zhang, Y., Liu, P., Pun, B., Seigneur, C., 2006a. A comprehensive performance evaluation of MM5-CMAQ for the Summer 1999

624   Southern Oxidants Study episode – Part I: Evaluation protocols, databases, and meteorological predictions. Atm. Environ. 40,

625   4825–4838.

626   Zhang, Y., Liu, P., Pun, B., Seigneur, C., 2006b. A comprehensive performance evaluation of MM5-CMAQ for the Summer 1999

627   Southern Oxidants Study episode – Part III: Diagnostic and mechanistic evaluations. Atm. Environ. 40, 4856–4873.

628   Zhang, Y., Liu, P., Queen, A., Misenis, C., Pun, B., Seigneur, C., Wu, S.-Y., 2006c. A comprehensive performance evaluation of

629   MM5-CMAQ for the Summer 1999 Southern Oxidants Study episode – Part II: Gas and aerosol predictions. Atm. Environ. 40,

630   4839–4855.

Tables

Table 1. Domain-wide statistics (including categorical statistics) for maximum daily running 8-hour mean O_3 and daily mean PM_{10} considering all predicted/observed pairs of values from all the sites in the Automatic Urban and Rural Network (AURN) for 2003. The metrics are defined in Appendix A. MB, ME, RMSE are expressed in unit of concentration, namely ppbv for O_3, and $\mu g\ m^{-3}$ for PM_{10}. NMB, MFB, NME, MFE, FO2, FOEX, A, H, and FAR are expressed in %

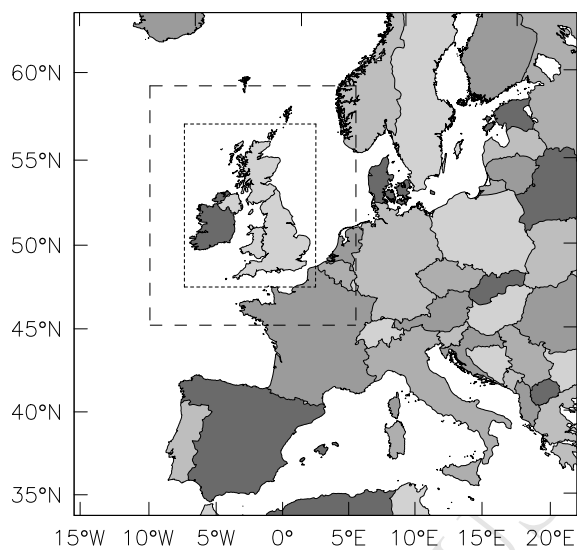| Metrics	O_3	PM_{10}
MB	1.65	−8.44
NMB	5.34	−34.00
FB	0.05	−0.41
MFB	12.22	−54.70
ME	7.69	13.12
NME	28.84	52.83
FE	0.24	0.64
MFE	28.71	67.60
RMSE	10.43	17.60
r	0.69	0.47
FO2	76.74	26.78
IA	0.97	0.87
FOEX	1.77	−40.79
A	96.41	91.91
B	0.03	0.69
H	0.72	16.54
FAR	79.31	75.91
a	23	479
b	6	152
c	22765	14008
d	825	767

Figure Captions

Fig. 1. Spatial coverage of the outer (coarser) domain used for the CMAQ simulation using a horizontal resolution of 45 km. The dashed and dotted polylines represent the areas of the nested domains using a horizontal resolution of 15 km and 5 km, respectively.

Fig. 2. Location and type (remote, rural, suburban, urban background, urban center, and urban industrial) of monitoring sites in the UK Automatic Urban and Rural Network (AURN, ●) and Acid Deposition Monitoring Network (ADMN, ○) used for the model evaluation. The numbers attributed to the ADMN sites are used as identifiers in the text. The displayed area corresponds to the innermost domain used for the CMAQ simulation using a horizontal resolution of 5 km (see Fig. 1).

Fig. 3. Time series of observed (●) and predicted (—) maximum daily running 8-hour mean O_3 for the year 2003: at (a) Ladybower, (b) Harwell, (c) Manchester Piccadilly, and (d) North Kensington (see Fig. 2 for the location of the sites). R and UB refer to rural and urban background types of site, respectively. The dashed lines represent the current limit value in Europe (*i.e.* European Union (EU) obligation of 60 ppbv) and the UK objective as defined by the UK National Air Quality Strategy (namely, 50 ppbv).

Fig. 4. Predicted versus observed maximum daily running 8-hour mean O_3 for the year 2003: at (a) Ladybower, (b) Harwell, (c) Manchester Piccadilly, and (d) North Kensington (see Fig. 2 for the location of the sites). R and UB refer to rural and urban background types of site, respectively. The dashed line indicates the 1:1 reference, while the solid lines indicate the 1:2 and 2:1 references. The dotted lines represent the current limit value in Europe (*i.e.* European Union (EU) obligation of 60 ppbv) and the UK objective as defined by the UK National Air Quality Strategy (namely, 50 ppbv). The letters a, b, c, and d denote all the exceedances that did not occur, exceedances that did occur, exceedances that were not predicted and not observed, and exceedances that were not predicted but observed, respectively (see § 3.3).

Fig. 5. (a) Mean bias and (b) root-mean square error when comparing predicted maximum daily running 8-hour mean O_3 mixing ratios with their observed counterparts for each site within the Automatic Urban and Rural Network (AURN) for 2003.

25

Fig. 6. Taylor diagrams of maximum daily running 8-hour mean O_3 (a) and daily mean PM_{10} (b) considering all predicted/observed pairs of values for each site within the Automatic Urban and Rural Network (AURN) for 2003.

Fig. 7. Normalized mean bias and error for acidifying and eutrophying gases and aerosols: (a) and (b) averaged over the sites within the Acid Deposition Monitoring Network (ADMN) for each month of the year 2003, and (c) and (d) for each site in the ADMN (see Fig. 2 for the location of the sites) for the year 2003.

Fig. 8. 'Bugle plots' for maximum daily running 8-hour mean O_3 and daily mean PM_{10} considering all predicted/observed pairs of values for each site within the Automatic Urban and Rural Network (AURN) during each season for 2003: (a) and (b) mean fractional bias; (c) and (d) mean fractional error.

Figures

Fig. 1. Spatial coverage of the outer (coarser) domain used for the CMAQ simulation using a horizontal resolution of 45 km. The dashed and dotted polylines represent the areas of the nested domains using a horizontal resolution of 15 km and 5 km, respectively.
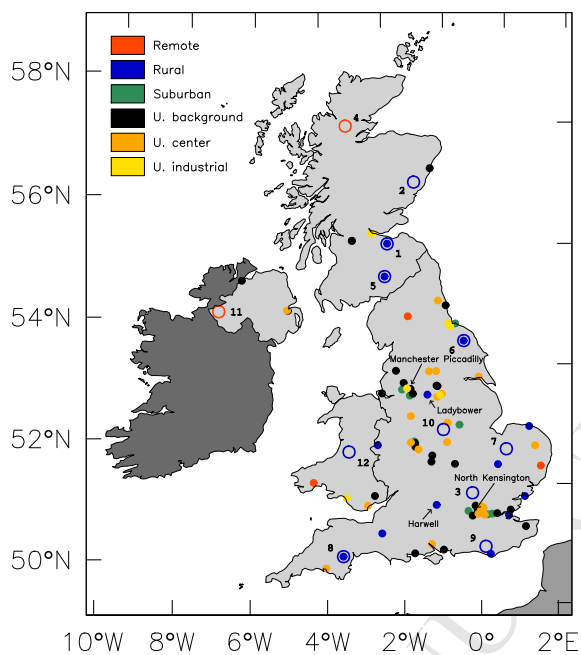
27

Fig. 2. Location and type (remote, rural, suburban, urban background, urban center, and urban industrial) of monitoring sites in the UK Automatic Urban and Rural Network (AURN, ●) and Acid Deposition Monitoring Network (ADMN, ○) used for the model evaluation. The numbers attributed to the ADMN sites are used as identifiers in the text. The displayed area corresponds to the innermost domain used for the CMAQ simulation using a horizontal resolution of 5 km (see Fig. 1).
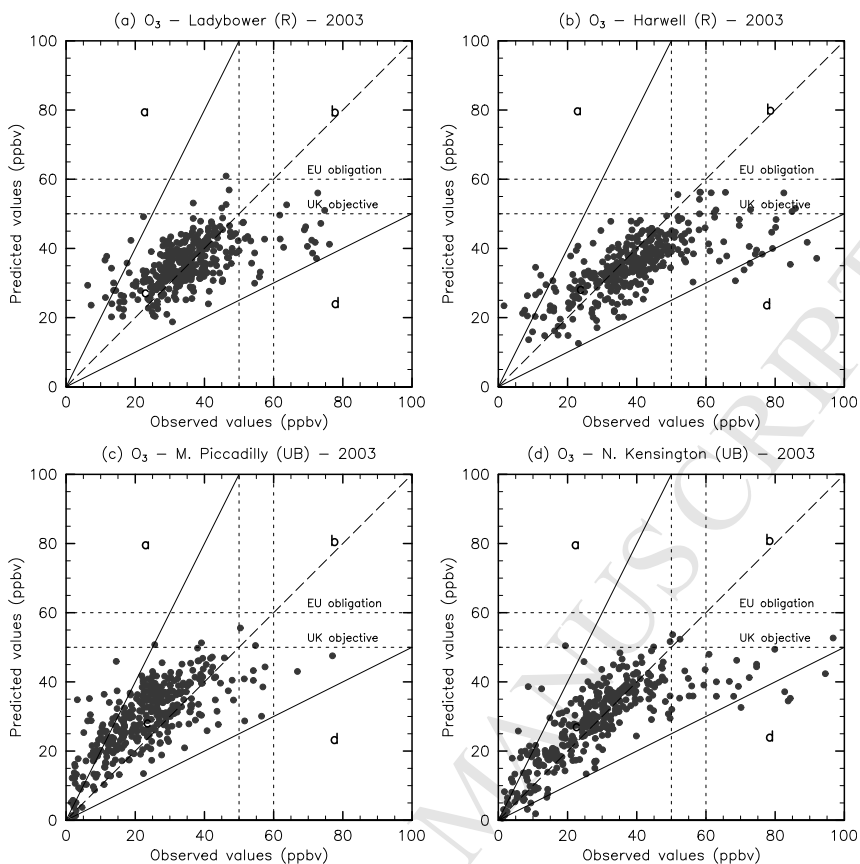
Fig. 3. Time series of observed (●) and predicted (—) maximum daily running 8-hour mean O_3 for the year 2003: at (a) Ladybower, (b) Harwell, (c) Manchester Piccadilly, and (d) North Kensington (see Fig. 2 for the location of the sites). R and UB refer to rural and urban background types of site, respectively. The dashed lines represent the current limit value in Europe (*i.e.* European Union (EU) obligation of 60 ppbv) and the UK objective as defined by the UK National Air Quality Strategy (namely, 50 ppbv).

Fig. 4. Predicted versus observed maximum daily running 8-hour mean O_3 for the year 2003: at (a) Lady-bower, (b) Harwell, (c) Manchester Piccadilly, and (d) North Kensington (see Fig. 2 for the location of the sites). R and UB refer to rural and urban background types of site, respectively. The dashed line indicates the 1:1 reference, while the solid lines indicate the 1:2 and 2:1 references. The dotted lines represent the current limit value in Europe (*i.e.* European Union (EU) obligation of 60 ppbv) and the UK objective as defined by the UK National Air Quality Strategy (namely, 50 ppbv). The letters a, b, c, and d denote all the exceedances that did not occur, exceedances that did occur, exceedances that were not predicted and not observed, and exceedances that were not predicted but observed, respectively (see § 3.3).
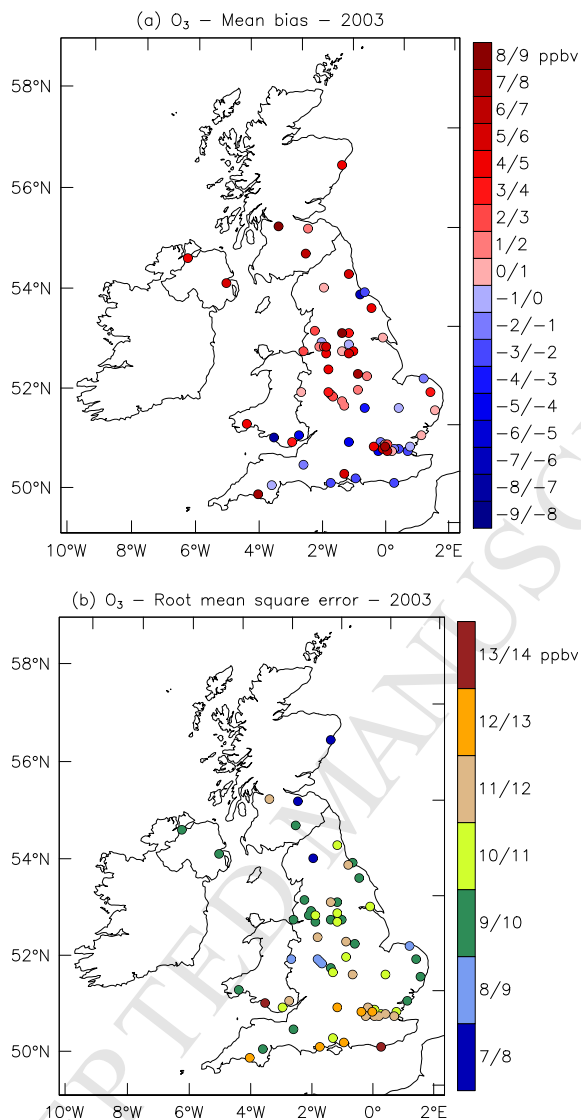
(a) O$_3$ – Mean bias – 2003

(b) O$_3$ – Root mean square error – 2003

Fig. 5. (a) Mean bias and (b) root-mean square error when comparing predicted maximum daily running 8-hour mean O$_3$ mixing ratios with their observed counterparts for each site within the Automatic Urban and Rural Network (AURN) for 2003.
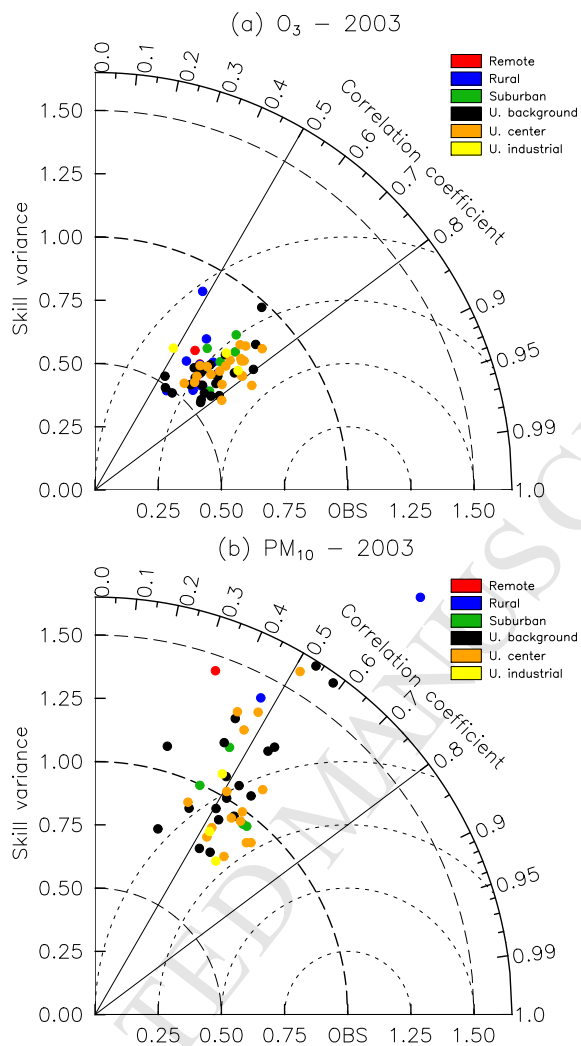
31

Fig. 6. Taylor diagrams of maximum daily running 8-hour mean O_3 (a) and daily mean PM_{10} (b) considering all predicted/observed pairs of values for each site within the Automatic Urban and Rural Network (AURN) for 2003.
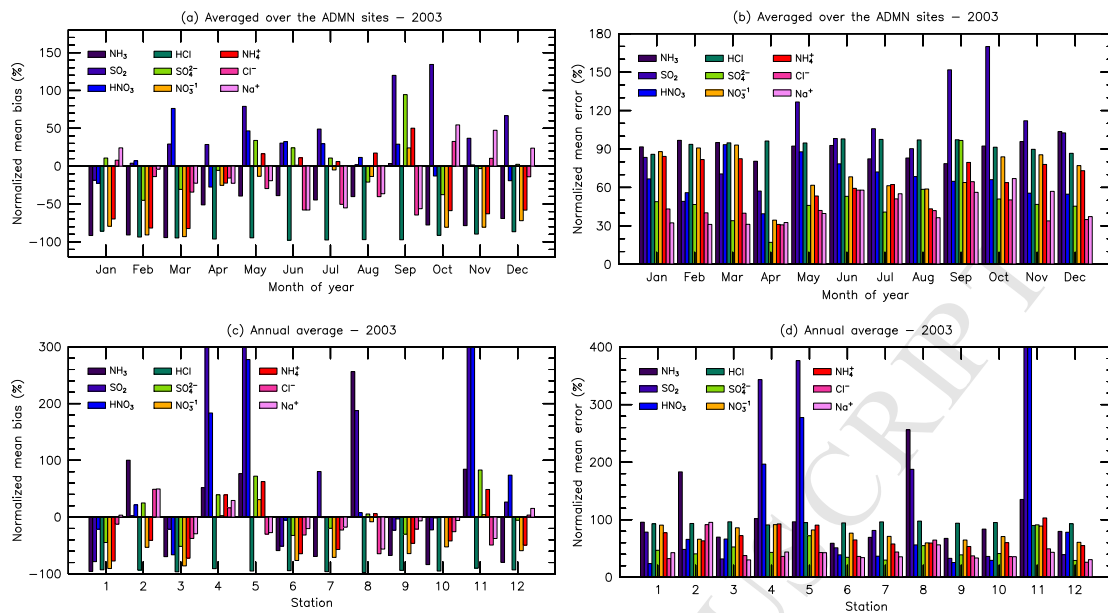
Fig. 7. Normalized mean bias and error for acidifying and eutrophying gases and aerosols: (a) and (b) averaged over the sites within the Acid Deposition Monitoring Network (ADMN) for each month of the year 2003, and (c) and (d) for each site in the ADMN (see Fig. 2 for the location of the sites) for the year 2003.
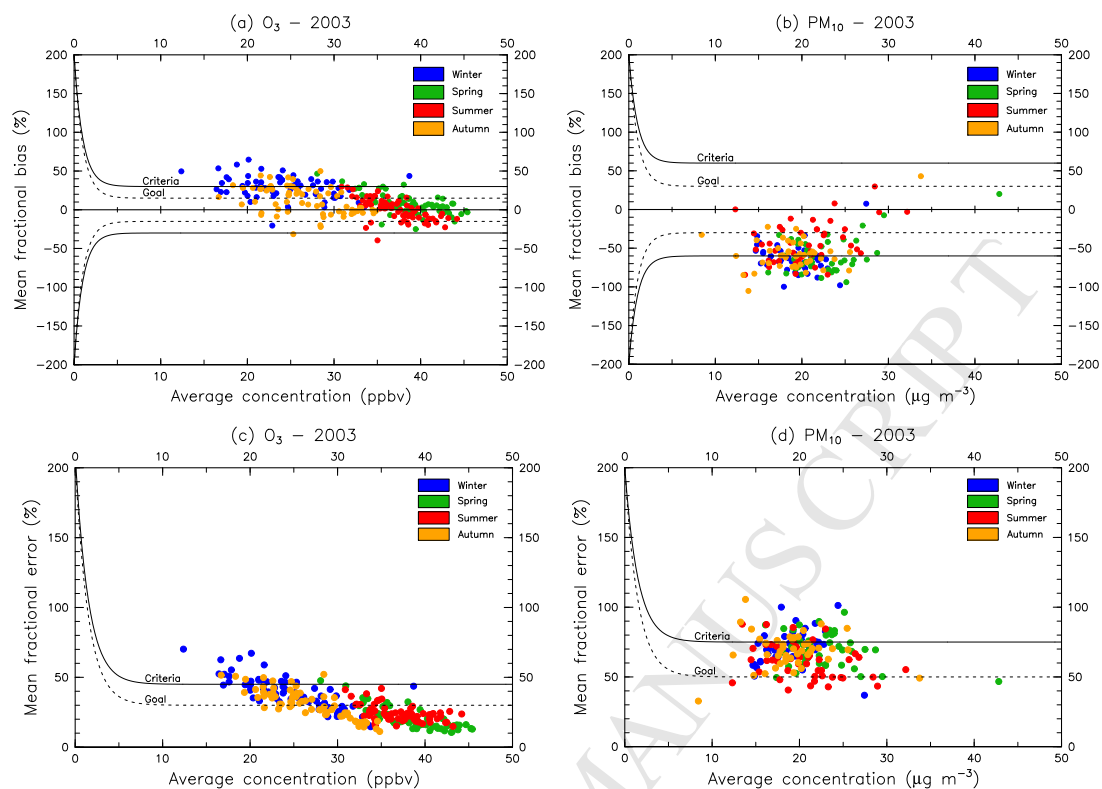
Fig. 8. 'Bugle plots' for maximum daily running 8-hour mean O_3 and daily mean PM_{10} considering all predicted/observed pairs of values for each site within the Automatic Urban and Rural Network (AURN) during each season for 2003: (a) and (b) mean fractional bias; (c) and (d) mean fractional error.