Geoscientific
Model Development

Open Access

Evaluating the capability of regional-scale air quality models to capture the vertical distribution of pollutants

E. Solazzo[1], R. Bianconi[2], G. Pirovano[3], M. D. Moran[4], R. Vautard[5], C. Hogrefe[6], K. W. Appel[6], V. Matthias[7], P. Grossi[2], B. Bessagnet[9], J. Brandt[8], C. Chemel[12,13], J. H. Christensen[8], R. Forkel[16], X. V. Francis[13], A. B. Hansen[8], S. McKeen[17], U. Nopmongcol[14], M. Prank[11], K. N. Sartelet[10], A. Segers[15], J. D. Silver[8], G. Yarwood[14], J. Werhahn[16], J. Zhang[4], S. T. Rao[6], and S. Galmarini[1]

[1]European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy
[2]Enviroware srl, Concorezzo (MB), Italy
[3]Ricerca sul Sistema Energetico (RSE SpA), Milan, Italy
[4]Air Quality Research Division, Science and Technology Branch, Environment Canada, Toronto, Canada
[5]Laboratoire des Sciences du Climat et de l'Environnement, IPSL Laboratoire CEA/CNRS/UVSQ, Orme des Merisier, 91191 Gif/Yvette Cedex, France
[6]Atmospheric Modelling and Analysis Division, Environmental Protection Agency, Research Triangle Park, NC, USA
[7]Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany
[8]Department of Environmental Science, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark
[9]NERIS, National Institute for Industrial Environment and Risks, Parc Technologique ALATA, Verneuilen Halatte, France
[10]CEREA, Joint Laboratory Ecole des Ponts ParisTech/ EDF R & D, Université Paris-Est, Paris, France
[11]Finnish Meteorological Institute, Helsinki, Finland
[12]National Centre for Atmospheric Science (NCAS), University of Hertfordshire, Hatfield, UK
[13]Centre for Atmospheric & Instrumentation Research (CAIR), University of Hertfordshire, Hatfield, UK
[14]ENVIRON International Corporation, 773 San Marin Drive, Suite 2115, Novato, CA 94998, USA
[15]Netherlands Organisation for Applied Scientific Research (TNO), Utrecht, the Netherlands
[16]Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany
[17]CIRES/University of Colorado, and NOAA/ESRL/CSD, 325 Broadway, Boulder, CO, USA

Correspondence to: S. Galmarini (stefano.galmarini@jrc.ec.europa.eu)

Abstract. This study is conducted in the framework of the Air Quality Modelling Evaluation International Initiative (AQMEII) and aims at the operational evaluation of an ensemble of 12 regional-scale chemical transport models used to predict air quality over the North American (NA) and European (EU) continents for 2006. The modelled concentrations of ozone and CO, along with the meteorological fields of wind speed (WS) and direction (WD), temperature (T), and relative humidity (RH), are compared against high-quality in-flight measurements collected by instrumented commercial aircraft as part of the Measurements of OZone, water vapour, carbon monoxide and nitrogen oxides by Airbus In-service airCraft (MOZAIC) programme. The evalua-tion is carried out for five model domains positioned around four major airports in NA (Portland, Philadelphia, Atlanta, and Dallas) and one in Europe (Frankfurt), from the sur-face to 8.5 km. We compare mean vertical profiles of mod-elled and measured variables for all airports to compute er-ror and variability statistics, perform analysis of altitudinal error correlation, and examine the seasonal error distribu-tion for ozone, including an estimation of the bias intro-duced by the lateral boundary conditions (BCs). The results indicate that model performance is highly dependent on the variable, location, season, and height (e.g. surface, planetary boundary layer (PBL) or free troposphere) being analysed. While model performance for T is satisfactory at all sites

(correlation coefficient in excess of 0.90 and fractional bias ≤ 0.01 K), WS is not replicated as well within the PBL (exhibiting a positive bias in the first 100 m and also underestimating observed variability), while above 1000 m, the model performance improves (correlation coefficient often above 0.9). The WD at NA airports is found to be biased in the PBL, primarily due to an overestimation of westerly winds. RH is modelled well within the PBL, but in the free troposphere large discrepancies among models are observed, especially in EU. CO mixing ratios show the largest range of modelled-to-observed standard deviations of all the examined species at all heights and for all airports. Correlation coefficients for CO are typically below 0.6 for all sites and heights, and large errors are present at all heights, particularly in the first 250 m. Model performance for ozone in the PBL is generally good, with both bias and error within 20 %. Profiles of ozone mixing ratios depend strongly on surface processes, revealed by the sharp gradient in the first 2 km (10 to 20 ppb km^{-1}). Modelled ozone in winter is biased low at all locations in the NA, primarily due to an underestimation of ozone from the BCs. Most of the model error in the PBL is due to surface processes (emissions, transport, photochemistry), while errors originating aloft appear to have relatively limited impact on model performance at the surface. Suggestions for future work include interpretation of the model-to-model variability and common sources of model bias, and linking CO and ozone bias to the bias in the meteorological fields. Based on the results from this study, we suggest possible in-depth, process-oriented and diagnostic investigations to be carried out next.

1 Introduction

Air quality (AQ) model evaluation studies are typically focused on the assessment of performance with respect to surface measurements since the primary goal of AQ models is to simulate the fate of pollutants to which humans, and the biosphere as a whole, are directly exposed, and also because regulatory limits are imposed at ground level only. Examples of model evaluation studies using only ground-level measurements are numerous (e.g. Van Loon et al., 2007; Vautard et al., 2009, 2012; Solazzo et al., 2012a, b; Appel et al., 2007, 2008, 2012). Evaluations of AQ models in the troposphere, from the ground to well above the planetary boundary layer (PBL), have been performed much less frequently than ground-level-only studies. Some specific experimental (e.g. Cros et al., 2004; Drobinski et al., 2007; Yu et al., 2007, 2010; Tang et al., 2011) or case studies (e.g. Emeis et al., 2011; Matthias et al., 2012) using upper-level measurements do exist, but comprehensive tropospheric evaluation and model inter-comparisons over long time periods are missing.

The development of policies designed to control and reduce air pollution requires an accurate knowledge about the sensitivity of atmospheric concentration to changes in anthropogenic emissions. This sensitivity is modulated by a number of factors, including atmospheric conditions and their variability, the state of the land surface, deposition, and concentrations of long-range transported pollutants, and primary emissions that produce a chemical environment in which secondary air pollutants develop. All of these factors influence, and are influenced by, the three-dimensional spatial distribution of pollutants and the variability of the latter over time. Accurate simulation of the troposphere is crucial from the perspective of emission control, and requires testing the capability of models to represent the vertical distribution of pollutants, the exchanges between the PBL and the free troposphere, and the horizontal fluxes between continental domains (Jonson et al., 2010; Gilge et al., 2010; Brunner et al., 2005). In addition, the pivotal role of the meteorological forcing in determining the fate of pollutant species cannot be underestimated, as errors in meteorology are inherited by AQ models, thereby producing errors in model-predicted pollutant concentrations, which can result in compounding or compensating errors. This is particularly true for long-lived species, whose concentrations are determined by the history of air parcels over long time periods during which meteorological errors may accumulate (e.g. Brunner et al., 2003; Blond and Vautard, 2004). Recent reports by the US National Research Council and the United Nations Task Force on Hemispheric Transport of Air Pollution (HTAP) (Szykman et al., 2012, and references therein) have urged the need for high-quality upper-air observational data to support model development and help address model deficiencies in the troposphere.

Examples of ensemble model evaluation in the troposphere do exist for global chemical transport models. Stevenson et al. (2006) analysed 26 global model simulations evaluated against ozonesonde profiles for the year 2000. The study suggested that the primary sources of ozone in the troposphere are chemical production and influx from the stratosphere (the latter was estimated as being about a tenth of the production term), and that removal is determined by chemical transformation and dry deposition. Furthermore, the authors showed that the ensemble mean of model results for ozone typically ranges within one standard deviation of the measurements over the entire depth of the troposphere. However, several significant discrepancies were detected, depending on location and season, as for example over-prediction throughout the northern tropics and overestimation of winter ozone at 30°–90° N. A further example of ensemble modelling of tropospheric ozone is reported by Jonson et al. (2010), who compared the ensemble mean from 12 global scale models against ozonesonde profiles in the framework of HTAP for the year 2001. The ensemble mean bias of the models was found to be smaller in winter and autumn, with large day-to-day variability among the models.

Generally, model performance was better at locations closer to major emission sources than in remote locations, with the best scores for Goose Bay (eastern North America (NA)) and the poorest scores for the Arctic stations. The modelled standard deviation was found to be low in the upper troposphere compared to the ozonesondes at all sites. The authors attributed this latter behaviour of the models to the difficulties resolving plumes at remote stations due to the coarse grid spacing (ranging from 1° to 4°) of global models. Most recently, Zyryanov et al. (2012) conducted a comparative three-dimensional analysis of six regional AQ models over Europe (EU) using MOZAIC (Measurements of OZone, water vapour, carbon monoxide and nitrogen oxides by Airbus In-service airCraft) observations, focusing on ozone during the summer months of 2008. The authors found that a large model-to-model variability existed in the upper troposphere, between 8 and 10 km height. Although a definitive reason for such variability could not be identified, the authors suggested that the differences in the model transport schemes, horizontal and vertical grid spacing, treatment of the top boundary conditions (BCs), and the way that vertical velocity is computed (diagnostically or as output of the meteorological model) could be at least partially responsible for the variability between the models. Data from MOZAIC were also used by Elguindi et al. (2010) to evaluate the performance of four chemical transport models against vertical profiles of CO mixing ratios for the year 2004. In addition, there are a number of case studies where individual regional-scale models are compared to three-dimensional concentration data collected during intensive measurement campaigns (e.g. Fisher et al., 2006; Tulet et al., 2002; Boynard et al., 2011).

However, none of the aforementioned studies is devoted to systematically evaluating an ensemble of regional-scale AQ models in the troposphere for extended periods and for several fields. Recently, a large effort within the air-quality community was established with the aim of evaluating many different regional-scale AQ modelling systems (Rao and Galmarini, 2011). The Air Quality Modelling Evaluation International Initiative (AQMEII) is the first consortium to provide a platform to evaluate the capability of AQ models to simulate pollutant transport and transformation processes throughout the PBL and the free troposphere. The specific goal of AQMEII is to perform an initial set of model evaluations and inter-comparisons on existing regional AQ model systems in NA and EU. To accomplish this goal, model simulations were conducted for the year 2006 for NA and EU model domains by independent modelling groups from both continents using state-of-the-science regional AQ modelling systems. Previous AQMEII analyses had focused on the evaluation of ensemble of AQ models at surface, using observational data provided by the regulatory ground-based networks in EU and NA for ozone, particulate matter, and meteorological fields (Solazzo et al., 2012a, b; Vautard et al., 2012) for the year of 2006.

In the present study we extend the evaluation analyses to include the vertical component. We operate in the framework of the four components of model evaluation proposed by Dennis et al. (2010), which is the main pillar of AQMEII (Rao et al., 2011). The primary goal of the analysis presented here is the operational evaluation of models in the troposphere, attempting to identify and quantify discrepancies between the models and measurements, and then suggesting hypotheses that should be investigated in detail as part of subsequent diagnostic evaluations. This study is therefore not intended to be an in-depth diagnostic analysis but rather an illustration of the potential of using upper-air measurements for regional-scale model evaluation, by introducing various approaches, metrics, and methods to analyse the complex upper-air data set, and set the stage for future work. Particular emphasis is put on ozone due to its importance on human health and climate, and also because a complementary model evaluation for ground-level ozone is available for AQMEII (Solazzo et al., 2012a)

2 Processing of four-dimensional observational data in the context of AQMEII

In this study we make use of the observational data for the year 2006 gathered by the MOZAIC programme at NA and EU airports. The MOZAIC project started in 1993 as a joint effort of European scientists, aircraft manufacturers and airlines to develop a better understanding of the natural and anthropogenic variability of the chemical composition of the atmosphere. The collection of chemical species and of meteorological drivers includes a large number of vertical profiles measured at several airports worldwide during the landing and take-off phases with a rate of data collection of 4 s, corresponding to approximately 50–100 m in the vertical. The limit of detection is rather low (2 ppbv for ozone) with an error of only 2 %, thus making the MOZAIC data set very accurate (Marenco et al., 1998; Thouret et al., 1998). Permission to access to the MOZAIC data set for the reference year of 2006 was granted to the AQMEII community, allowing analysis of over 2000 vertical profiles measured by instrumented commercial aircraft landings and take-offs from 12 selected airports in NA (i.e. Portland, Vancouver, Atlanta, Dallas, Charlotte, Boston, Chicago, Montreal, New York, Philadelphia, Toronto, and Washington DC) and three in EU (i.e. Frankfurt, Munich, and Vienna). These profiles contain the following species and meteorological fields: ozone, CO, temperature (T), relative humidity (RH), wind speed (WS) and direction (WD). Ozonesonde measurements are also used for further comparison in special cases.

For the purposes of AQMEII, the "MOZAIC domains" that contained all the trajectories up to the altitude of approximately 13 km over the year of 2006 are identified around each airport. This was particularly useful in the case, for example, of the Northeast US where trajectories of aircrafts

landing in Washington DC passed over New York airports too. Similarly, in EU, all the trajectories relating to Vienna also passed over Munich and Frankfurt. After we identified the MOZAIC domains, 13 vertical levels above the ground were selected at 0, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7500, and 8500 m to be relevant to the analysis of the model results.

MOZAIC data were gathered during take-off and landing phases, with larger occurrences of early-to-mid-afternoon (corresponding to landings in NA) and mid-to-late afternoon (corresponding to take-offs in NA) local time flights, approximately 70 % of which occurred during spring and summer. Table 1 gives the list of airports within each simulation domain and the number of MOZAIC flights associated with each airport during 2006. To simplify the analysis, the airport with the largest number of flights is selected and used for the analysis. In total, five airports (four in NA and one in EU) are identified for analysis in this study. The Portland and Philadelphia airports are selected to represent the west and east coasts of NA, respectively, while the Dallas and Atlanta airports represent the two other areas of NA. Frankfurt (with over 1200 hourly flights) had the best yearly coverage and represents the central EU area. The locations of the airports and flight areas are shown in Fig. 1.

In addition to the MOZAIC measurements, ozonesonde data were extracted from the World Meteorological Organization (WMO) World Ozone and Ultraviolet Radiation Data Centre in Toronto (Canada) and made available to the AQMEII community. These measurements report vertical profiles of ozone partial pressure at several vertical pressure levels. Further details on these data are given in Schere et al. (2012).

2.1 Pairing models and observational data

The groups participating in AQMEII were requested to deliver their model results for the species, time, location, and altitude for which observational data were available. The model data corresponding to the three-dimensional grid enclosing the MOZAIC trajectories at the aforementioned 13 altitudes were subsequently delivered to European Commission's Joint Research Centre in Ispra (Italy) by each modelling group, with a common horizontal grid resolution for the whole period of 2006 with one-hour time resolution. Hosting the data is the ENSEMBLE system (Galmarini et al., 2004, 2012), a web-interfaced data hub allowing model and observed data to be harmonized and paired in time and space to facilitate the model evaluation. Once a specific airport, date, and flight are selected, the ENSEMBLE system automatically extracts the model data from the domain volume and couples them with the MOZAIC profiles in space and time. Since measurements and model results are stored in ENSEMBLE at the same heights, the trajectory analysis module in ENSEMBLE works by looping through the measurement points along each three-dimensional trajectory.

3 Participating models

The participating research groups and modelling systems used are reported in Table 2. In total, four to five (depending on the variable) groups delivered data to compare against the MOZAIC profiles for NA, whereas data from eight to nine groups were available for EU.

The emissions and chemical BCs used by the various AQMEII groups are summarised in Table 2. AQMEII provided a set of reference time-varying gridded emissions (referred to as the "standard" emissions) for each continent. These inventories have been extensively discussed in other AQMEII publications (Pouliot et al., 2012; Solazzo et al., 2012a, b). Although the standard emissions were used by the vast majority of the participating AQMEII groups, there were still a number of degrees of freedom (e.g. fire emissions or biogenic emissions) which were chosen by each group independently. Model results generated with other emissions inventories were also submitted. AQMEII also provided a set of time-dependent chemical concentrations at the lateral boundaries of the EU and NA domains, referred to as the "standard" BCs, extracted from the Global and regional Earth-system Monitoring using Satellite and in-situ data (GEMS) re-analysis product provided by the European Centre for Medium-range Weather Forecast (ECMWF) (Schere et al., 2012). These standard BCs were used by the majority of the modelling groups, although other BCs for ozone were also used, which were based on satellite measurements assimilated within the Integrated Forecast System (IFS). Models were driven by different meteorological simulations, which were described and evaluated in Vautard et al. (2012). Details on the model configurations are given in Table 2.

3.1 Participating models and settings

3.1.1 WRF-CMAQ

The model configurations used by the US Environmental Protection Agency (US EPA) and the University of Hertfordshire are similar for NA and EU, with both simulations utilizing the Community Multiscale Air Quality (CMAQ) model version 4.7.1 (Byun and Schere, 2006; Foley et al., 2010). The NA simulation uses 34 vertical layers and 12 km horizontal grid spacing covering the continental United States, southern Canada and northern Mexico, while the EU simulation uses 34 vertical layers and 18 km horizontal grid spacing covering most of EU. Other model options employed that are common to both simulations include the CB05 chemical mechanism with chlorine chemistry extensions (Yarwood et al., 2005), the AERO5 aerosol module (Carlton et al., 2010), and the Asymmetric Convective Model 2 (ACM2) vertical mixing scheme (Pleim, 2007a, b). Additional details regarding the WRF-CMAQ simulations for AQMEII can be found in Appel et al. (2012) and references therein.

Table 1. Airport domains and number of hourly flights during 2006 for each variable. The airports selected for analyses are highlighted in bold.

| Domain Airports	IATA Code	Num. of Avail. flights	Ozone	CO	T	RH	WS	WD
West Coast								
Portland	**KPDX**	**142**	**126**	**135**	**142**	**142**	**142**	**142**
Vancouver	CYVR	72	62	65	72	72	72	72
Atlanta								
Atlanta	**KATL**	**142**	**126**	**129**	**142**	**141**	**142**	**142**
Charlotte	KCLT	2	2	2	2	2	2	2
East Coast								
Boston	KBOS	30	27	26	30	30	30	30
Chicago	KORD	4	4	4	4	4	4	4
Montreal	CYUL	2	2	2	2	2	2	2
New York	KJFK	15	10	15	15	15	15	15
Philadelphia	**KPHL**	**110**	**102**	**102**	**109**	**109**	**110**	**110**
Toronto	CYYZ	90	72	72	89	89	89	89
Washington DC	KIAD	62	28	54	61	61	62	62
Dallas								
Dallas	**KDFW**	**124**	**124**	**114**	**121**	**124**	**124**	**124**
Central Europe								
Frankfurt	**EDDF**	**1214**	**1088**	**1134**	**1134**	**1135**	**1135**	**1026**
Munich	EDDM	6	4	6	6	6	6	6
Vienna	LOWW	374	291	374	360	374	374	249

3.1.2 ECMWF-SILAM

The SILAM (Sofiev et al., 2010) model uses a transport algorithm based on the non-diffusive Eulerian advection scheme of Galperin (1999, 2000) and the adaptive vertical diffusion algorithm of Sofiev (2002). The use of sub-grid variables in these schemes allows the model to be used with thick vertical layers. A detailed description of SILAM can be found in Sofiev et al. (2008). SILAM includes a meteorological pre-processor for diagnosing the basic features of the PBL and the free troposphere (e.g. diffusivities) from the meteorological fields provided by meteorological models (Sofiev et al., 2010). Only horizontal wind components are taken from the meteorological input, with the vertical component computed from the continuity equation. SILAM is a terrain-following height coordinate models, and for the AQMEII simulations the model consisted of nine layers up to approximately 10 km.

3.1.3 MM5-DEHM

The Danish Eulerian Hemispheric Model (DEHM) (Christensen, 1997; Frohn et al., 2002; Brandt et al., 2012) is a 3-D long-range atmospheric chemistry-transport model with a horizontal domain covering the Northern Hemisphere with 150 km horizontal grid spacing. For AQMEII, the model includes two two-way nested domains, one covering EU and one covering NA, and both with 50 km horizontal grid spacing. The vertical grid is defined using the σ-coordinate system, with 29 vertical layers extending up to 100 hPa. The horizontal advection is solved numerically using the higher order accurate space derivatives scheme, applied in combination with a Forester filter (Forester, 1977). The vertical advection, as well as the dispersion sub-models, is solved using a finite elements scheme for the spatial discretisation. For the temporal integration of the dispersion, the method is applied and the temporal integration of the three-dimensional advection is carried out using a Taylor series expansion to third order. The three-dimensional wind fields are derived from the MM5 model, corrected in DEHM to ensure mass conservation. DEHM also includes a module for diagnostically calculating the vertical wind. For ozone, the initial and BCs (both lateral and top) are based on ozonesonde measurements, interpolated to global monthly three-dimensional values with a grid spacing of $4° \times 5°$ (Logan, 1999). A thorough description of DEHM and model set-up for AQMEII is given in Brandt et al. (2012).
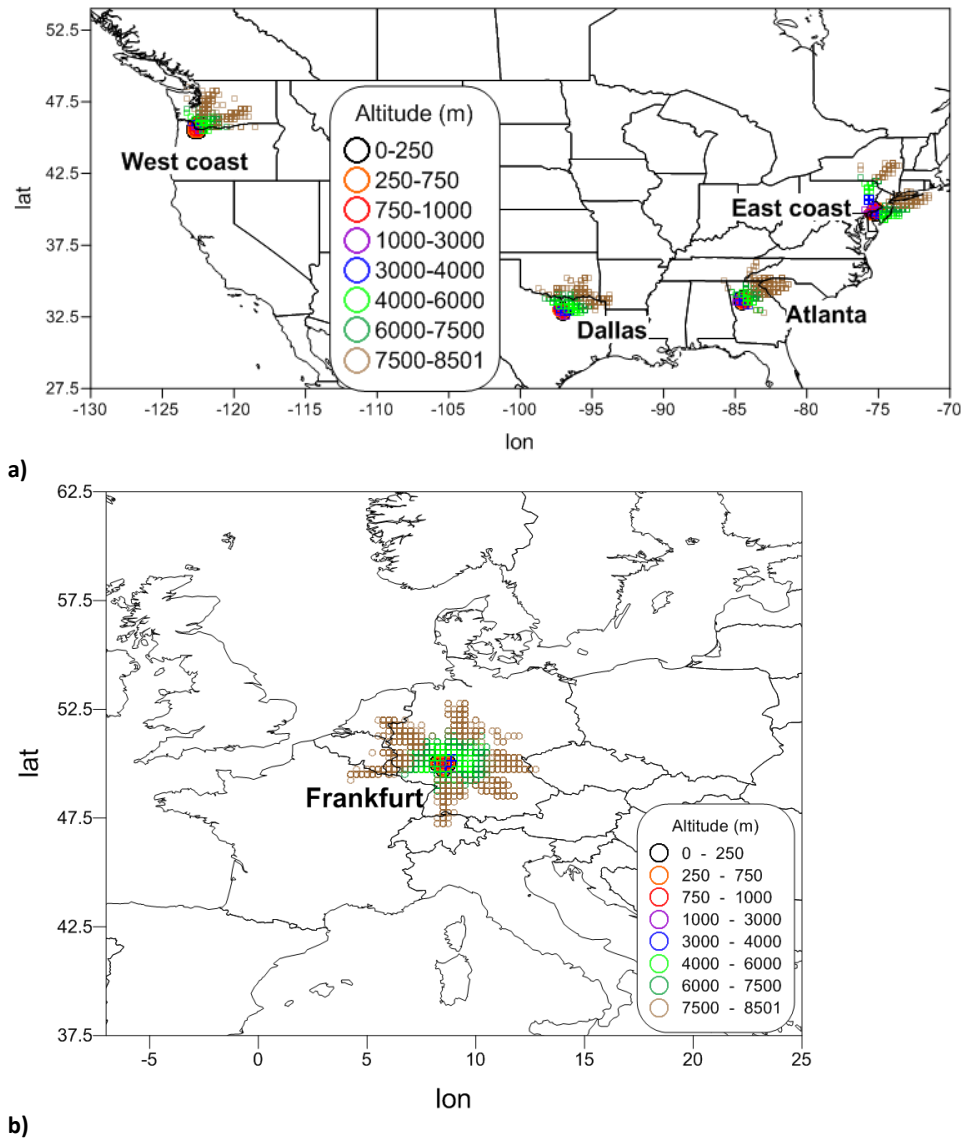
Fig. 1. Flight trajectories by height for the selected airports in **(a)** North America and **(b)** Europe.

3.1.4 MM5-CHIMERE

The CHIMERE model (Bessagnet et al., 2004) applied over EU uses a 0.25° (∼25 km) horizontal grid spacing and nine vertical layers expressed in a hybrid σ-pressure coordinate system between surface and 500 hPa. The first near-surface layer height was 20 m. All meteorological fields are interpolated from the driver meteorology (MM5), with the vertical velocity recalculated to preserve mass balance. Turbulence in the PBL is represented using a diffusivity coefficient following the parameterization of Troen and Mahrt (1986) without the counter-gradient term. The second-order Van Leer scheme (Van Leer, 1979) is used for the horizontal transport, and horizontal diffusion is neglected. Top BCs are ob-

tained from the GEMS re-analysis at 3 h interval and interpolated in three dimensions. The depth of the PBL is taken directly from MM5. Full model documentation is available at http://euler.lmd.polytechnique.fr/chimere

3.1.5 ECMWF-LOTOS-EUROS

The LOTOS-EUROS model (Schaap et al., 2008) simulated pollutant concentrations over EU at a regular horizontal grid spacing of approximately 25 km. In the vertical, the model is defined on four layers: a 25 m surface layer, the PBL, and 2 residual layers with a top at 3.5 km (or higher in mountainous areas). The height of the PBL, as well as the other meteorological input, is taken from ECMWF meteorological fields (short range forecasts over 0–12 h at 3 h interval). Advection

Table 2. Participating models and features.

Met	AQ model	Resolution (km)	Vertical (bottom-top)	Emission	BC
North America					
GEM	AURAMS	15 × 15	28 (14 in the first 2 km)	Standard[b]	Climatology
MM5	DEHM	50 × 50	29 (100 hPa)	Global emission database/ EMEP[d]	Estimated from hemispheric domain
WRF	CMAQ	12 × 12	34 (25 m to 50 hPa)	Standard	Standard
WRF	CAM×	12	23 from 30 m	Standard	Standard
Cosmo-CLM (CCLM)	CMAQ	24 × 24	30 (60 m to 100 hPa)	Standard[a]	Standard
Europe					
MM5	DEHM	50 × 50	29 (100 hPa)	Global emission database/EMEP[d]	Estimated from hemispheric domain
ECMWF	SILAM	24	9 (10 km)	Standard	Standard
MM5	CHIMERE	25	9 (20 m to 500 hPa)	Standard	Standard, MEGAN
ECMWF	LOTOS-EUROS	25	4	Standard[a]	Standard
WRF	CMAQ	18	34 (20 m to 50 hPa)	Standard[a]	Standard
MM5	CAM ×	15	23 (30 m to 12 km)	Standard	Standard
WRF	WRF-Chem1,2	22.5	35 (16 m to 50 hPa)	Supplied by WRF/Chem (constant background values)	Standard
Cosmo-CLM (CCLM)	CMAQ	24 × 24	30 (36 m to 100 hPa)	Standard[a]	Standard
MM5	Polyphemus	24	9 (up to 1200 m)	Standard[a,c]	Standard

[a] Standard anthropogenic emission and biogenic emission derived from meteorology (temperature and solar radiation) and land use distribution implemented in the meteorological driver (Guenther et al., 1994; Simpson et al., 1995). [b] Standard anthropogenic inventory but independent emissions processing, exclusion of wildfires, and different version of BEIS (v3.09) used [c] Emissions include biomass burning, biogenic organic compounds of SOA: α-pinene, limonene, sesquiterpene, and hydrophilic isoprene.
[d] Global: IPCC RCP 3-PD Lamarque et al. (2010); Bond et al. (2007); Smith et al. (2001); ships: Corbett and Fischbeck (1997); GEIA natural emissions (Graedel et al., 1993); wildfires as in Schultz et al. (2008). Europe: EMEP (Vestreng and Støren, 2000)

of tracers is implemented in all three dimensions following Walcek (2000), where vertical mass fluxes are derived from the mass balance. Vertical mixing is primarily determined by the growth of the second model layer following the rise of the PBL during the day, which leads to mixing of concentrations from the residual layer to the lower layer. Additional vertical mixing is implemented following standard K-diffusion (Louis, 1979). In the horizontal direction, no explicit diffusivity was added, apart from the numerical diffusion implied by the advection on a discrete grid.

3.1.6 MM5-Polyphemus

The Polyphemus AQ modelling platform is used with the Polair3D Eulerian chemistry transport model (Sartelet et al., 2012). Over EU, the horizontal resolution is 0.25° (∼25 km). Polyphemus used terrain-following height coordinates with nine vertical layers ranging from 20 m to 9 km for the AQMEII simulation. The reactive-transport equations are solved using operator splitting (sequence: advection, diffusion, chemistry and aerosol). The advection scheme is a direct space–time third-order scheme with a Koren flux limiter.

Diffusion and chemistry are solved with a second-order Rosenbrock method. The vertical eddy-diffusion coefficient is parameterized following Louis (1979), except in the unstable convective PBL where the coefficients are calculated using the parameterization of Troen and Mahrt (1986). The horizontal diffusion coefficient is constant and equal to $10\,000\,m^2\,s^{-1}$. In the AQMEII simulation, meteorological fields are interpolated from MM5, except for the eddy-diffusion coefficient and the vertical velocity, which are derived from the continuity equation.

3.1.7 GEM-AURAMS

The AURAMS model (Gong et al., 2006; Smyth et al., 2009) applied for AQMEII on the NA domain uses a secant polar-stereographic map projection true at 60° N, with 45 km horizontal grid spacing. In the vertical, 29 layers are used with a terrain-following modified Gal-Chen vertical coordinate and relatively thin layers near the surface (the first five layers were located at 0, 14, 55, 120, and 196 m) increasing monotonically in thickness to a model top at approximately 22 km. Horizontal and vertical advection are both calculated using a semi-Lagrangian advection scheme (e.g. Pudykiewicz et al., 1997) with wind components provided by the GEM meteorological model (Côté et al., 1998a, b; Mailhot et al., 2006). The vertical diffusion operator is solved using an implicit first-order Laasonen scheme, where the vertical eddy-diffusion coefficient is parameterized in GEM using a turbulence kinetic energy scheme (Benoit et al., 1989; Bélair et al., 1999). Horizontal diffusion is neglected. Model output is only saved for the first 20 layers, so vertical profiles are only available for the AURAMS model up to 5 km.

3.1.8 WRF-WRF/Chem

The only model with atmospheric chemistry coupled online with the meteorology applied within this study is the community WRF/Chem mode (Grell et al., 2005). The two instances of WRF/Chem applied for EU are the same in all aspects except that only one (referred to as WRF/Chem1) includes feedback of the direct and indirect aerosol radiative effects to meteorology. The other simulation that does not include the radiative feedback effects is referred to as WRF/Chem2. Both simulations are set up with a horizontal grid spacing of 22.5 km, 36 vertical layers, and identical physics and chemistry options as described by Forkel et al. (2012). Gas-phase/aerosol chemistry and non-hydrostatic physics within WRF/Chem are tightly coupled, with fifth-order horizontal, third-order vertical, and a third-order Runge–Kutta time integration scheme (Skamarock et al., 2008). YSU PBL physics (Hong et al., 2006) are used for vertical mixing within the PBL. Boundary conditions for all species are derived from the default WRF/Chem configuration, which are designed to be representative of clean, mid-latitude Pacific Ocean conditions. In spite of major differences in simulated solar radi-

ation for cloudy conditions, only small differences between the two model versions were found, since simulated T, mixing ratio, and wind are nudged to the driving global analysis above the PBL for the simulations discussed here (Forkel et al., 2012).

3.1.9 Cosmo-CLM-CMAQ

At the Helmholtz-Zentrum Geesthacht (HZG) Institute, the CMAQ model is set up using 24 km horizontal grid spacing and 30 vertical layers for both continents. Eleven of the 30 layers are below 1000 m, with the lowest layer top at approximately 36 m. Version 4.6 of the CMAQ (Matthias et al., 2010; Aulinger et al., 2011) model was used for the EU domain, while version 4.7.1 of the model was used for NA (the same version as the other CMAQ participants). Horizontal and vertical advection schemes in CMAQ use a modification of the piecewise parabolic method (PPM; Colella and Woodward, 1984). At each grid cell, a vertical velocity component is derived from the horizontal advection that satisfies the continuity equation using the density from driving meteorological model. In CMAQ 4.7.1, this scheme is further modified by adjusting the vertical velocities by the ratio of upwind fluxes to PPM-calculated fluxes. Vertical diffusion is based on the asymmetric convective model (ACM; Pleim and Chang, 1992). In both CMAQ 4.6 and 4.7.1, version 2 of the ACM (ACM2; Pleim, 2007, 2007a) is implemented. The minimum value for the eddy diffusivity depends on the land use of the individual grid cells and varies between $0.5\,m^2\,s^{-1}$ in grid cells with no urban areas and $2\,m^2\,s^{-1}$ in grid cells that contain only urban area. Zero flux BCs are used at the horizontal borders and at the top of the domain.

3.1.10 WRF/MM5-CAMx

Simulations for both the NA and EU continents use CAMx version 5.21 (ENVIRON, 2010), with the CB05 chemical mechanism (Yarwood et al., 2005). The EU domain has 0.125° by 0.25° ($\sim 12\,km \times 25\,km$) latitude/longitude grid spacing and 23 vertical layers that follow the WRF pressure levels between the surface and 100 hPa. The MM5 vertical domain has 32 vertical layers with a 30 m deep surface layer. CAMx employed fewer vertical layers (23) than MM5 to reduce the computational burden of the AQ simulations. The CAMx vertical layers exactly match those used in MM5 for the lowest 14 layers (up to ~ 1800 m), and above this altitude were aggregates of several MM5 layers. The top CAMx layer placed at 12 km height was 3.4 km deep. The NA domain has 12 km horizontal grid spacing and 26 layers that follow MM5 pressure levels between the surface and 50 hPa. The BCs and emissions provided by AQMEII are used, with the exception of biogenic emissions for EU, which are estimated using MEGAN (Guenther et al., 2006). The CAMx vertical transport scheme is described by Emery et al. (2011) with vertical advection solved by a backward-Euler (time)

hybrid centred/upstream (space) scheme and vertical diffusion solved using K-theory.

4 Model performance in the troposphere

4.1 Mean vertical profiles

Figure 2 presents the comparison of modelled versus observed seasonally averaged mean profiles of ozone and annually averaged mean profiles of CO, T, RH and WS. Observations are the grey symbols. While interpreting the results, it should be kept in mind that a source of model underestimation may derive from the incommensurability of comparing point measurements over airports to grid cell model values (Elguindi et al., 2010).

All models show similar mean profiles of wind above the PBL and T from the ground to higher level and in agreement with the observed profiles, most likely due to the nudging techniques applied to these fields. The MM5-DEHM is the only model with a positive bias for WS, which increases with altitude at all sites, possibly due to the grid spacing, which is two to three times coarser than the other models. Relative humidity profiles have the lowest bias within the first approximately 2 km, but significantly diverge above that height at all sites. It should be noted that in regions where the water vapour mixing ratio is very low (e.g. in the diverging zone), small differences in T can give rise to significant differences in RH. The RH peaks at 1000 m at all airports, although with different magnitudes. The peak is generally well captured by the models at all airports, with the exception of Dallas, where the bias at 1000 m ranges between 7 and 15 %.

The observed annual mean profiles of CO peak at ground level and diminish monotonically with altitude. The rate of decay is faster in the first 1.5–2.5 km (dependant on the airport) due to enhanced mixing near the surface, and is smoother aloft. This feature is common to all sites. The value of the observed CO peak ranges between a minimum of 166 ppb at Dallas and a maximum of 250 ppb at Frankfurt, with the remaining airports all in the range of approximately 200 ppb. Since CO is a directly emitted pollutant, its concentration is primarily determined by local emissions and BCs for CO. The modelled profiles of CO are significantly biased, typically with the largest bias near the ground, with few exceptions. While the DEHM model is biased low throughout the altitudinal range at all sites, the other models tend to have high biases near the ground, the exception being the CCLM-CMAQ model, which performs the best in general, possibly due to a more effective vertical mixing scheme employed by the CCLM driver compared to the WRF driver for the other CMAQ simulations.

Mean profiles of observed CO show a steep decrease in mixing ratio in the first 2 km, with an average gradient ranging between approximately −24 ppb km (Dallas and Philadelphia) and −37 ppb km (Atlanta), while Frankfurt is an outlier at −57 ppb km. Between 2 and 8 km, the rate of decrease in CO mixing ratio is considerably less, with a gradient of approximately 5–6 ppb km^{-1} at all airports. Averaged profiles of modelled CO show that ground-level mixing ratios can differ by up to a factor of 2, although differences are smaller above the PBL, but still significant, as discussed in detail with the aid of Taylor diagram in Sect. 4.2.4 (the large range of the horizontal axis in Fig. 2 has the effect of clustering the observed and modelled profiles at any given altitude, thus deceptively reducing the bias). A possible reason for the difference might be related to the horizontal grid spacing of the AQ model (e.g. Brunner et al., 2003). As shown in Table 2, the DEHM model, which has the coarsest horizontal grid resolution, simulated the lowest mixing ratios, whereas the CAMx and CMAQ models, with much smaller horizontal grid spacing, simulated the highest mixing ratios. Another possible reason could be related to the influence of the chemical BCs to CO and ozone. At the NA airports that are closer to the boundaries (i.e. Portland and Philadelphia), the models that used GEMS BCs (i.e. CMAQ and CAMx) simulated CO mixing ratios that are closer in magnitude to each other than the DEHM and AURAMS models, which are driven by different BCs (model sensitivity to BCs is further discussed in Sect. 6).

The seasonally averaged ozone has the lowest mixing ratios in winter (all airports) and the typical maximum in spring and summer. The mean ozone mixing ratio increases with altitude in the first 1000 m, as near-ground depletion by deposition and titration reduce the ozone availability (e.g. Chevalier et al., 2007). The effect of surface processes on ozone is also evident by the strong gradient in the first 2 km of the troposphere, ranging on average between 10 and 20 ppb km^{-1} at all sites. Modelled ozone mixing ratio in winter is typically biased low at all locations, with the exception of the DEHM model (both NA and EU). The CCLM-CMAQ, MM5-CAMx, and SILAM for EU are all biased high above approximately 6 km throughout the year, likely due to inadequate representation of the tropopause in the model, which results in too much mixing between the high stratospheric ozone mixing ratios in the BCs and the troposphere in the model. Ozone within the PBL is generally overestimated in the summer (NA only), while in the autumn the simulated mixing ratios are closer to the observed values, especially for Frankfurt. Layered ozone structures are most notable in summer months in the lowest levels close to the surface as result of the photochemistry and thermal mixing. At Portland, by contrast, the observed ozone in June–August peaks at 3 km and is underestimated by all models. A relative minimum of RH at the same height (annual average), which corresponds to the plateau values ($\sim$45 %) of RH above 3 km, might indicate a downward intrusion of ozone from the upper
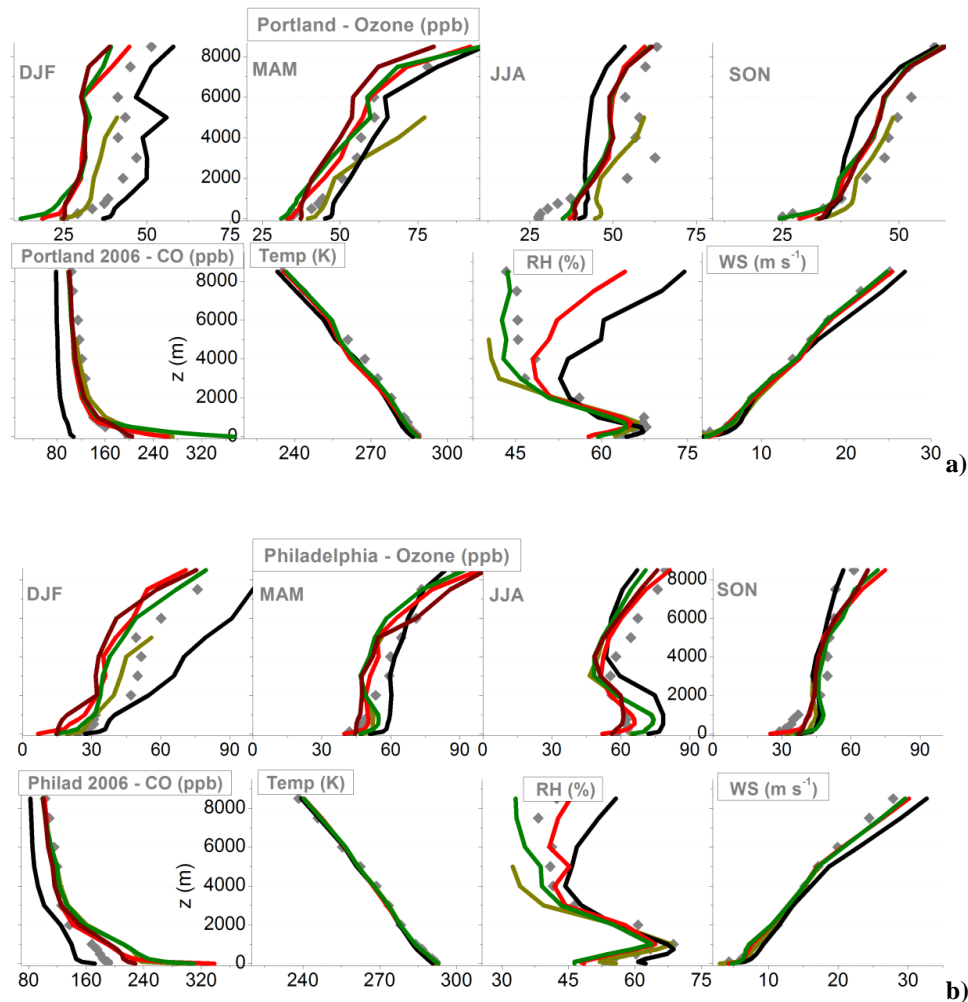
Fig. 2. Seasonally averaged mean profiles of ozone (top rows) and annually averaged mean profiles of CO, temperature, relative humidity (RH) and wind speed (WS) (bottom rows) (year 2006) for **(a)** Portland, **(b)** Philadelphia, **(c)** Atlanta, **(d)** Dallas, and **(e)** Frankfurt. The grey symbols are the observations.

troposphere (Tulet et al., 2002). The lack of seasonal variation of ozone for WRF/Chem in the upper air is due to the use of constant background values of ozone as BCs.

4.2 Operational statistics and variability

In this section, operational statistics are presented with the aid of Taylor diagrams (Taylor, 2001), which simultaneously show error, standard deviation and Pearson correlation coefficient (PCC). In a Taylor diagram, the observed field is represented by a point at a distance from the origin along the abscissa that is equal to the variance. All other points on the plot area represent values for the simulated fields and are positioned such that the variance of the modelled fields is the radial distance from the origin, the correlation coefficient of the two fields is the cosine of the azimuthal angle (desired value: 1), and the RMSE is the distance to the observed point

(desired value: 0). In practical terms, the closer a model point is to an observed point, the better the model performance.

To synthesise the discussion, we present annually averaged statistics (Figs. 3–6). At any given height, the number of paired observation–model data is the same as the number of MOZAIC flights for all of 2006 reported in Table 2. Because flight paths differ between ascents and descents and can also vary due to meteorological conditions, the horizontal spatial location of the various measurements for a given altitude will also vary, and computing a temporal average across all data points also implies computing a spatial average. The spatial coverage of the virtual "horizontal plane" containing the individual flight trajectories over which the spatial averaging is calculated depends on the height. While the $z = 0$ level has the spatial coverage of a point (having the airport's coordinates), that of the $z = 8.5$ km level is much larger, containing all the trajectories (ascending and descending) to and from
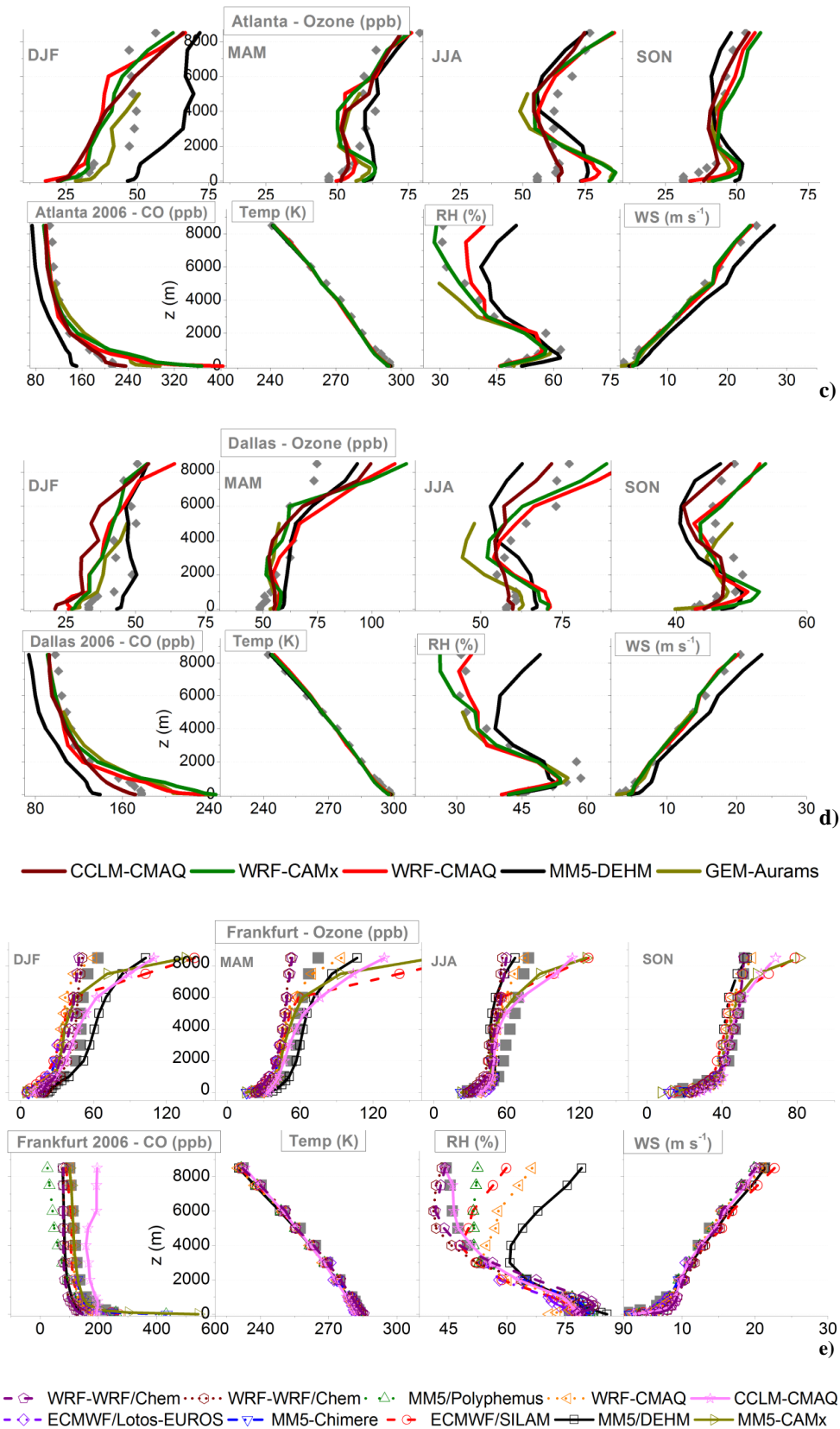
Fig. 2. Continued.

the airport. This aspect has an influence on the spread of the data. Coefficients of variation (CoVs), defined as the ratio between the standard deviation and the average of the annual distribution at any height, have been calculated in order to assess the model's capability of reproducing the observed variability. For synthesis, CoVs are shown for Frankfurt only (Fig. 7), but are discussed for all airports.

Performance for T is considered to be satisfactory for all models, at all domains for at least the first 5 km (PCC in excess of 0.90 and fractional bias ≤ 1 %). For this reason, analysis for T is not shown.

4.2.1 Wind speed

For all the airports, all models show significant errors in the modelled WS, and poor correlation with the observed values in the first 100 m. Although the sign of the bias cannot be deduced from a Taylor diagram, further analysis (not shown) indicates that the models are biased high. Smaller errors (still positive bias) are also observed at 500 and 1000 m. Poor correlation and moderate to high error within the PBL are common to all models at all sites, thus indicating that the meteorological models are not reproducing WS in the PBL precisely, which is arguably the most important region of the atmosphere in relation to AQ processes. This result supports the findings by Vautard et al. (2012) for the ground-based measurements. The poor model performance in the first 100 m at all sites is also confirmed by the analysis of the CoV, with modelled values well below the observations.

A feature emerging is the clustering of performance based on altitude rather than on modelling system. For heights below the PBL, all models show the same deficiencies, indicating that incorrect simulation of WS near the surface is a common problem. There may be several reasons for this. First, near-surface winds are sensitive to several driving factors other than the synoptic circulation, such as land cover and surface energy exchange processes, which have been found to exhibit large model-to-model variability in global and regional models (e.g. see de Noblet-Ducoudré et al., 2012; Stegehuis et al., 2012). Second, unlike upper-air winds which are measured by rawinsondes and are incorporated in reanalysis products, surface winds are not used for producing reanalyses. Therefore, even with a strong nudging of the wind field to the reanalyses, a lower model skill is expected for surface winds as compared to upper-air winds. This is demonstrated by the improvement of model performance above 500 m. For all models at all sites, and more markedly at Portland and Philadelphia, PCC, error and variability improve constantly with height. In general, the best performance by all models is observed for $z \geq 3$ km, with PCC > 0.8 at all NA sites, with the exception of the WRF/Chem model (PCC ~ 0.65 at $z = 3$ km), and at Frankfurt.

4.2.2 Wind direction

Although of pivotal importance for the transport of the polluted air masses, evaluation of the modelled WD is often overlooked. In this section we summarise the results of evaluating the AQ models of Table 2 using observed WD data from ground to 8.5 km. To simplify the analysis we have binned the frequency counts of the occurrences of WD angle using a 20° interval. Results are discussed in terms of normalised bias (modelled minus observed count, and divided by the observed count). In general, we find that the bias in the PBL for the Frankfurt airport is smaller and decreases faster with height than the NA airports. For the NA airports, in general there is a large bias within the PBL, which is most pronounced on the west coast (Portland).

For the Portland airport, WRF-CMAQ and GEM show similar biases from 0 to 500 m: positive for north-west and negative for south-east (~ 100 %). The other models are biased low in the eastern sector. From $z = 500$ to 1000 m, the bias increases and all models show a clear tendency to overestimate the occurrences of both westerly and easterly winds. Above the 1000 m height, the bias decreases and the models better reproduce the frequency distribution of observed WD. On the east coast of NA (Philadelphia airport), isolated cases of large bias are present at the 500 m and 4 km heights, where the two WRF models are biased high in the north and south sectors, respectively. For the inland airports of Atlanta and Dallas, some occurrences of very large bias are observed. For the Atlanta airport, all models are biased high in the west and north-west directions (50 %) in the first 250 m. From 500 to 1000 m, the models are biased high (50 to 100 %) in the north-eastern and south-western directions. For the Dallas airport, the east-west positive bias (~ 20 %) present for all of the models in the first 250 m turns clockwise at 500 m, directed in the north-west direction. GEM and WRF-CAMx models also acquire a south-western bias component (about 70 %) at 1000 m. The bias progressively decreases above 2 km. At the Frankfurt airport, the models share similar bias patterns, at the ground, with a tendency to underestimate occurrences of northerly winds (all models) and overestimate the westerly components (with different magnitude by all models).

Finally, we note that for the NA airports, the two simulations that used WRF do not always provide the same counts at all heights and at all locations. These two fields were derived from the same underlying WRF simulation. Therefore, the differences in this analysis were introduced when the two modelling groups independently interpolated these fields to the common horizontal and vertical analysis grid defined for the MOZAIC analysis under AQMEII and transferred these processed fields to ENSEMBLE. Thus, these differences highlight the difficulties in attributing differences in model performance to differences in model options, in particular for quantities with significant horizontal and vertical gradients such as wind direction.
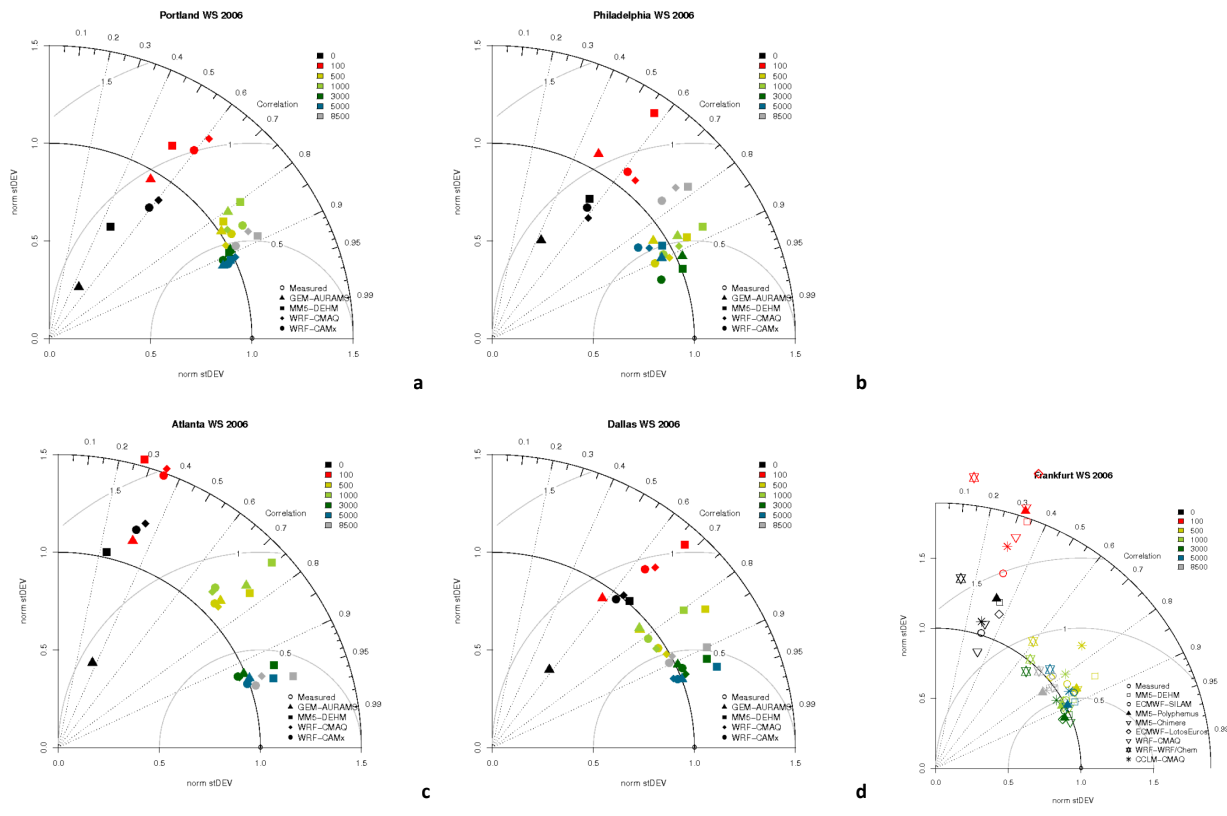
Fig. 3. Taylor diagrams for wind speed at (**a**) Portland, (**b**) Philadelphia, (**c**) Atlanta, (**d**) Dallas and (**e**) Frankfurt. Height by colours, model by symbols.
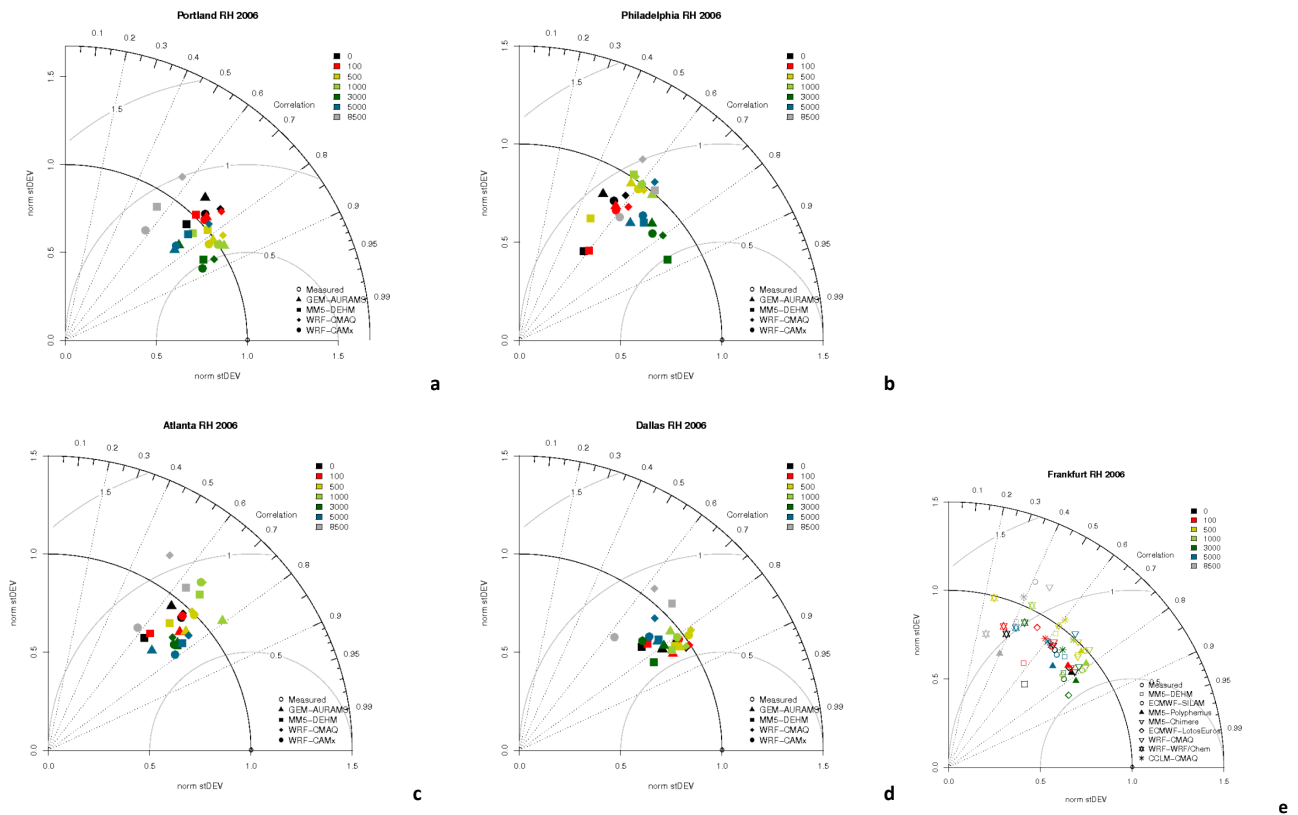
4.2.3 Relative humidity

In contrast to WS, modelled values of RH exhibit a smaller increase in variability with height and vary more by site than WS (Fig. 4). The variability is underestimated by all models, as shown by the normalised standard deviation of Fig. 4, most predominantly below unity at all airports, with the exception of Portland where the variability in the first 1000 m is close to the observed variability. The best model performance for variability occurs between 500 and 1000 m, and then deteriorates with height. As stated earlier, this could be due to very small water vapour mixing ratio values aloft, which makes RH computation less reliable at higher altitudes. The analysis of the CoV reveals that the models tend to underestimate the observed variability in the free troposphere, especially for $z > 6$ km. Model performance is in general grouped by height, although with a few exceptions (such as the WRF/Chem at Frankfurt, Fig. 4e). The minimum model errors and maximum PCC occur for altitudes in the range of 1 to 3 km, although with some exceptions (most notably Atlanta) and not consistently for all models. For example, the Taylor diagram for Frankfurt (Fig. 4e) reveals a large scatter of data independent of height and model, with the same model performing well in some instances (LOTOS-

EUROS at 3 km) but very poorly in others (the same model is off-scale at the surface). Further investigation is needed to determine whether the modelled variation of RH is not ill-treated, or in other words whether the modelling systems are getting the "right answer for the right reason", which is one of the main objectives of the diagnostic model evaluation component defined in Dennis et al. (2010).

4.2.4 CO

Profiles of modelled CO show PCC values typically below 0.6 at all heights and airports, with a limited number of exceptions (e.g. WRF-CMAQ at Frankfurt, whose PCC above 3 km exceeds 0.7) (Fig. 5a). All of the models have difficulty reproducing the temporal variability of CO mixing ratios, possibly due to the influence of local sources close to the airport, whose temporal variation is not captured by the models due to too coarse grid spacing. Analysis of the CoV demonstrates that the variability is largely overestimated in the PBL at all NA airports (up to a factor of two at Dallas, Atlanta, and Philadelphia). CoV estimates improve slightly with height above the PBL, with the best agreement in the range of 2 to 4 km. This may be due to a stronger dependence on the large-scale transport from the BCs combined with the

Fig. 4. Taylor diagrams for relative humidity at **(a)** Portland, **(b)** Philadelphia, **(c)** Atlanta, **(d)** Dallas and **(e)** Frankfurt. Height by colours, model by symbols.

nudging of winds at these altitudes, while in the lower troposphere CO mixing ratios are the result of a more complex mix between transport and emission.

The analysis for CO was not separated by season, but as there are more measurements available in summer than in winter, a bias to summer months is expected in the Taylor diagrams of Fig. 5a. CO concentrations are typically lower in the spring and summer than in winter at all altitudes with a minimum in July and maximum in February to April (Gilge et al., 2010; Elguindi et al., 2010). As a result, the annual mean deviation of the modelled to the observed profiles in Fig. 2 and the error in Fig. 5a are primarily associated with lower CO mixing ratios. Because model performance for CO is similar across all models, this may indicate problems with the inputs that were shared among the models. In particular, the fact that at the ground the majority of models overestimate observed mixing ratios suggests that emissions from certain source categories, such as mobile sources or biomass burning, are overestimated. Another reason could be the emission profiles used. If CO is primarily emitted in the lowest layer, mixing ratios at the surface might be overestimated. This hypothesis could be explored further by comparing the vertical profiles of other more long-lived species, such as aerosols. However, it is beyond the scope of this study to

conclusively determine the reasons for the poor model performance near the surface.

It is important to consider that biased meteorological drivers might have a bigger impact on long-lived species such as CO. Any error in the input wind field (speed and direction) can accumulate over time, possibly exacerbating errors introduced by imperfect emission inventories. To test this hypothesis for CO, we calculated the correlation between the error in the modelled CO mixing ratio and the error in the WS at each given altitude and airport. Let $corr(f_{z,s1}; g_{z,s2})$ be the operator returning the PCC of any two fields f and g at height z for the species s_1 and s_2. Then $corr(e_{z,WS}; e_{z,CO})$ is the PCC between the error of the WS at height z and the error of CO at the same height, whose profiles are reported in Fig. 5b. From the graph it can be seen that there are instances for the NA airports where the two fields are largely positively correlated (up to values of ~ 0.3): CAMx, AURAMS, and DEHM models for the Philadelphia airport for $z = 500$ m and the WRF-CMAQ model at $z = 3$ km; the CCLM-CMAQ model for the Dallas airport for height up to 4 km and all models for the Dallas airport for z between 3 and 4 km. For these cases the errors of the two fields have the same trend. While a general trend is difficult to identify, there does appear to be a correlation between the fields for isolated
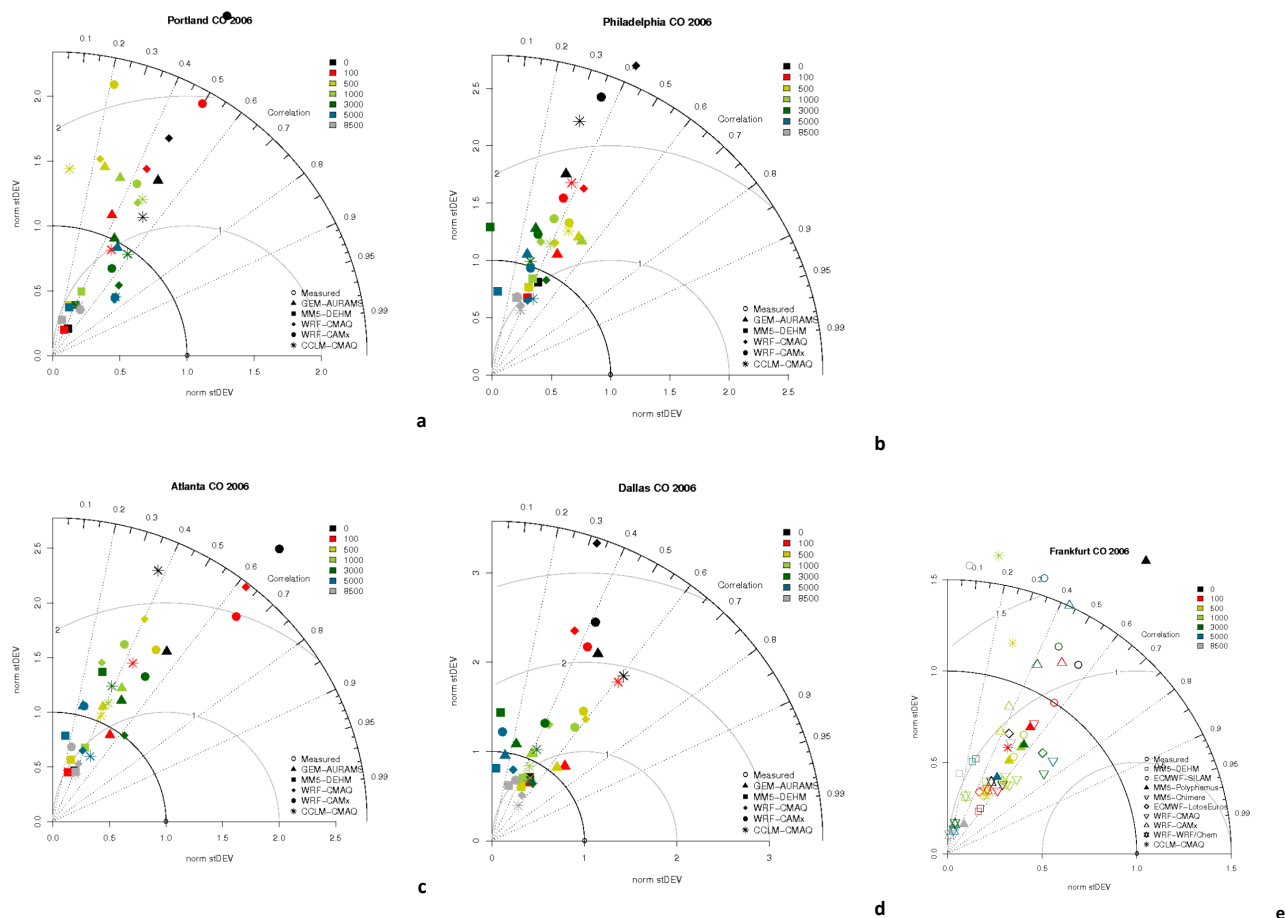
Fig. 5a. Taylor diagrams for CO at **(a)** Portland, **(b)** Philadelphia, **(c)** Atlanta, **(d)** Dallas and **(e)** Frankfurt. Height by colours, model by symbols.

model/airport/height combinations. Nonetheless, this analysis demonstrates that an association exists in certain cases, which deserve further investigations. We also note that instances of anti-correlation (negative PCC) are also numerous. In these cases, a decrease in WS error corresponds to an increase in CO error and vice versa, but it is difficult to determine whether they are associated with vertical intrusion or should just be treated as uncorrelated. Finally, WS and CO errors at Frankfurt do not show any association and can be considered independent.

4.2.5 Ozone

Due to its adverse effects on human health, climate, and ecosystems, surface ozone is among the most extensively studied atmospheric trace gases. Solazzo et al. (2012a) have shown that overall the AQ modelling systems participating in AQMEII perform satisfactorily in reproducing ozone mixing ratios at ground level (showing better skill in NA than in EU). For most models and airports, the yearly averaged ozone scores of Fig. 6 indicate better performance near the

ground ($z \leq 1000\,\mathrm{m}$) than aloft. At Portland the scores under and above the PBL are overall comparable, and at Frankfurt several models have scores higher for altitude between 500 and 1000 m (PCC in excess of 0.7) than for $z < 500\,\mathrm{m}$. The 1000 m height tends to mark a transition from the PBL regime, characterised by well-mixed dynamics and active photochemistry, to a tropospheric regime characterised by a more marked model-to-model variability. Similar conclusions were drawn by Chevalier et al. (2007).

No clear factor emerges for grouping the ozone skill scores in Fig. 6. The two simulations using CMAQ over NA (driven by WRF and CCLM) do not produce similar scores at any of the airports, although they share the same BCs and emissions. By contrast, CMAQ and CAMx, the two AQ models driven by WRF, have rather similar results, indicating that the meteorological driver may have a substantial weight in determining the skill for ozone (in particular PBL dynamics). However, this latter hypothesis is not confirmed by the Taylor plots for Frankfurt (WRF/Chem and CMAQ are both driven by WRF) though, as the scores are primarily grouped
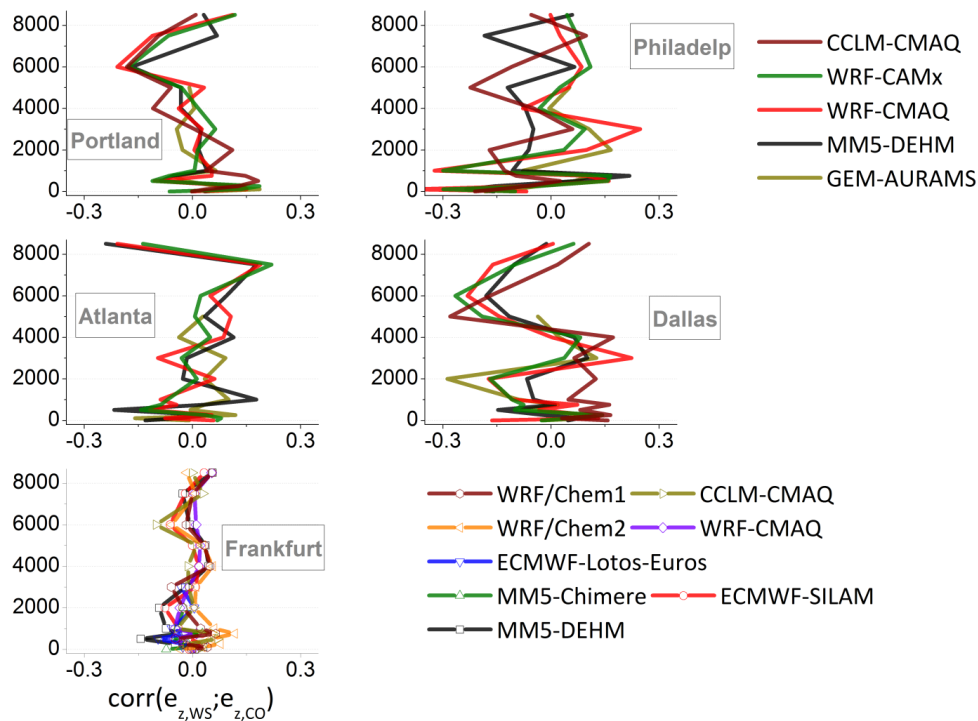
Fig. 5b. Profiles of correlation between the WS and CO errors for the airports of Portland, Philadelphia (first row), Atlanta, Dallas (middle row) and Frankfurt (lower panel).

by AQ modelling system, with the two simulations of CMAQ producing similar scores.

The standard deviation and the CoV are generally overestimated within the PBL and underestimated aloft, although with several exceptions depending on the site, height, and model used, especially at the Atlanta, Dallas, and Frankfurt airports where the variability score at high altitudes is comparable to that in the PBL. WRF/Chem at Frankfurt shows the lowest variability at all heights, and close to zero for $z \geq 3$ km, possibly because at these heights the simulated values are already influenced to some extent by the WRF/Chem BCs (constant values), which are too low, particularly in the spring and summer, at and above that height (see Fig. 2). The DEHM model underestimates the CoV at all NA airports, possibly due to the use of RCP emissions and different set of BCs (see Table 2).

The variability and errors above 3 km and up to 8.5 km reveal some model deficiencies. The mean seasonal profiles in Fig. 2 indicate that the models tend to underestimate ozone between 3 and 6 km, but they overestimate ozone mixing ratios above that height (see Frankfurt for example), though not systematically for all locations and all seasons. This could be due to an underestimation of the exchange between the upper troposphere and the lower stratosphere and/or diffusive model errors related to coarse vertical gridding. By contrast, for the case of global-scale chemistry models, Brunner et al. (2003) reported a tendency of models to overestimate

ozone mixing ratios over NA between 5.5 and 8 km throughout the year (based on the period 1995–1998). However, while those results were averaged over the entire NA continent, here we are comparing regional-scale models at considerably smaller scales, and the different behaviour could simply be attributed to the different grid sizing as well as the different time period.

5 Altitudinal correlations

In this section we investigate the degree of association between the error at the ground and the error aloft ($0 < z \leq 8.5$ km), measured by the PCC for each of the variables. With the terminology introduced in Sect. 4.2.4, the association is expressed as $corr(e_{0,s}, e_{z,s})$, where s is the variable under investigation (either ozone, CO, WS, RH) and z identifies the altitude in the range defined above (between 0 and 8500 m). The intent of this analysis is to investigate whether errors in each field stem from sources and processes common to the surface and upper air. If errors at the ground and aloft are caused by independent processes, we would expect the respective time series of errors to be uncorrelated. For example, if errors near the surface are primarily due to the treatment of PBL processes (photochemistry, deposition, emissions, turbulent mixing), these errors should not be correlated with errors above the PBL. Conversely, in the case where errors are
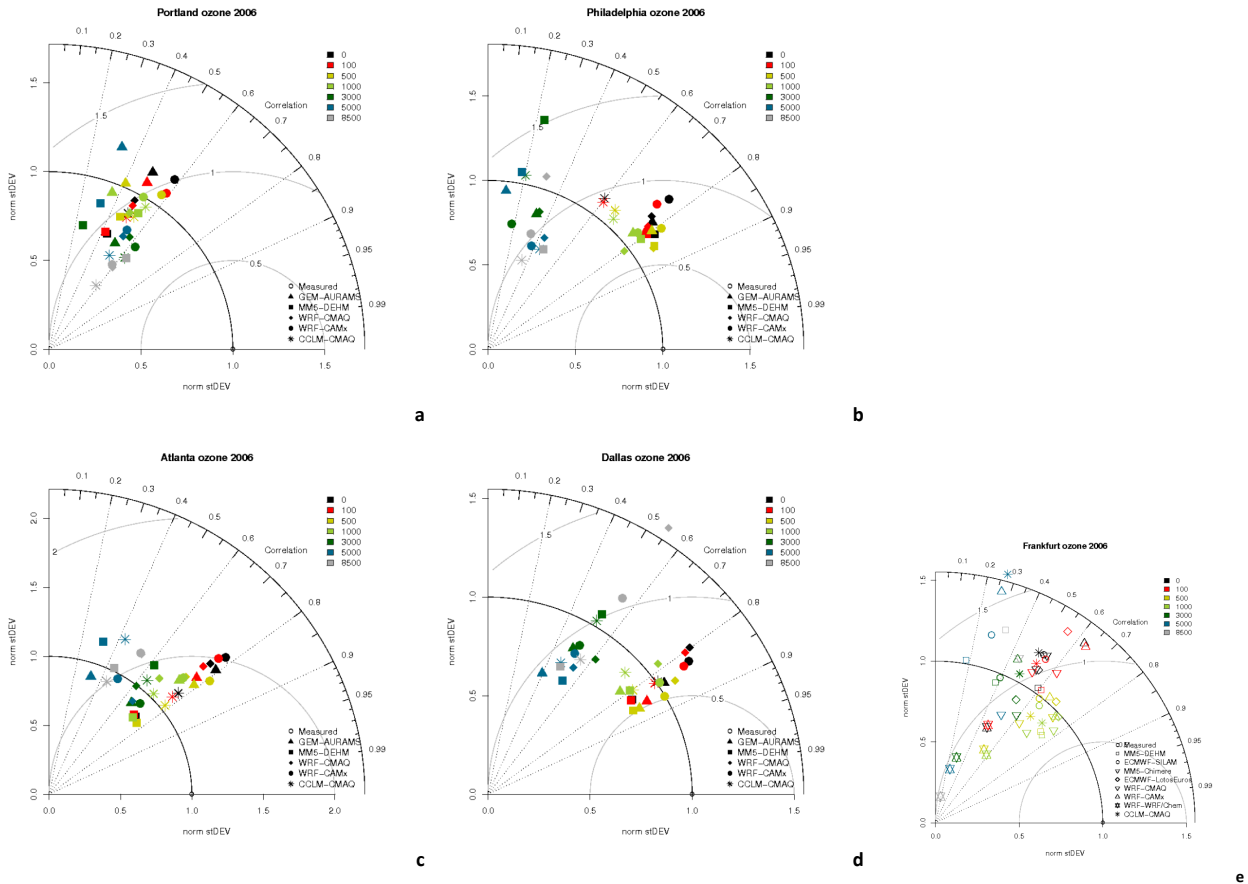
Fig. 6. Taylor diagrams for ozone at **(a)** Portland, **(b)** Philadelphia, **(c)** Atlanta, **(d)** Dallas and **(e)** Frankfurt. Height by colours, model by symbols.

contributed to the large-scale circulation, we would expect that under- or over-estimation is observed simultaneously at different altitudes, and therefore correlated.

Figure 8 reports how the fractional difference (FD) at the ground is correlated with the FD aloft for all investigated airports. FD is defined as

$$FD = \frac{Mod - Obs}{0.5\,(Mod + Obs)}, \qquad (1)$$

where Mod and Obs are N-length vectors denoting the modelled and the observed field value, respectively. Here N is the number of hourly flights available for the period being analysed (reported in Table 1). The correlations are calculated between the N-length FD vector at the ground and the N-length FD vector at other levels.

The profiles of FD correlations in Fig. 8 demonstrate that for the investigated fields (with some exceptions for CO as discussed below) positive correlations are limited to the PBL, and biases within the PBL are poorly correlated with those of the free troposphere at all airports. The gradient of the correlation profiles drops sharply with height for both WS

and RH. For these latter fields in particular, we observe the strongest vertical gradient as well as the most homogeneous behaviour among the different models, suggesting that the proper simulation of WS and RH in the PBL is principally determined by near-surface processes (the only exception is WRF/WRF-Chem at Frankfurt).

Correlation of FD profiles for CO exhibits a less homogeneous behaviour. PCC varies aloft between 0.1–0.2 and 0.4–0.6, this latter value primarily associated with the MM5-DEHM modelling system, for which the hypothesis of large-scale error might explain the correlation between bias at the ground and aloft (note that DEHM results reflect the strong vertical mixing shown by the mean vertical profiles of Fig. 2). Correlation of FD profiles for ozone is more homogeneous than CO among the different models, supporting the proposition that the vertical ozone profile is the result of an interplay between an ozone reservoir in the upper troposphere and ground-level mixing ratios (Zhang and Rao, 1999; Godowitch et al., 2011). Overall, errors produced in the PBL for ozone remain contained within the PBL (all models show similar patterns) and are weakly connected with those above,
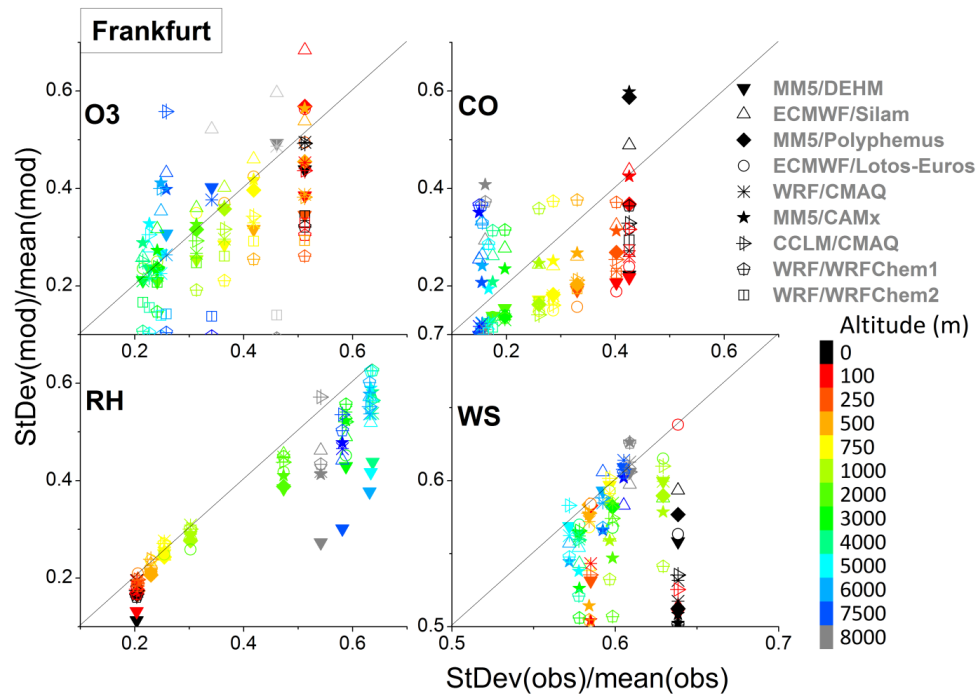
Fig. 7. Modelled (vertical axis) versus observed (horizontal axis) coefficients of variation for Frankfurt (year 2006) for (clockwise from top-left) ozone, CO, WS, and RH. The straight line is the perfect agreement. Height by colours, model by symbols.

although for Frankfurt this pattern is less pronounced. Above the PBL, and especially for the two CMAQ simulations at Portland and WRF/Chem at Frankfurt, the FD correlation profiles are more varied, typically low for CAMx (close to zero between 3 and 4 km at all airports) and scattered between PCC values of 0 to 0.3 for the other models.

Although the altitudinal correlation analysis is very informative about how far the errors at the ground propagate in the vertical, it cannot be considered exhaustive as it does not provide detailed insight into the modelling systems' ability to represent specific processes. Instead, it reveals the need for a more quantitative evaluation of vertical mixing. This could be accomplished by comparing the modelled PBL height with measurements; however, these data were not collected within AQMEII.

6   Altitudinal error: the case of ozone

6.1   Aggregate seasonal error

The seasonal error associated with the model's capability of estimating ozone mixing ratios in the vertical is investigated for all airports and models. The error is expressed as fractional absolute difference (FAD):

$$\text{FAD} = \frac{|\text{Mod} - \text{Obs}|}{0.5(\text{Mod} + \text{Obs})} \qquad (2)$$

(the elements in Eq. 2 have been defined in Eq. 1). Results are presented for the winter months of December, January, and February (DJF, Fig. 9) and for the summer months of June, July, and August (JJA, Fig. 10) in terms of cumulative FAD. The bars in Figs. 9 and 10 represent the sum of the errors produced by each model for that height. Since there are about twice the number of models for EU as for NA (9 vs. 5), one would expect the cumulative FAD of EU to be about twice that of NA. However, this is not the case for winter at Frankfurt, where errors at the ground are more than twice those of the NA airports. Almost every model has a large FAD at the ground, suggesting some common problem among the models (possibly biased emissions). With height the errors of CAMx and LOTOS-EUROS decrease, while SILAM's remains large (Frankfurt, winter). By contrast, in the summer, cumulative FAD at Frankfurt is about twice that of Philadelphia and Dallas, but approximately the same as Portland and Atlanta, where all of the models exhibit larger errors than at Philadelphia and Dallas (the AURAMS model has the largest FAD). The marked decrease of FAD with height for summer (Fig. 10) indicates that model discrepancies are mainly related to surface processes. Particularly, errors could stem from an overestimation of emissions in the surrounding area, giving rise to an excess of ozone production from primary precursors. As for ozone, the underestimation of deposition processes could also be a source of error. Horizontal advection and photochemistry schemes are also possible causes of model-to-model discrepancies (Zyryanov et al., 2012).
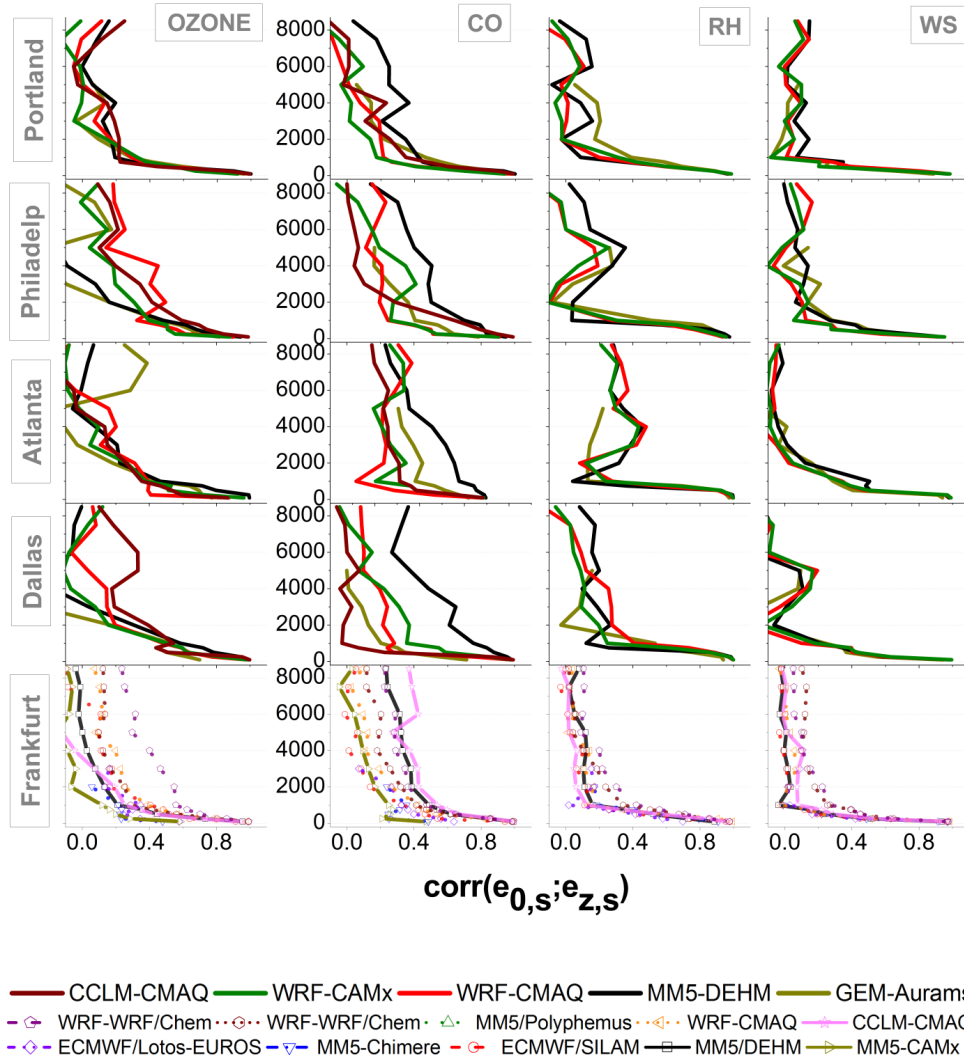
Fig. 8. Correlation coefficient between the fractional difference (FD) at any height and the FD at the ground ($z = 0$) for ozone, CO, RH, and WS (columns) and for the airports of Portland, Philadelphia, Atlanta, Dallas, and Frankfurt (rows) (year 2006). The first line of the legend refers to the NA models, while the other two lines to Frankfurt.

Furthermore, the error of the WRF-CMAQ model at the ground at Philadelphia in winter is larger than the other instance of CMAQ driven by CCLM, likely due to a higher titration effect. Figure 2 also shows that WRF-CMAQ exhibits the highest CO mixing ratio, suggesting that differences between the two CMAQ simulations may be due to the different vertical diffusion scheme used. Moreover, the distribution of FAD from the CCLM-CMAQ model (negligible at ground and larger above 500 m) at Portland in winter and at Philadelphia and Dallas in summer suggests that CCLM-CMAQ is able to capture the surface layer equilibrium between NOx and ozone better than the other models. It is also worth noting that the performance of CCLM-CMAQ aloft is comparable with those of the other models, indicating that surface-level processes are not strongly influenced by ozone

aloft. Finally, the large error at 8.5 km at Frankfurt is due to model top boundary being close to that height, as for example the SILAM model.

Overall the diversity of modelling errors both in the PBL (particularly in summer) and in the free troposphere (particularly in winter) indicates a large range of differences in the AQ models, particularly at Frankfurt in both seasons. These differences relate to turbulence, transport, and photochemistry in the PBL (the emissions are largely shared), and to grid spacing, top and lateral boundaries, and the advection schemes used in the free troposphere.
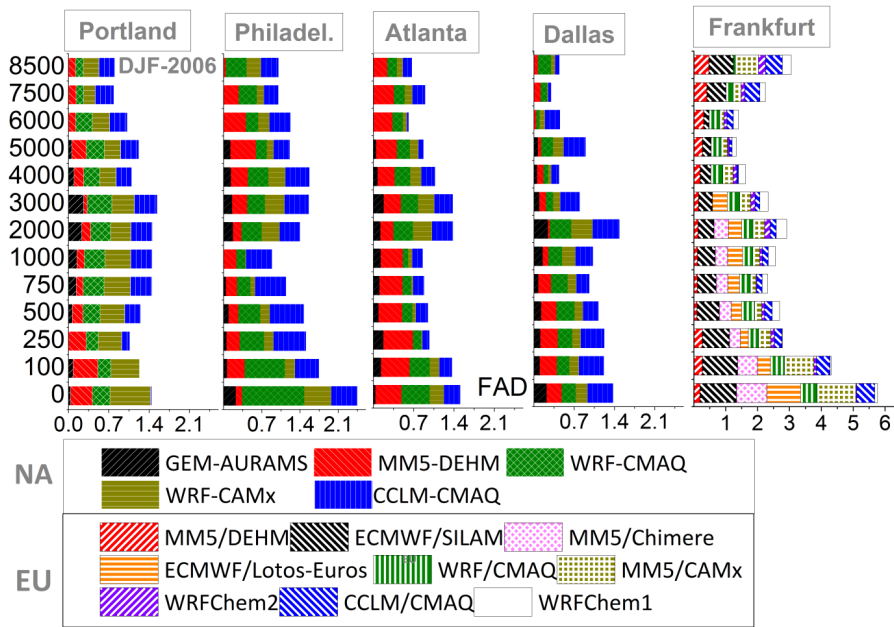
Fig. 9. Cumulative fractional absolute difference (FAD) of modelled ozone for the months of December, January, and February 2006.
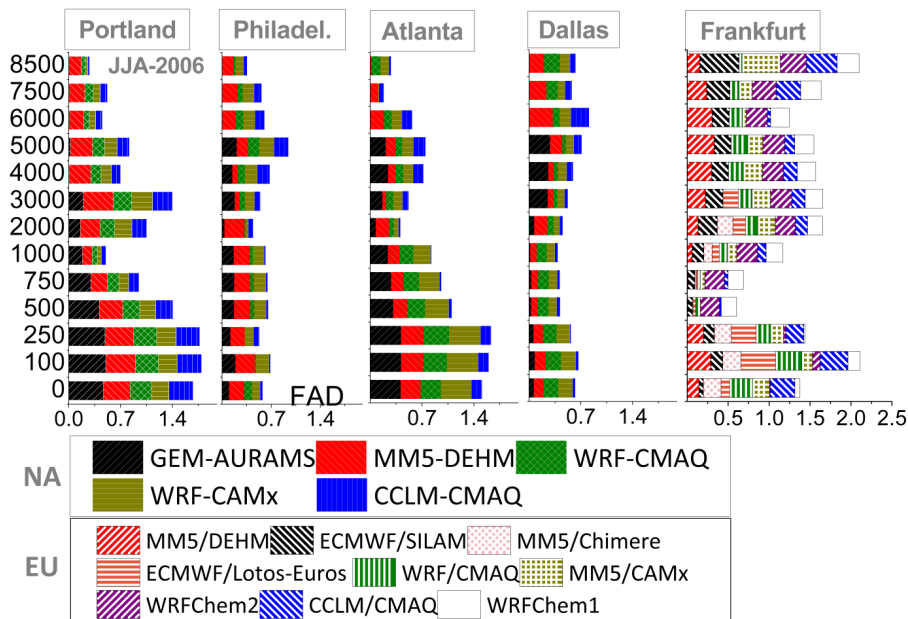

Fig. 10. Cumulative fractional absolute difference (FAD) of modelled ozone for the months of June, July, and August 2006.

6.2 Bias due to boundary conditions

In this section we expand the model evaluation by providing comparison between the lateral BCs for ozone and the vertical profiles of ozone measured on-board MOZAIC aircraft. We show the analysis for NA only, as the proximity of the airports to the edges of the modelled domain allows testing the influence of the BCs more effectively. Details about

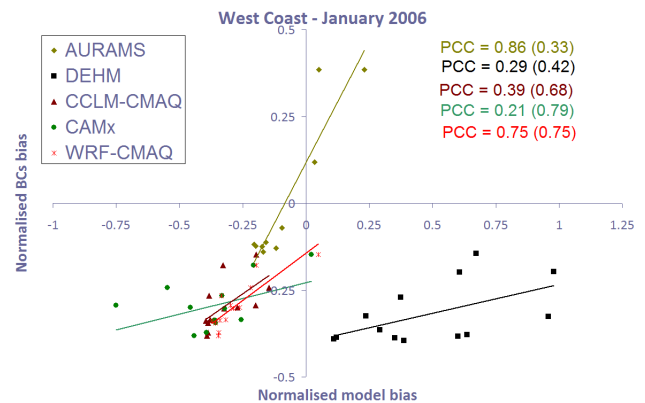the preparation of the standard BCs for AQMEII are given in Schere et al. (2012).

For the airports of Vancouver and Portland (west coast) and Philadelphia (east coast), we analyse the following hourly vertical profiles averaged over the months of January and August:
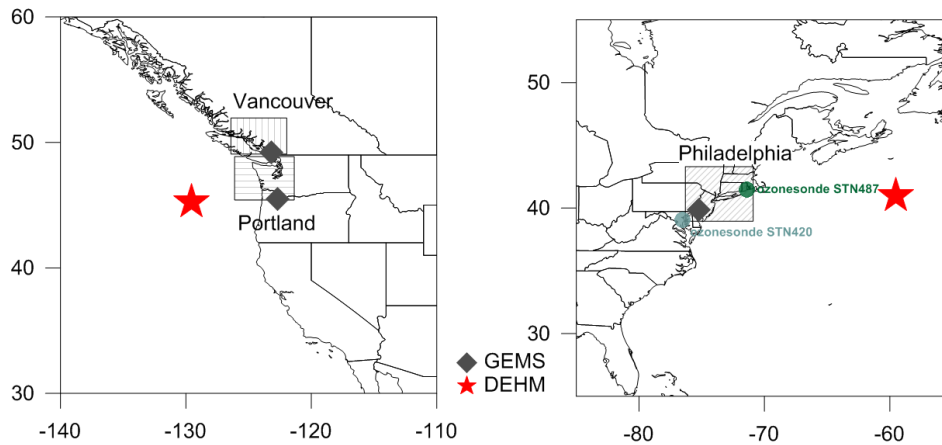
Fig. 11. Positions where BC data are provided for GEMS and DEHM (position for AURAMS is the same as for GEMS). For the east coast of NA, the locations of two ozonesondes used in the analysis are also shown. The shaded areas indicate the extension of the domain around the airport
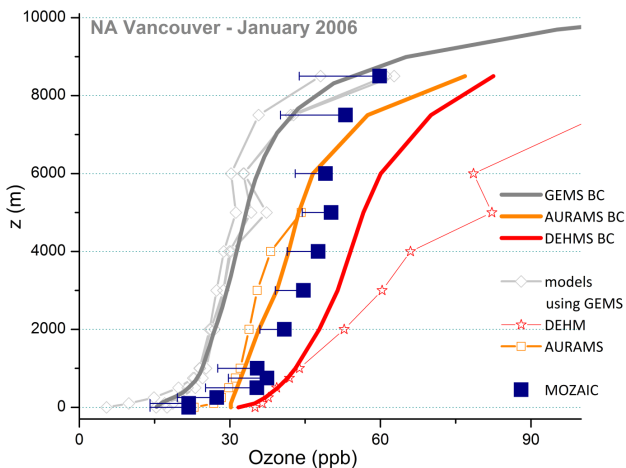
Fig. 12. Monthly averaged vertical profiles of ozone for Vancouver, January 2006: BC providers (thick lines), AQ model outputs (thin lines), MOZAIC (filled squares). The horizontal lines by the measurements are the standard deviations (symmetric about the mean point and only the left portion is shown).

– monthly ozone BCs adopted by the AQ models;

– monthly profiles of modelled ozone mixing ratio over the selected airports;

– monthly MOZAIC measurements of ozone at the selected airports and the observed standard deviation;

– ozonesonde data collected at rural locations falling within the MOZAIC domain for Philadelphia for the month of August 2006.

Results of the comparison are shown in Figs. 12 to 16. The GEMS BCs were adopted by the WRF-CMAQ,

Fig. 13. Normalised modelled vs. BC bias with respect to MOZAIC for the month of January 2006 for the Portland airport. The two CMAQ instances and CAMx use the same BCs. The continuous lines are derived from the linear regression fit, with the squared correlation coefficients reported on the top right corner (colour coded). In parentheses are the squared coefficients obtained by correlating the bias excluding the data in the first 250 m (i.e. by including data from 500 m to 8.5 km).
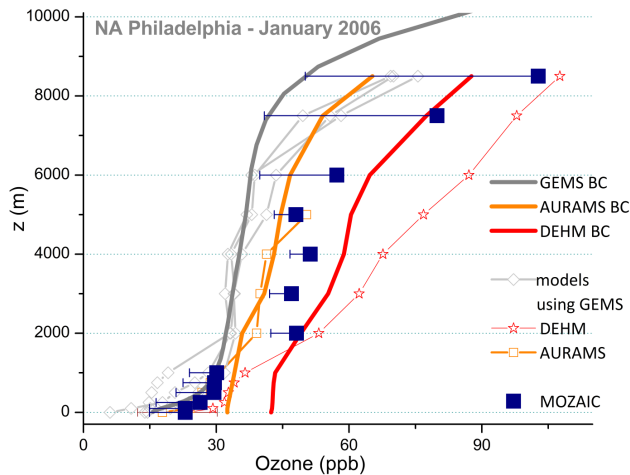
CCLM-CMAQ, and CAMx models (Table 2); the AURAMS group used its own BCs, while ozone BCs for DEHM were extracted from climatology and satellite measurements and applied to the model's outer domain, which encompasses the entire Northern Hemisphere. The concentration profiles at the DEHM's inner domains (one centred in NA and one in EU), which reflect the dynamics of the hemispheric domain, are taken here for comparison (the cells where the profiles were extracted are shown in Fig. 11).

During winter, the absence of strong photochemical activity amplifies the influence of BCs. For $z > 2$ km, the ozone monthly profiles for both US coasts overlap to a large extent
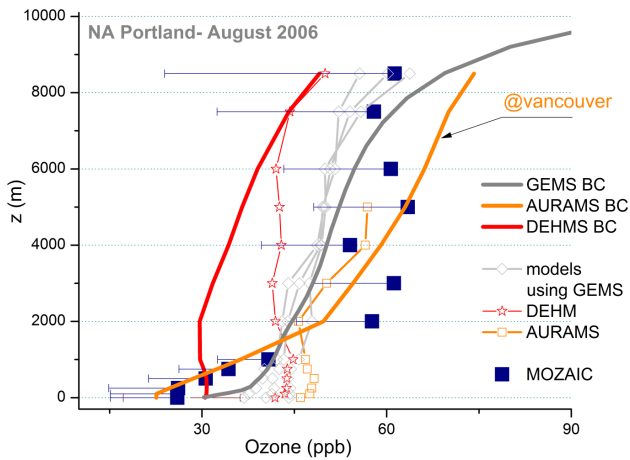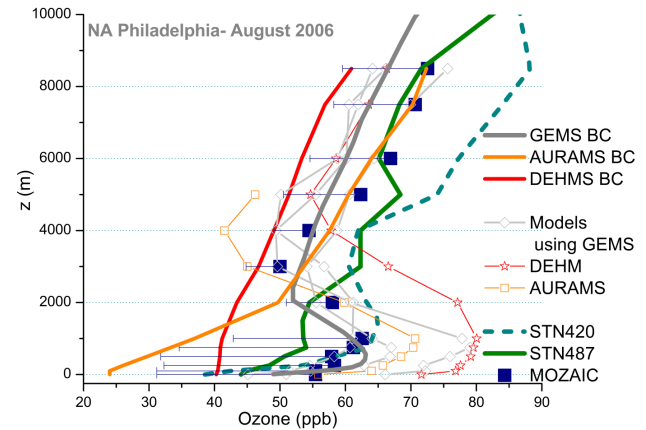
Fig. 14. Monthly averaged vertical profiles of ozone for Port-land, August 2006. The BC for the AURAMS model (thick orange line) was provided at Vancouver.

Fig. 16. Monthly averaged vertical profiles of ozone for Philadel-phia, August 2006. Monthly averaged ozonesonde profiles from two nearby rural locations (see Fig. 11) are also reported.

Fig. 15. Monthly averaged vertical profiles of ozone for Philadel-phia, January 2006.

the corresponding boundary profiles (Figs. 13 and 15). The only exception is for DEHM, which shows a significant de-parture with altitude from the boundary profile, and, conse-quently, the bias with respect to the observed MOZAIC pro-files also reflects the error of the BCs for this model. For January at Vancouver and Philadelphia, the modelled profiles have sharper gradients within the PBL than aloft due to titra-tion, but the modelled ozone mixing ratios clearly show the influence of BCs acting as an offset to surface mixing ratios, confirming that during winter the mean bias in ground-level mixing ratios is mostly driven by the large-scale background mixing ratios. This is further demonstrated by the corre-lation analysis in Fig. 13, where we show the normalised bias of the BCs and modelled ozone mixing ratios with re-spect to MOZAIC measurements. The regression analysis

demonstrates the large degree of association between the bias of two fields (especially for AURAMS and WRF-CMAQ, with PCC of 0.86 and 0.75 respectively). When including only the bias from levels above 500 m in the analysis, the PCC values are even larger for all models, except for AU-RAMS, which has fewer data available as the top of the model domain extends to only 5 km. This confirms the find-ings of Schere et al. (2012) regarding the penetration of the bias induced by the BCs for ozone far into the interior of the model simulation domain, which they found to be espe-cially true for winter months and for rural areas away from major emission sources. Systematic underestimation of tro-pospheric ozone by the global modelling systems that were used to derive the BCs for the AQMEII modelling caused the regional-scale models to frequently underestimate near-surface ozone mixing ratios in NA.

Different behaviour is observed in summer. At Port-land (Fig. 14) all models underestimate the observed mixing ratio above 2 km, with the GEMS BCs also underestimat-ing the MOZAIC measurements. Below 1000 m, all models overestimate the observed mixing ratios by approximately 20–25 ppb, possibly due to an overestimation of vertical mix-ing in the PBL. It should be further noted that the distance between Portland and the coast is approximately 100 km, and therefore the grid cell over the airport does not include any ocean area (model grid spacings are well below 100 km); thus there should not be any gradient between deposition over ocean and over land leading to overestimation of modelled ozone. At 1000 m height, the bias between the models and MOZAIC is smallest before turning positive at 2 km, with the modelled profiles unable to replicate the sharp gradient ob-served in the MOZAIC data. Such a gradient is possibly due to titration of NO_X surface emissions together with surface deposition. Although the aircrafts fly rapidly away from/to the airport, the influence from the large urban area around

the airport is expected to increase the ozone gradient in the lowest levels due to fast titration by NO. This effect is expected to vanish aloft (Chevalier et al., 2007). Differences in the aloft BCs are also seen in the aloft modelled values, and in turn in the ground-level mixing ratios. This is not the case for DEHM, where once again the vertical mixing appears to play the dominant role. Overall, the BCs do not appear to be the primary driver for the bias on the west NA coast during summer.

At the Philadelphia airport, the models tend to slightly underestimate (5 to 10 ppb) ozone mixing ratios for August 2006 (Fig. 16) above 4 km, while there is an overestimation of 6 to 20 ppb below 1000 m. The models driven by the BCs provided by GEMS and, to a lesser extent, AURAMS replicate very well the observed profile between 1.5 and 3 km, while DEHM is biased high. GEMS profiles are biased low above 4 km, but are in agreement with MOZAIC observations from the ground up to 1.5 km. The models driven by the GEMS BCs (the two instances of CMAQ and CAMx) show some differences within the first 3 km, though they cluster around the corresponding boundary value in the upper levels. The bias at these altitudes is possibly influenced by the GEMS BCs. The spread among the models is likely driven by photochemistry, the effect of which is stronger in Philadelphia than in Portland, being that the east coast has higher NOx emissions than the western areas (e.g. see Appel et al., 2012). Ground-level ozone mixing ratios simulated from models driven by GEMS range between 45 and 65 ppb, consistent with the notion that local production enhances summer ozone mixing ratios in polluted areas more than the large-scale background mixing ratios. This result is also shared by DEHM, which exhibits the highest surface mixing ratios, although it had the lowest boundary mixing ratio values. Profiles of ozone mixing ratios measured by ozonesondes for August were available from two rural sites (Fig. 11) within the Philadelphia domain, and are reported in Fig. 16. Measurements at the site STN487 (~370 km northeast of the airport, close to the coast) were collected daily for August at 18:00 GMT, while 12 h measurements from the STN420 (~160 km south-west of the airport) were collected between 04:00 GMT and 20:00 GMT. Profiles from STN420 are similar to those of the MOZAIC in the first 1000 m but diverge from the modelled mixing ratios more than the MOZAIC above 1000 m. The profiles from this rural ozonesonde station therefore suggest that the GEMS BCs may be biased low at this site (also found by Schere et al., 2012). Ozone mixing ratios from station STN487 also support this inference, as they are also close to the MOZAIC values in the upper levels. Finally, ozonesonde profiles from both sites confirm that the modelled ozone mixing ratios are biased high in the first 1000 m.

7 Summary and conclusions

This study has been conducted in the framework of the AQMEII activity and presents an evaluation of regional-scale AQ models in the troposphere supported by measurements gathered in the MOZAIC campaign. Fifteen regional-scale modelling groups from EU and NA have produced three-dimensional data of ozone, CO, WS, WD, RH and T with hourly resolution for the full year of 2006 at selected areas around major airports in EU and NA. These data have been paired with MOZAIC vertical profiles by the ENSEMBLE platform.

The scope of this study is the operational evaluation of the AQ modelling systems, i.e. the comparison of model results to observed data, supported by error metrics and variance statistics to gauge model performance in an overall sense. Results are thus based on the operational statistics obtained by comparing modelled against measured profiles. Several cases exhibiting significant differences between observations and model predictions have been identified as potential subjects of more in-depth diagnostic analysis to be undertaken in future.

From the study of the annual average mean vertical profiles of WS, we conclude the following:

– Above the PBL, there is little model-to-model variability at all airports, with all models showing similar mean profiles, most likely due to observational nudging. MM5-DEHM is the only model with a positive bias for WS, which increases with altitude at all sites, and is likely due to the coarser grid spacing used.

– For winds within the PBL (which are not nudged), all models (at all airports) show significant errors (predominantly as a positive bias) and poor correlation with the observed WS values in the first 100 m, and generally up to 500 m, with smaller errors between 500 and 1000 m. In addition, most models underestimated the variability within the first 100 m at all airports, indicating that the meteorological models are still weak in reproducing WS in the PBL. Factors that could contribute to the poor performance include sensitivity of the near-surface winds to influences other than the synoptic circulation (e.g. land cover), and processes driving the exchange of energy at surface for which global and regional models are known to largely differ from one another. For all models and at all airports, PCC, error and variability improve consistently with altitude. In general, the best performance by all models was observed for $z \geq 3$ km with a PCC > 0.8 at all NA airports and, with the exception of the WRF/Chem model, at Frankfurt as well.

We also investigated the model-to-model bias in predicting the frequency of WD counts when the whole wind rose is binned into 20° intervals, at all heights. We find the following:

– The bias in the PBL at Frankfurt is smaller and decreases faster with height with respect to the NA airports, for which larger biases are found, most pronounced at Portland (west NA coast), where a positive bias of 80 % (south-east sectors) is found for $0 \le z \le 500 \,\mathrm{m}$ for the GEM and WRF/Chem models.

– Overall, the bias at NA airports tends to decrease above 3 km, but it is significant in the first 250 to 500 m (depending on the airport) and is predominantly associated with an over-estimation of westerly winds.

With regards to RH, we find the following:

– RH profiles are layered with height and agree satisfactorily with the observations within approximately the first ~2 km, but they significantly diverge above that height at all sites. We attribute such discrepancy to the low values of water vapour mixing ratio at those heights for which small differences in T can give rise to significant differences in RH. The peak in RH at 1000 m is generally well captured by the models at all airports, with the exception of Dallas, where the model bias at 1000 m ranges between 7 and 15 %.

– Model performance for RH is found to be site-dependant, and a general trend could not be identified, although PCC values are above 0.5 at all NA sites, with the exception of WRF/Chem and the Frankfurt airport.

– Variability tends to be underestimated by all models (especially for $z > 6 \,\mathrm{km}$) with the exception of Portland, where the variability in the first 1000 m is close to the observed variability.

Model performance for T is satisfactory for all models and airports for the first 5 km (PCC in excess of 0.90 and fractional bias ≤ 1 %).

Modelled CO had the largest range of modelled-to-observed standard deviations at all heights and for all airports in this study. Correlation for CO is typically below 0.6 at all airports and all heights, with very few exceptions. Errors are also large at all altitudes and mainly in the first 250 m. Because the errors are common to all models, it is possible that one or more of the inputs shared by the models is biased. Performance improves slightly with height, giving the best scores between 2 and 4 km. With respect to the impact of WS errors on CO errors, there are instances for the NA airports where the WS and CO errors are correlated and have the same general trend. However, the correlations only exist for certain models at particular airports and heights.

Ozone mixing ratio profiles are strongly dependent on surface processes, as revealed by the strong gradient in the first 2 km of the troposphere, ranging on average between 10 and 20 ppb km^{-1} at all sites. Modelled ozone in winter is typically biased low at all locations, with the exception of DEHM for both NA and EU. Ozone within the PBL is generally overestimated in summer (NA only), with better performance in autumn, especially for Frankfurt. Layered ozone profiles are most notable in the summer months close to the surface, likely as result of the photochemistry and thermal mixing. Overall, model performance for ozone in the PBL can be considered satisfactory (with bias and error within 20 %).

The impact of BCs on the ozone has been examined by comparing observed ozone profiles to model profiles at NA airports close to the model boundary, where the effect from the boundary would be the largest. While the BCs contribute heavily to the bias of the modelled profiles for winter on both coasts, for summer the near-surface effects (e.g. photochemistry, transport, emissions, and deposition) dominate the ozone mixing ratios, dampening the impact from the BCs. However, on the east coast for the summer, a non-negligible influence from the BCs on the free-tropospheric ozone cannot be excluded. Finally, an analysis of the correlation between the bias at the ground and that aloft conducted for all variables demonstrates that, with the possible exception of CO, the errors produced in the PBL are weakly associated with those aloft, indicating that errors near the surface are primarily due to the treatment of PBL processes. Suggestions for future work include interpretation of the model-to-model variability and common sources of model bias, and linking CO and ozone bias to the bias in the meteorological fields.

Edited by: R. Sander

References

Appel, K. W., Gilliand, A. B., Sarwar, G., and Gilliam, R. C.: Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance; Part I – ozone, Atmos. Environ., 41, 9603–9615, 2007.

Appel, K. W., Bhave, P. V., Gilliland, A. B., Sarwar, G., and Roselle, S. J.: Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance; Part II – particulate matter, Atmos. Environ., 42, 6057–6066, 2008.

Appel, W., Chemel, C., Roselle, S. J., Francis, X. V., Hu, R.-M., Sokhi, R. S., Rao, S. T., and Galmarini, S.: Examination of the Community Multiscale Air Quality (CMAQ) model performance for North America and Europe for the AQMEII project, Atmos. Environ., 53, 142–155, 2012.

Aulinger, A., Matthias, V., and Quante, M.: An approach to temporally disaggregate Benzo (a) pyrene emissions and their application to a 3-D Eulerian atmospheric chemistry transport model, Water Air Soil Pollut. 216, 643–655, 2011.

Bélair, S., Mailhot, J., Strapp, J. W., and MacPherson, J. I.: An examination of local versus nonlocal aspects of a TKE-based boundary-layer scheme in clear convective conditions, J. Appl. Meteor., 38, 1499–1518, 1999.

Benoit, R., Côté, J., and Mailhot, J.: Inclusion of TKE boundary-layer parameterization in the Canadian regional finite-element model, Mon. Wea. Rev. 117, 1726–1750, 1989.

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

Bessagnet, B., Seigneur, C., and Menut, L.: Impact of dry deposition of semi-volatile organic compounds on secondary organic aerosols, Atmos. Environ., 44, 1781–1787, ISSN 1352-2310, doi:10.1016/j.atmosenv.2010.01.027., 2010.

Blond, N. and Vautard, R.: Three-dimensional ozone analyses and their use for short-term ozone forecasts, J. Geophys. Res., 109, D17, doi:10.1029/2004JD004515, 2004.

Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., Streets, D. G., Fernandes, S., and Trautmann, N.: Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850–2000, Glob. Biogeochem. Cy., 21, GB2018, doi:10.1029/2006GB002840, 2007.

Boynard, A., Beekmann, M., Foret, G., Ung, A., Szopa, S., Schmechtig, C., and Coman, A.: An ensemble assessment of regional ozone model uncertainty with an explicit error representation, Atmos. Environ., 45, 784–793, 2011.

Brandt, J., Silver, J. D., Frohn, L. M., Geels, C., Gross, A., Hansen, A. B., Hansen, K. M., Hedegaard, G. B., Skjøth, C. A., Villadsen, H., Zare, A., and Christensen, J. H.: An integrated model study for Europe and North America using the Danish Eulerian Hemispheric Model with focus on intercontinental transport, Atmos. Environ., 53, 156–176, 2012.

Brunner, D., Staehelin, J., Rogers, H. L., Köhler, M. O., Pyle, J. A., Hauglustaine, D., Jourdain, L., Berntsen, T. K., Gauss, M., Isaksen, I. S. A., Meijer, E., van Velthoven, P., Pitari, G., Mancini, E., Grewe, G., and Sausen, R.: An evaluation of the performance of chemistry transport models by comparison with research aircraft observations – Part 1: Concepts and overall model performance, Atmos. Chem. Phys., 3, 1609–1631, doi:10.5194/acp-3-1609-2003, 2003.

Brunner, D., Staehelin, J., Rogers, H. L., Köhler, M. O., Pyle, J. A., Hauglustaine, D. A., Jourdain, L., Berntsen, T. K., Gauss, M., Isaksen, I. S. A., Meijer, E., van Velthoven, P., Pitari, G., Mancini, E., Grewe, V., and Sausen, R.: An evaluation of the performance of chemistry transport models – Part 2: Detailed comparison with two selected campaigns, Atmos. Chem. Phys., 5, 107–129, doi:10.5194/acp-5-107-2005, 2005.

Byun, D. and Schere, K. L.: 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, 2006.

Carlton, A. G., Bhave, P. V., Napelenok, S. L., Edney, E. O., Sarwar, G., Pinder, R. W., Pouliot, G. A., Houyoux, M.: Model representation of secondary organic aerosol in CMAQv4.7, Environ. Sci. Technol., 44, 8553–8560, 2010.

Chevalier, A., Gheusi, F., Delmas, R., Ordóñez, C., Sarrat, C., Zbinden, R., Thouret, V., Athier, G., and Cousin, J.-M.: Influence of altitude on ozone levels and variability in the lower troposphere: a ground-based study for western Europe over the period 2001–2004, Atmos. Chem. Phys., 7, 4311–4326, doi:10.5194/acp-7-4311-2007, 2007.

Christensen, J. H.: The Danish Eulerian Hemispheric Model – a three-dimensional air pollution model used for the Arctic, Atmos. Environ., 31, 4169–4191, 1997.

Colella, P. and Woodward, P. R.: The Piecewise Parabolic Method (PPM) for gas-dynamical simulations, J. Comput. Phys., 54, 174–201, 1984.

Corbett, J. J. and Fischbeck, P. S.: Emissions from ships, Science, 278, 823–824, 1997.

Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC/MRB Global Environmental Multiscale (GEM) model – Part I: Design considerations and formulation, Mon. Wea. Rev., 126, 1373–1395, 1998a.

Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC/MRB Global Environmental Multiscale (GEM) model – Part II: Results, Mon. Wea. Rev., 126, 1397–1418, 1998b.

Cros, B., Durand, P., Frejafon, E., Kottmeier, C., Perros, P., Peuch, V.-H., Ponche, J.-L., Robin, D., Said, F., Toupance, G., and Wotham, H.: The ESCOMPTE program: an overview, Atmos. Res., 69, 241–279, 2004.

De Noblet-Ducoudré, N., Boissier, J.-P., Pitman, A., Bonan, G. B., Brovkin, V., Cruz, F., Delire, C., Gayler, V., van den Hurk, B., Lawrence, P. J., van der Mollen, M. K., Müller, C., Reick, C. H., Strengers, B. J., and Voldoire, A.: Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: results from the first set of LUCID experiments, J. Climate, 25, 3261–3281, 2012.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S. T., Scheffe, R., Schere, K., Steyn, D., and Venkatram, A.: A framework for evaluating regional-scale numerical photochemical modelling systems, Environ. Fluid Mech., 10, 471–489, doi:10.1007/s10652-009-9163-2, 2010.

Drobinski, P., Saïd, F., Ancellet, G., Arteta, J., Augustin, P., Bastin, S., Brut, A., Caccia, J. L., Campistron, B., Cautenet, S., Colette, A., Cros, B., Corsmeier, U., Coll, I., Dabas, A., Delbarre, H., Dufour, A., Durand, P., Guénard, V., Hasel, M., Kalthoff, N., Kottmeier, C., Lemonsu, A., Lohou, F., Masson, V., Menut, L., Moppert, C., Peuch, V. H., Puygrenier, V., Reitebuch, O., and Vautard, R.: Regional transport and dilution during high pollution episodes in southeastern France: summary of findings from the ESCOMPTE experiment, J. Geophys. Res., 112, D13105, doi:10.1029/2006JD007494, 2007.

Elguindi, N., Clark, H., Ordóñez, C., Thouret, V., Flemming, J., Stein, O., Huijnen, V., Moinat, P., Inness, A., Peuch, V.-H., Stohl, A., Turquety, S., Athier, G., Cammas, J.-P., and Schultz, M.: Current status of the ability of the GEMS/MACC

models to reproduce the tropospheric CO vertical distribution as measured by MOZAIC, Geosci. Model Dev., 3, 501–518, doi:10.5194/gmd-3-501-2010, 2010.

Emeis, S., Forkel, R., Junkermann, W., Schäfer, K., Flentje, H., Gilge, S., Fricke, W., Wiegner, M., Freudenthaler, V., Groß, S., Ries, L., Meinhardt, F., Birmili, W., Münkel, C., Obleitner, F., and Suppan, P.: Measurement and simulation of the 16/17 April 2010 Eyjafjallajökull volcanic ash layer dispersion in the northern Alpine region, Atmos. Chem. Phys., 11, 2689–2701, doi:10.5194/acp-11-2689-2011, 2011.

Emery C. A., Tai, E., Yarwood, G., and Morris, R.: Investigation into approaches to reduce excessive vertical transport over complex terrain in a regional photochemical grid model, Atmos. Environ., 45, 7341–7351, doi:10.1016/j.atmosenv.2011.07.052, 2011.

ENVIRON: User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx) version 5.20, available at: http://www.camx.com (last access: 9 November 2012), 2010.

Fischer, H., Lawrence, M., Gurk, Ch., Hoor, P., Lelieveld, J., Hegglin, M. I., Brunner, D., and Schiller, C.: Model simulations and aircraft measurements of vertical, seasonal and latitudinal O_3 and CO distributions over Europe, Atmos. Chem. Phys., 6, 339–348, doi:10.5194/acp-6-339-2006, 2006.

Foley, K. M., Roselle, S. J., Appel, K. W., Bhave, P. V., Pleim, J. E., Otte, T. L., Mathur, R., Sarwar, G., Young, J. O., Gilliam, R. C., Nolte, C. G., Kelly, J. T., Gilliland, A. B., and Bash, J. O.: Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7, Geosci. Model Dev., 3, 205–226, doi:10.5194/gmd-3-205-2010, 2010.

Forester, C. K.: Higher order monotonic convective difference schemes, J. Comput. Phys., 23, 1–22, 1977.

Forkel, R., Werham, J., Hansen, A. B., McKeen, S., Peckam, S., Grell, G., and Suppan, P.: Effect of aerosol-radiation feedback on regional air quality – a case study with WRF/Chem, Atmos. Environ., 53, 202–211, 2012.

Frohn, L. M., Christensen, J. H., and Brandt, J.: Development of a high resolution nested air pollution model – the numerical approach, J. Comput. Phys., 179, 68–94, 2002.

Galmarini, S., Bianconi, R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J. C., Bellasio, R., Buckley, R., Champion, H., Chino, M., D'Amours, R., Davakis, E., Eleveld, H., Glaab, H., Manning, A., Mikkelsen, T., Pechinger, U., Polreich, E., Prodanova, M., Slaper, H., Syrakov, D., Terada, H., and Van der Auwera, L.: Ensemble dispersion forecasting – Part II: Application and evaluation. Atmos. Environ., 38, 4619–4632, 2004.

Galmarini, S., Bianconi, R., Appel, W., Solazzo, E., Mosca, S., Grossi, P., Moran, M., Schere, K., and Rao, S. T.: ENSEMBLE and AMET: two systems and approaches to a harmonized, simplified and efficient facility for air quality models development and evaluation, Atmos. Environ., 53, 51–59, 2012.

Galperin, M. V.: Approaches for improving the numerical solution of the advection equation, in Large-Scale Computations in Air Pollution Modelling, edited by: Zlatev, Z., Brandt, J., Builtjes, P. J. H., Carmichael, G., Dimov, I., Dongarra, J., van Dop, H., Georgiev, K., Hass, H., and San Jose, R., Kluwer Acad., Dordrecht, Netherlands, 161–172, 1999.

Galperin, M. V.: The approaches to correct computation of airborne pollution advection, in: Problems of Ecological Monitoring and Ecosystem Modelling (in Russian), 17, Gidrometeoizdat, St. Petersburg, Russia, 54–68, 2000.

Gazdag, J.: Numerical convective schemes based on accurate computation of space derivatives, J. Comput. Phys., 13, 100–113, 1973.

Gilge, S., Plass-Duelmer, C., Fricke, W., Kaiser, A., Ries, L., Buchmann, B., and Steinbacher, M.: Ozone, carbon monoxide and nitrogen oxides time series at four alpine GAW mountain stations in central Europe, Atmos. Chem. Phys., 10, 12295–12316, doi:10.5194/acp-10-12295-2010, 2010.

Godowitch, J. M., Gilliam, R. C., and Rao, S. T.: Diagnostic evaluation of ozone production and horizontal transport in a regional photochemical air quality modeling system, Atmos. Environ., 45, 3977–3987, 2011.

Gong, W., Dastoor, A. P., Bouchet, V. S., Gong, S., Makar, P. A., Moran, M. D., Pabla, B., Ménard, S., Crevier, L.-P., Cousineau, S., and Venkatesh, S.: Cloud processing of gases and aerosols in a regional air quality model (AURAMS), Atmos. Res., 82, 248–275, 2006.

Graedel, T. E., Bates, T. S., Bouwman, A. F., Cunnold, D., Dignon, J., Fung, I., Jacob, D. J., Lamb, B. K., Logan, J. A., Marland, G., Middleton, P., Pacyna, J. M., Placet, M., and Veldt, C.: A compilation of inventories of emissions to the atmosphere, Global Biogeochem. Cy., 7, 1–26, 1993.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled online chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, 2005.

Guenther, A., Zimmerman, P., and Wildermuth, M.: Natural volatile organic compound emission rate estimates for US woodland landscapes, Atmos. Environ., 28, 1197–1210, 1994.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181–3210, doi:10.5194/acp-6-3181-2006, 2006.

Hong, S. Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Wea. Rev., 134, 2318–2341, 2006.

Jonson, J. E., Stohl, A., Fiore, A. M., Hess, P., Szopa, S., Wild, O., Zeng, G., Dentener, F. J., Lupu, A., Schultz, M. G., Duncan, B. N., Sudo, K., Wind, P., Schulz, M., Marmer, E., Cuvelier, C., Keating, T., Zuber, A., Valdebenito, A., Dorokhov, V., De Backer, H., Davies, J., Chen, G. H., Johnson, B., Tarasick, D. W., Stübi, R., Newchurch, M. J., von der Gathen, P., Steinbrecht, W., and Claude, H.: A multi-model analysis of vertical ozone profiles, Atmos. Chem. Phys., 10, 5759–5783, doi:10.5194/acp-10-5759-2010, 2010.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017–7039, doi:10.5194/acp-10-7017-2010, 2010.

Logan, J. A.: An analysis of ozonesonde data for the troposphere: recommendations for testing 3-D models, and development of a gridded climatology for tropospheric ozone, J. Geophys. Res., 104, 115–116, 1999.

Louis, J.: A parametric model of vertical eddy fluxes in the atmosphere, Bound.-La. Meteorol., 17, 197–202, 1979.

Mailhot, J., Bélair, S., Lefaivre, L., Bilodeau, B., Desgagné, M., Girard, C., Glazer, A., Leduc, A.-M., Méthot, A., Patoine, A., Plante, A., Rahill, A., Robinson, T., Talbot, D., Tremblay, A., Vaillancourt, P., Zadra, A., and Qaddouri, A.: The 15-km version of the Canadian regional forecast system, Atmos. Ocean, 44, 133–149, 2006.

Marenco, A., Thouret, V., Nédélec, P., Smit, H., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C., and Cook, T.: Measurement of ozone and water vapor by Airbus in-service aircraft: the MOZAIC airborne program, an overview, J. Geophys. Res., 103, 670–694, 1998.

Matthias, V., Bewersdorff, I., Aulinger, A., and Quante, M.: The contribution of ship emissions to air pollution in the North Sea regions, Environ. Pollut., 158, 2241–2250, 2010.

Matthias, V., Aulinger, A., Bieser, J., Cuesta, J., Geyer, B., Langmann, B., Serikov, I., Mattis, I., Minikin, A., Mona, L., Quante, M., Schumann, U., and Weinzierl, B.: The ash dispersion over Europe during the Eyjafjallajökull eruption – "Comparison of CMAQ simulations to remote sensing and air-borne in-situ observations", Atmos. Environ., 48, 184–194, 2012.

Pepper, D. W., Kern, C. D., and Long, P. E.: Modeling the dispersion of atmospheric pollution using cubic splines and chapeau functions, Atmos. Environ., 13, 223–237, 1979.

Pleim, J. E. and Chang, J. S.: A nonlocal closure-model for vertical mixing in the convective boundary-layer, Atmos. Environ. A, 26, 965–981, 1992.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer – Part I: Model description and testing, J. Appl. Meteorol. Clim., 46, 1383–1395, 2007a.

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer – Part II: Application and evaluation in a mesoscale meteorological model, J. Appl. Meteorol. Clim., 46, 1396–1409, 2007b.

Pouliot, G., Pierce, T., van der Gon, H. D., Schaap, M., Moran, M., and Nopmongcol, U.: Comparing emission inventories and model-ready emission datasets between Europe and North America for the AQMEII project, Atmos. Environ., 53, 4–14, 2012.

Pudykiewicz, J. A., Kallaur, A., and Smolarkiewicz, P. K.: Semi-Lagrangian modelling of tropospheric ozone, Tellus B, 49, 231–248, 1997.

Rao, S. T., Galmarini, S., and Puckett, K.: Air Quality Model Evaluation International Initiative (AQMEII), B. Am. Meteor. Soc., 92, 23–30, doi:10.1175/2010BAMS3069.1, 2011.

Sartelet, K., Couvidat, F., Seigneur, C., and Roustan, Y.: Impact of biogenic emissions on air quality over Europe and North America, Atmos. Environ., 53, 131–141, 2012.

Schaap, M., Sauter, F., Timmermans, R. M. A., Roemer, M., Velders, G., Beck, J., and Builtjes, P. J. H.: The LOTOS-EUROS model: description, validation and latest developments, Int. J. Environ. Pollut., 32, 270–290, 2008.

Schere, K., Flemming, J., Vautard, R., Chemel, C., Colette, A., Hogrefe, C., Bessagnet, B., Meleux, F., Mathur, R., Roselle, S., Hu, R.-M., Sokhi, R. S., Rao, S. T., and Galmarini, S.: Trace gas/aerosol concentrations and their impacts on continental-scale AQMEII modelling sub-regions, Atmos. Environ., 53, 38–50, 2012.

Schultz, M., Heil, A., Hoelzemann, J., Spessa, A., Thonicke, K., Goldammer, J., Held, A., Pereira, J., and van het Bolscher, M.: Global wildland fire emissions from 1960 to 2000, Global Biogeochem. Cy., 22, GB2002, doi:10.1029/2007GB03031, 2008

Simpson, D., Guenther, A., Hewitt, C. N., Steinbrecher, R.: Biogenic emissions in Europe, 1. estimates and uncertainties, J. Geophys. Res., 100, 22875–22890, 1995.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X. Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR technical note NCAR/TN-475+STR, available at www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf, June 2008.

Smith, S. J., Pitcher, H., and Wigley, T. M. L.: Global and regional anthropogenic Sulfur Dioxide emissions, Glob. Planet. Change, 29, 99–119, 2001.

Smyth, S. C., Jiang, W., Roth, H., Moran, M. D., Makar, P. A., Yang, F., Bouchet, V. S., and Landry, H.: A comparative performance evaluation of the AURAMS and CMAQ air quality modelling systems, Atmos. Environ., 43, 1059–1070, 2009.

Sofiev, M.: Extended resistance analogy for construction of the vertical diffusion scheme for dispersion models, J. Geophys. Res., 107, 4159, doi:10.1029/2001JD001233, 2002.

Sofiev, M., Galperin, M., and Genikhovich, E.: Construction and evaluation of Eulerian dynamic core for the air quality and emergency modelling system SILAM, in: Air Pollution Modelling and Its Application XIX, NATO Sci. Peace Security Ser. C: Environ. Security, edited by: Borrego, C. and Miranda, A. I., 699–701, Springer, New York, 2008.

Sofiev, M., Genikhovich, E., Keronen, P., and Vesala, T.: Diagnosing the surface layer parameters for dispersion models within the meteorological-to-dispersion modelling interface, J. Appl. Meteorol. Climatol., 49, 221–233, doi:10.1175/2009JAMC2210.1, 2010.

Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., van der Gon, H. D., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Jericevic, A., Kraljevic, L., Miranda, A. I., Nopmongcol, U., Pirovano, G., Prank, M., Riccio, A., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., and Galmarini, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII, Atmos. Environ., 53, 60–74, 2012a.

Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Moran, M. D., Appel, K. W., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Hogrefe, C., Miranda, A. I., Nopmongco, U., Prank, M., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., and Galmarini, S.: Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII, Atmos. Environ., 53, 75–92, 2012b.

Stegehuis, A., Vautard, R., Ciais, P., Teuling, R., Jung, M., and Yiou, P.: Summer temperatures in Europe and land heat fluxes in observation-based data and regional climate model simulations, Clim. Dynam., doi:10.1007/s00382-012-1559-x, online first, 2012.

Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J.-F., Lawrence, M. G., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K., and Szopa, S.: Multimodel ensemble simulations of present day and near future tropospheric ozone, J. Geophys. Res., 111, D08301, doi:10.1029/2005JD006338, 2006.

Szykman, J., Solazzo, E., Cooper, O., Silverman, S., Trepte, C., Newchurch, M., Cammas, J.-P., and Volz-Thomas, A.: Profiles and remote sensing observation datasets for regional-scale model evaluation under the AQMEII North American and European perspectives, Environmental Manager, July 2012, 21–29, 2012.

Tang, W., Cohan, D. S., Morris, G. A., Byun, D. W., and Luke, W. T.: Influence of vertical mixing uncertainties on ozone simulation in CMAQ, Atmos. Environ., 45, 2898–2909, doi:10.1016/j.atmosenv.2011.01.057., 2011.

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

Thouret, V., Marenco, A., Logan, J. A., Nedelec, P., and Grouhel, C.: Comparison of ozone measurements from the MOZAIC airborne program and the ozone sounding network at eight locations, J. Geophys. Res., 103, 695–720, 1998.

Troen, I. and Mahrt, L.: A simple model of the atmospheric boundary layer: sensitivity to surface evaporation, Bound.-Layer Meteorol., 37, 129–148, 1986.

Tulet, P., Suhre, K., Mari, C., Solomon, F., and Rosset, R.: Mixing of boundary layer upper tropospheric ozone during a deep convective event over Western Europe, Atmos. Environ., 36, 4491–4501, 2002.

Van Leer, B.: Towards the ultimate conservative difference scheme. A second order sequel to Godunov's method, J. Comput. Phys., 32, 101–136, 1979.

Van Loon, M., Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Graff, A., Jonson, J. E., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrasón, L., Thunis, P., Vignati, E., White, L., and Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble average, Atmos. Environ., 41, 2083–2097, 2007.

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Foltescu, V., Graf, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouïl, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., and Wind, P.: Skill and uncertainty of a regional air quality model ensemble, Atmos. Environ., 43, 4822–4832, 2009.

Vautard, R., Moran, M. D., Solazzo, E., Gilliam, R. C., Matthias, V., Bianconi, R., Chemel, C., Ferreira, J., Geyer, B., Hansen, A. B., Jericevic, A., Prank, M., Segers, A., Silver, J. D., Werhahn, J., Wolke, R., Rao, S. T., and Galmarini, S.: Evaluation of the meteorological forcing used for the Air Quality Model Evaluation International Initiative (AQMEII) air quality simulations, Atmos. Environ., 53, 15–37, 2012.

Vestreng V. and Støren E.: Analysis of the UNECE/EMEP Emission Data, MSC-W Status Report 2000, Norwegian Meteorological Institute: Blindern, Oslo, 2000.

Walcek, C. J.: Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection, J. Geophys. Res., 105, 9335–9348, 2000.

Yarwood, G., Rao, S., Yocke, M., and Whitten, G.: Updates to the Carbon Bond Chemical Mechanism: CB05, US EPA report RT-0400675, available at: http://www.camx.com (last access: 9 November 2012), 2005.

Yu, S. C., Mathur, R., Kang, D., Schere, K., Pleim, J., and Otte, T. L.: A detailed evaluation of the Eta-CMAQ forecast model performance for O_3, its related precursors, and meteorological parameters during the 2004 ICARTT study, J. Geophys. Res., 112, D12S14, doi:10.1029/2006JD007715, 2007.

Yu, S., Mathur, R., Sarwar, G., Kang, D., Tong, D., Pouliot, G., and Pleim, J.: Eta-CMAQ air quality forecasts for O_3 and related species using three different photochemical mechanisms (CB4, CB05, SAPRC-99): comparisons with measurements during the 2004 ICARTT study, Atmos. Chem. Phys., 10, 3001–3025, doi:10.5194/acp-10-3001-2010, 2010.

Zhang, J. and Rao, S. T.: The role of vertical mixing in the temporal evolution of ground-level ozone concentrations, J. Appl. Meteorol., 38, 1674–1691, 1999.

Zyryanov, D., Foret, G., Eremenko, M., Beekmann, M., Cammas, J.-P., D'Isidoro, M., Elbern, H., Flemming, J., Friese, E., Kioutsioutkis, I., Maurizi, A., Melas, D., Meleux, F., Menut, L., Moinat, P., Peuch, V.-H., Poupkou, A., Razinger, M., Schultz, M., Stein, O., Suttie, A. M., Valdebenito, A., Zerefos, C., Dufour, G., Bergametti, G., and Flaud, J.-M.: 3-D evaluation of tropospheric ozone simulations by an ensemble of regional Chemistry Transport Model, Atmos. Chem. Phys., 12, 3219–3240, doi:10.5194/acp-12-3219-2012, 2012.