Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone

Ulas Im[a], Roberto Bianconi[b], Efisio Solazzo[a], Ioannis Kioutsioukis[a], Alba Badia[c], Alessandra Balzarini[d], Rocío Baró[e], Roberto Bellasio[b], Dominik Brunner[f], Charles Chemel[g], Gabriele Curci[h], Johannes Flemming[i], Renate Forkel[j], Lea Giordano[f], Pedro Jiménez-Guerrero[e], Marcus Hirtl[k], Alma Hodzic[l], Luka Honzak[m], Oriol Jorba[c], Christoph Knote[l], Jeroen J.P. Kuenen[n], Paul A. Makar[o], Astrid Manders-Groot[n], Lucy Neal[p], Juan L. Pérez[q], Guidio Pirovano[d], George Pouliot[r], Roberto San Jose[q], Nicholas Savage[p], Wolfram Schroder[s], Ranjeet S. Sokhi[g], Dimiter Syrakov[t], Alfreida Torian[r], Paolo Tuccella[h], Johannes Werhahn[j], Ralf Wolke[s], Khairunnisa Yahya[u], Rahela Zabkar[m,v], Yang Zhang[u], Junhua Zhang[o], Christian Hogrefe[r], Stefano Galmarini[a*]

a. Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy.
b. Enviroware srl, Concorezzo (MB), Italy.
c. Earth Sciences Department, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain.
d. Ricerca sul Sistema Energetico (RSE SpA), Milano, Italy
e. University of Murcia, Department of Physics, Physics of the Earth. Campus de Espinardo, Ed. CIOyN, 30100 Murcia, Spain.
f. Laboratory for Air Pollution and Environmental Technology, Empa, Dubendorf, Switzerland.
g. Centre for Atmospheric & Instrumentation Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK.
h. Department of Physical and Chemical Sciences, Center of Excellence for the forecast of Severe Weather (CETEMPS), University of L'Aquila, L'Aquila, Italy.
i. ECMWF, Shinfield Park, RG2 9AX Reading, United Kingdom.
j. Karlsruher Institut für Technologie (KIT), Institut für Meteorologie und Klimaforschung, Atmosphärische Umweltforschung (IMK-IFU), Kreuzeckbahnstr. 19, 82467 Garmisch-Partenkirchen, Germany.
k. Section Environmental Meteorology, Division Customer Service, ZAMG - Zentralanstalt für Meteorologie und Geodynamik, 1190 Wien, Austria.
l. National Center for Atmospheric Research, Boulder, CO, US.
m. Center of Excellence SPACE-SI, Ljubljana, Slovenia.
n. Netherlands Organization for Applied Scientific Research (TNO), Utrecht, The Netherlands.
o. Air Quality Research Section, Atmospheric Science and Technology Directorate, Environment Canada, 4905 Dufferin Street, Toronto, Ontario, Canada.
p. Met Office, FitzRoy Road, Exeter, EX1 3PB, United Kingdom.
q. Environmental Software and Modelling Group, Computer Science School - Technical University of Madrid, Campus de Montegancedo - Boadilla del Monte-28660, Madrid, Spain.

r. Emissions and Model Evaluation Branch, Atmospheric Modeling and Analysis Division/NERL/ORD, Research Triangle Park, North Carolina, USA.
s. Leibniz Institute for Tropospheric Research, Permoserstr. 15, D-04318 Leipzig, Germany.
t. National Institute of Meteorology and Hydrology, Bulgarian Academy of Sciences, 66 Tzarigradsko shaussee Blvd., Sofia 1784, Bulgaria.
u. Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, USA.
v. University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia.

* Corresponding author: S. Galmarini (Stefano.galmarini@jrc.ec.europa.eu)

Highlights

- Sixteen modeling groups from EU and NA simulated O_3 for 2010 under AQMEII phase 2
- A general model underestimation of surface O_3 over both continents up to 22%
- Models tend to over/under estimate surface O_3 in all regions during autumn/winter
- Boundary conditions influence O_3 predictions especially during winter and autumn
- Models tend to under-predict high O_3 values that are of concern for policy

Keywords: AQMEII, on-line coupled models, performance analysis, ozone, Europe, North America

ABSTRACT

The second phase of the Air Quality Model Evaluation International Initiative (AQMEII) brought together sixteen modeling groups from Europe and North America, running eight operational online-coupled air quality models over Europe and North America on common emissions and boundary conditions. With advent of online-coupled models providing new capability to quantify the effects of feedback processes, the main aim of this study is to compare the response of coupled air quality models to simulate levels of O_3 over the two continental regions. The simulated annual, seasonal, continental and sub-regional ozone surface concentrations and vertical profiles for the year 2010 have been evaluated against a large observational database from different measurement networks operating in Europe and North America. Results show a general model underestimation of the annual surface ozone levels over both continents reaching up to 18% over Europe and 22% over North America. The observed temporal variations are successfully reproduced with correlation coefficients larger than 0.8. Results clearly show that the simulated levels highly depend on the meteorological and chemical configurations used in the models, even within the same modeling system. The seasonal and sub-regional analyses show the models tendency to overestimate surface ozone in all regions during autumn and underestimate in winter. Boundary conditions strongly influence ozone predictions especially during winter and autumn whereas during summer local production dominates over regional transport. Daily maximum 8-hour averaged surface ozone levels below 50-60 $\mu g\ m^{-3}$ are overestimated by all models over both continents while levels over 120-140 $\mu g\ m^{-3}$ are underestimated, suggesting that models have a tendency to severely under-predict high O_3 values that are of concern for air quality forecast and control policy applications.

1. Introduction

Tropospheric ozone (O_3) is an important secondary air pollutant produced by photochemical oxidation of volatile organic compounds (VOC) and carbon monoxide (CO) in the presence of nitrogen oxides (NOx). It has implications on climate and health and therefore its levels are subject to regulatory monitoring in Europe (EU) and North America (NA). The regulatory O_3 levels are still exceeded in a number of cities and are especially a concern in growing urban areas (European Environmental Agency, 2013). Air quality models (AQMs) are valuable tools to investigate the complex and dynamic interactions between meteorology and chemistry leading to O_3 pollution episodes at multiple temporal and spatial scales. In the last decade, the AQMs started shifting from off-line-coupled models where the meteorological forcing for chemistry was produced off-line by a separate meteorological model, to fully-coupled online models, which are able to simulate the feedbacks between chemistry and meteorology, taking the advantage of increased computational power (Zhang, 2008; Baklanov et al., 2014). The use of on-line models for O_3 predictions is beneficial as O_3 not only depend on emissions and chemistry but also on regional transport, clouds, photolysis and vertical mixing in the boundary layer, all of which can be more realistically represented in an on-line model. The wide use of regional AQMs for supporting policy, abatement strategies and forecasting justifies the increased need for online models, which can simulate feedback mechanisms, and especially account for the effect of aerosols on the radiative balance and photolysis (e.g. Hodzic et al., 2007).

The Air Quality Model Evaluation International Initiative (AQMEII) served to promote policy-relevant research on regional air quality model evaluation across the atmospheric modeling communities in Europe and North America through the exchange of information on current practices and the identification of research priorities (Galmarini and Rao, 2011). As part of this collaboration, standardized observations and model outputs were made available through the ENSEMBLE system (http:// http://ensemble2.jrc.ec.europa.eu/public/) that is hosted at the Joint Research Centre (JRC). This web-interface allows temporal and spatial analyses of individual models as well as their ensemble operators (Bianconi et al., 2004; Galmarini et al., 2012). The first phase of AQMEII was focused on the evaluation of off-line coupled atmospheric modelling systems against large sets of monitoring observations over Europe and North America for the year 2006 (Solazzo et al., 2012a,b; Vautard et al., 2012; Solazzo et al., 2013; Hogrefe et al., 2014). As summarized in Schere et al. (2012), the intercomparison model results for O_3 suggested a strong influence of chemical boundary conditions for ozone, whose bias extends far into the interior of the modelling domains, especially during winter months. The observed variance as well as the daily ozone cycle was underestimated by the majority of models. Night-time, overcast, and stable conditions led to poor model skill in reproducing ozone mixing ratios over both continents.

The second phase of AQMEII extends this model assessment to on-line-coupled air quality models. In this study, we analyze O_3 concentrations provided by eight on-line-coupled models, which have been run by sixteen independent groups from Europe and North America (while a companion study is devoted to the analysis of particulate matter, Im et al., 2014). The

models made use of the same input emissions and chemical boundary conditions, in an effort to eliminate errors in the interpretation of the model results. The goal of the study is to evaluate the performances of widely used operational on-line coupled models in Europe and North America in simulating O_3 levels on a sub-regional and seasonal basis employing an experimental set up with common anthropogenic emission and boundary conditions. The surface levels and vertical profiles simulated by the individual models as well as their ensemble mean and median are compared with the observational data provided by the ENSEMBLE system.

2. Materials and Methods

2.1. Participating models

In the context of AQMEII2, twelve modeling groups from EU and four modeling groups from NA (Table 1) have applied their modeling systems to simulate hourly O_3 concentrations for the year 2010 over the EU and NA continental scale domains (Fig.1). Among all participants, seven groups from EU and two groups from NA applied the same model system (WRF-CHEM), but with different settings such as different shortwave radiation schemes, gas-phase chemical mechanisms and aerosol modules. The WRF-CHEM community applied a common horizontal grid spacing of 23 km over Europe and 36 km over North America. Other modeling groups applied different grid spacings, ranging from 12×12 km^2 to ~50×25 km^2 as seen in Table 1. The simulations were conducted for continental-scale domains of Europe and North America covering continental U.S., southern Canada and northern Mexico (Fig.1). To facilitate the cross-comparison between models, the participating groups interpolated their model output to a common grid with 0.25˚ resolution for both continents. Model values at observation locations were extracted from the original model output files for comparison to observations (described below).

2.2. Emissions and boundary conditions

For the EU domain, the recently updated anthropogenic emissions for the year 2009 (http://www.gmes-atmosphere.eu/; Kuenen et al., 2014; Pouliot et al., 2014) were applied by all modelling groups and are based on the TNO-MACC-II (Netherlands Organization for Applied Scientific Research, Monitoring Atmospheric Composition and Climate – Interim Implementation) framework. Annual emissions of methane (CH_4), carbon monoxide (CO), ammonia (NH_3), total non-methane volatile organic compounds (NMVOC), nitrogen oxides (NO_X), particulate matter (PM_{10}, $PM_{2.5}$) and sulfur dioxide (SO_2) from ten activity sectors are provided on a latitude/longitude grid of 1/8º×1/16º resolution. Emission inventories for the NA domain were provided by US EPA and Environment Canada. The 2008 National Emission Inventory (http://www.epa.gov/ttn/chief/net/2008inventory.html) and the 2008 Emission Modeling Platform (http://www.epa.gov/ttn/chief/emch/index.html#2008) with year specific updates for 2006 and 2010 were used for the US portion of the modeling domain. Canadian emissions were derived from the Canadian National Pollutant Release Inventory (http://www.ec.gc.ca/inrp-npri/) and Air Pollutant Emissions Inventory

(http://www.ec.gc.ca/inrp-npri/donnees-data/ap/index.cfm?lang=En) values for the year 2006. These included updated spatial allocations for Canadian mobile emissions (Zhang et al, 2012) for the emissions of NH_3 (Makar et al, 2009), as well as other updates (Sassi et al, 2010). Mexican emissions were 2008 projected forward from a 1999 inventory (Wolf et al, 2009). Seven pollutants (CO, NO_x, NH_3, SO_2, $PM_{10,}$ $PM_{2.5,}$ and VOC) were used to develop the model ready emission inventory. Further details and analysis of the anthropogenic emissions used in both domains are provided in Pouliot et al. (2014). Annually-integrated anthropogenic emissions for both domains are presented in Table 2 while the spatial distribution of NO_x emissions for the EU and NA domains are depicted in Fig. 1. Table 2 shows that anthropogenic emissions per km^2 in EU are larger than those in NA, except for PM_{10}. Particularly NO_x and NH_3 emissions in EU are more than a factor of two larger than those in NA. Consistent temporal profiles (diurnal, day-of-week, seasonal) and vertical distributions were also made available to maintain consistency among different groups. NMVOC speciation factors were applied by all groups individually with a recommendation to follow the NMVOC speciation profiles for EU by Visschedijk et al. (2007). The temporal profiles for the EU anthropogenic emissions were provided from Schaap et al. (2005). Chemical and temporal profiles for the EPA anthropogenic emissions were based on the 2007v5 modeling platform (http://www.epa.gov/ttn/chief/emch/index.html#2008).

Each modeling group used their own biogenic emission module as detailed in Table 1. The majority of the models used the online MEGAN2 model (Model of Emissions of Gases and Aerosols from Nature version 2; Guenther et al., 2006), two groups used the BEIS v3.14 model (Biogenic Emission Inventory System; Schewede et al., 2005) and one group (NL2) used the Beltman et al. (2013) biogenic model. It should be noted that UK4 group used the off-line simulated biogenic emissions provided by the Beltman et al. (2013) model. In addition to the biogenic emissions algorithm used in the models, they may also differ in the databases used for vegetation. Feedbacks may have a significant influence on biogenic emissions; reductions in biogenic isoprene emissions of 20% were found with the introduction of the aerosol indirect effect (Makar et al., 2014a). The biogenic isoprene emissions calculated on-line by each group show a large variability as shown in Table 2 that may lead to large differences in the simulated O_3 levels. Curci et al. (2009) showed that different biogenic emission models may lead to a factor of 2 difference in domain-integrated isoprene emissions over Europe while difference can be up to a factor of 5-6 locally. They estimated that these differences on average may lead to an increase of 2.5 ppb in domain-mean surface O_3 levels and up to 10-15 ppb locally in the Mediterranean. Hourly biomass burning emissions were provided by the Finnish Meteorological Institute (FMI) fire assimilation system (http://is4fires.fmi.fi/; Sofiev et al., 2009). More details on the fire emissions and their uncertainties are discussed in Soares et al. (2014). The fire assimilation system provides only data for total PM emissions. Emissions of other species (CO, NO, NH_3, SO_2, NMVOC) were therefore deduced based on mass ratios relative to PM following Andreae and Merlet (2001). NMVOC speciation followed Wiedinmeyer et al. (2011) combined with the mapping to different chemical mechanisms proposed by Emmons et al. (2010). Note that the ES2a model does not include biomass burning emissions and as it does not contain aerosols leading to a lack of effect of aerosols on photolysis rate calculations and therefore producing

overestimated O_3 within the fire plumes (Badia and Jorba, 2014). Lightning NO_x is included in the UK4 model (O'Connor et al., 2014) as well as in the global MACC model used for the boundary conditions as described below.

3-D daily chemical boundary conditions were taken from the MACC re-analysis (Inness et al, 2013). The MACC re-analysis (referred to as MACC hereafter) has been produced by assimilating satellite observations of O_3, CO and NO_2 in the coupled system IFS-MOZART (Flemming et al., 2009). As pointed out in Inness et al. (2013), the assimilation of satellite-corrected O_3 greatly improved the ozone total columns and stratospheric profiles but did not change significantly the surface levels because of the limited signal from this region in the assimilated satellite observations. The chemical species available in the reanalysis included O_3, NO_x, CO, CH_4, SO_2, NMVOCs, sea-salt, dust, organic matter, black carbon and sulfate. NMVOC species had to be lumped or disaggregated according to the individual models' chemical speciation and particulate matter size discretization.

2.3. Observations

Measurements of hourly surface O_3 concentrations for the year 2010 in EU were taken from the European Monitoring and Evaluation Programme (EMEP; http://www.emep.int/) and the European Air Quality Database (AirBase; http://acm.eionet.europa.eu/databases/airbase/) and in NA from Aerometric Information Retrieval Systems (AIRS; http://www.epa.gov/air/data/aqsdb.html), National Air Pollution Surveillance (NAPS; http://www.ec.gc.ca/rnspa-naps/) and the Canadian National Atmospheric Chemistry (NAtChem) Database and Analysis Facility operated by Environment Canada (http://www.ec.gc.ca/natchem/) that contains measurements from the Canadian National Air Pollution Surveillance Network (http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx), the Canadian Air and Precipitation Monitoring Network (http://www.ec.gc.ca/natchem/), the U.S. Clean Air Status and Trends Network (http://java.epa.gov/castnet/clearsession.do), the U.S. Interagency Monitoring of Protected Visual Environments Network (http://views.cira.colostate.edu/web/DataWizard/), and the U.S. Environmental Protection Agency's Air Quality System database for U.S. air quality data (http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm). In the AQMEII2, rural, urban and suburban background stations were extracted from the EMEP and AirBase networks. Given the coarse native grid resolutions used in different models (Table 1), data from only rural background stations was used in the comparisons. Stations that have more than 90% data availability have been selected for the comparisons. Regarding the whole simulation domains, hourly surface O_3 observations were provided by 510 and 200 stations in EU and NA, respectively. A geographical break-down into four sub-regions for each continent has also been defined based on the climatological and source characteristics. The geographical break-down of these stations overlaid with the annually-averaged anthropogenic NO_x emissions is shown in Fig.1. Model evaluation statistics were computed for the four sub-regions separately. The European sub-region EU1 is characterized by north-western European sources with a transition climate between marine and continental and hosts 102 stations. Sub-region EU2 covers the north-eastern and central Europe sources as well as Germany with 277 monitoring stations. Sub-regions EU3 and EU4 are characterized by a Mediterranean type

climate. Sub-region 3 covers south-western sources including Italy (30 stations) while sub-region 4 covers the East Mediterranean with 101 stations. The North American sub-region 1 (NA1) covers the western U.S. and south western Canada with 80 stations. It includes large emission sources along the coast as well as polluted hot spots like Los Angeles that are characterized by poor air quality. NA2 consists of U.S. plains and covers 36 monitoring stations and is characterized by a continental and humid climate. NA3 consists of north eastern NA and south central Canada and is characterized by the largest emissions in North America and contains 60 monitoring stations. Finally NA4 covers the south eastern part of U.S., consisting of 24 monitoring stations.

To evaluate the capability of the modeling systems to simulate the tropospheric distribution of O_3 concentrations, comparisons against O_3 soundings provided by the World Ozone and Ultraviolet Radiation Data Centre (WOUDC: http://www.woudc.org/) have been carried out. Ozone concentration data from nine stations in EU and six stations in NA have been used for the comparisons. For an optimal comparison with observations, model profiles were computed by averaging only over the available observation hours. The participants were required to provide their data at fixed heights up to 18 km above the ground in order to be comparable. However, due to the coarse vertical resolution of some models in the upper troposphere and not simulating the stratospheric chemistry, the analyses are performed only for the first 9 km above ground.

2.4. Statistical analysis

To score the individual model performances as well as those of the ensemble mean and median, the following statistical parameters have been calculated: Pearson's correlation coefficient (*PCC*: Eq.1), root mean square error (*RMSE*: Eq.2); normalized mean standard error (*NMSE*: Eq.3) and normalized mean bias (*NMB*: Eq.4).

$$PCC = \left[\frac{\frac{1}{N}\sum_{i=1}^{N}\left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)}{\sigma_O \sigma_P} \right] \qquad \text{(Eq. 1)}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right)^2} \qquad \text{(Eq. 2)}$$

$$NMSE = \frac{\sum_{i=1}^{N}\left(P_i - O_i\right)^2}{N \times \overline{P} \times \overline{O}} \times 100 \qquad \text{(Eq. 3)}$$

$$NMB = \frac{\sum_{i=1}^{N}\left(P_i - O_i\right)}{\sum_{i=1}^{N} O_i} \times 100 \qquad \text{(Eq. 4)}$$

where *P* and *O* denote model predictions and observations, respectively. The *PCC* is a measure of associativity and allows gauging whether trends are captured, and it is not sensitive to bias; *RMSE* is a measure of accuracy and, because it is squared, is sensitive to large departures. *NMSE* and *NMB* are normalised operators, useful for comparing scores coming from time series of different lengths, as those produced over different areas and/or with different time span. The comparison is performed individually for the two domains and their sub-regions for the whole year of 2010 and on a seasonal basis, in order to identify which regions and/or seasons lead to systematic errors.

3. Results and Discussion

3.1. Surface ozone analysis

Observed and simulated diurnal cycles of surface O_3 concentrations averaged over the whole simulation period (2010) are shown in Fig. 2a,d for EU and NA, respectively. Models are labelled by the ID of the respective modeling group, with each ID corresponding to a member of the overall model ensemble. In the same figures, the MACC IFS-MOZART global model (MACC) and the ensemble mean and median are also shown. Note that the MACC model is not considered in the ensemble calculations.

3.1.1. Europe

Most models capture reasonably well the shape of the annual diurnal cycle over Europe as seen in Fig.2. The temporal variations on all time scales were captured successfully as seen in Table 3 (*PCC*>0.80), although the predicted O_3 levels are generally underestimated by up to 18%. Only one group (UK4) slightly overestimates the yearly-averaged observed surface O_3 levels by 2% while the other groups have underestimations up to 18%. The largest underestimations are calculated for IT2 (by 16%) and CH1 (by 18%) groups. Other groups have mean normalized biases within the ±5% to 15% range suggested by Russell and Dennis (2000). Fig.2a shows that the underestimations generally occur both during day and night hours, which is expected to some extent given the coarse horizontal resolution (Qian et al., 2010). The exceptions are AT1, DE4, SI1 and UK4 that overestimate the night time levels. The MACC model underestimates the nighttime levels as also reported in Innes et al. (2013). Overestimation of nighttime O_3 levels can be due to the overestimation of NO_2 concentrations under low-NO_x conditions leading to overestimated O_3 concentrations (e.g. DE4). It should be noted that the ES2a model does not include anthropogenic aerosols and secondary aerosol formation, leading to a more oxidized atmosphere due to higher photolysis when aerosols are not present as well as more hydroxyl radical (OH) consumption by VOCs, compared to other models (Badia and Jorba, 2014). As a consequence, the ES2a model overestimated the annual domain-mean NO_2 levels by 15% while the rest of the models underestimate NO_2 by 9% to 45%. The overestimation of surface O_3 levels by the ES2a model can also partly be due to the coarser vertical resolution of its first layer (45 m) compared to other models (Table 1). The general underestimation may be partly attributed to biases in meteorological variables, including an overestimation of surface wind speeds by all models by up to 60% and a general slight underestimation of surface temperatures by less than 1 K (Brunner et al., 2014). Such a

small temperature bias, however, will affect ozone levels by no more than a few ppb (Sillman and Samson, 1995). A common feature of all groups is that the daily maximum is simulated earlier than the observed maximum. Differences in O_3 predictions between the WRF-CHEM models suggest that the choice of the chemical mechanism plays an important role in the model performance. WRF-CHEM runs using RADM chemical mechanism (AT1, ES1 and SI1) produced higher concentrations than runs using RACM (IT2) and CBMZ (ES3 and IT1) mechanisms (Baro et al., 2014). These differences may partly be attributed to VOC emission preprocessing. WRF-CHEM is designed to ingest VOC emissions for RADM2 and then, in case of other mechanisms, the emissions are chemically specified to the final scheme, possibly leading to a degradation of the reactivity in the VOC mixture. There are also differences in the microphysics schemes among the different WRF-CHEM configurations used, leading to different cloudiness and therefore to different temperature and radiation acting on the O_3 production (Brunner et al., 2014; Baro et al., 2014). Makar et al. (2014a) and Wang et al. (2014) found that models including the simulation of indirect effects tended to have lower O_3 concentrations during the summer production period than those with the direct effect only, or those with no feedbacks. This is due to the reduction of NO_2 mixing ratios during daytime and near-surface temperatures, resulting from the reduction of solar radiation (Wang et al., 2014). Dry deposition of O_3 is also investigated for the models that provided deposition data (CH1, DE3, DE4, ES1, ES2a, ES3, IT2, NL2 and SI1) in order to explain the differences in simulated O_3 levels among the models (Table 3).The results show a negative relation between underestimation and dry deposition; i.e. the underestimation increases with decreasing deposition, suggesting that other terms aside from deposition were controlling the O_3 concentrations (chemistry, vertical diffusion etc.).

The model performances are also assessed against the observed variability in box-and-whisker plots of Fig.2b and e. The plot shows the frequency distribution of observed and simulated surface O_3 mixing ratios. The spread of the data in the European case is largest in CH1, ES2a and UK4 (Fig. 2b). The majority of other models show a much lower spread, which also tends to be lower than the observed spread. Data from MACC are associated with a larger spread compared to the observations in both domains, suggesting a better representation of local processes by regional models as well as an indication of an exaggerated seasonal cycle simulated by the MACC model. The larger spread in some models as compared to others is partially related to the amplitude of the diurnal ozone cycle, which tends to be larger in models simulating a more stable and shallow nocturnal PBL such as the global MACC model (Innes et al., 2013). A larger amplitude may also be expected for models with a higher vertical resolution. The *NMB* vs *NMSE* plot (also known as the soccer diagram) for EU (Fig. 2c) shows that the models have mean biases below 30% and mostly below 15%. The geographical analyses for the EU domain presented in Fig.3 shows that for the majority of models, the underestimation is mainly originating from sub-region EU2 (north Eastern Europe) while in sub-region EU4 (East Mediterranean), most models overestimate the observed mean. The underestimation, particularly in EU1 and EU2 can be partly due to the chemical boundary conditions (Fig.3) as discussed in more detail in Sect. 3.3.

3.1.2. North America

The hourly O_3 temporal variability over the whole simulation period is also well captured (*PCC*>0.78) by all groups for the NA domain (Table 3). The CA2f model overestimates the nighttime surface O_3 concentrations and underestimates the daytime levels with a slight overall overestimation of 2% while other groups underestimate the nighttime levels (Fig.2d). *NMSE* values are below 10% for all the groups while NMB values are within ±15% except for the US8 model, which underestimates the surface O_3 levels by 22%. The box plots for the NA case (Fig. 2e) shows that the MACC model has the highest variability while CA2f is characterized with the smallest spread. In the NA case, according to the soccer diagrams (Fig. 2f), all groups and sub-regions are characterized with biases lower than 25% except for US8. The geographical break down presented in Fig.4 shows that the US8 model underestimates in all sub-regions. The MACC model also shows a general underestimation in all sub-regions except for NA4. Regarding the dry deposition of O_3 (Table 3), the results suggest that the large underestimation by US8 can be partly due to the relatively large O_3 dry deposition simulated by the model, acting as a significant sink. As analyzed in Yahya et al. (2014a,b) and Wang et al. (2014), other factors that contribute to underpredictions of O_3 by the US8 model include large underpredictions of afternoon temperatures, low MACC boundary conditions of O_3, the overpredictions of the NO_x titration effects on O_3 during nighttime, possible underestimates in biogenic VOCs and wildfire emissions, and the inclusion of aerosol indirect effects. The lower spread in CA2f seems to be due to overpredicting the lower end of the O_3 range compared to the observations, in regions NA3 and NA4.

3.2. Seasonal vs. geographical surface ozone variations

3.2.1. Europe

Intra-seasonal variations of surface O_3 concentrations are analyzed for each sub-region in order to understand how the model bias varies depending on the region and season. The results for the EU domain are depicted in Fig. 5. The temporal variability in Europe is better captured in all models in summer and autumn (*PCC*=0.8-0.9) than in winter and spring (PCC=0.6-0.8). There is a systematic overestimation of the observed concentrations in autumn by up to 35%, particularly by the DE4 model. In winter (Fig. 5a), O_3 mixing ratios in EU2 are underestimated by more than 50% by three groups (CH1, ES2a and UK4), which also underestimate systematically in other sub-regions, probably due to the bias from the boundary conditions from the MACC model. The MACC model underestimates by largest during winter (by 8% to 55%) and overestimates by largest in autumn (by 8% to 25%). Regarding EU1, all groups are within the 30% bias range. Spring and summer O_3 mixing ratios (Fig. 5b,c) in all EU sub-regions are similarly reproduced by all groups, with error below 30%. In autumn, the majority of the models are biased high. In northern Europe (EU1 and EU2), the majority of the models underestimate O_3 levels in all seasons with the DE4, UK4, and ES2a models overestimating during summer. There is a general overestimation in autumn in the EU1 sub-region by all models except for CH1 and IT2. The models NL2, DE4, UK4 and ES2a overestimate the summertime O_3 levels in southern Europe. The East Mediterranean region (EU4) is characterized by overestimated O_3 levels, in particular during autumn. The results show that the largest underestimations were calculated for the EU2

region, which is characterized with large anthropogenic emissions in the Eastern Europe that may lead to overestimated O_3-titration by NO_x.

3.2.2. North America

Intra-seasonal and geographical variations of the models performances in NA are presented in Fig.6. US8 underestimates the observations in all seasons and in particular in winter and spring, and much larger compared to other models. In sub-region NA1, US6 overestimates by up to 9% while US8 underestimates by up to 22% in all seasons. CA2f slightly overestimates the winter and autumn O_3 levels by 3% and 5%, respectively. In the sub-regions NA2 and NA3, there is a general underestimation of all O_3 in winter and spring and a general overestimation in summer and autumn except for the US8 model. The winter and spring underestimates may be the result of underpredictions of afternoon temperatures and excessive O_3 titration by NO_x as NA3 can be characterized by the largest emission sources in NA. In NA4, summertime O_3 levels are overestimated by all models including the US8 model. Slightly lower correlation coefficients (*PCC*=0.7-0.9) are calculated for winter in NA while other seasons are simulated with *PCC* values of ~0.8-0.9, with slightly lower PCC values calculated for US7 (not shown).

3.3. Influence of chemical boundary conditions

The influence of the chemical boundary conditions on the simulated surface O_3 levels has also been investigated on a seasonal basis. The analysis is carried out for the EU2 (north Eastern Europe) sub-region for Europe assuming that it is the least affected by the dominant westerly transport and having large anthropogenic emissions, suggesting that O_3 levels are more strongly controlled by local processes than regional transport, compared to the other sub-regions. Following the same rationale, sub-region NA3 was selected for the NA domain. The results presented in Fig.7a show that in winter, all models underestimate O_3 levels along with the MACC model that provides the boundary conditions suggesting that large scale circulation and chemistry dominates over the local O_3 production. In spring and in summer (Fig.7b,c), the regional production is more important than transport due to increased photochemical activity. In autumn (Fig.7d), transport becomes more effective over local production. The MACC model slightly overestimates the summer levels (*NMB*=1%), and slightly underestimates the autumn levels (NMB=-5%) while it underestimates the winter and spring levels 55% and 21%, possibly leading to the systematic overestimation of the regional models in autumn. The impact of large-scale transport over NA is less pronounced compared to Europe (Fig.8). The impact is the smallest during summer when photochemical production is the largest (Fig.8c). At the same time, it is interesting to note that the MACC results in the winter for NA1 are the lowest of the models shown in Fig.8a, with a deficit of 8 ppb relative to the observations at 0 LST. The implication is that local chemistry, physics, model resolution and/or emissions relative to the global model all account for an increase in the winter O_3 levels for region NA1 of 8 ppb (28.5%), and these local effects are captured by the suite of regional models. This may be compared to findings from the HTAP experiment, which suggest a 20% reduction in emissions in Europe, South Asia and East Asia would result in a 0.9 ppb reduction in O_3 in North America (Reidmiller et al, 2009): here, model O_3 levels

seem to be much more sensitive to the local O_3 chemistry than to the boundary conditions associated with long-range transport (winter being the dominant season for long-range transport effects). Over both continents, the nighttime differences in all seasons are particularly large, with the MACC model largely underestimating the nighttime O_3. Similar results were reported by Solazzo et al. (2012 and 2013a) for the first phase of the AQMEII project. A more detailed analysis of the influence of the MACC boundary conditions on a range of simulated species is presented in Giordano et al. (2014).

3.4. Multi-model mean and median

The combination of concentrations simulated by several models can enhance the skill when compared to those from individual models (Galmarini et al., 2004a,b), which has also been demonstrated by Solazzo et al. (2012) in the first phase of the AQMEII project. In the present study, we provide simple multi-model mean and median analyses. Therefore, the calculated multi-model mean and median presented in Table 3-5 and in Fig.2-11 can only provide a basic distribution of all models with respect to the observations and should not be treated as multi-model ensemble analyses as they represent the bias originating from each individual model. As shown in Solazzo et al. (2012, 2013b) and Kioutsioukis and Galmarini (2014), introducing correlated biases into ensembles and analysis of the redundancy of the datasets is essential. As detailed multi-model ensemble analysis is not the scope of this paper, further analysis have been performed by Kioutsioukis et al. (2014) for the EU case using the multi-model data presented in the present paper.

3.5. Regulatory analysis based on 8-hour maximum surface O_3

Observed and simulated daily maximum 8-hour averaged surface O_3 levels during the O_3 season (May-September), which is a regulatory metric used in EU and NA, are compared in order to understand how the model biases vary with O_3 levels. The results are shown in Fig.9. Over EU (Fig.9a: note that observed concentrations are presented by /10), all models overestimate O_3 concentrations below 50 μg m^{-3} by ~40% to ~80% while they underestimate values above 140 μg m^{-3} except for the UK4 model that overestimates the levels above 160 μg m^{-3}. Most models follow the MACC model up to a concentration of 200 μg m^{-3} with increasing variability towards higher concentrations. NL2 and UK4 models overestimate the 230-240 μg m^{-3} concentration bin where the spread is also largest among other models. The UK4 model defines the upper boundary while IT2 defines the lower boundary of the envelope until 100 μg m^{-3} while above that, the highest differences are calculated for IT1. The CH1 model, which together with the IT2 model showed the largest negative biases in annual mean values, is more consistent with other models when considering 8-hour maximum values. Above a concentration of 70 μg m^{-3}, ES2a, NL2 and UK4 models are associated with positive deviations from the MACC model while other models are below the MACC-simulated levels. Results show that depending on the station, there are underestimations by up to >200 μg m^{-3}.

Over NA (Fig.9b), the biases are lower compared to EU. Note that for NA the values are reported in volume mixing ratios (ppb) rather than concentrations (μg/m^3). The surface O_3 levels below 30 ppb are overestimated by all models by ~15-25% and levels above 60 ppb are

underestimated by all models by up to ~80%.The largest biases are calculated for US8 except for the 120-130 ppb bin where US7 has the largest bias. US8 has the smallest bias below 50 ppb. The results show that models have a tendency to severely underpredict high O_3 values which are of concern for air quality forecast and control policy applications. Further improvement of model treatments (e.g., O_3 dry deposition and processes affecting afternoon temperature predictions) and inputs (e.g., boundary conditions, biogenic VOCs and wildfire emissions) as well as a better understanding of interplays among on-line-coupled atmospheric processes (e.g., the impact of aerosol indirect effects on O_3 formation) are urgently needed.

3.6. Vertical ozone profiles

The model results from each group as well as the ensemble mean and median are compared with O_3 soundings obtained from WOUDC for the EU and NA domains up to 9 km height above the ground. Figs.10 and 11 show the observed and simulated vertical O_3 levels at fixed heights over the EU and NA domains, respectively while Tables 4 and 5 present the normalized mean bias (*NMB*) for all the models and ensemble mean and median. On average, most models underestimate the observed vertical profiles by up to 22% over EU. The DE4 model generally has smaller biases compared to other groups except for the station STN156 where it overestimates by ~12% (Fig.10). The ensemble mean/median improves the results considering the majority of the models depending on the station. The ensemble mean results in smaller biases compared to the median. Over NA (Fig.11), the CA2f model underestimates the vertical O_3 levels at all stations by10-17% (Table 5). US6 and US7 have the smallest biases in most stations but with overestimations of 14% and 5%, respectively, at STN457. The US8 model underestimates at all stations by 4-15% but overestimates at STN457 by 2%. The ensemble mean and median lead to improved results compared to CA2f at all stations above ~1000-2000 m and to US8 at STN107 and STN456 below 2000-3000 m. Over Europe, among others, STN318 station (Valentia Observatory, Ireland) can be considered as a site that is largely impacted by long-range transport and is associated with the largest underestimation (*NMB* =-11%) by the MACC model (not shown), suggesting that boundary conditions can partly contribute to the underestimated vertical profiles by majority of the models. Results also show that the tropospheric biases in the MACC model (Fig.10,11) are less pronounced than the surface bias as also shown by Inness et al. (2013).

4. Summary and Conclusions

An operational evaluation of simulated ozone (O_3) levels over Europe (EU) and North America (NA) in 2010 using eight different on-line-coupled air quality models from sixteen groups has been conducted in the context of the AQMEII project. Seven groups from EU and two groups from NA applied the WRF-CHEM model, but with different settings. Anthropogenic emissions and chemical boundary conditions were prescribed while biogenic emissions were calculated online by each individual group. All groups interpolated their model output to a common output grid and a common set of receptor locations and uploaded the data to the ENSEMBLE system. The results are evaluated against surface and sounding

observations, which are provided by operational over EU and NA, at continental and sub-regional levels on annual and seasonal basis.

All models capture reasonably well the shape of the domain-averaged annual diurnal cycle of O_3 over both domains while the sub-regional temporal variability are simulated from moderate to good depending on the season and the sub-region that the particular model is configured for. There is a general underestimation of the annual surface O_3 by up to 18% and 22% over EU and NA, respectively. Differences in performance among models can be attributed partly to the chemical mechanism used in the models, partly to VOC preprocessing and different biogenic emissions, and partly to the differences in the microphysics, leading to different cloudiness and therefore to different photolysis, temperature and radiation acting on the O_3 production. The sub-regional analyses highlight the influence of the anthropogenic emissions while the seasonal analyses show a strong tendency to overestimate the autumn surface levels. The temporal variation and magnitudes are much better captured during summer compared to other seasons. The winter and spring underestimations may be resulting from underprediction of afternoon temperatures, excessive O_3 titration by too much NO_x as well as biases from the chemical boundary conditions. Boundary condition analyses show that wintertime levels are mostly driven by transport rather than local production due to limited photochemistry. The global MACC model providing the boundary conditions to the regional models largely underestimate the surface ozone levels particularly in winter, leading to a negative bias in the regional model simulations, while in most sub-regions, it largely overestimates the autumn O_3 levels in winter, leading to the systematic overestimations of surface autumn O_3 levels by the regional models. The inclusion of aerosol indirect effects in some online-coupled models also contributes in part to the underpredictions of O_3 mixing ratios. On average, most models underestimate the observed vertical profiles by up to 22% over EU and up to 17% over NA.

Comparison of observed and simulated daily maximum 8-hour averaged surface O_3 levels during the O_3 season (May-September), which is a regulatory metric used in EU and NA, show that over Europe, O_3 concentrations below 50 $\mu g\ m^{-3}$ are overestimated by up to 80% while levels above 140 $\mu g\ m^{-3}$ are underestimated. Over NA the surface O_3 levels below 30 ppb are overestimated by all models by up to 25% and levels above 60 ppb are underestimated by all models by up to 80%. This has implications for air quality forecast and policy applications.

Overall, the results show a slight improvement in the surface ozone level predictions over EU by the models participating to the second phase of AQMEII compared to those participating to the first phase. The *NMB* calculated for the whole domain and simulation period in the first phase ranged from -24% to 9% while in this second phase, the *NMB* range was calculated to be -18% to 2%. On the other hand over NA, there is a significant change between the two phases of the project: the overestimation of 3% to 22% in the first phase shifted to a *NMB* range of -22% to 3%. These results, however, should not be considered as solely the difference between on-line and off-line models as different simulation years, different emissions, different sets of models, particularly for the NA case, and different boundary condition data should be taken into account. Additionally, as the results presented in this

paper are temporally and spatially averaged, cases where feedback mechanisms are of importance must be further studied and evaluated in order to draw more solid conclusions on on-line versus off-line model applications.

Acknowledgements

We gratefully acknowledge the contribution of various groups to the second air Quality Model Evaluation international Initiative (AQMEII) activity: U.S. EPA, Environment Canada, Mexican Secretariat of the Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales-SEMARNAT) and National Institute of Ecology (Instituto Nacional de Ecología-INE) (North American national emissions inventories); U.S. EPA (North American emissions processing); TNO (European emissions processing); ECMWF/MACC project & Météo-France/CNRM-GAME (Chemical boundary conditions). Ambient North American concentration measurements were extracted from Environment Canada's National Atmospheric Chemistry Database (NAtChem) PM database and provided by several U.S. and Canadian agencies (AQS, CAPMoN, CASTNet, IMPROVE, NAPS, SEARCH and STN networks); North American precipitation-chemistry measurements were extracted from NAtChem's precipitation-chemistry data base and were provided by several U.S. and Canadian agencies (CAPMoN, NADP, NBPMN, NSPSN, and REPQ networks); the WMO World Ozone and Ultraviolet Data Centre (WOUDC) and its data-contributing agencies provided North American and European ozonesonde profiles; NASA's AErosol RObotic NETwork (AeroNet) and its data-contributing agencies provided North American and European AOD measurements; the MOZAIC Data Centre and its contributing airlines provided North American and European aircraft takeoff and landing vertical profiles; for European air quality data the following data centers were used: EMEP European Environment Agency/European Topic Center on Air and Climate Change/AirBase provided European air- and precipitation-chemistry data. The Finish Meteorological Institute is acknowledged for providing biomass burning emission data for Europe. Data from meteorological station monitoring networks were provided by NOAA and Environment Canada (for the US and Canadian meteorological network data) and the National Center for Atmospheric Research (NCAR) data support section. Joint Research Center Ispra/Institute for Environment and Sustainability provided its ENSEMBLE system for model output harmonization and analyses and evaluation. The co-ordination and support of the European contribution through COST Action ES1004 EuMetChem is gratefully acknowledged. The views expressed here are those of the authors and do not necessarily reflect the views and policies of the U.S. Environmental Protection Agency (EPA) or any other organization participating in the AQMEII project. This paper has been subjected to EPA review and approved for publication. C. Knote was supported by the DOE grant DE-SC0006711. The UPM authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the Centro de Supercomputación y Visualización de Madrid (CESVIMA) and the Spanish Supercomputing Network (BSC). G. Curci and P. Tuccella were supported by the Italian Space Agency (ASI) in the frame of PRIMES project (contract n.I/017/11/0). The Centre of Excellence for Space Sciences and Technologies SPACE-SI is an operation partly financed by the European Union,

European Regional Development Fund and Republic of Slovenia, Ministry of Higher Education, Science, Sport and Culture. Y. Zhang acknowledges funding support from the NSF Earth System Program (AGS-1049200) and high-performance computing support from Yellowstone by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation and Stampede, provided as an Extreme Science and Engineering Discovery Environment (XSEDE) digital service by the Texas Advanced Computing Center (TACC). The technical assistance of Bert van Ulft (KNMI) and Arjo Segers (TNO) in producing the results of the RACMO2-LOTOS-EUROS system is gratefully acknowledged. L. Giordano was supported by the Swiss SERI COST project C11.0144.

REFERENCES

Andreae, M.O., Merlet, P., 2001. Emission of trace gases and aerosols from biomass burning. Global Biogeochemical Cycles, 15, 4, 955-966.

Badia, A. and Jorba, O., 2014. Gas-phase evaluation of the online NMMB/BSC-CTM model over Europe for 2010 in the framework of the AQMEII-Phase2 project. Atmospheric Environment, Accepted.

Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Moussiopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., Zhang, Y., 2014. Online coupled regional meteorology chemistry models in Europe: current status and prospects. Atmospheric Chemistry and Physics, 14, 317-398.

Baró, R., Jiménez-Guerrero, P., Balzarini , A., Curci, G., Forkel, R., Hirtl, M., Honzak, L., Im, U., Lorenz, C., Pérez, J.L., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., Žabkar, R., 2014. Sensitivity analysis of the microphysics scheme in WRF-Chem contributions to AQMEII phase 2. Atmospheric Environment, In preparation.

Beltman, J.B., Hendriks, C., Tum, M., Schaap, M., 2013. The impact of large scale biomass production on ozone air pollution in Europe. Atmospheric Environment, 71, 352-363.

Bianconi, R., Galmarini, S., Bellasio, R., 2004.Web-based system for decision support in case of emergency: ensemble modelling of long-range atmospheric dispersion of radionuclides. Environmental Modelling and Software 19, 401-411.

Binkowski, F.S., Arunachalam, S., Adelman, Z., Pinto, J., 2007. Examining photolysis rates with a prototype on-line photolysis module in CMAQ. Journal of Applied Meteorology and Climatology. 46, 1252-1256.

Brunner, D., Jorba, O., Savage, N., Eder, B., Makar, P., Giordano, L., Badia, A., Balzarini, A., Baro, R., Bianconi, R., Chemel, C., Forkel, R., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Im, U., Knote, C., Kuenen, J.J.P., Makar, P.A., Manders-Groot, A., Neal, L.,

Perez, J.L., Pirovano, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A., Werhahn, K., Wolke, R., van Meijgaard, E., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., Galmarini, S., 2014. Evaluation of the meteorological performance of coupled chemistry-meteorology models in phase 2 of the Air Quality Model Evaluation International Initiative. Atmospheric Environment, to be submitted.

Curci, G., Beekman, M., Vautard, R., Smiatek, G., Steinbrecher, R., Theloke, J., Friedrich, R., 2009. Modelling study of the impact of isoprene and terpene biogenic emissions on European ozone levels. Atmospheric Environment, 43, 1444-1455.

Dave, J.V., 1972. Development of programs for computing characteristics of ultraviolet radiation, Final Report under Contract NAS5-21680, NASA Report CR-139134, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, NTIS # N75-10746/6SL, 27 pp.

European Environmental Agency, 2013. Air pollution by ozone across Europe during summer 2012. Overview of exceedances of EC ozone threshold values for April–September 2012. EEA Technical report, No 3/2013, 52 pp.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J. J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., Kloster, S., 2010. Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). Geoscientific Model Development, 3, 43–67.

Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G., Stein, O., 2009. Coupling global chemistry transport models to ECMWF's integrated forecast system. Geoscientific Model Development, 2, 253–265.

Forkel , R., Balzarini, A., Baró, R., Curci, G., Jiménez-Guerrero, P., Hirtl, M., Honzak, L., Im, U., Lorenz, C., Pérez, J.L., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., Žabkar, R., 2104. Analysis of the WRF-Chem contributions to AQMEII phase2 with respect to aerosol radiative feedbacks on meteorology and pollutant distribution. Atmospheric Environment, Submitted.

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., Van der Auwera, L., 2004a. Ensemble dispersion forecasting, Part II: application and evaluation. Atmospheric Environment, 38 (28), 4619-4632.

Galmarini, S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J.C., Bellasio, R., Bompay, F., Buckley, R., Bouzom, M., Champion, H., D'Amours, R., Davakis, E., Eleveld, H., Geertsema, G.T., Glaab, H., Kollax, M., Ilvonen, M., Manning, A., Pechinger, U., Persson, C., Polreich, E., Potemski, S., Prodanova, M., Saltbones, J., Slaper, H., Sofiev, M.A., Syrakov, D., Sørensen, J.H., Van der Auwera, L., Valkama, I., Zelazny, R., 2004b. Ensemble dispersion forecasting, Part I: concept, approach and Indicators. Atmospheric Environment, 38 (28), 4607-4617.

Galmarini, S., Rao, S.T., 2011. The AQMEII two-continent Regional Air Quality Model evaluation study: Fueling ideas with unprecedented data. Atmospheric Environment, 45, 2464.

Galmarini, S., Bianconi, R., Appel, W., Solazzo, E., et al., 2012. ENSEMBLE and AMET: two systems and approaches to a harmonised, simplified and efficient assistance to air quality model developments and evaluation. Atmospheric Environment, 53, 51-59.

Giordano, L., Brunner, D., Flemming, J., Im, U., Hogrefe, C., Bianconi, R., Badia, A., Balzarini, A., Baro, R., Belassio, R., Chemel, C., Curci, G., Forkel, R., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J.J.P., Makar, P.A., Manders-Groot, A., Neal, L., Perez, J.L., Piravano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A., Werhahn, K., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Galmarini, S., 2014. Assessment of the MACC/IFS-MOZART model and its influence as boundary conditions on AQMEII phase 2 modelling domains. Atmospheric Environment, Submitted.

Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., Eder, B., 2005. Fully coupled “online” chemistry within the WRF model. Atmospheric Environment, 39, 6957-6975.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P.I., Geron, C., 2006. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature. Atmospheric Chemistry and Physics, 6, 3181-3210.

Guenther, A.B., Zimmerman, P.R., Harley, P.C., Monson, R.K., Fall, R., 1993. Isoprene and monoterpene rate variability: model evaluations and sensitivity analyses. Journal of Geophysical Research, 98, D7, 12609-12617.

Hogrefe, C., Roselle, S., Mathur, R., Rao, S.T., Galmarini, S., 2014. Space-time analysis of the Air Quality Model Evaluation International Initiative (AQMEII) Phase 1 air quality simulations. Journal of Air Waste Management Association, 64, 388-405.

Hodzic, A., Madronich, S., Bohn, B., Massie, S., Menut, L., Wiedinmyer, C., 2007. Wildfire particulate matter in Europe during summer 2003: meso-scale modeling of smoke emissions, transport and radiative effects. Atmospheric Chemistry and Physics, 7 (15), 4043-4064.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Denier van der Gon, H.A.C., Flemming, J., Forkel, R., Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Makar, P.A., Manders-Groot, A., Neal, L., Perez, J.L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, K., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., Galmarini, S., 2014. Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate Matter. Atmospheric Environment, Submitted.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão,

J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team, 2013. The MACC reanalysis: an 8 yr data set of atmospheric composition. Atmospheric Chemistry and Physics, 13, 4073-4109.

Jorba, O., Dabdub, D., Blaszczak-Boxe, C., Pérez, C., Janjic, Z., Baldasano, J. M., Spada, M., Badia, A., Gonçalves, M., 2012. Potential Significance of Photoexcited NO2 on Global Air Quality with the NMMB/BSC Chemical Transport Model. Journal of Geophysical Research, 117.

Karl, M., Dorn, H.-P., Holland, F., Koppmann, R., Poppe, D., Rupp, L., Schaub, A. & Wahner, A., 2006. Product study of the reaction of OH radicals with isoprene in the atmosphere simulation chamber SAPHIR. Journal of Atmospheric Chemistry 55 (2), 167-187.

Kioutsioukis, I., Galmarini, S., 2014. De praeceptis ferendis: good practice in multi-model ensembles. Atmospheric Chemistry and Physics Discussions, Submitted.

Knote, C., Hodzic, A., Jimenez, J. L., Volkamer, R., Orlando, J. J., Baidar, S., Brioude, J., Fast, J., Gentner, D. R., Goldstein, A. H., Hayes, P. L., Knighton, W. B., Oetjen, H., Setyan, A., Stark, H., Thalman, R., Tyndall, G., Washenfelder, R., Waxman, E., Zhang, Q., 2013. Simulation of semi-explicit mechanisms of SOA formation from glyoxal in a 3-D model. Atmospheric Chemistry and Physics Discussions, 13, 26699-26759.

Kuenen, J.J.P., Visschedijk, A.J.H., Jozwicka, M., Denier van der Gon, H.A.C., 2014. TNO_MACC_II emission inventory: a multi-year (2003-2009) consistent high-resolution European emission inventory for air quality modelling. Atmospheric Chemistry and Physics Discussions, 14, 5837-5869.

Lurmann, F. W., l.,loyd, A. C. Atkinson, R., 1986. A chemical mechanism for use in long-range transport/acid deposition computer modeling. Journal of Geophysical Research 91, 10 905-10 936.

Makar, P.A., Gong, W., Hogrefe, C., Zhang, Y., Curci, G., Zabkar, R., Milbrandt, J., Im, U., Galmarini, S., Balzarini A., Baro, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, M., Honzak, L., Hou, A., Jimenez-Guerrero, P., Langer M., Moran, M.D., Pabla, B., Perez, P.L., Pirovano, G., San Jose, R., Tuccella, P., Werhahn, J., Zhang, J., 2014a. Feedbacks between Air Pollution and Weather, Part 1: Effects on Chemistry. Atmospheric Environment, Submitted.

Makar, P.A., Gong, W., Hogrefe, C., Zhang, Y., Curci, G., Zabkar, R., Milbrandt, J., Im, U., Galmarini, S., Balzarini A., Baro, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, M., Honzak, L., Hou, A., Jimenez-Guerrero, P., Langer M., Moran, M.D., Pabla, B., Perez, P.L., Pirovano, G., San Jose, R., Tuccella, P., Werhahn, J., Zhang, J., 2014b. Feedbacks between Air Pollution and Weather, Part 2: Effects on Weather. Atmospheric Environment, Submitted.

Makar, P.A., Moran, M.D., Zheng, Q., Cousineau, S., Sassi, M., Duhamel, A., Besner, M., Davignon, D., Crevier, L.-P., Bouchet, V.S., 2009. Modelling the impacts of ammonia emissions reductions on North American air quality. Atmospheric Chemistry and Physics, 9 (18), 7183-7212.

O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., Folberth, G. A., Sanderson, M. G., Telford, P. J., Voulgarakis, A., Young, P. J., Zeng, G., Collins, W. J., Pyle, J. A., 2014. Evaluation of the new UKCA climate-composition model – Part 2: The Troposphere. Geoscientific Model Development, 7, 41-91.

Poppe, D., Andersson-Sköld, Y., Baart, A., Builtjes, P.J.H., Das, M., Fiedler, F., Hov, O., Kirchner, F., Kuhn, M., Makar, P.A., Milford, J.B., Roemer, M.G.M., Ruhnke, R., Simpson, D., Stockwell, W.R., Strand, A., Vogel, B., Vogel, H., 1996. Gas-phase reactions in atmospheric chemistry and transport models: a model intercomparison. Eurotrac report. ISS, Garmisch-Partenkirchen.

Pouliot, G., Pierce, T., Denier van der Gon, H., Schaap, M., Moran, M., Nopmongcol, U., 2012. Comparing emissions inventories and model-ready emissions datasets between Europe and North America for the AQMEII Project. Atmospheric Environment 53, 75-92.

Pouliot, G., Denier van der Gon, H., Kuenen, J., Makar, P., Zhang, J., Moran, M., 2014. Analysis of the Emission Inventories and Model-Ready Emission Datasets of Europe and North America for Phase 2 of the AQMEII Project. Atmospheric Environment, Submitted.

Qian, Y., Gustafson, Jr. W.I., Fast, J.D., 2010. An investigation of the sub-grid variability of trace gases and aerosols for global climate modeling. Atmospheric Chemistry and Physics, 10, 6917-6946.

Reidmiller, D.R., Fiore, A.M., Jaffe, D.A., Bergmann, D., Cuvelier, C., Dentener, F.J., Duncan, B.N., Folberth, G., Gauss, M., Gong, S., Hess, P., Jonson, J.E., Keating, T., Lupu, A., Marmer, E., Park, R., Schultz, M.G., Shindell, D.T., Szopa, S., Vivanco, M.G., Wild, O., Zuber, A., 2009. The influence of foreign vs. North American emissions on surface ozone in the US. Atmospheric Chemistry and Physics, 9, 5027-5042.

Russell, A., Dennis, R., 2000. NARSTO critical review of photochemical models and modelling. Atmospheric Environment 34, 2283-2324.

Sarwar, G., Appel, K. W., Carlton, A. G., Mathur, R., Schere, K., Zhang, R., Majeed, M. A., 2011. Impact of a new condensed toluene mechanism on air quality model predictions in the US. Geoscientific Model Development, 4, 183-193.

Sassi, M., Chen, J., Samaali, M., Davignon, D., Moran, M.D., Taylor, B., Zheng, Q., 2010. 2006 Canadian emissions for air quality modelling, 19th International Emission Inventory Conference, 27-30 September, 2010, San Antonio, Texas, US.

Sauter, F., van der Swaluw, E., Manders-Groot, A., Wichink Kruit, R, Segers, A., Eskes, H., 2012. LOTOS-EUROS v1.8 Reference Guide. TNO report TNO-060-UT-2012-01451, TNO.

Savage, N. H., Agnew, P., Davis, L. S., Ordóñez, C., Thorpe, R., Johnson, C. E., O'Connor, F. M., Dalvi, M., 2013. Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation, Geoscientific Model Development, 6, 353-372.

Schaap, M., M. Roemer, F. Sauter, G. Boersen, R. Timmermans, P.J.H. Builtjes, LOTOS-EUROS: Documentation, TNO report B&O-A, 2005-297, Apeldoorn 2005

Schwede, D., Pouliot, G., Pierce, T., 2005. Changes to the Biogenic Emissions Inventory System version 3 (BEIS3). In: 4th CMAS Models-3 Users' Conference, Chapel Hill, NC, 26–28 September 2005.

Schere, K., Vautard, R., Solazzo, E., Hogrefe, C., Galmarini S. Results and lessons learned from Phase 1 of AQMEII. Environmental Manager, July 2012.

Sillman, S., Samson, P.J., 1995. Impact of temperature on oxidant photochemistry in urban, polluted rural and remote environments. Journal of Geophysical Research: Atmospheres, 100, D6, 11497–11508.

Soares, J., Sofiev, M., Prank, M, San Jose, R., Perez, J.L., 2014. On uncertainties of wild-land fires emission in AQMEII case study. Atmospheric Environment, In preparation.

Sofiev, M., Vankevich, R., Lotjonen, M., Prank, M., Petukhov, V., Ermakova, T., Koskinen, J., Kukkonen, J., 2009. An operational system for the assimilation of the satellite information on wild-land fires for the needs of air quality modelling and forecasting. Atmospheric Chemistry and Physics, 9, 6833-6847.

Solazzo, E., Riccio, A., Kioutsioukis, I., Galmarini, S., 2013b. Pauci ex tanto numero: reduce redundancy in multi-model ensembles. Atmospheric Chemistry and Physics, 13, 8315-8333.

Solazzo, E., Bianconi, R., Pirovano, G., Moran, M. D., Vautard, R., Hogrefe, C., Appel, K. W., Matthias, V., Grossi, P., Bessagnet, B., Brandt, J., Chemel, C., Christensen, J. H., Forkel, R., Francis, X. V., Hansen, A. B., McKeen, S., Nopmongcol, U., Prank, M., Sartelet, K. N., Segers, A., Silver, J. D., Yarwood, G., Werhahn, J., Zhang, J., Rao, S. T., Galmarini, S., 2013a. Evaluating the capability of regional-scale air quality models to capture the vertical distribution of pollutants. Geoscientific Model Development, 6, 791-818.

Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., 5 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., Galmarini, S., 2012. Ensemble modelling of surface level ozone in Europe and North America in the context of AQMEI. Atmospheric Environment, 53, 60–74.

Stockwell, W. R., Kirchner, F., Kuhn, M., Seefeld, S., 1997. A new mechanism for regional atmospheric chemistry modeling. Journal of Geophysical Research, 102, 25847-25879.

Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang, 1990. The second generation regional acid deposition model chemical mechanism for regional air quality modeling. Journal of Geophysical Research, 95, 16343-16367.

Tie, X., Madronich, S., Walters, S., Zhang, R., Rasch, P., Collins, W., 2003. Effects of clouds on photolysis and oxidants in the troposphere. Journal of Geophysical Research, 108, D20, 4642.

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., Galmarini, S., 2012. Evaluation of the meteorological forcing used for AQMEII air quality simulations, Atmospheric Environment, 53, 15–37.

Visschedijk, A.J.H., Zandveld, P., Denier van der Gon, H.A.C., 2007. A High Resolution Gridded European Emission Database for the EU Integrated Project GEMS. TNO report 2007-A-R0233/B.

Vogel, B., Vogel, H., Baumer, D., Bangert, M., Lundgren, K., Rinke, R., Stanelle, T., 2009. The comprehensive model system COSMO-ART – Radiative impact of aerosol on the state of the atmosphere on the regional scale. Atmospheric Chemistry and Physics, 9, 8661–8680.

Wang, K., Yahya, K., Zhang, Y., Wu, S.-Y., Grell, G., 2014. Implementation and Initial Application of A New Chemistry-Aerosol Option in WRF/Chem for Simulation of Secondary Organic Aerosols and Aerosol Indirect Effects. Atmospheric Environment, Under review.

Whitten, G. Z., Heo, G., Kimura, Y., McDonald-Buller, E., Allen, D., Carter, W. P. L., Yarwood, G., 2010. A new condensed toluene mechanism for Carbon Bond: CB05-TU. Atmospheric Environment, 44, 5346–5355, 2010.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., Soja, A. J., 2011. The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625-641.

Wild, O., Zhu, X., Prather, M.J., 2000. FAST-J: accurate simulation of in- and below-cloud photolysis in tropospheric chemical models. Journal of Atmospheric Chemistry 37, 245–282.

Wolf, M.E., Fields, P.G., Manne, G.K., Villegas, M.T.L., Bravo, V.G., Gomez, R.I., 2009. Developing Mexico National Emissions Inventory Projections for the Future Years of 2008, 2012, and 2030, 18th International Emission Inventory Conference, 14-17 August, 2009, Baltimore, Maryland.

Wolke, R., Schroder, W., Schrodner, R., Renner, E., 2012. Influence of grid resolution and meteorological forcing on simulated European air quality: a sensitivity study with the modeling system COSMO–MUSCAT. Atmospheric Environment 53, 110–130.

Wong, D. C., Pleim, J., Mathur, R., Binkowski, F., Otte, T., Gilliam, R., Pouliot, G., Xiu, A., Young, J. O., Kang, D., 2012. WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results. Geoscientific Model Development, 5, 299-312.

Yahya, K., Wang, K., Gudoshava, M., Glotfelty, T., Zhang, Y., 2014a. Application of WRF/Chem over the continental U.S. under the AQMEII Phase II: Comprehensive Evaluation of 2006 Simulation. Atmospheric Environment, Under review.

Yahya, K., Wang, K., Zhang, Y., Kleindienst, T. E., 2014b. Application of WRF/Chem over the continental U.S. under the AQMEII Phase II: Comprehensive Evaluation of 2010

Simulation and Responses of Air Quality and Meteorology-Chemistry Interactions to Changes in Emissions and Meteorology from 2006 to 2010. Atmospheric Environment, Under review.

Yardwood, G., Rao, S., Yocke, M., Whitten, G. Z., 2005. Updates to the Carbon Bond chemical mechanism: CB05. Final Report to the US EPA, RT-0400675, 8 December 2005.

Zaveri, R.A., Peters, L.K., 1999. A new lumped structure photochemical mechanism for large-scale applications. Journal of Geophysical Research, 104, D23, 30387-30415.

Zhang, J., Zheng, Q., Moran, M.D., Gordon, M., Liggio, J., Makar, P.A., Taylor, B., Stroud, C., 2012. Improvements to SMOKE processing of Canadian On-Road mobile emissions, 20th International Emission Inventory Conference, 13-16 Aug. 2012, Tampa, Florida.

Zhang, Y., 2008. Online Coupled Meteorology and Chemistry models: History, Current Status, and Outlook. Atmospheric Chemistry and Physics, 8, 2895-2932.

Table 1. Modelling systems participated to AQMEII2 and their configurations

	Groups	Domain	Model	Grid Spacing	First layer height (m)	Biogenic Model	Gas Phase	Photolysis	Model Reference
M1	AT1	EU	WRF-CHEM	23 km	24	MEGAN	RADM2[1]	Fast-J[13]	Grell et al., 2005
M2	CH1	EU	COSMO-ART	0.22°	20	Gunter et al., 1998	RADM2K[2]	GRAALS+STAR[2]	Vogel et al., 2009
M3	DE3	EU	COSMO-MUSCAT	0.25°	20	Gunther et al., 1993	RACM-MIM2[3]	Fast-J	Wolke et al., 2012
M4	DE4	EU	WRF-CHEM	23 km	24	MEGAN	RADM2 modified[4]	Fast-J	Grell et al., 2005; Forkel et al., 2014
M5	ES1	EU	WRF-CHEM	23 km	24	MEGAN	RADM2	Fast-J	Grell et al., 2005
M6	ES2a	EU	NMMB-BSC-CTM	0.20 °	45	MEGAN	CB-V[5]	Fast-J	Jorba et al., 2012
M7	ES3	EU	WRF-CHEM	23 km	24	MEGAN	CBMZ[6]	Fast-J	Grell et al., 2005
M8	IT1	EU	WRF-CHEM	23 km	24	MEGAN	CBMZ	Fast-J	Grell et al., 2005
M9	IT2	EU	WRF-CHEM	23 km	24	MEGAN	RACM[7]	Fast-J	Grell et al., 2005
M10	NL2	EU	RACMO LOTOS-EUROS	0.5 ° ×0.25°	25	Beltman et al., 2013	CB-IV modified[8]	Poppe et al., 1996	Sauter et al., 2012
M11	SI1	EU	WRF-CHEM	23 km	25	MEGAN	RADM2	Fast-J	Grell et al., 2005
M12	UK4	EU	MetUM-UKCA RAQ	0.22 °	20	TNO	UKCA RAQ[9]	Fast-J	Savage et al., 2013
M13	CA2f	NA	GEM-MACH	15 km	20.66	BEIS	ADOM-II[10]	Dave, 1972	Makar et al., 2014a,b
M14	US6	NA	WRF-CMAQ	12 km	19	BEIS3.14	CB-V-TU[11]	Binkowski et al., 2007	Wong et al., 2012
M15	US7	NA	WRF-CHEM	36 km	55-60	MEGAN	MOZART[12]	fTUV[14]	Grell et al., 2005
M16	US8	NA	WRF-CHEM	36 km	38	MEGAN	CB-V	fTUV	Grell et al., 2005; Wang et al., 2014

1 Stockwell et al. (1990); 2 Vogel et al. (2009); 3 Karl et al. (2006); 4 Forkel et al. (2014); 5 Yardwood et al. (2005); 6 Zaveri et al. (1999); 7 Stockwell et al. (1997); 8 Sauter et al. (2012); 9 Savage et al. (2013); 10 Lurmann et al. (1986); 11 Whitten et al., 2010; Sarwar et al., 2011; 12 Emmons et al. (2010); Knote et al. (2013); 13 Wild et al., 2000; 14 Tie et al., 2003

Table 2. Annual anthropogenic emissions (ktons km^{-2} yr^{-1}) provided by TNO-MACC-II inventory and biogenic isoprene emissions (ktons-C km^{-2} yr^{-1}) integrated over the EU and NA domains.

Species	EU	NA
CO	614	478
NO_x	277	120
NMVOC	230	85
NH_3	109	31
SO_2	109	70
$PM_{2.5}$	49	29
PM_{10}	69	76
ISOP*	2.1-22.0	0.02-7.1

* The groups that provided isoprene emissions are AT1, CH1, DE3, IT2, NL2 and UK4 for the EU domain and CA2f, US6 and US7 for the NA domain.

Table 3. Statistical comparisons of observed and simulated annual domain-mean hourly surface O_3 and domain- and annually-integrated O_3 dry deposition over EU and NA in 2010.

Members	*r*	*NMSE (%)*	*NMB (%)*	*RMSE*[*]	*Dry Deposition (Tg km^{-2})*
M1/AT1	0.86	2.66	-4.92	9.57	*NA*
M2/CH1	0.82	8.03	-18.30	15.42	0.28
M3/DE3	0.68	6.37	-2.12	15.02	0.13
M4/DE4	0.83	3.17	-1.64	10.62	2.24
M5/ES1	0.86	4.08	-11.41	11.44	2.18
M6/ES2a	0.83	6.37	-7.71	14.59	2.79
M7/ES3	0.86	4.29	-12.07	11.69	1.82
M8/IT1	0.85	4.57	-12.45	12.03	*NA*
M9/IT2	0.84	6.21	-15.80	13.76	1.77
M10/NL2	0.89	2.83	-4.34	9.90	0.14
M11/SI1	0.87	2.38	-3.78	9.10	1.91
M12/UK4	0.85	7.88	2.30	17.08	*NA*
EU Mean	0.86	3.22	-7.70	10.37	
EU Median	0.86	3.23	-8.69	10.33	
M13/CA2f	0.85	1.45	2.43	4.02	0.09
M14/US6	0.84	2.15	1.14	4.85	0.10
M15/US7	0.78	4.36	-4.56	6.72	0.15
M16/US8	0.88	8.11	-22.36	8.26	3.05
NA Mean	0.83	3.70	-11.98	5.94	
NA Median	0.87	2.62	-9.51	5.07	

[*] *RMSE* is in units of μg m^{-3} for EU and ppb for NA.

Table 4. *NMB* calculated for vertical O_3 profiles for each model group and ensemble mean and median for the WOUDC stations in EU.

Stations	Station Name	Country	Lat/Lon	AT1	CH1	DE3	DE4	ES1	ES2a	ES3	IT1	IT2	NL2	SI1	UK4	Mean	Median
STN043	Lerwick	United Kingdom	60.1/-1.2	-8.40	-12.14	-27.80	-2.39	-9.82	-11.46	-7.86	-6.91	-9.05	-3.32	-8.16	-11.13	-10.40	-7.40
STN053	UCCLE	Belgium	50.8/4.4	-4.11	-10.09	-14.08	3.80	-6.02	-7.46	-4.14	-4.53	-7.23	-1.96	-3.95	-4.86	-5.50	-3.56
STN099	Hohenpeissenberg	Germany	47.8/11.0	-10.65	-21.94	-23.98	-2.04	-12.15	-11.98	-8.96	-9.55	-11.47	0.17	-10.39	-8.43	-11.62	-9.63
STN156	Payerne	Switzerland	46.5/6.6	1.18	-10.06	-11.71	11.77	-0.63	1.84	2.51	2.43	0.51	2.70	1.44	3.94	0.64	2.52
STN242	Praha	Czech Rep.	50.0/14.5	-8.55	-16.18	-26.48	-1.72	-11.38	-8.82	-6.98	-6.97	-8.68	-4.77	-7.86	-5.06	-9.50	-7.35
STN308	Barajas	Spain	40.5/-3.7	-6.02	-14.29	-9.83	1.91	-7.72	-4.77	-6.72	-6.32	-7.83	0.21	-5.67	-1.61	-5.95	-5.01
STN316	De Bilt	Netherlands	52.1/5.2	-4.57	-5.83	-9.82	3.62	-6.14	-4.29	-4.99	-5.08	-7.29	1.15	-4.37	-0.59	-4.23	-3.76
STN318	Valentia	Ireland	51.9/-10.3	-6.51	-10.56	-15.49	-0.44	-8.01	-9.30	-6.00	-2.93	-6.35	-5.74	-6.43	-5.97	-7.01	-5.04
STN348	Ankara	Turkey	40.0/32.9	-11.48	-16.13	-12.94	5.76	-13.38	-4.28	-10.95	-11.12	-15.24	0.55	-11.36	2.41	-8.66	-9.74

Table 5. *NMB* calculated for vertical O_3 profiles for each model group and ensemble mean and median for the WOUDC stations in NA.

Stations	Station Name	Country	Lat/Lon	CA2f	US6	US7	US8	Mean	Median
STN021	Stony Plain	Canada	53.4/-114.1	-9.82	1.58	-2.29	-4.71	-3.81	-2.85
STN107	Wallops Island	USA	37.9/-75.5	-10.19	1.77	-1.17	-13.52	-5.78	-6.30
STN338	Bratts Lake	Canada	50.2/-104.8	-14.29	0.27	-3.26	-9.46	-6.68	-4.47
STN456	Egbert	Canada	44.2/-79.8	-16.78	-1.40	-3.95	-15.01	-9.28	-8.54
STN457	Kelowna	Canada	49.9/-119.4	-10.09	13.61	4.95	-0.62	1.96	2.05
STN458	Yarmouth	Canada	43.9/-66.1	-17.76	-1.17	-5.95	-15.27	-10.04	-10.20

Figure Captions

Fig.1. Annual NO_x emissions overlaid with the rural monitoring stations used for model performance evaluation in EU (a) and in NA (b). The red circles show EU1/NA1, yellow diamonds show EU2/NA2, green squares show EU3/NA3 and black triangles show EU4/NA4.

Fig.2. Observed and simulated annual mean diurnal profiles (a,d), box plots (b,e) and soccer diagrams (c,f) for surface levels ozone mixing ratios in EU (upper panel) and NA (lower panel). EU0 and NA0 represent the two respective continents. Note the differences in scales.

Fig.3. Geographical distributions of observed and simulated annual surface level ozone mixing ratios in EU. Note the differences in scales.

Fig.4. Geographical distributions of observed and simulated annual surface level ozone mixing ratios in NA. Note the differences in scales.

Fig.5. Soccer diagrams for the seasonal and geographical model performances in EU: a) winter, b) spring, c) summer and d) autumn. Mn and Md represent the mean and median ensembles, respectively. EU0 and NA0 represent the continental levels. Note the differences in scales.

Fig.6. Soccer diagrams for the seasonal and geographical model performances in NA: a) winter, b) spring, c) summer and d) autumn. Mn and Md represent the mean and median ensembles, respectively. EU0 and NA0 represent the continental levels. Note the differences in scales.

Fig.7. Observed and simulated seasonal diurnal O_3 profiles in a) winter, b) spring, c) summer and d) autumn over EU2.

Fig.8. Observed and simulated seasonal diurnal O_3 profiles in a) winter, b) spring, c) summer and d) autumn over NA3.

Fig.9. Observed surface O_3 concentration bins against mean bias for the EU and NA domains for the O_3 season (May-September).

Fig.10. Observed and simulated (models, mean and median) vertical O_3 profiles averaged over 2010 in the EU domain.

Fig.11. Observed and simulated (models, mean and median) vertical O_3 profiles averaged over 2010 in the NA domain.

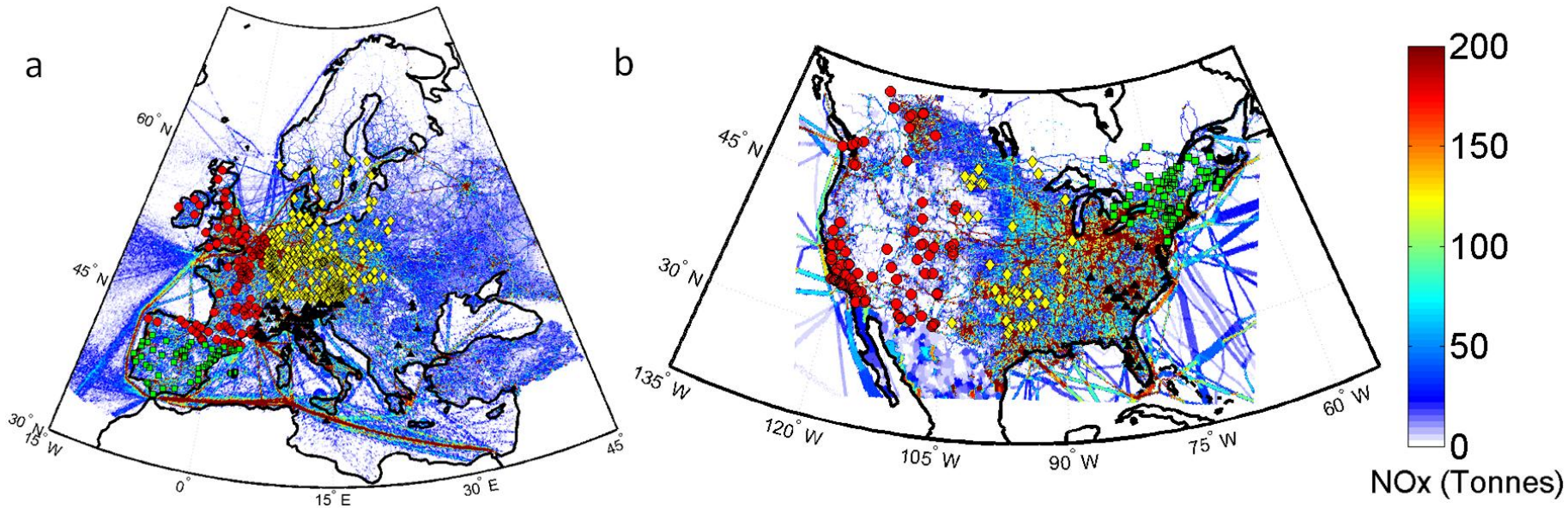
a
60° N
45° N
30° N
15° W
0°
15° E
30° E
45°
b
45° N
30° N
135° W
120° W
105° W
90° W
75° W
60° W
200
150
100
50
0
NOx (Tonnes)

Fig.1.

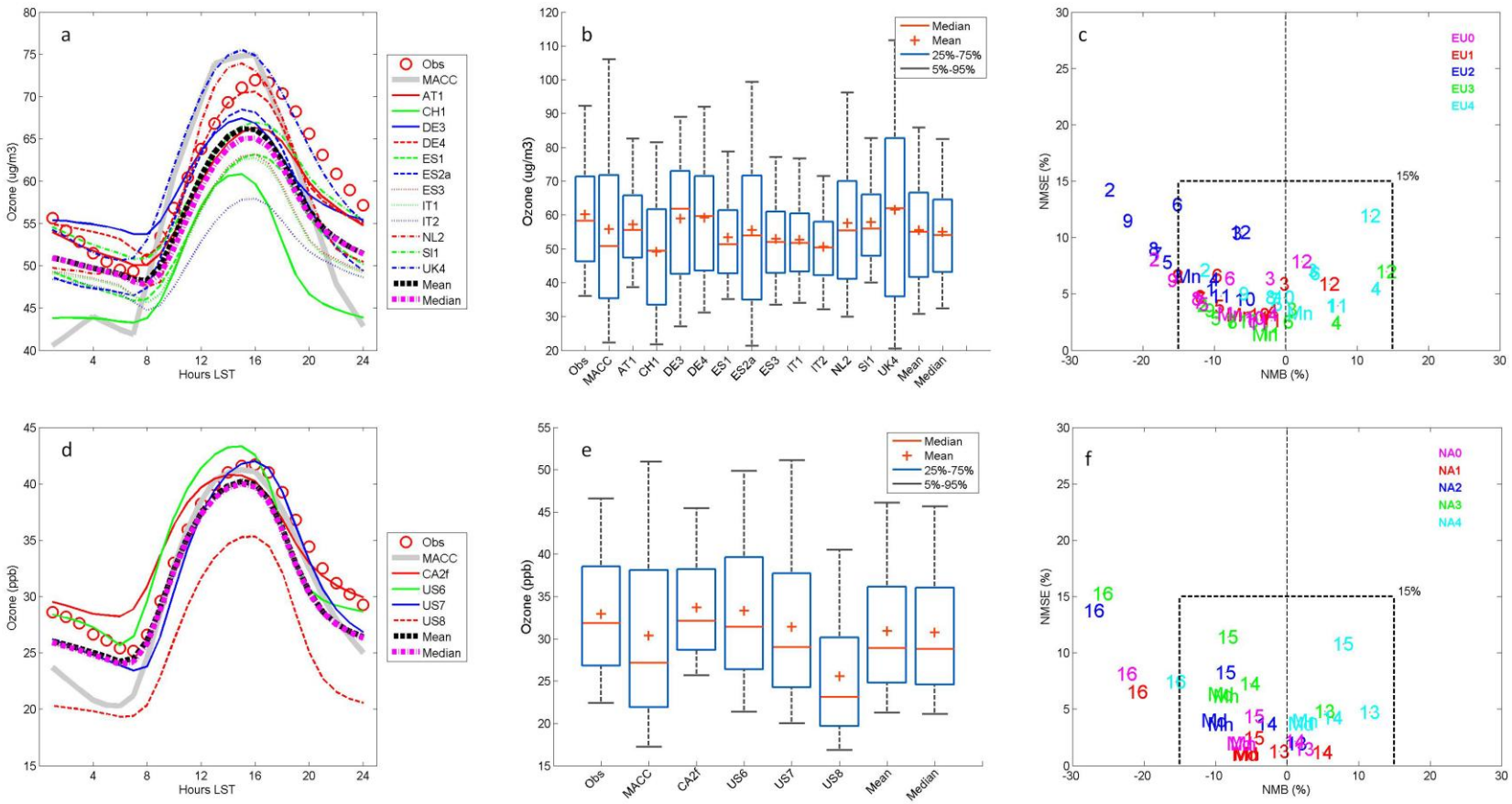
a
Ozone (ug/m3)
Hours LST
Obs
MACC
AT1
CH1
DE3
DE4
ES1
ES2a
ES3
IT1
IT2
NL2
SI1
UK4
Mean
Median
b
Ozone (ug/m3)
Median
Mean
25%-75%
5%-95%
c
NMSE (%)
NMB (%)
15%
EU0
EU1
EU2
EU3
EU4
d
Ozone (ppb)
Hours LST
Obs
MACC
CA2f
US6
US7
US8
Mean
Median
e
Ozone (ppb)
Median
Mean
25%-75%
5%-95%
f
NMSE (%)
NMB (%)
15%
NA0
NA1
NA2
NA3
NA4

Fig.2.

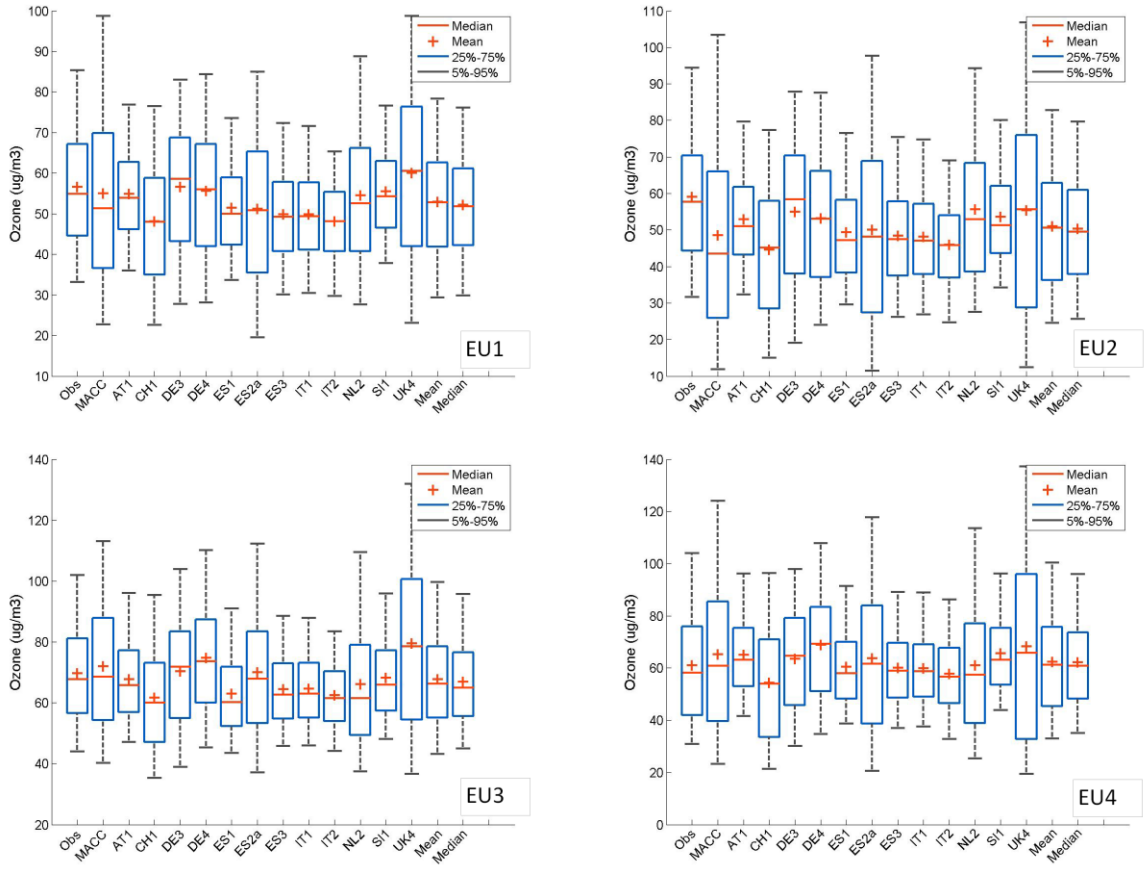
Median
Mean
25%-75%
5%-95%
Ozone (ug/m3)
EU1
EU2
EU3
EU4
Obs
MACC
AT1
CH1
DE3
DE4
ES1
ES2a
ES3
IT1
IT2
NL2
SI1
UK4
Mean
Median

Fig.3.

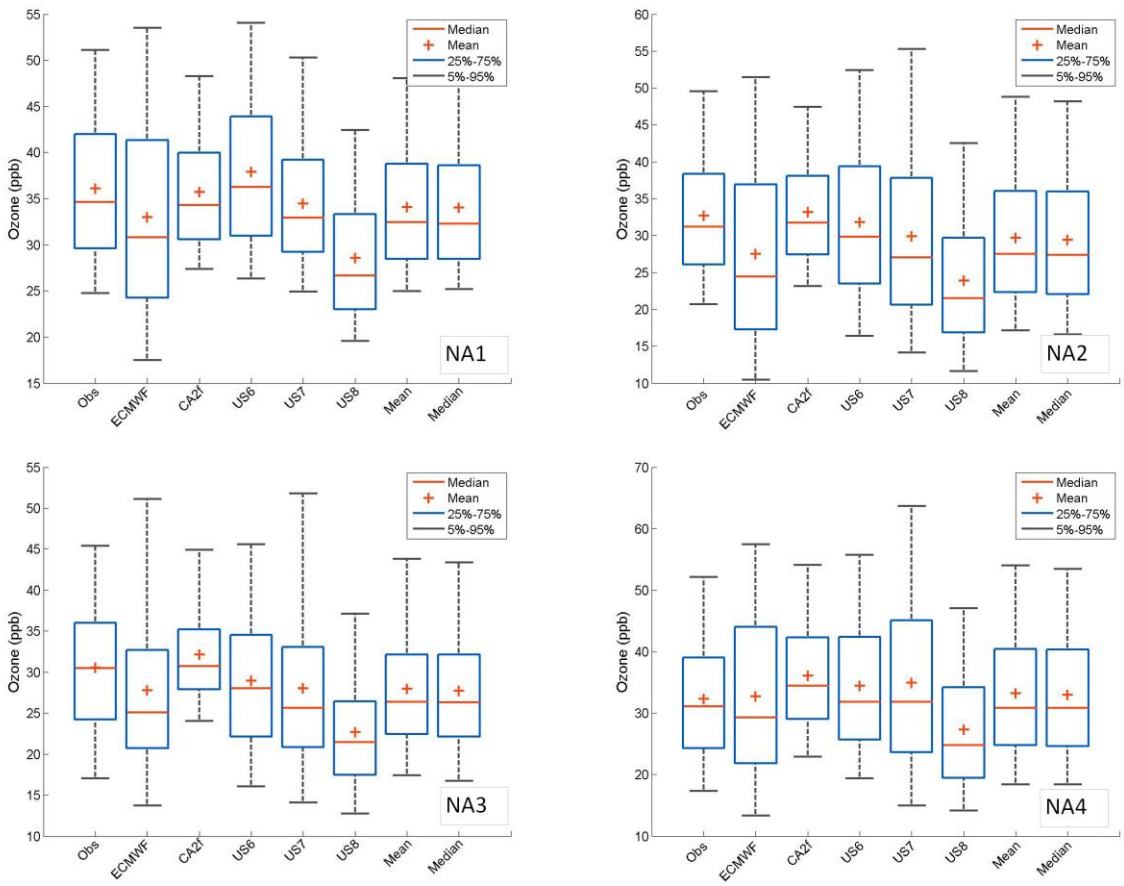
Median
Mean
25%-75%
5%-95%
Ozone (ppb)
NA1
NA2
NA3
NA4
Obs
ECMWF
CA2f
US6
US7
US8
Mean
Median

Fig.4.

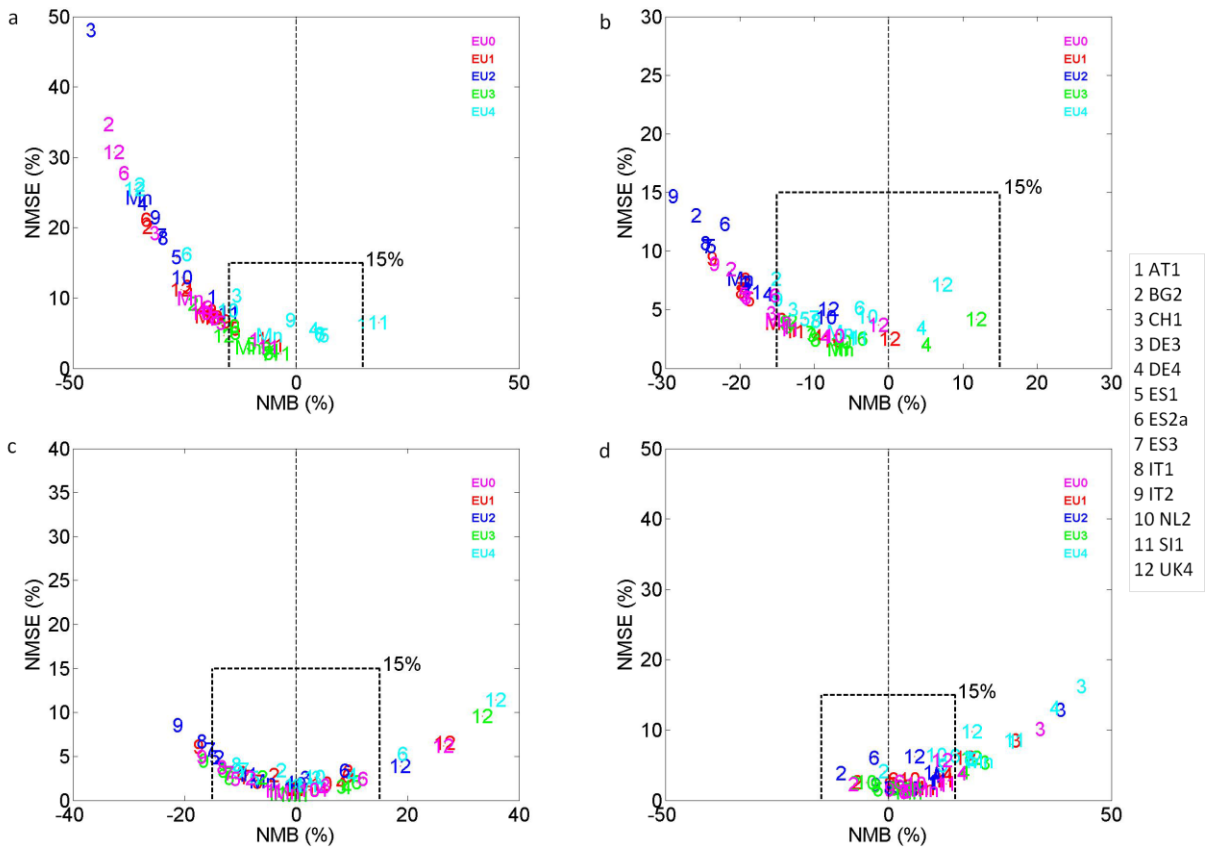
a
b
c
d
NMSE (%)
NMB (%)
15%
EU0
EU1
EU2
EU3
EU4
1 AT1
2 BG2
3 CH1
3 DE3
4 DE4
5 ES1
6 ES2a
7 ES3
8 IT1
9 IT2
10 NL2
11 SI1
12 UK4

Fig.5.

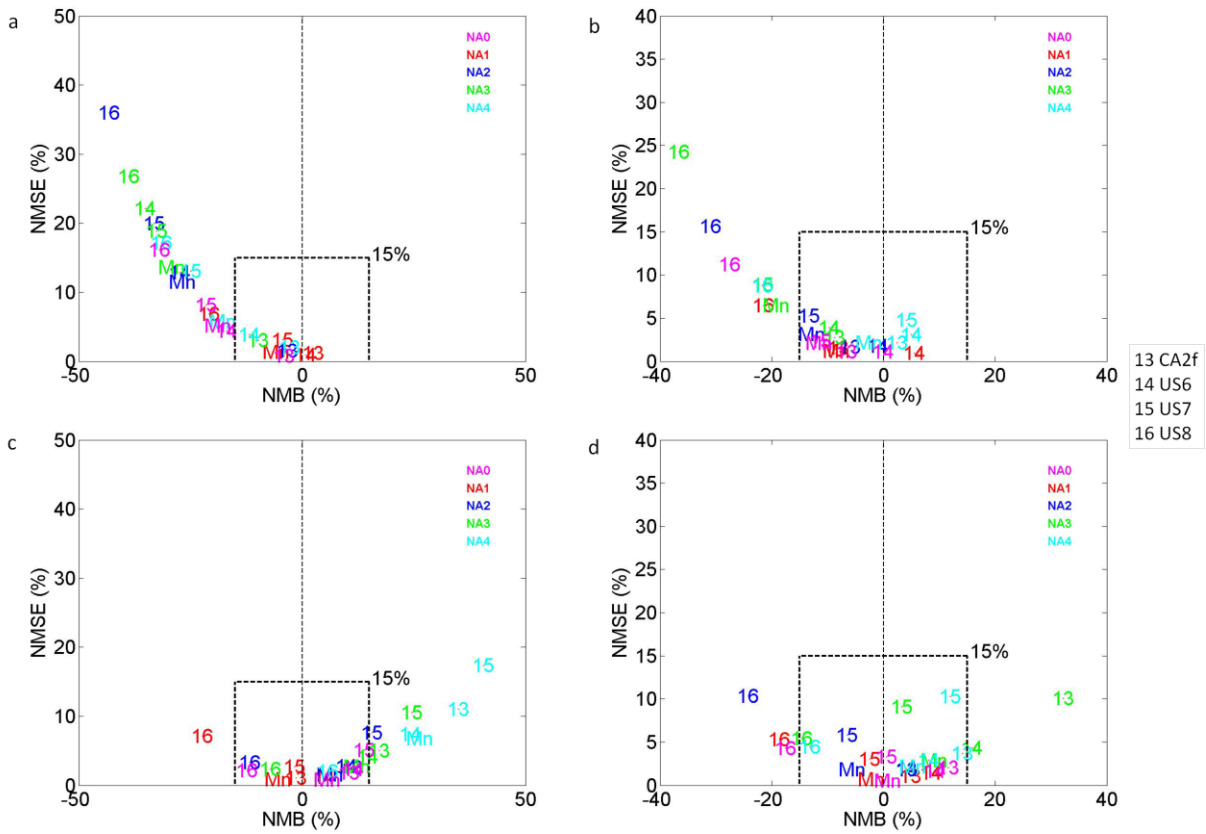

Fig.6.

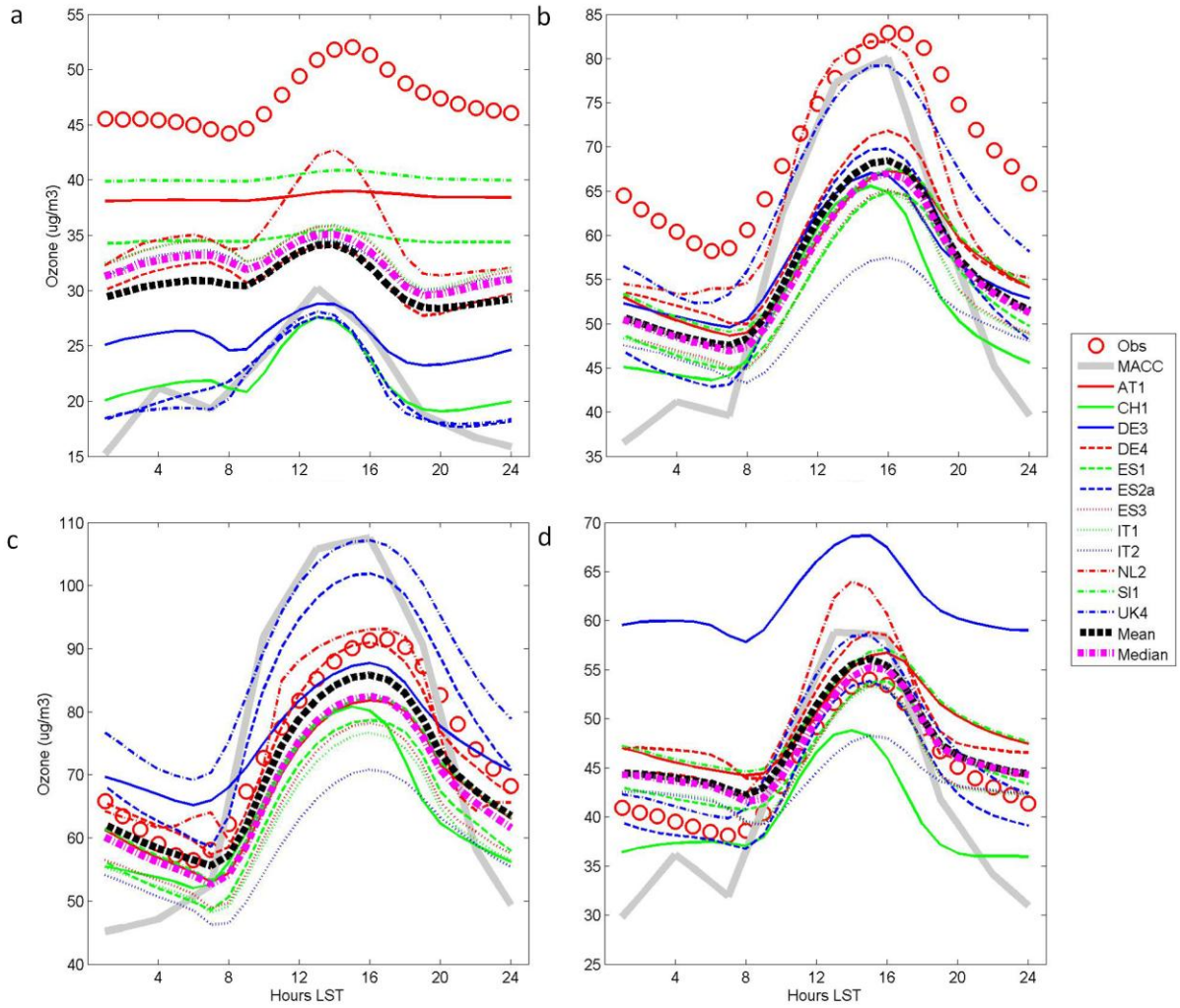
a
b
c
d
Ozone (ug/m3)
Hours LST
Obs
MACC
AT1
CH1
DE3
DE4
ES1
ES2a
ES3
IT1
IT2
NL2
SI1
UK4
Mean
Median

Fig.7.

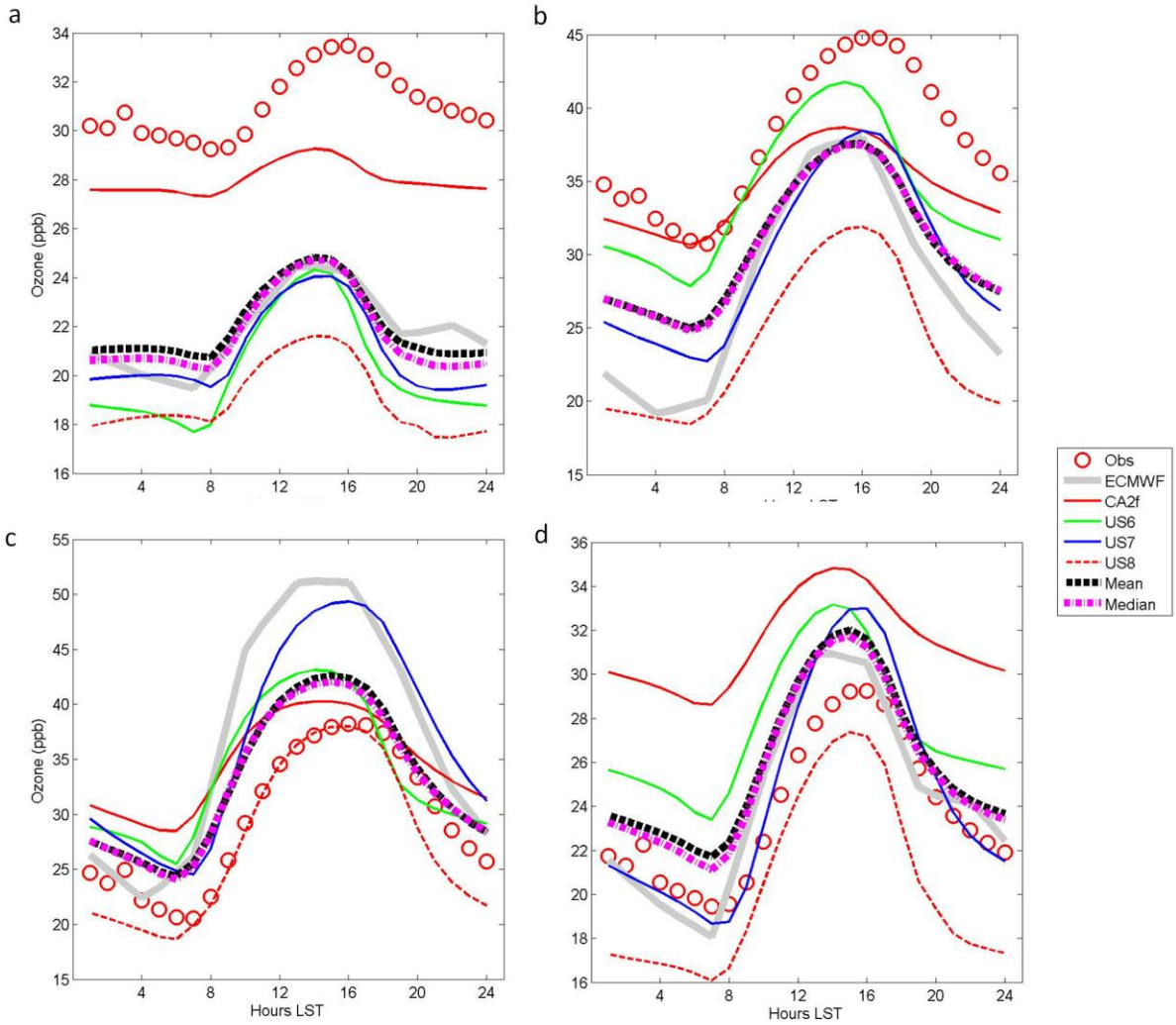
a
b
c
d
Ozone (ppb)
Hours LST
Obs
ECMWF
CA2f
US6
US7
US8
Mean
Median

Fig.8.

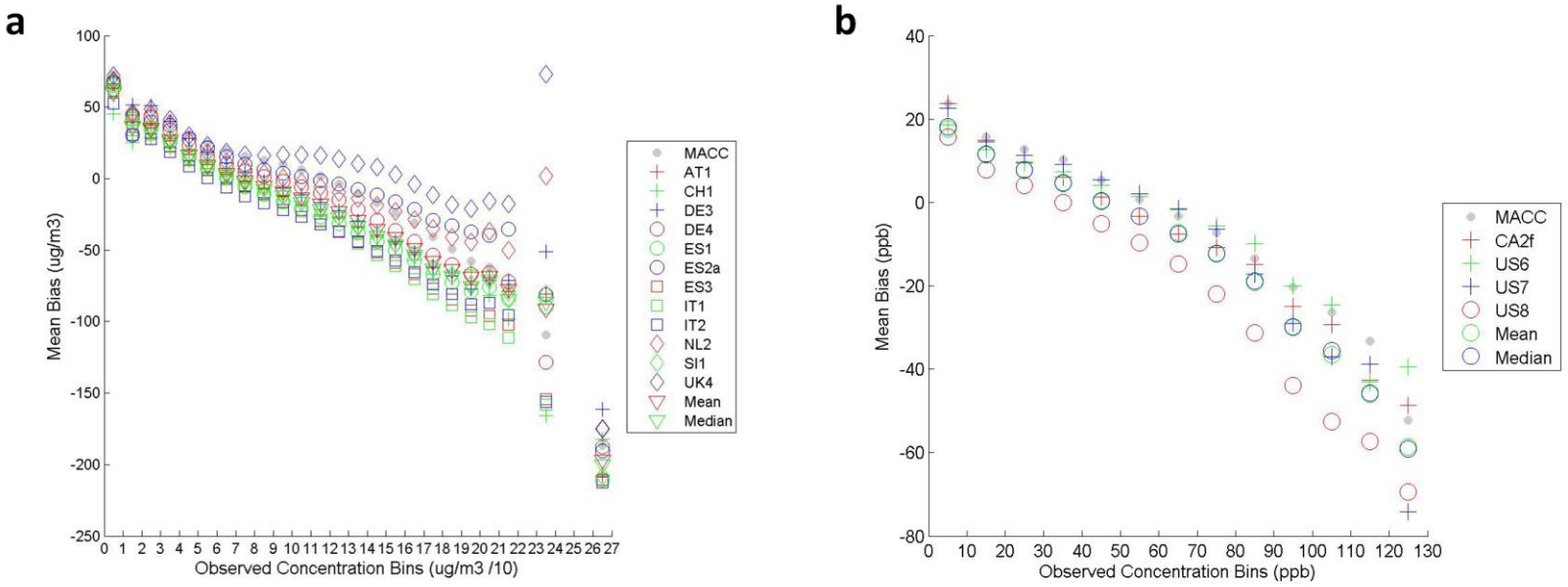
a
Mean Bias (ug/m3)
Observed Concentration Bins (ug/m3 /10)
MACC
AT1
CH1
DE3
DE4
ES1
ES2a
ES3
IT1
IT2
NL2
SI1
UK4
Mean
Median
b
Mean Bias (ppb)
Observed Concentration Bins (ppb)
MACC
CA2f
US6
US7
US8
Mean
Median

Fig.9.

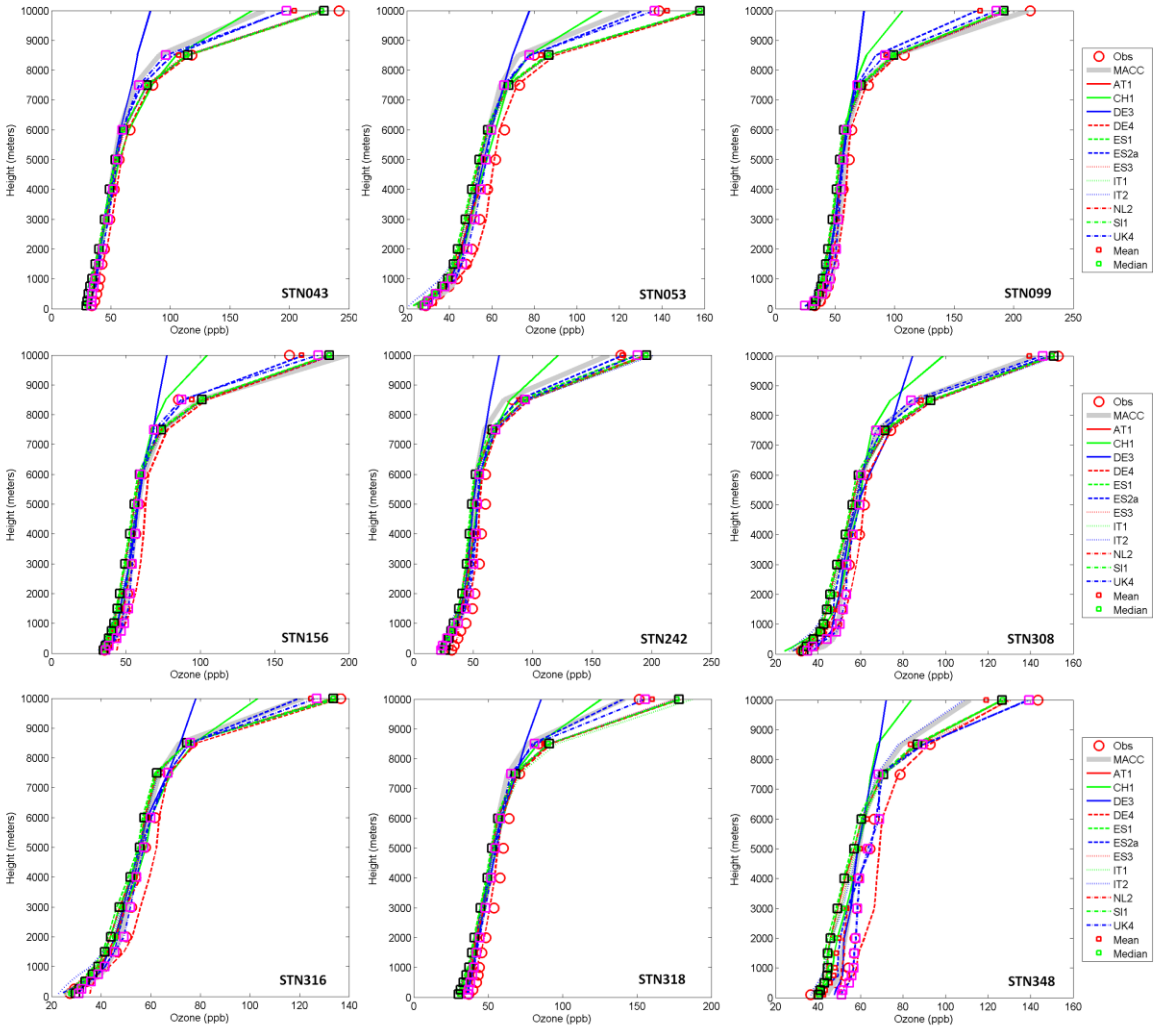
STN043
STN053
STN099
STN156
STN242
STN308
STN316
STN318
STN348
Height (meters)
Ozone (ppb)
Obs
MACC
AT1
CH1
DE3
DE4
ES1
ES2a
ES3
IT1
IT2
NL2
SI1
UK4
Mean
Median

Fig.10.

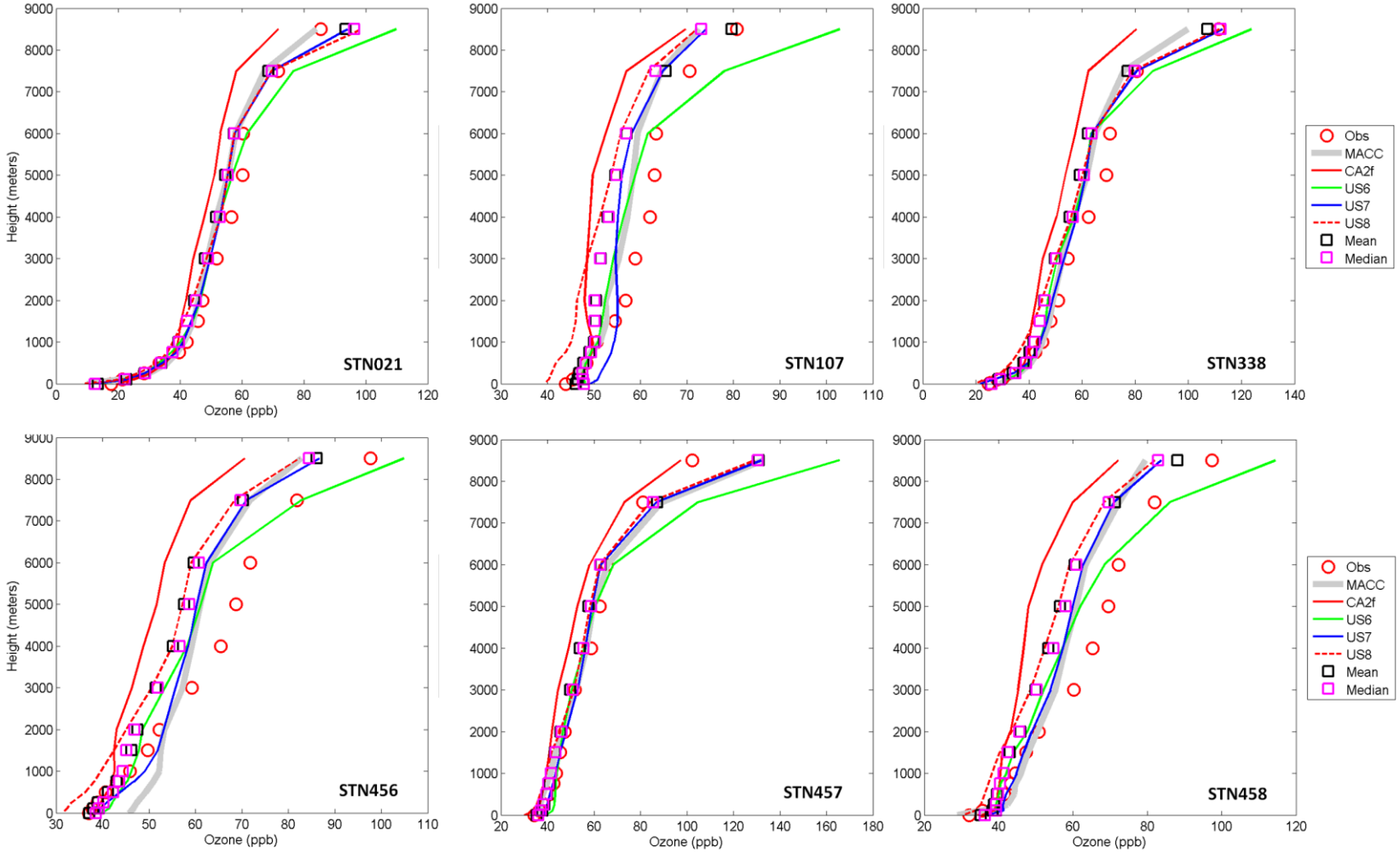
STN021
STN107
STN338
STN456
STN457
STN458
Height (meters)
Ozone (ppb)
Obs
MACC
CA2f
US6
US7
US8
Mean
Median

Fig.11.