Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII

Efisio Solazzo[1,], Roberto Bianconi[2], Guido Pirovano[6,7], Volker Matthias [17], Robert Vautard[3], Michael D. Moran[14], K. Wyat Appel[9], Bertrand Bessagnet[6], Jørgen Brandt[16], Jesper H. Christensen[16], Charles Chemel[11,12], Isabelle Coll[15], Joana Ferreira[8], Renate Forkel[10], Xavier V. Francis[12], George Grell[18], Paola Grossi[2], Ayoe B. Hansen[16], Christian Hogrefe[9], Ana Isabel Miranda[8], Uarporn Nopmongcol[4], Marje Parnk[19], Karine N. Sartelet[5], Martijn Schaap[20], Jeremy D. Silver[16], Ranjeet S. Sokhi[12], Julius Vira[19], Johannes Werhahn[10], Ralf Wolke[13], Greg Yarwood[4], Junhua Zhang[14], S.Trivikrama Rao[9], Stefano Galmarini[1]*

[1]Institure for Environment and Sustainability, Joint Research Centre, European Commission, ISPRA, Italy;
[2]Enviroware srl, Concorezzo (MB), Italy
[3]IPSL/LSCE Laboratoire CEA/CNRS/UVSQ
[4]Environ International Corporation, Novato CA, USA
[5]CEREA, Joint Laboratory Ecole des Ponts ParisTech/ EDF R & D, Université Paris-Est, France
[6]Ineris, Parc Technologique Halatte, France
[7]Ricerca sul Sistema Energetico (RSE SpA), Milano, Italy
[8]CESAM & Department of Environment and Planning, University of Aveiro, Aveiro, Portugal
[9]Atmospheric Modelling and Analysis Division, Environmental Protection Agency, NC, USA
[10]IMK-IFU, Institute for Meteorology and Climate Research-Atmospheric Environmental Division, Germany
[11]National Centre for Atmospheric Science (NCAS), University of Hertfordshire, Hatfield, UK
[12]Centre for Atmospheric & Instrumentation Research (CAIR), University of Hertfordshire, Hatfield, UK
[13]Leibniz Institute for Tropospheric Research, Leipzig, Germany
[14]Air Quality Research Division, Science and Technology Branch, Environment Canada, Toronto, Canada
[15]IPSL/LISA UMR CNRS 7583, Université Paris Est Créteil et Université Paris Diderot
[16] Department of Environmental Science, Faculty of Science and Technology, Aarhus University, Denmark
[17]Institute of Coastal Research, Helmholtz-Zentrum Geesthacht,Geeshacht, Germany
[18]CIRES-NOAA/ESRL/GSD National Oceanic and Atmospheric Administration Environmental Systems Research Laboratory Global Systems Division Boulder, Colorado USA
[19]Finnish Meteorological Institute, Helsinki, Finland
[20]Netherlands Organization for Applied Scientific Research (TNO), Utrecht, The Netherlands

[*] Author for correspondence: E.Solazzo. Email: efisio.solazzo@jrc.ec.europa.eu

Abstract. Ten state-of-the-science regional air quality (AQ) modeling systems have been applied to continental-scale domains in North America and Europe for full-year simulations of 2006 in the context of Air Quality Model Evaluation International Initiative (AQMEII), whose main goals are model inter-comparison and evaluation. Standardised modelling outputs from each group have been shared on the web-distributed ENSEMBLE system, which allows statistical and ensemble analyses to be performed. In this study, the one-year model simulations are inter-compared and evaluated with a large set of observations for ground-level particulate matter (PM_{10} and $PM_{2.5}$) and its chemical components. Modelled concentrations of gaseous PM precursors, SO_2 and NO_2, have also been evaluated against observational data for both continents. Furthermore, modelled deposition (dry and wet) and emissions of several species relevant to PM are also inter-compared. The unprecedented scale of the exercise (two continents, one full year, fifteen modelling groups) allows for a detailed description of AQ model skill and uncertainty with respect to PM.

Analyses of PM_{10} yearly time series and mean diurnal cycle show a large underestimation throughout the year for the AQ models included in AQMEII. The possible causes of PM bias, including errors in the emissions and meteorological inputs (e.g., wind speed and precipitation), and the calculated deposition are investigated. Further analysis of the coarse PM components, $PM_{2.5}$ and its major components (SO_4, NH_4, NO_3, elemental carbon), have also been performed, and the model performance for each component evaluated against measurements. Finally, the ability of the models to capture high PM concentrations has been evaluated by examining two separate $PM_{2.5}$ episodes in Europe and North America. A large variability among models in predicting emissions, deposition, and concentration of PM and its precursors during the episodes has been found. Major challenges still remain with regards to identifying and eliminating the sources of PM bias in the models. Although $PM_{2.5}$ was found to be much better estimated by the models than PM_{10}, no model was found to consistently match the observations for all locations throughout the entire year.

Keywords: AQMEII, regional air quality model, particulate matter, model evaluation, $PM_{2.5}$ speciation

1. Introduction

Particulate matter (PM) is a worldwide environmental concern that threatens both human health and ecosystems (Manders et al., 2009; Aan de Brugh et al., 2011). Human exposure to high PM concentrations is associated with respiratory disease and shortened life expectancy (Amann et al., 2005). PM also contributes to acid rain, visibility degradation, and modification of the Earth's surface energy balance, and thus contributes to short-term climate forcing (Forster, 2007; Mebust et

al., 2003; Appel et al., 2008; Smyth et al., 2009; Boylan and Russel, 2006). Recent studies have suggested that long-term changes in aerosol concentrations, primarily due to decreasing use of coal for energy production, have significantly influenced regional warming rates (Vautard et al., 2009; Philipona et al., 2009; Yiou et al., 2011). Although major efforts are being made in Europe (EU) and North America (NA) (United States and Canada) to reduce anthropogenic emissions of primary PM and PM precursors, PM levels remain problematic and their adverse effects are expected to persist (Klimont et al., 2009). The characterisation of PM sources is an area of active research as many gaps in our knowledge of the chemical speciation of PM sources, the spatial and temporal distribution of airborne particles, and the physical and chemical transformations need to be filled. This is particularly true for regional air quality (AQ) models, which incorporate a wide range of PM physics and chemistry and consider a large variety of PM emissions sources. The problem is especially difficult when simulating long temporal periods and large spatial scales due to the variety of sources involved and the chemical and physical transformations of some species that can occur over long time periods (e.g., Mathur et al., 2008).

PM is a conglomerate of many different types of chemical components (i.e., elemental and organic carbon, ammonium, nitrates, sulphates, mineral dust, trace elements, and water) with varying physical and chemical properties. Particles are either emitted directly from a source or formed from the chemical and/or physical transformation of precursor species, which depend, among other factors, on particle size. Furthermore, given its composite nature, high PM concentrations might be observed at any time during the year and under a large variety of atmospheric conditions (unlike elevated ozone mixing ratios, which are typically associated with hot and stagnant conditions). A widely accepted classification of PM is based on particle size, with PM_{10} indicating those particles with an aerodynamic diameter between 0 and 10 μm, while $PM_{2.5}$ indicates particles with an aerodynamic diameter less than 2.5 μm (note that PM_{10} includes $PM_{2.5}$). This classification is dictated by the fact that the mechanisms for the generation, transformation, removal, chemical composition, and optical properties of the two classes of particles are notably different. They also behave differently in the human respiratory track, with $PM_{2.5}$ penetrating deeper into the lungs (see, e.g., Seinfeld and Pandis (2006) for a detailed description of particle properties). In the past decade, $PM_{2.5}$ has attracted considerably more attention than PM_{10} due to its greater potential to cause adverse effects on public health. As a result, AQ model development for PM has focused primarily on modelling $PM_{2.5}$. This development is assisted by the availability of comprehensive $PM_{2.5}$ measurements, which allows model performance to be evaluated for individual PM chemical components as well as total mass, which in turn allows deductions to be made about different

aspects of model performance (e.g., the relationships between emissions, dispersion, chemistry, and deposition) (e.g., Yu et al., 2007, 2008).

The analysis presented in this paper is part of the Air Quality Model Evaluation International Initiative (AQMEII). The main objective of the project is to assess the state-of-science in current regional-scale AQ models in order to improve the ability of models to accurately characterize the spatial and temporal features embedded in air quality observations (Rao et al., 2011). Within AQMEII, standardised modelling outputs have been shared on the web-distributed ENSEMBLE system, allowing statistical and ensemble analyses to be performed (Bianconi et al., 2004; Galmarini et al., 2012). A cooperative exercise was launched for modelling groups to use their AQ models retrospectively to simulate the entire year 2006 for the continents of EU and NA. The primary goal of AQMEII is to evaluate the ability of regional AQ models to reconstruct (i.e., hindcast) atmospheric pollutant concentrations and not to forecast air quality.

While there have been a number of other model inter-comparison studies for PM (McKeen et al., 2007; Smyth et al., 2009; Stern et al., 2008; Vautard et al., 2009; Hayami et al. 2008), the scale of the model evaluation and inter-comparison presented in this study is unprecedented given the number of models evaluated, the spatial extent of the model domains, and the amount of observational data collected over the two continents.

In this paper we focus on the evaluation of PM in EU and NA using simulated results from ten different state-of-the-science regional AQ models run by 15 independent groups from both continents. In addition to the model results, observational PM data have also been gathered and made available on the ENSEMBLE system. A companion paper by Solazzo et al. (2012) has reported on a related AQMEII multi-model evaluation for ozone.

2. AQ models and monitoring

Predictions from the regional AQ models participating in the AQMEII exercise are compared to observations over the full year of 2006. Modelling groups have provided gridded surface concentrations of PM_{10}, $PM_{2.5}$, and selected gaseous compounds (e.g., SO_2 and NO_2) for two continental areas: 15°W to 35°E and 35°N to 70°N for EU; and 130°W to 58.5°W and 23.5°N to 59.5°N for NA. In addition to the gridded surface concentrations, the modelling groups also provided interpolated model concentrations at the monitoring locations for the purposes of model evaluation. To assist the analysis and interpretation of model performance, AQMEII participants

also provided emissions and deposition data for several species. To facilitate the discussion of the results, three sub-regions for each continent have been selected for analysis (see Section 3 for details), following the methodology adopted by two other AQMEII collective studies by Vautard et al. (2012) and Solazzo et al. (2012).

2.1 Participating models

Table 1 summarises the AQ models and the meteorological drivers providing PM concentrations at receptor sites for EU and NA. These are:

- CHIMERE (Bessagnet et al., 2004);
- POLYPHEMUS (Sartelet et al., 2007; Mallet et al., 2007);
- CAMx (Environ., 2010);
- MUSCAT (Wolke et al., 2004; Renner and Wolke, 2010);
- SILAM (Sofiev et al., 2006);
- DEHM (Brandt et al., 2007);
- CMAQ (Foley et al., 2010);
- LOTOS-EUROS (Schaap et al., 2008);
- AURAMS (Gong et al., 2006; Smyth et al., 2009)
- WRF/Chem (Forkel et al., 2012)

The CHIMERE, CAMx, CMAQ, and DEHM models were applied over both continents, the POLYPHEMUS, MUSCAT, SILAM, WRF/Chem, and LOTOS-EUROS models were applied for EU only, and the AURAMS model was applied for NA only. Meteorological drivers for these models are also listed in Table 1. Four AQ model simulations for EU (CHIMERE, POLYPHEMUS, CAMx, DEHM) used meteorological fields generated by different versions of the 5th Generation Mesoscale Model (MM5; Dudhia, 1993). The SILAM and LOTOS-EUROS models used meteorological data provided by the European Centre for Medium-range Weather Forecasting (ECMWF), while the Weather Research and Forecasting (WRF) v3.1 (Skamarock et al., 2008) meteorological model was used to provide meteorological input data the CMAQ simulations for EU and NA, the CAMx and CHIMERE simulations for NA, and the WRF/Chem simulations for EU. The COSMO model provided meteorological input for the EU runs of the MUSCAT and CMAQ models (for CMAQ, the CLM version of the COSMO model was used; http://www.clm-community.eu/). Finally, meteorological data from the Global Environmental Multiscale (GEM) model was used as input for AURAMS for NA. A more detailed description and assessment of model performance for the various meteorological models used can be found in Vautard et al. (2012).

The regional AQ models use different approaches in estimating PM concentrations, with key physical and chemical mechanisms handled in different ways by different models. Table 1 lists the treatments used by the models for a number of these mechanisms. For instance, the number of modes or sectional bins used for representing the particle size distribution varies between one (LOTOS-EUROS) and 12 (AURAMS), with the majority of models utilizing two size bins (PM_{10} and $PM_{2.5}$). The ISORROPIA scheme (Nenes et al., 1998) is used by the majority of the AQ models to parameterize inorganic thermodynamic equilibrium. Dry deposition of gases and particles is modelled using the resistance analogy described by Seinfeld and Pandis (2006), whereas wet deposition is taken into account by various modifications of the scavenging parameter approach (e.g., Wang et al., 2010). More details regarding individual model parameterizations are given in Table 1 and references therein. Horizontal resolution was not harmonised within AQMEII, with participants applying their own settings, ranging from 12 km (CAMx over EU and CMAQ over NA) up to 50 km (the hemispheric model DEHM). The number of vertical layers used by each model ranges from 36 layers in the WRF/Chem simulations to four for LOTOS-EUROS model simulations (adjusting their position to the height of the boundary layer). The majority of the models used between nine and 30 vertical layers, with the greatest number of layers situated in the first 2 km of the troposphere.

The emissions and chemical boundary conditions used by the various AQMEII groups are also reported in Table 1. The AQMEII organizers provided a set of time-varying gridded emissions files (referred to as the "standard" emissions) for both continents to all participants, with the goal of minimizing the contribution of differences in emissions inputs to differences in AQ model predictions. Standard emissions for EU were provided by TNO (Netherlands Organization for Applied Scientific Research) as a gridded emissions database for the years 2005 and 2006, which was partly developed in the framework of the European MACC project (http://www.gmes-atmosphere.eu/) and which is an update of an earlier TNO emissions database prepared for the GEMS project (http://gems.ecmwf.int). This inventory consists of annual anthropogenic emissions from ten SNAP (Selected Nomenclature for Air Pollution) sectors (Visschedijk et al., 2007) on a 1/16° by 1/8° latitude-longitude grid. TNO also provided SNAP-specific PM speciation profiles to allocate PM_{10} to sulphate (SO_4), elemental carbon (EC), primary organic carbon (POC), sodium, other fine PM, and other coarse PM chemical components. Biomass burning emissions for 2006 provided by the Finnish Meteorological Institute (FMI) for AQMEII were also used by some groups. The TNO inventory for EU does not include biogenic emissions, and thus different

approaches were used by different groups to provide biogenic emissions, as summarised in Table 1. The NA standard emissions were provided on a 12-km Lambert conformal grid and are based on the 2005 U.S. National Emissions Inventory (NEI), 2006 Canadian national emissions inventory, and 1999 Mexican BRAVO inventory. Biogenic emissions were provided by the BEISv3.14 model, while daily estimates of fire emissions were provided by the HMS fire detection and SMARTFIRE system (year 2006). In-stack emissions measurements for many U.S. power plants were provided by Continuous Emissions Monitoring data for the year 2006. The AQMEII inventory did not include wind-blown dust or lightning-generated NO. Full details regarding the standard emissions used for EU and NA are given in Pouliot et al. (2012). The standard emissions were used by the vast majority of the participating AQMEII groups (Table 1). Model results generated with other emissions inventories have also been submitted, however, which provides a useful comparison in interpreting the results.

AQMEII also made available a set of time-varying chemical concentrations for 2006 at the lateral boundaries of the EU and NA domains (referred to as the "standard" chemical boundary conditions). They were extracted from the Global and regional Earth-system Monitoring using satellite and in-situ data (GEMS) re-analysis product provided by European Centre for Medium-range Weather Forecast (Schere et al., 2012). The majority of modelling groups used the standard chemical boundary conditions. However, LMDZ-INCA, which couples the general circulation model Laboratoire de Meteorologie Dynamique and the Interaction with Chemistry and Aerosol model (Hauglustaine et al., 2004) was used by CHIMERE in one set of simulations (NA simulations), with another CHIMERE model simulation using the standard AQMEII boundary conditions (Table 1). The DEHM, AURAMS, and WRF/Chem groups also used non-standard chemical boundary conditions.

2.2 Receptor observations for particulate matter

The locations of the monitoring stations making PM_{10} and $PM_{2.5}$ measurements (hourly and daily) in EU and NA are displayed in Fig.1 for the three chosen sub-regions for each continent (referred to as EU1 to EU3 and NA1 to NA3).

PM measurement data for EU were assembled starting from hourly and daily data of total $PM_{2.5}$ and PM_{10} collected by the AirBase (European AQ database; see http://acm.eionet.europa.eu/databases/airbase/) network and the EMEP (European Monitoring and Evaluation Programme; see http://www.emep.int/) network. A total of 863 stations with valid data

for 2006 were made available in the ENSEMBLE database for Europe, which includes urban, sub-urban and rural stations. The number of EU stations providing daily (hourly) data was 1318 (547) for PM_{10} and 203 (74) for $PM_{2.5}$. Note that the number of EU stations measuring $PM_{2.5}$ is rather limited because the primary PM regulatory standard for EU is based on PM_{10}. Because of the small number of AirBase stations with $PM_{2.5}$ speciation data for 2006 these data have not been included in the ENSEMBLE database. More details about the networks and their QA/QC procedures can be found on the AirBase and EMEP websites.

PM measurement data for NA were assembled starting from data collected by the Aerometric Information Retrieval Systems (AIRS: see http://www.epa.gov/air/data/aqsdb.html) network and the Interagency Monitoring of Protected Visual Environments (IMPROVE; see http://vista.cira.colostate.edu/improve/) network in the U.S. and the National Air Pollution Surveillance (NAPS; see http://www.ec.gc.ca/rnspa-naps/) network in Canada. A total of 1902 stations with valid data were available for the U.S. and Canada for 2006; these include urban, suburban, and rural stations. Not all NA stations provided data with the same frequency (daily or hourly), nor are daily $PM_{2.5}$ speciation data available at all sites for all species. The number of NA stations providing daily (hourly) data was 989 (210) for PM_{10} and 1000 (149) for $PM_{2.5}$. For speciated PM data (daily average), there are 218 sites with SO_4 data, 216 sites with NH_4 data, 361 sites with NO_3 data, and 360 with EC data. More details about the networks and the QA/QC procedure for NA can be found in Mebust et al. (2003) and Appel et al. (2008).

3. PM_{10} evaluation and model cross-comparison

In the following sections, the different AQ model simulations are denoted by the labels Mod1 to Mod10 for EU and Mod12 to Mod18 for NA. In some cases the same model, but with a different configuration, was run for both continents. Such is the case for Mod3 and Mod18, Mod4 and Mod13, and Mod10 and Mod17. Note that no direct correspondence exists between these model labels and the list of models provided in Table 1 so as to preserve anonymity.

In this section, model simulations and observations are compared for PM_{10} for the three sub-regions of each continent shown in Fig. 1. These sub-regions have been selected (a) based on availability of observational data and (b) to match, when possible, the regions used by two other AQMEII multi-model studies, one focused on the meteorological drivers (Vautard et al., 2012) and the other on ozone predictions (Solazzo et al., 2012). For EU, sub-region EU1 encompasses the northwest Atlantic region, including the UK, France, Belgium, and northern Spain. Sub-region EU2, which

covers central Europe, has a continental climate with marked seasonality, many large cities, and large emissions sources. Sub-region EU3 covers the northern Mediterranean basin and includes the east coast of Spain, southern Italy, Greece, and the Mediterranean islands, along with some large urban areas (Barcelona, Rome, Athens). It was selected for its hot climate with strong solar insolation. For NA, sub-region NA1 consists of the arid southwestern part of the U.S. bounded on the east by the western slope of the Rocky Mountains. Sub-region NA2 covers the more humid southeastern U.S. Sub-region NA3, which consists of northeastern NA, including parts of Canada and three of the North American Great Lakes, has a marked seasonal cycle, the highest emissions sources in NA, and a number of large cities (e.g., New York City, Philadelphia, Toronto, Montreal). The selected NA sub-regions are characterised by rather homogeneous biogenic emissions (based on the standard NA inventory), with sub-region NA2 also characterised by the highest SOA concentrations (in summer months) of the entire continent (Sartelet et al., 2012).

3.1 Model cross-comparison

3.1.1 Emissions of PM precursors and primary PM

Analysis of emissions and deposition of several species (both PM precursors and primary PM) can aid understanding of each model's internal mass balance and chemical transformations and equilibria. The stacked distribution of quarterly accumulated emissions for five compounds is presented in Fig. 2a for EU and Fig. 2b for NA. Each segment of the bars represents emissions over a quarter of year (i.e., three months, from January to March, April to June, etc.), with the full bar reflecting total emissions over the full year. The majority of participating models delivered summed totals of the detailed emissions data used as input to the AQ models to the ENSEMBLE system. Emissions data were required to be reported by the AQMEII participants in order to check whether differences in the modelled pollutant concentrations were influenced by differences in the input emissions. Too often this check is ignored, assuming that models' sharing a common set of emissions files actually translates into models inputting exactly the same emissions. This is not always the case, as shown in Fig. 2.

Accumulated emissions for the EU domain shows that, even with the exception of Mod4 (which employed a different emissions inventory) and Mod6 (which reported only surface emissions data, neglecting plume rise and volumetric sources although they were included in the EU standard emissions), there are some differences amongst the remaining models, all of which used the standard emissions) for SO_2, aNO_x (anthropogenic NO_x), and NH_3. The fluctuations of accumulated emissions for these models may be due at least in part to the interpolation step in going from the

inventory grid to the AQ model grids. The differences do not seem to depend on season; instead, they are equally distributed over the year. The largest differences in emissions are noted for primary PM. PM emissions vary among models, with a high variability in both the coarse and the fine components, with discrepancies reaching ~550 kg km^{-2} for PM_{10} between Mod1 and Mod7 and ~300 kg km^{-2} for $PM_{2.5}$ between Mod1 and Mod6. Such large differences in emissions are attributable to the PM species included within each model, in particular the sea-salt emissions, which are not included by all models. By taking into consideration only the land-based PM emissions, the differences in emissions between the models are much smaller (not shown).

Accumulated emissions for the NA domain also exhibit a degree of variability, especially for PM_{10}, with Mod13 showing lower emissions for all quarters. This difference is due to Mod13 using an emissions inventory (same as EU Mod4) that is different from the standard AQMEII emissions inventory for NA. Overall, though, with the further exception of low SO_2 emissions reported by Mod18 (same issue as for EU Mod6 discussed above), reported domain-wide emissions for NA are more homogeneous than for EU, including smaller primary PM emissions differences amongst models.

3.1.2 Deposition of PM precursors and PM pollutants

Seasonal accumulated wet and dry deposition for each model for the EU and NA domains are presented in Fig. 3. Dry deposited substances such as $PM_{2.5}$ and selected PM chemical components exhibit the largest differences between the model simulations for both continents (Figs. 3a and 3c). While investigating the reasons leading to the differences in dry deposition is beyond the scope of this study, we note that although the dry deposition module (Table 1) is similar for all models (i.e., based on the resistance analogy scheme: Zhang et al., 2001; Seinfeld and Pandis, 2006), large deviations among models seem to indicate that the detailed parameterisations of this scheme are quite different. This may be due to the sensitivity of dry deposition to the characterization of surface properties and to near-surface meteorological conditions (e.g., surface roughness, wind shear, temperature, and radiation). Unfortunately, knowledge of dry-deposition processes, as well as the availability of dry deposition measurements, is limited (e.g., Zhang et al., 2001; 2002). In addition, dry deposition schemes are coupled with chemistry and vertical diffusion schemes as well as with the treatment of the atmospheric surface layer, and these are treated differently by each model. In the context of AQMEII, Nopmongcol et al. (2012) evaluated the sensitivity of the CAMx AQ model to two different dry deposition schemes, concluding that, at least for PM_{10}, the model is very sensitive to the deposition module (the fractional bias of the PM_{10} seasonal concentration varied by

as much as 60%). However, given the uncertainties in the PM emissions, the authors did not draw any definitive conclusions about the superiority of either of the tested schemes. Note that large differences in $PM_{2.5}$ dry deposition are mostly due to the PM sea-salt component being included in the Mod4, Mod5, and Mod10 simulations (EU) and the Mod12, Mod13, and Mod17 simulations (NA). Spatial maps of $PM_{2.5}$ dry deposition for these models (not shown) reveal that most of the deposition occurs over the ocean, while over land predicted dry deposition is comparable with the other models. EU Mod4 and Mod3 (which are essentially the same models as Mod13 and Mod18 for NA) exhibit PM-NO_3 dry deposition values that are higher than the other models, with Mod3 also having the highest PM-SO_4 dry deposition values of any model, which may be due to the inclusion of a sulphate fraction in marine sea-salt emissions that cannot be distinguished from anthropogenic fine-mode sulphate.

Seasonal accumulated wet deposition for the soluble ions SO_4^{2-}, NO_3^-, NH_4^+ are shown in Figs. 3b (EU) and 3d (NA). Prediction of the wet deposition of these ions is a challenge since it depends on the ability of models to predict both (a) the amount, duration, and type of precipitation, and (b) in-cloud and below-cloud concentrations of PM-SO_4 and SO_2, PM-NO_3 and HNO_3, and PM-NH_4 and NH_3. Vautard et al. (2012), in the context of AQMEII, analysed meteorological model predictions of precipitation over NA for 2006 and concluded that the models have a tendency to underestimate the occurrence of dry conditions and extreme precipitation events but overestimate the occurrence of light to moderate precipitation conditions, possibly leading to an overestimation of wet removal of particles and water-soluble gases, especially in areas of more frequent convection. This would result in an overestimation of the wet removal of SO_4^{2-}, NO_3^-, NH_4^+, at least for NA (discussed further in Section 4.2). Looking at the model results for wet deposition, Mod4 and Mod5 tend to have higher wet deposition amounts for all species for EU, whereas Mod17 has very low NO_3^- wet deposition values for NA. In particular, Mod5 has the highest wet deposition of SO_4^{2-}, NO_3^- and NH_4^+ and the smallest dry deposition of the same species among all the models, which might be attributed to the internal parameterisations of the model itself.

3.2 PM_{10} – model skill

Time series of monthly continental mean PM_{10} surface concentration are presented in Fig. 4 for EU and NA and their sub-regions (only receptors providing daily observations have been used). A persistent underestimation of PM_{10} by the models is common to both continents and all sub-regions. For NA in particular, the underestimation is systematic across all models, though in sub-region NA2 one model slightly overestimates PM_{10} for the October to January period. For sub-region

NA1, the model underestimation is severe (~20 μg m^{-3}), with a greater underestimation in the summer and winter. The large underestimation may be due to a lack of wind-blown dust in the models, which can be an important source of PM_{10} in this region (Yin et al., 2005; Park et al., 2010), but which is not accounted for in the emissions inventory. In sub-regions NA2 and NA3, on the other hand, the underestimation is small for some models but more significant for others, with the worst performing model, Mod13, exhibiting a bias of ~20 μg m^{-3} for both sub-regions NA2 and NA3, possibly due to lower PM emissions (Fig. 2b). For the EU continent and EU sub-regions, the majority of models also underestimate compared to the measurements, with Mod9 showing the highest bias for sub-regions EU1 and EU2. Nopmongcol et al. (2012) attributed part of the severe PM_{10} bias to an underestimation of coarse PM emissions. Mod17 in sub-regions NA2 and NA3 (fall and winter months) and Mod1 and Mod6 in EU (all sub-regions) are exceptions and estimate PM_{10} concentrations of the same magnitude or higher than the measurements. Mod10 (the same as Mod17 for NA) agrees satisfactorily with observed values of PM_{10} in sub-regions EU1 and EU2 in the fall and winter, as observed for NA. However, as discussed in Section 4.1, these same models have a tendency to overestimate $PM_{2.5}$ during the same months, thus indicating that the enhanced performance in estimating PM_{10} might be achieved at the expense of $PM_{2.5}$.

It is worth noting that the low in the observed PM_{10} concentrations occurs during the month of August for sub-regions EU1 and EU2, which all models simulated with varying degrees of success. The relatively good performance of the models in the summer has been noted in previous studies (e.g., Hodzic et al., 2005), while the poorer model performance in the winter (as compared to summer) may be due to the meteorological models having difficulty in properly simulating the occurrence of strong stable conditions, which are often responsible for elevated PM concentrations in the winter (e.g., Stern et al., 2008). It should also be noted that in the EU3 sub-region (Mediterranean area) and all of the NA sub-regions, the highest PM_{10} concentrations occur in summer, likely due to greater secondary organic aerosol production under more ideal photochemical conditions and/or an increase in the contribution from wind-blown dust (Putaud et al., 2004).

Figure 5 displays the diurnal PM_{10} cycles averaged over the entire year for the same sub-regions as Fig. 4. The amplitude of the diurnal cycle is generally underestimated by the majority of the models for EU and NA. Mod1 and Mod6 are closest to the observations in terms of mean concentration and trend, with all other models biased low compared to observations. The models underestimate

observed PM_{10} during the day-time hours for NA, with Mod17 able to simulate the magnitude of PM_{10} concentrations during night-time hours reasonably well for sub-region NA3.

Values of the root mean square error (RMSE), Pearson correlation coefficient (PCC), arithmetic mean, and spread (Std Dev) for each model simulation (entire continent, full year) are provided in Table 2 (only the rural receptors providing daily observations have been included). PCC values for the simulations vary greatly and are generally higher for EU than NA, ranging for EU from a minimum of 0.2 for Mod6 to a maximum of 0.7 for Mod4, Mod7 and Mod8, and for NA from 0.15 for Mod17 to 0.4 for Mod16. The models severely underestimate mean PM_{10} concentrations, with the exception of Mod1 and Mod6 for EU, which are the only models whose mean concentration is greater than the observed concentration. The variability of the observations for NA is underestimated by the models by a factor of two on average, indicating that the models are unable to simulate the same range of variability as the measurements. While the standard deviation for the EU observed data (7.3 μg m^{-3}) is larger than that of NA (5.4 μg m^{-3}), the majority of the models underestimate the observed variability, with the exception of Mod5, which agrees well with the observed variability. Such large differences among models depend strongly on PM composition and on the chemical components of PM_{10} included in each model's chemistry module.

3.3 Model bias

In this section we examine the error statistics for PM_{10} further and for the gaseous precursors SO_2 and NO_2, which were also made available within AQMEII. The mean fractional bias (MFB) and the mean fractional error (MFE), defined as:

$$MFB = \frac{1}{N}\sum_{1=1}^{N}\frac{(C_m - C_o)}{0.5(C_m + C_o)} \quad (1)$$

$$MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{|C_m - C_o|}{0.5(C_m + C_o)} \quad (2)$$

where C_m and C_o are the model-predicted and observed PM_{10} concentrations, respectively. Boylan and Russel (2006) have suggested that model performance goals (desired level of model accuracy) for PM should be set to MFE < 0.50 and MFB < ±0.30 (internal box of Fig. 6) and the minimum model performance criteria (the acceptable level of model accuracy) for PM should be set to MFE ≤ 0.75 and MFB ≤ ±0.60 (external box in Fig. 6). For this analysis, only receptors providing daily measurements have been used. Note that for rural receptors in the EU sub-regions, the performance criteria are met by the majority of models, with only a few models straying beyond the bounding criteria box. Conversely, for EU urban sites the MFE and MFB values for several models are higher

and fail to meet the target criteria. This not entirely surprising, as regional AQ models are not designed to capture the exposure of roadside and urban receptors to high emission fluxes (due to the spatial resolution of the models being too coarse). Mod1, Mod6, and Mod10 (three different AQ models driven by different meteorological inputs) achieve the target performance goals in all EU sub-regions, displaying generally uniform performance across the continent. Note again that the observed-modelled data pairs have not been averaged in space or time prior to the analysis presented in Fig. 6. Therefore, the clustering of performance by model rather than by sub-region (most notable for rural EU) is driven by the model itself (e.g., internal formulations, chemical modules, etc.) rather than by the model inputs, such as emissions and meteorology. For NA, the models meet the target performance goals for rural sites more successfully than for EU, although all models perform poorly at rural sites for sub-region NA1. As noted in the discussion of Fig. 4, the poor performance may be partially due to a lack of natural wind-blown dust emissions in the emissions inventory, which can contribute significantly to the PM in the rural locations in sub-region NA1.

For rural receptors of SO_2 and NO_2 (Fig. 7), which are precursors of secondary inorganic aerosol, model performance appears worse relative to PM_{10}, with larger minimum errors at all sub-regions for both continents. This is likely due to the use of hourly data for this analysis without performing any spatial or temporal averaging, whereas daily data were used for PM_{10}. Note that for NO_2, model performance is similar between regions both continents, whereas for SO_2, model performance is more similar between models, indicating that the model results are influenced more heavily by the inputs (e.g., meteorology and emissions) for NO_2 than for SO_2. The differences in model performance for NO_2 and SO_2 are not surprising however, as SO_2 is typically emitted by elevated point sources whose plumes are not easily modelled with the current resolution of chemistry transport models, whereas NO_2 is often emitted at ground level by large area and mobile sources, as well as by elevated sources. Nopmongcol, et al. (2012) attributed the tendency of the CAMx AQ model to underestimate daytime NO_x to insufficient NO_x emissions, too much dilution during the daytime, or monitors that are located near large NO_X sources (e.g., roadside monitors) that the models cannot resolve spatially. Note that the EU models that used the vertical distribution of emissions from the EMEP inventory might have too high an emissions spread for point sources and other elevated sources, which is particularly important for SO_x, as it is largely emitted by elevated point sources (e.g., Bieser et al., 2011).

3.3.1 Meteorological fields contributing to PM_{10} bias

Errors in the meteorological drivers can have a large impact on AQ model simulations, leading to errors in the dispersion of pollutants and eventually in the concentration estimates of pollutants. Vautard et al. (2012) showed that the meteorological models participating in AQMEII have a tendency to overestimate the 10-m wind speed (WS), especially in EU, which could translate into a negative bias for concentration estimates due to too much mixing and transport. The meteorological models also tend to overestimate precipitation frequency, as noted in Section 3.1.2 (for NA only, since no precipitation data were available for EU for a similar evaluation). Identifying the fraction of the total PM_{10} bias attributable to errors in the meteorological fields would require dedicated model sensitivity tests, which is beyond the scope of this study. Nonetheless, several inferences regarding the impact that errors in the meteorological inputs have on the AQ models can be made by pairing the MFB of the modelled PM_{10} with the MFB of the modelled WS and precipitation. Fig. 8 shows the daily PM_{10} bias compared to the WS bias for the months of August and November (Fig. 8a-d), along with the PCC values calculated for each sub-region. Note that the analysis presented in these figures is only valid in a spatially-averaged sense, as observed WS and PM_{10} concentrations are not co-located.

There is a systematic overestimation of the WS in the EU sub-regions, which tends to correspond to the largest underestimations of PM_{10} (although there are some exceptions in sub-region EU3). The correlation between WS overestimation and PM_{10} underestimation is more evident for EU than for NA, which supports the results of Vautard et al. (2012), who noted a larger WS bias in EU than NA. The tendency of the meteorological models to overestimate precipitation frequency in NA (not shown) would generally result in increased removal of PM by wet deposition, also resulting in lower PM concentrations. However, the removal of PM by wet deposition is more important above the surface and therefore may not impact surface PM concentrations as significantly as the bias in WS.

4. $PM_{2.5}$ and $PM_{2.5}$ components

4.1 Time series and statistics

Time series of monthly, continental mean $PM_{2.5}$ surface concentrations, based on stations providing daily measurements, are shown in Fig. 9 for EU and NA. Compared to PM_{10} (Fig. 4), model bias is much lower for both continents, demonstrating an enhanced capability of the AQ models to simulate $PM_{2.5}$. For EU, the models underestimate the monthly mean $PM_{2.5}$ surface concentrations for all sub-regions, with several exceptions. In particular, the models that overestimated PM_{10} concentrations (Mod1, Mod6) also overestimate $PM_{2.5}$ concentrations. The majority of the NA

models also tend to underestimate $PM_{2.5}$ concentrations (especially for sub-regions NA2), and only Mod17 overestimates $PM_{2.5}$ concentrations throughout the year fort sub-regions NA2 and NA3. Mod16 and Mod18 underestimate $PM_{2.5}$ in the summer, but overestimate $PM_{2.5}$ during the other seasons, while the other four models underestimate $PM_{2.5}$ throughout the entire year. Model 13 (the same as Mod4 for EU) and Mod12 are among the models most severely biased low and also have among the highest deposition of, respectively, PM components (dry and wet), and total $PM_{2.5}$, for NA (Fig. 3c-d). Looking at the individual sub-regions, several models perform well for short periods, closely following the observations for a season, such as Mod18 for sub-region NA1 between October and December, Mod17 for sub-region NA2 for April-June, and Mod12, Mod15 and Mod14 for sub-region NA3 for October-December. For sub-region NA3 in particular, all models predict the July peak and the April and October lows, although the amplitude is not well captured.

The ratio of observed and modelled daily PM_{10} to $PM_{2.5}$ concentrations has also been computed to provide an indication of the proportion of the fine and coarse components of PM predicted by each model and how that ratio compares to the observed ratio. The analysis is based on co-located PM_{10} and $PM_{2.5}$ rural monitoring stations (134 and 50 in NA and EU, respectively), with availability of data exceeding 50% (over the year) and an altitude below 1000 m. Results for PCC, mean, and Std Dev for the continent-wide domains are presented in Table 3. The mean observed value, higher for NA than EU, is underestimated by all of the models, with the largest underestimation for NA due to the large underestimation of PM_{10}, and only Mod1 for EU overestimating the observed mean. The overall underestimation is consistent with the larger underestimation of PM_{10} compared to $PM_{2.5}$. The standard deviation of the $PM_{10}/PM_{2.5}$ ratio (0.27 for both continents) is underestimated by ~40 to 90% for both continents, with the only exception being the large overestimation by Mod1 for EU. Because correlation coefficients for PM_{10} and $PM_{2.5}$ individually are higher than those of their ratio, this seems to indicate that the coarse fraction (difference between PM_{10} and $PM_{2.5}$) is too low in most models. Since most coarse PM is related to primary emissions, this points to problems with the emissions inventories, as already discussed in the previous sections.

4.2 $PM_{2.5}$ - Model Skill

Figs. 10 and 11 present Taylor diagrams (Taylor, 2001) for EU and NA, respectively, to help assess the skill of each model in simulating the day-to-day variability of daily mean $PM_{2.5}$ concentration at urban and rural sites. In a Taylor plot, 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 (PCC or r) 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 practice, the closer a model point is to an observed point, the better the model performance. The ensemble mean of all available models is also provided for comparison. No prior averaging was performed on the data for this analysis (i.e., all available daily data pairs were used).

For EU, the amplitude of daily $PM_{2.5}$ variability is underestimated by all models for sub-region EU2 for both rural and urban stations. The PCC ranges between 0.55 and 0.75 for the majority of models, with better results for the rural sites and for sub-region EU3 (PCC above 0.8 for the ensemble mean). The variance observed in sub-region EU2 is a factor of two higher than any other EU or NA sub-region, possibly due to the occurrence of large regional PM episodes (an example of which is examined in the next section) that are not well reproduced by EU model simulations. The modelled variance is significantly underestimated for this region, while for sub-regions EU1 and EU3, the variance is reproduced satisfactorily by models, with ensemble mean values of 4.6 and 4.5 μg m^{-3} providing the best agreement with the observed variances of 4.7 and 4.3 μg m^{-3} for the same sub-regions, respectively.

Model-to-measurement correlation is slightly higher for the NA sub-regions (total $PM_{2.5}$, Fig. 11a,b), with the majority of points between 0.6 and 0.8, indicating that the daily variability of $PM_{2.5}$ is better reproduced for NA, possibly due to differences in the PM emission datasets (Pouliot et al., 2012). The only exception is Mod18, which has PCC values below 0.6 for all NA sub-regions. The observed Std Dev is rather well reproduced for sub-region NA2 and to a slightly lesser extent sub-region NA3, which is supported by the ensemble mean scores for these regions, while model performance for sub-region NA1 is notably worse than for the other two NA sub-regions.

Fig. 12 presents Taylor diagrams of four speciated $PM_{2.5}$ components, specifically NH_4, SO_4, NO_3, and EC, for the NA sub-regions. The analysis for the inorganic aerosols and elemental carbon confirms the systematic underestimation of the Std Dev for sub-region NA1. By contrast, model performance for sub-regions NA2 and NA3 varies depending on the speciated PM component being considered. Day-to-day variability of PM-SO_4 is well reproduced for both sub-regions NA2 and NA3, as indicated by the high correlation values (exceeding 0.7), and the model variance, which closely matches that of the observations. This agreement can possibly be attributed to SO_4 chemistry being less complex than that of other PM components (Boylan and Russell, 2006).

Mod13 and Mod17 had the highest dry deposition of SO_4 (Fig. 3c) and are the two models with the highest SO_4 errors (Fig. 12 b). Furthermore, Mod 18, which had a large deposition of NO_3 , also has high error for NO_3 (Fig. 12 c). Although a direct relationship between these two aspects cannot be inferred here, these associations are worth further investigation. While NO_3 is overestimated for sub-region NA3 and to a lesser extent for sub-region NA2, NH_4 is underestimated for both regions. In most cases, model performance for sulphate and nitrate is mutually compensating, with the underestimations in SO_4 related to overestimations in NO_3. The only exception is Mod15, which overestimates both SO_4 and NO_3. Finally, EC Std Dev is well reproduced for sub-region NA3, while it is underestimated for sub-region NA2, suggesting that EC emissions may be underestimated for portions of NA.

Overall, model performance for sub-region NA1 is worse than for the other NA sub-regions. The systematic underestimation of the Std Dev for all species in this region indicates there may be large emission sources missing in the emissions inventory for western NA. In addition, western NA is also a challenging region to model due to its complex terrain and the close proximity to the western boundary of the modelling domain, which makes the region particularly sensitive to errors in the prescribed meteorological and chemical boundary conditions.

Finally, it is worth noting that the models showed more homogenous performance across each sub-region for the speciated compounds than for total $PM_{2.5}$ mass. This result might suggest that, while AQ models are reliable for simulation of inorganic aerosol, there are still gaps in the representation of some processes other than inorganic aerosol chemistry that strongly influence $PM_{2.5}$ concentration.

5. Two episodes with elevated PM concentrations

Two episodes with elevated PM concentrations, one in EU and one in NA, have been selected for a more detailed investigation of the models' performance. It is important to determine that the AQ models not only capture the average PM concentrations correctly, but that they also reproduce the peak values as well.

For EU, a period of 16 days between 13 and 28 April 2006 was chosen for detailed analysis. During this period, elevated $PM_{2.5}$ concentrations were observed at several stations in central EU. For the evaluation of the AQMEII model results, a region between 49° and 56° North latitude and between 0° and 14° East longitude was selected. Daily average $PM_{2.5}$ values were available at seven stations,

four in Germany, and one each in Denmark, Belgium, and Great Britain. All of these stations are classified as rural stations, which are generally better represented by regional AQ models.

Fig. 13 presents time series of observed and modeled mean daily $PM_{2.5}$ concentrations for these stations. It is evident that the model values scatter considerably around the observations. All of the models, with the exception of Mod4 and Mod6, show an increase in $PM_{2.5}$ concentrations from April 13 to April 25. The observations show a secondary peak on April 15 and April 16 that is not well reproduced by the models, but the high values between April 23 and April 26 are captured. PCC values are between 0.47 (Mod3) and 0.62 (Mod2 and Mod9) for all models except Mod6, which has a low PCC of 0.14. The mean bias varies between -8.8 to 12.2 μg m^{-3}, with only five models having mean standard deviations less than 3 μg m^{-3}. The mean observed $PM_{2.5}$ concentration is 15.6 μg m^{-3}.

In NA, a more detailed analysis is possible, as a large number of stations are available with hourly $PM_{2.5}$ measurements. However, $PM_{2.5}$ speciation data are only available as daily average values every three days. Regardless, the availability of hourly measurements and speciation measurements allows additional insights into the possible reasons for discrepancies between the modeled and observed concentrations. The region chosen for the intensive investigation was in the eastern US between 32° and 45° North latitude and between 72° and 92° West longitude. Hourly data from 18 receptor stations classified as either rural or suburban were available, and results from six different AQ models were used for the evaluation.

Between 14 July and 29 July 2006, elevated $PM_{2.5}$ concentrations greater than 20 μg m^{-3} were observed on several days (Fig. 14). On other days, the concentrations generally averaged around 5 μg m^{-3}. These abrupt changes in $PM_{2.5}$ concentrations are mostly driven by transport phenomena, which the models seem to capture quite well. However, the models do have some difficulty simulating the timing of the episodes. Consider for example the peak in observed $PM_{2.5}$ on July 18, which is seen a bit later in the model simulations, and another day of high $PM_{2.5}$ concentrations that are modeled near the end of the period on July 28, but that are not supported by the observed data. PCC values for hourly predictions are between 0.53 (Mod12) and 0.66 (Mod18) for all models. PCC values for the aggregate daily concentration are between 0.64 (Mod13) and 0.92 (Mod16), and are thus higher than the EU case.

The model biases for $PM_{2.5}$ are between -5.2 μg m^{-3} (-39%) and +3.8 μg m^{-3} (28%), with a mean observed value of 13.3μg m^{-3}. These biases are smaller than those for the episode chosen for

Europe. However, these values are only representative of this one short period and will certainly be different for other episodes. It is interesting to note that the results from the EU groups were all biased low (-5.2 to -4.3 μg m^{-3}) while the results for the NA groups matched the observed values better (biases between -3.4 and +3.8 μg m^{-3}).

The analysis of the $PM_{2.5}$ chemical composition is based on 19 stations located in the same analysis region. At each station, between three and five observations were available within the 16-days period, which allows for only a limited statistical analysis for the period. The major contribution to $PM_{2.5}$ comes from SO_4, whose mean value was 6.0 μg m^{-3}. The EU groups underestimated SO_4 by 7 to17 %, whereas the bias for the NA groups was between -11% and +21%. As expected in summer, NO_3 concentrations were much lower than SO_4, with an observed mean value of 0.5 μg m^{-3} and much greater scatter around the observations (biases ranged from -54% to + 61%). NH_4 had a mean observed concentration of 1.8 μg m^{-3}. Since NH_4 is closely linked to SO_4 when NO_3 concentrations are low, model biases are in the same direction as for SO_4, ranging from -36% to +30%. Again, the European groups calculated lower concentrations than the North American groups.

In summary, the $PM_{2.5}$ chemical components SO_4, NO_3 and NH_4, were better simulated by the models than total $PM_{2.5}$. This suggests that other components of PM, such as organic aerosols and unspeciated PM mass, which can make up a significant portion of the total PM mass, are simulated less well than the inorganic components. Although it is not possible to identify the main reasons for such behavior, it seems that the simulations performed by the NA groups were better adapted to simulate PM concentrations than the EU groups for their respective continents.

6. Conclusions

Annual AQ model simulations in the context of AQMEII have been inter-compared and evaluated. The focus has been put on surface concentrations of PM, both fine and coarse. Given the scale of the project - involving ten AQ models that were run over two continents (EU and NA) for one entire year, 2006 – the available model results allow for a comprehensive analysis.

We have analysed predictions of PM_{10} and $PM_{2.5}$ in several sub-regions of the continental domains, quantifying bias and model performance with the aid of a number of statistical indicators. We conclude that a large variability exists among models (and even among different simulations performed using the same model), especially for PM_{10}, with model concentrations varying by up to a factor of seven between the different model simulations. Because most of the models shared

common emissions and chemical boundary condition data sets, reasons for the large prediction spread need to be determined through future studies. Here, the model outputs have been analyzed in terms of emissions and dry and wet deposition of several species relevant to PM, with the conclusion that the internal parameterisations of models play a pivotal role, even though the native schemes in the models are often similar. The models severely underestimate PM_{10} concentrations over the entire year and for all sub-regions, often with mean fractional errors exceeding 0.75 for both continents. The tested AQ models also show a tendency to under-predict concentrations of two PM precursors (NO_2, SO_2) in both continents throughout the year. While typical explanations of PM_{10} underestimation, such as missing PM_{10} sources in the emission inventory (e.g., anthropogenic and natural dust), no doubt contribute to the PM_{10} bias in the model simulations, other causes for the large PM_{10} underestimation were investigated. For EU, areas with large biases in wind speeds were found to have a larger bias in PM_{10} concentrations. Therefore, at least part of the model bias for PM_{10} in EU can be attributed to errors in the PM emissions and meteorological inputs, such as wind speed.

Model performance for $PM_{2.5}$ was greatly improved over that of PM_{10}, with PCC values often exceeding 0.7 (higher on average for NA than EU). While $PM_{2.5}$ time series reveal that some models perform better than others in some areas and during some periods of the year (e.g., single seasons), this behaviour is not uniform in time and space. Model skill in estimating some of the major chemical components of $PM_{2.5}$ (i.e., SO_4, NO_3, NH_4 and EC) was found to be more homogenous than for total $PM_{2.5}$ mass for NA. This suggests that while the models do relatively well in simulating the inorganic aerosol species, large uncertainty remains in the simulation of other components of $PM_{2.5}$ (e.g., secondary organic aerosols and unspeciated $PM_{2.5}$).

Finally, analysis of two high $PM_{2.5}$ concentration episodes in EU and NA revealed that while there is a considerable scatter of model results around the observed concentration, the models were able to capture the buildup, peak, and eventual decrease in $PM_{2.5}$ concentrations during elevated PM episodes, especially for NA. Investigation of the individual $PM_{2.5}$ species showed that SO_4, NO_3 and NH_4, are better reproduced by the models than total $PM_{2.5}$ mass, indicating that large errors exist in the simulation of the other components of $PM_{2.5}$ such as organic aerosols and unspeciated $PM_{2.5}$ mass.

Acknowledgments

The work carried out with the DEHM model was supported by The Danish Strategic Research Program on Sustainable Energy under contract no 2104-06-0027 (CEEH). Homepage: www.ceeh.dk. The RSE contribution to this work has been partially financed by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE and the Italian Ministry of Economic Development (Decree of March 19th, 2009).

References

Aan de Brugh, J.M.J., Schaap, M., Vignati, E., Dentener, F., Kahnert, M., Sofiev, M., Huijnen,V., Krol, M.C., 2011. The European aerosol budget in 2006. Atmos. Chem. Phys 11, 1117-1139.

Ackermann, I.J., Hass, H., Memmesheimer, M., Ebel, A.,Binkowski, F.S., Shankar, U., 1998. Modal aerosol dynamics model for Europe: development and first applications. Atmospheric Environment 32 (17), 2981–2999.

Amann, M., Bertok, I., Cofala, J., Gyarfas, F., Heyes, C., Klimon, Z., 2005. Baseline scenarios for the Clean Air for Europe (CAFÉ) Programme. Final Report, International Institute for applied systems analysis, Schlossplatz 1, A-2361 Laxenburg, Austria.

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

Berge, E., 1997. Transboundary air pollution in Europe. In: MSC-W Status Report 1997, Part 1 and 2, EMEP/MSC-W Report 1/97, The Norwegian Meteorological Institute, Oslo.

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

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.

Biesier, J., Aulinger, A., Matthias, V., Quante, M., Denier van Der Gon, H.A.C., 2011. Vertical Emission profiles for Europe based on plume rise calculations. Environmental Pollution, in press.

Binkowski, F.S. and S.J. Roselle, 2003: Models-3 Community Multiscale Air Quality (CMAQ) Model Aerosol Component 1. Model Description, J. Geophys. Res., 108(D6), 4183, doi:10.1029/2001JD001409.

Binkowski, F.S., Shankar, U., 1995. The regional particulate matter model, 1. Mode description and preliminary results. Journal of Geophysical Research 100, 26191–26209.

Bond, T.C., E. Bhardwaj, R. Dong, R. Jogani, S. Jung, C. Roden, D.G. Streets, S. Fernandes, and N. Trautmann (2007), Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850-2000, Glob. Biogeochem. Cyc., 21, GB2018

Boylan, J.W., Russell, A.G., 2006. PM and light extinction model performance metrics, goal, and criteria for three-dimensional air quality models. Atmospheric Environment 40, 4946-4959.

Brandt, J., J. H. Christensen, L. M. Frohn, C. Geels, K. M. Hansen, G. B. Hedegaard, M. Hvidberg and C. A. Skjøth, 2007. THOR – an operational and integrated model system for air pollution forecasting and management from regional to local scale. Proceedings of the 2nd ACCENT Symposium, Urbino (Italy), July 23-27, 2007

Byun, D. and K.L. Schere, 2006: Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. Appl. Mech. Rev., 59, 51-77.

Chin, M., Ginoux, P., Kinne, S., Holben, B.N., Duncan, B.N., Martin, R.V., Logan, J.A., Higurashi, A., Nakajima, T., 2002. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sunphotometer measurements. Journal of Atmospheric Science 59, 461-483.

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

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

Dudhia, J. 1993 A nonhydrostatic version of the PennState/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. Monthly Weather Review 121, 1493-1513.

ENVIRON, 2010 . User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx). Version 5.2.Available at: http://www.camx.com.

Erisman, J.W., Draaijers, G.P.J., Mennen, M.G., Hogenkamp, J.E.M., van Putten, E., Uiterwijk, W., Kemkers, E., Wiese, H., Duyzer, J.H., Otjes, R. and Wyers, G.P., 1996, Towards development of a deposition monitoring network for air pollution in Europe, RIVM Report 722108014, RIVM, Bilthoven, The Netherlands.

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

Forkel, R., Werham, J., Hansen,A.B., McKeen, S., Peckam, S., Grell, G., Suppan, P., 2012. Effect of aerosol-radiation feedback on regional air quality – A case study with WRF/Chem. Atmospheric Environment doi:10.1016/j.atmosenv.2011.10.009

Forster, PM; Ramaswamy, V; et al. 2007. Changes in Atmospheric Constituents and in Radiative Forcing , In: Solomon; S; Qin, D; Manning, M; Chen, Z; Marquis, M; Averyt, KB; Tignor, M; Miller, HL (Ed) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press.

Galmarini, S., Bianconi, R., Appel, W., Solazzo, E., and 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 doi:10.1016/j.atmosenv.2011.08.076

Gong, S.-L., Barrie, L.A., Blanchet, J.-P., 1997. Modeling sea-salt aerosols in the atmosphere. Part 1: Model development. J. Geophys. Res. 102, 3805-3818.

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

Graedel, T.E., T.S. Bates, A.F. Bouwman, D. Cunnold, J. Dignon, I. Fung, D.J. Jacob, B.K. Lamb, J. A. Logan, G. Marland, P. Middleton, J.M. Pacyna, M. Placet, and C. Veldt (1993): A Compilation of Inventories of Emissions to the Atmosphere. Global Biogeochemical Cycles. 7, 1-26.

Guelle, W., Schulz, M., Balkanski, Y., Dentener, F., 2001. Influence of the source formulation on meddling the atmospheric global distribution of sea salt aerosol. Journal of Geophysical Research 106, 27509-27524.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., Fall, R., 1993. Isoprene and monoterpene emission rate Variability: Model Evaluations and Sensitivity Analyses. Journal of Geophysical Research 98 (D7), 12609-12617.

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

Guenther, A., Karl, T., Harley, P., Wiedinmeyer,C., Palmer, Geron, C., 2006. Estimates of global terrestrial isoprene emissions using MEGAN. Atmos. Chem. Phys 6, 3181-3210

Hayami, H., Sakurai, T., Han, Z., Ueda, H., Carmichael, G.R., Streets, D., Holloway, T., Wang, Z., Thongboonchoo, N., Engardt, M., Bennet, C., Fung, C., Chang, A., Park, S.U., Kajino, M., Sartelet, K., Matsuda, K., Amann, M., 2008. MICS-Asia II: Model intercomparison and evaluation of particulate sulfate, nitrate and ammonium. Atmospheric Environment 42, 3510-3527.

Hodzic, A., Vautard, R., Bessagnet, B., Lattuati, M., Moreto, F., 2005. On the quality of long-term urban particulate matter simulation with the CHIMERE model. Atmospheric Environment, 39, 5851-5864.

Jiang, W., 2003. Instantaneous secondary organic aerosol yields and their comparison with overall aerosol yields for aromatic and biogenic hydrocarbons. Atmospheric Environment 37, 5439–5444.

Kelly, J.T., P.V. Bhave, C.G. Nolte, U. Shankar, and K.M. Foley, 2010: Simulating emission and chemical evolution of coarse sea-salt particles in the Community Multiscale Air Quality (CMAQ) model, Geosci. Model Dev., 3, 257-273.

Kim, Y., Couvidat, F., Sartelet, K., and Seigneur, C.: Comparison of different gas-phase mechanisms and aerosol modules for simulating particulate matter formation, J. Air Waste Manage. Assoc, in press, 2011.

Klimont, Z., et al. 2009. Projections of SO2, NOx and carbonaceous aerosols emissions in Asia. Tellus, 61B, 602-617.

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., Van Vuuren, D.P., 2010. Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. Atmospheric Chemistry and Physics Discussions 10, 4963-5019.

Loosmore, G., Cederwall, R., 2004. Precipitation scavenging of atmospheric aerosols for emergency response applications: testing an updated model with new real-time data. Atmospheric Environment 38, 993–1003.

Makar, P.A., Bouchet, V.S., Nenes, A., 2003: Inorganic chemistry calculations using HETV – a vectorized solver for the SO_4^{2-}–NO_3^-–NH_4^+ system based on the ISORROPIA algorithms. Atmos. Environ. 37, 2279-2294.

Mallet V., Quélo D., Sportisse B., Ahmed de Biasi M., Debry E., Korsakissok I., Wu L., Roustan Y., Sartelet K., Tombette M., and Foudhil H., 2007. Technical Note: The air quality modeling system Polyphemus. Atmos. Chem. Phys., 7, 5479-5487.

Manders, A.M.M., Schaap, M., Hoogerbrugge, R., 2009. Testing the capability of the chemistry transport model LOTUS-EUROS to forecast PM10 levels in the Netherlands. Atmospheric Environment 43, 4050-4059.

Mårtensson E.M., Nilsson E.D., de Leeuw G., Cohen L.H and Hansson H-C, Laboratory simulations and parameterisation of the primary marine aerosol production, J. Geophys. Res., 108(D9), 4297, doi:10.1029/2002JD002263, 2003.

Mathur, R., Yu, S., Kang, D., Schere, K.L., 2008. Assessment of the wintertime performance of developmental particulate matter forecasts with the Eta-Community Multiscale Air Quality modeling system. Journal of Geophysical Research D: Atmospheres, 113, D02303.

McKeen, S., Chung, S.H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G.R., Mathur, R., Yu, S., 2007. Evaluation of several real-time PM2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study, J. Geophys. Res., 112, D10S20, doi:10.1029/2006JD007608, 20 pp.

Mebust, M.R., Eder, B.K., Binkowski, F.S., Roselle, S.J., 2003. Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component. 2. Model evaluation. Journal of Geophysical Research 108, D6, 4184, doi:10.1029/2001JD001410.

Monahan, E.C., Spiel, D.E., Davidson, K.L., 1986. In: Monahan, E.C., Mac Niocaill, G. (Eds.), Oceanic Whitecaps. Riedel, Norwell, MA, pp. 167–174.

Nenes, A., Pilinis, C., and Pandis, S., 1998. ISORROPIA: A new thermodynamic equilibrium model for multicomponent inorganic aerosols, Aquat. Geochem., 4, 123-152.

Nopmongcol, U., Koo, B., Tai, E., Jung, J., Piyachaturawat, P. Modeling Europe with CAMx for the Air Quality Model Evaluation International Initiative (AQMEII). Atmospheric Environment 10.1016/j.atmosenv.2011.11.023

Park, S. H., Gong, S. L., Gong, W., Makar, P. A., Moran, M. D., Zhang, J., Stroud, C. A., 2010. Relative impact of windblown dust versus anthropogenic fugitive dust in $PM_{2.5}$ on air quality in North America, J. Geophys. Res., 115, D16210, doi:10.1029/2009JD013144.

Philipona, R., Behrens, K., Ruckstuhl, C., 2009. How declining aerosols and rising greenhouse gases forced rapid warming in Europe since the 1980s. Geophys. Res. Lett. 36:doi:10.1029/2008GL036350

Pouliot, G., Pierce, T, Denier van der Gon, H., Schaap, M., Moran, M., Nopmongcol, U., Comparing Emissions Inventories and Model-Ready Emissions Datasets between Europe and North America for the AQMEII Project. Atmospheric Environment, in this issue

Pun, B., C. Seigneur and K. Lohman (2006) Modeling secondary organic aerosol via multiphase partitioning with molecular data, Environ. Sci. Technol., 40, 4722-4731.

Putaud, J.-P., Raes, F., Van Dingenen, Bruggemann, E., and et al., 2004. A European aerosol phenomenology – 2 : Chemical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe. Atmospheric Environment 38, 2579-2595.

Rao, S.T., Galmarini, S., Puckett, K., 2011. Air quality model evaluation international initiative (AQMEII). Bulletin of the American Meteorological Society 92, 23-30. DOI:10.1175/2010BAMS3069.1

Renner, E., Wolke, R., 2010. Modelling the formation and atmospheric transport of secondary inorganic aerosols with special attention to regions with high ammonia emissions. Atmospheric Environment 44(15), 1904–1912.

Sartelet K., Debry E., Fahey K., Roustan Y., Tombette M., Sportisse B., 2007. Simulation of aerosols and gas-phase species over Europe with the Polyphemus system. Part I: model-to-data comparison for 2001. Atmospheric Environment, 41 (29), 6116-6131

Sartelet K., Couvidat F., Seigneur C., Roustan Y., 2012. Impact of biogenic emissions on air quality over Europe and North America. Atmospheric Environment doi:10.1016/j.atmosenv.2011.10.046

Schaap, M., Timmermans, R.M.A., Roemer, M., Boersen, G.A.C., Builtjes, P.J.H. Sauter, F.J., Velders, G.J.M. and Beck, J.P. (2008) The LOTOS–EUROS model: description, validation and latest developments', Int. J. Environment and Pollution, Vol. 32, No. 2, pp.270–290.

Schaap, M., Timmermans, R.M.A., Sauter, F.J., Roemer, M., Velders, G.J.M., and et al. 2008. The LOTOS-EUROS model: description, validation and latest developments. Int. J. environ. Pollut. 32, 270-290.

Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A., 2001. Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. Journal of Geophysical Research 106, 28275e28293.

Schere, K., Flemming, J., Vautard, R., Chemel, C., et al. Trace Gas/Aerosol concentrations and their impacts on continental-scale AQMEII modelling sub-regions. Atmospheric Environment doi:10.1016/j.atmosenv.2011.09.043

Seinfeld, J., and S. Pandis, Atmospheric Chemistry and Physics, 1326 pp., Wiley, New York, 2006.

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

Simpson, D., Fagerli, H., Jonson, J. E., Tsyro, S., Wind, P., and Tuovinen J-P 1-8-2003, Transboundary Acidification, Eutrophication and Ground Level Ozone in Europe, PART I, Unified EMEP Model Description. pp. 104.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J.G., 2008. A description of the Advanced Research WRF Version 2. NCAR technical note NCAR/TND468+STR.

Smith, S. J., Pitcher, H., and Wigley, T.M.L., 2001. Global and Regional Anthropogenic Sulfur Dioxide Emissions. Global and Planetary Change, 29, pp 99-119

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

Sofiev, M.., 2000. A model for the evaluation of long-term airborne pollution transport at regional and continental scales, Atmos. Env., 34(15), pp. 2481-2493.

Sofiev, M., Siljamo, P., Valkama, I., Ilvonen, M., Kukkonen, J., 2006. A dispersion modeling system SILAM and its evaluation against ETEX data. Atmospheric Environment 40, 674-685

Sofiev, M., Soares, J., Prank, M., de Leeuw, G., Kukkonen, J., 2011. A regional-to-global model of emission and transport of sea salt particles in the atmosphere, submitted for publication in Journal of Geophysical Research – Atmospheres.

Solazzo, E., R. Bianconi, R. Vautard, K.W. Appel, M.D. Moran, C. Hogrefe, et al., 2012. Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. Atmospheric Environment doi:10.1016/j.atmosenv.2012.01.003

Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic,A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., Kerschbaumer,A., 2008. A model inter-comparison study focussing on episodes with elevated PM10 concentrations. Atmospheric Environment 42, 4567-4588

Stohl, A., Williams, E., Wotawa, G., Kromp-Kolb, H., 1996. A European inventory of soil nitric oxide emissions and the effect of these emissions on the photochemical formation of ozone, Atmospheric Environment 30, 3741-3755.

Strand, A., Hov, O., 1996. The impact of man-made and natural NOx emissions on upper tropospheric ozone: A two-dimensional model study. Atmospheric Environment 8, 1291-1303.

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

Vautard, R., Moran, M.D., Solazzo, E., Gilliam, R.C., Matthias, V., et al. 2012. Evaluation of the meteorological forcing used for AQMEII air quality simulations. Accepted for publication to Atmospheric Environment (this issue)

Vautard, R., Yiou, P., and G. J. van Oldenborgh, 2009. Decline of fog, mist and haze in Europe over the past 30 years, Nature Geoscience, 2, 115-119.

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

Visschedijk, A.J.H., Zandveld, P., Denier van der Gon, H.A.C.: A high resolution gridded European emission database for the EU integrated project GEMS, TNO report 2007-A-R0233/B, 2007

Wang, X., Zhang, L., Moran, M.D., 2010. Uncertainty assessment of current size-resolved parameterizations for below-cloud particle scavenging by rain. Atmos. Chem. Phys., 10, 5685-5705, doi:10.5194/acp-10-5685-2010.

Wesley, M.L., 1989. Parameterization of surface resistance to gaseous dry deposition in regional numerical models. Atmospheric Environment 16, 1293–1304.

Wolke, R., Knoth, O., Hellmuth, O., Schröder, W., Renner, E., 2004. The parallel model system LM-MUSCAT for chemistry-transport simulations: Coupling scheme, parallelization and applications. In: Parallel Computing, Elsevier, 363–370.

Yin, D., Nickovic, S., Barbaris, B., Candy, B., Sprigg, W.A., 2005. Modeling wind-blown desert dust in the southwestern United States for public health warning: A case study. Atmospheric Environment 39, 6243–6254.

Yiou, P., G.-J. Van Oldenborgh and R. Vautard, 2011. Trends of fog, mist and haze over the past 30 years and regional warming, Atmospheric Environment, submitted.

Yu, S., Bhave, P.V., Dennis, R.L., Mathur, R., 2007. Seasonal and regional variations of primary and secondary organic aerosols over the continental United States: Semi-empirical estimates and model evaluation. Environmental Science and Technology, 41, 4690-4697.

Yu, S., Mathur, R., Schere, K., Kang, D., Pleim, J., Young, J., Tong, D., Pouliot, G., McKeen, S.A., Rao, S.T., 2008. Evaluation of real-time PM2.5 forecasts and process analysis for PM2.5 formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study. Journal of Geophysical Research, 113, D06204.

Zhang, L., Gong, S., Padro, J., Barrie, L., 2001. A size-segregated particle dry deposition scheme for an atmospheric aerosol module. Atmospheric Environment 549–560.

Zhang, L., Moran, M.D., Makar, P.A., Brook, J.R., Gong, S., 2002. Modelling gaseous dry deposition in AURAMS: a unified regional air-quality modeling system. Atmospheric Environment 36, 537-560

Captions

Figures

Figure 1. Continental maps of *(a)* Europe and *(b)* North America showing PM receptor locations by sub-regions (stars, circles, squares for sub-regions 1, 2, and 3, respectively). Solid and empty symbols indicate daily PM_{10} and hourly PM_{10}, respectively. Symbols with crosses indicate daily and/or hourly $PM_{2.5}$.

Figure 2. Quarterly accumulated emissions of five species for the *(a)* EU and *(b)* NA continental domains.

Figure 3. Quarterly accumulated *(a)* dry and *(b)* wet deposition amounts for the EU continental domain and quarterly accumulated *(c)* dry and *(d)* wet deposition amounts for the NA continental domain.

Figure 4. Time series of monthly average daily PM_{10} concentrations for the EU (left column) and NA (right column) domains (top row) and sub-regions 1 to 3 (second to fourth rows). Monthly average observed values are represented by the filled diamonds.

Figure 5. Hourly PM_{10} concentrations averaged by hour showing the daily PM_{10} cycle for the EU (left column) and NA (right column) domains (top row) and sub-regions 1 to 3 (second to fourth rows). Hourly average observed values are represented by the filled diamonds. Model colours as in the legend of Fig. 4.

Figure 6. MFB versus MFE for daily PM_{10} concentrations in EU (top row) and NA (bottom row) by sub-region at urban (left column) and rural (right column) sites (no time or space averaging of data). The number of monitoring stations is indicated in the legend for each sub-region. Model colour as in the legend of Fig. 4.

Figure 7. MFB versus MFE for hourly concentrations of NO_2 (left column) and SO_2 (right column) for the EU (top row) and NA domains (bottom row) by sub-region at rural sites (no time or space averaging of data). The number of monitoring stations is indicated in the legend for each sub-region. Model colours as in the legend of Fig. 4.

Figure 8. PM_{10} concentration MFB vs WS MFB for the months of August and November for EU (top panels) and NA (lower panels) (triangles, circles, and squares for sub-regions 1, 2, and 3, respectively). The best-fit lines for each sub-region are shown with the PCC in the legend. Model colours as in the legend of Fig. 4.

Figure 9. Time series of monthly average daily $PM_{2.5}$ for the EU (left column) and NA (right column) domains (top row) and sub-regions 1 to 3 (second to fourth rows).

Figure 10. Taylor plots of $PM_{2.5}$ for the EU sub-regions (triangles, circles, and squares for sub-regions 1, 2, and 3, respectively) at *a)* urban and *b)* rural sites. Horizontal and vertical axes indicate the variance ($\mu g\ m^{-3}$) and the arcs represent the curves of ideal variance for each sub-region. The straight lines indicate the values of the PCC.

Figure 11. Taylor plots of $PM_{2.5}$ for 2006 for the NA sub-regions (triangles, circles, and squares for sub-regions 1, 2, and 3, respectively) at *a)* urban and *b)* rural sites. Horizontal and vertical axes indicate the variance ($\mu g\ m^{-3}$) and the arcs represent the curves of ideal variance for each sub-region. The straight lines indicate PCC values.

Figure 12. Taylor plots of the $PM_{2.5}$ speciated components *a)* NH_4 *b)* SO_4 *c)* NO_3 and *d)* EC for 2006 for the NA sub-regions (triangles, circles, and squares for sub-regions 1, 2, and 3 respectively). Horizontal and vertical axes indicate the variance ($\mu g\ m^{-3}$) and the arcs represent the curves of ideal variance for each sub-region. The straight lines indicate PCC values.

Figure 13. Time series of daily average $PM_{2.5}$ concentrations at seven receptor sites in central EU for 13 – 28 April 2006. Observed concentrations are indicated by the solid gray diamonds and model-predicted concentrations by the colored lines.

Figure 14. Time series of hourly average $PM_{2.5}$ concentrations at 18 receptor sites in the eastern U.S. for 14 – 29 July 2006. Observed concentrations are indicated by the gray line and model-predicted concentrations by the colored lines.

Tables

Table 1. Summary of model main modules, resolution and data providers.

Table 2. Statistical analysis for PM_{10} *a)* EU and *b)* NA – whole continents (RMSE: root mean square error; PCC: Pearson correlation coefficient; Std Dev: standard deviation).

Table 3. $PM_{2.5}/PM_{10}$ mean, standard deviation and Pearson correlation coefficient based on daily data for the continent-wide domain of *a)* EU and *b)* NA.

Table 1

Model	Aerosol size	Inorganic module	Organic module	Sea salt module	Wet dep.	Dry dep.	Emi	BC	Horizontal and vertical resolution	Other
WRF (NA) MM5 (EU) Chimere	40 nm to 10 µm (8 bins)	ISORROPIA[2]	Pun et al. (2006); SOA isoprene chemistry	Monahan et al. (1986)	In cloud and sub cloud scavenging for gases and particles	Resistive (Seinfeld and Pandis, 2006)	standard	Standard LMDZ-INCA	25 km; 9 levels from 20m to 500 hPa)	Biogenic emissions from MEGAN[6] model
MM5 Polyphemus	0.01µm to 10µm (5 bins)	ISORROPIA[2]	SuperSorgam (Kim et al., 2011)	Monahan et al. (1986)	Washout coefficient (with parameters from Loosmore and Cederwall, 2004)	Zhang et al. (2001)	standard[1,3]	standard	24 km; 9 levels from surface to 1200	
WRF (NA) MM5 (EU) CAMx	$PM_{2.5}$ $PM_{2.5-10}$	ISORROPIA[2]	SOAP semi-volatile scheme with aging	Emission input (Environ, 2010)	CAMx rainout and washout	Zhang et al. (2001) scheme for gases and particles	standard	standard	15 (12) km NA (EU); 26 (23) first: 40 (30) m NA (EU)	CB05 Chemical Mechanism. Biogenic emissions from MEGAN[6] model
COSMO Muscat	$PM_{2.5}$ $PM_{2.5-10}$	Renner and Wolke (2010)	SORGAM (Schell et al., 2001)	Guelle et al. (2003)	Berge (1997)	Resistive (Seinfeld and Pandis, 2006)	standard[1,4]	standard	24 km; 22 layers (first: 60 m)	Wolke et al. (2004)

[1] Biogenic emission based on land-use and meteorology (Guenther et al., 1994; Simpson et al., 1995).
[2] Nenes et al. (1998).
[3] Emissions Includes: Biomass burning; Biogenic organic compounds of SOA:a-pinene, limonene, sesquiterpene, hydrophilic isoprene.
[4] Fire from FMI; Biogenic as in Guenther et al. (1993) and Stohl et al. (1996).

ECMWF SILAM	$PM_{2.5}$ and PM_{10}; sea salt – 5 bins from 0.01 to 30 mkm	Updated DMAT model (Sofiev, 2000), includes SO_4, NO_3, NH_4	None	SILAM native scheme (Sofiev et al., 2011)	Scavenging coefficient parameterization	Resistive (Seinfeld and Pandis, 2006)	standard	Standard; own boundaries for sea salt	24 km; 9 layers with varying thickness; top: 10 km	In-house biogenic emission inventory
MM5 DEHM	2 size bins (1-2.5, 6-10 µm diameter). No particle dynamics.	Modified version of Strand and Hov (1994)	similar to the EMEP model (Simpson et al., 2003).	1 size bin mean 6 µm diameter	In cloud, below cloud scavenging coefficients (Christensen, 1997)	Resistive (Seinfeld and Pandis, 2006)	Several sources[5]	Nested in a hemispheric model	50 km; 29 layers, top: 100 hPa	Two-way nesting between hemispheric and NA/EU domains
WRF CMAQ	Trimodal size distribution (Binkowski and Roselle, 2003)	ISORROPIA[2]	Foley et al. (2010)	Kelly et al. (2010)	Derived from that of RADM (see Byun and Schere, 2006)	Resistance law concept (Byun and Schere, 2006)	EU: standard anthropogenic[1] NA: standard	standard	12 km; 34 layers, first: 25 m, top: 50hPa	
ECMWF LOTOS-EUROS	1 size bin (0-2.5um); also 2.5-10um for sea-salt	ISORROPIA[1]	None	Monahan et al. (1986); Martensson et al. (2003)	Simpson et al. (2003)	Resistance approach; aerosol resistances following EDAC (Erisman et al., 1994)	standard[1]	Standard	25 km; 4 layers (surface, boundary layer, 2 residual layers)	Forest fire emissions using data from FMI/Silam (Sofiev, 2011) Model description in (Schaap et al., 2008)
GEM AURAMS	0.01 µm to 41 µm (diameter; 12 bins)	HETV (Makar et al., 2003) based on ISORROPIA[2]	Jiang (2003); 7 lumped VOC precursor species	Gong et al. (1997)	In-cloud and below-cloud scavenging for gases and particles (Gong et al., 2006)	Zhang et al. (2001)	quasi-standard (same inventories, own processing); BEIS v3.09	Climatological profiles	45 km; 29 levels (first layer: 15m; 14 in the first 2km)	Explicit treatment of nucleation, condensation, coagulation, swelling, activation;

[5] Global: IPCC RPC 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).

[6] Model of emission of Gases and Aerosol from Nature (Guenther et al., 2006).

										ADOM-2 gas-phase mechanism
COSMO -CLM CMAQ	Trimodal size distribution (Binkowski and Roselle, 2003)	ISORROPIA[2]	SORGAM (Schell et al., 2001)	Wind speed driven (Monahan et al., 1986)	Byun and Schere, (2006)	Resistive (Seinfeld and Pandis, 2006)	Standard[1]	standard	24 km; 30 layers, first 36 m, top: 100hPa	CBIV gas phase chemistry for Europe, and CB05 for NA. BEIS v3 model to calculate biogenic emissions
WRF WFR/Chem	3 modes (MADE, Ackermann et al., 1998)	Based on MARS and SCAPE (Binkowsi and Shankar, 1995)	SORGAM (Schell et al., 2001)	Wind speed driven (GOCART model, Chin et al., 2002)	In cloud and sub cloud scavenging for gases and particles	Resistance approach (Wesely, 1989)	Standard[1]	Standard WRF/Chem BCs	22.5km 36 layers	Online coupled with aerosol radiation feedback and indirect cloud effect. No fire emission

Table 2a

		RMSE ($\mu g\ m^{-3}$)	PCC*	Mean ($\mu g\ m^{-3}$)	Std Dev ($\mu g\ m^{-3}$)
EU PM_{10} daily averaged hourly data	**Mod1**	7.3	0.4	22.9	5.4
	Mod2	14.2	0.5	8.9	2.8
	Mod3	11.2	0.5	12. 5	5.3
	Mod4	15.2	0.7	7.5	2.2
	Mod5	10.8	0.4	14.4	7.7
	Mod6	9.8	0.2	22.3	6.6
	Mod7	9.8	0.7	13.2	4.6
	Mod8	14.5	0.7	8.0	3.2
	Mod9	13.9	0.4	9.3	3.9
	Mod10	7.7	0.5	17.7	5.3
	Observation			21.5	7.3

***** Statistically significant at >99% confidence level

Table 2b

		RMSE ($\mu g\ m^{-3}$)	PCC*	Mean ($\mu g\ m^{-3}$)	Std Dev ($\mu g\ m^{-3}$)
NA PM_{10} daily averaged hourly data	**Mod12**	20.9	0.3	10.5	2.0
	Mod13	25.6	0.3	5.7	1.6
	Mod14	17.2	0.2	14.5	3.0
	Mod15	17.5	0.2	14.2	3.2
	Mod16	19.9	0.4	11.5	2.2
	Mod17	5.3	0.15	20.1	2.9
	Observation			30.7	5.4

***** Statistically significant at >99% confidence level

Table 3a

		PCC	Mean	Std Dev
EU Daily $PM_{10}/PM_{2.5}$ (N=365)	**Mod1**	<0.10^{+}	1.80	0.63
	Mod2	0.20*	1.04	0.03
	Mod3	<0.10^{+}	1.22	0.15
	Mod4	<0.10^{+}	1.11	0.06
	Mod5	<0.10^{+}	1.10	0.04
	Mod6	0.15*	1.29	0.13
	Mod7	0.54*	1.37	0.17
	Mod8	0.46*	1.16	0.07
	Mod9	0.23*	1.36	0.14
	Mod10	<0.10^{+}	1.08	0.03
	Observation		1.54	0.26

*correlation significance at >99% confidence level
$^{+}$correlation significance at >10% confidence level

Table 3b

		PCC	Mean	Std Dev
NA Daily $PM_{10}/PM_{2.5}$ (N=365)	**Mod12**	0.34*	1.67	0.17
	Mod13	<0.10^{+}	1.06	0.03
	Mod14	<0.10^{+}	1.79	0.15
	Mod15	0.19*	1.92	0.14
	MOd16	<0.10^{+}	1.15	0.06
	Mod17	0.24*	1.40	0.13
	Mod18	0.21*	1.20	0.08
	Observation		2.05	0.27

*correlation significance at >99% confidence level
$^{+}$correlation significance at >10% confidence level

Figures

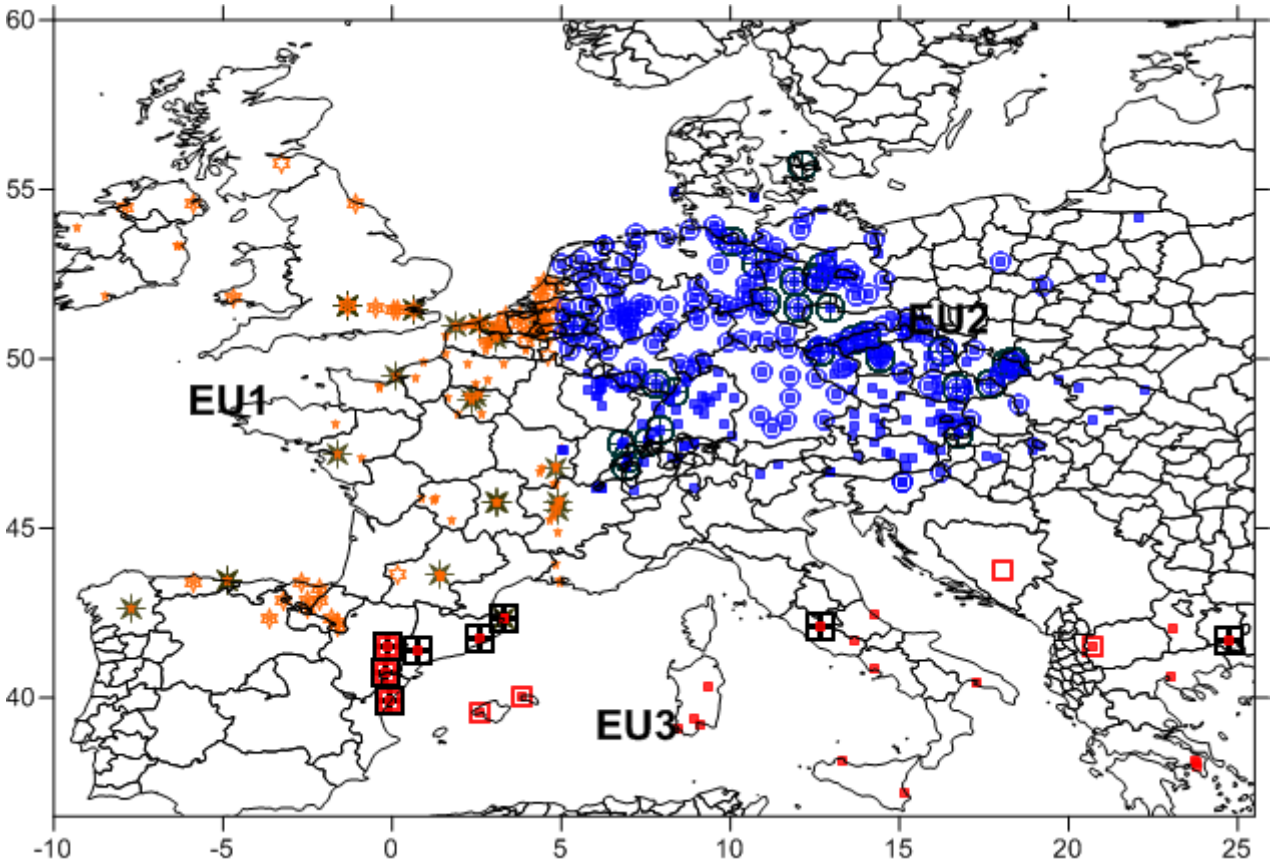

Figure 1a.

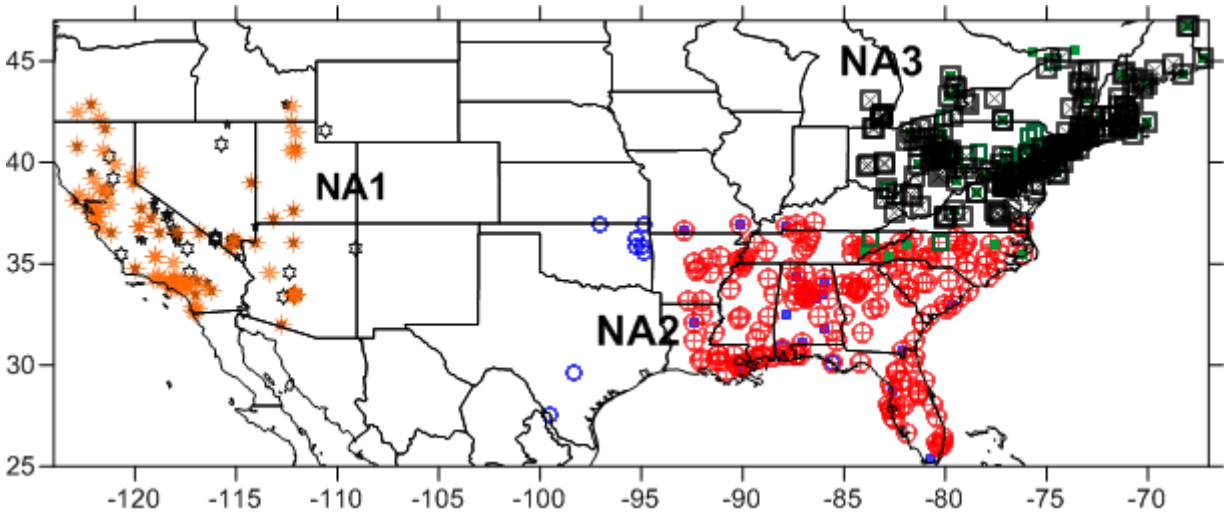

Figure 1b.

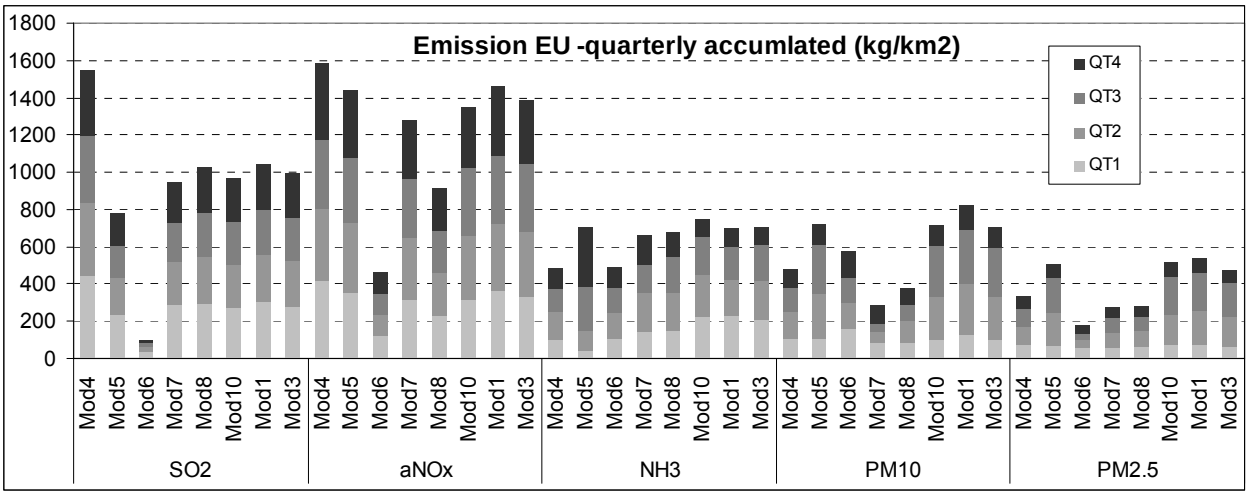

Figure 2a

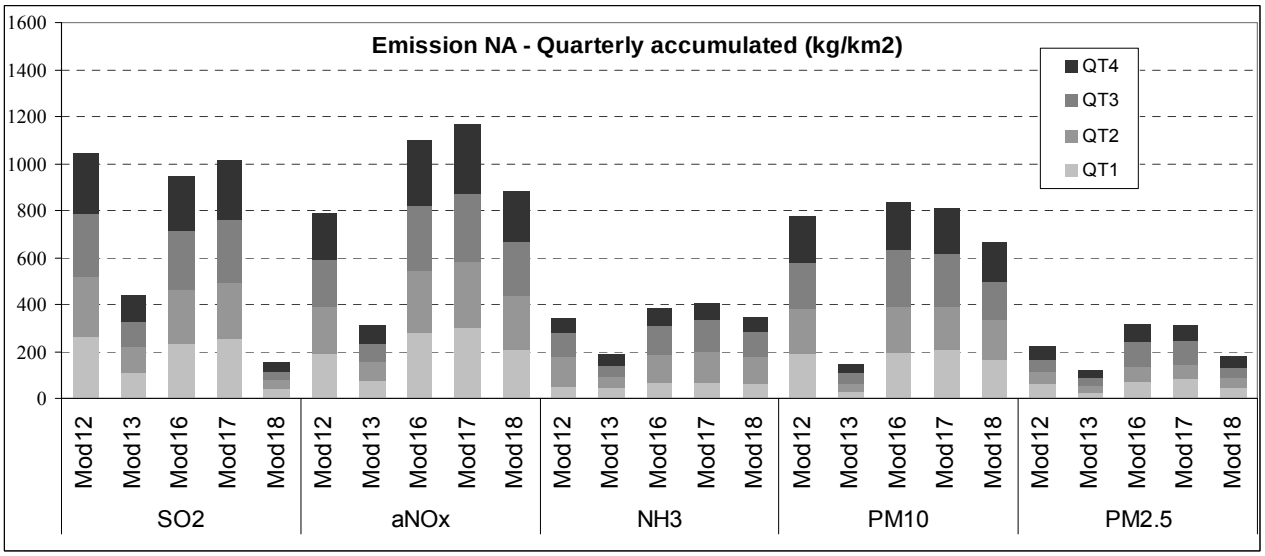

Figure 2b

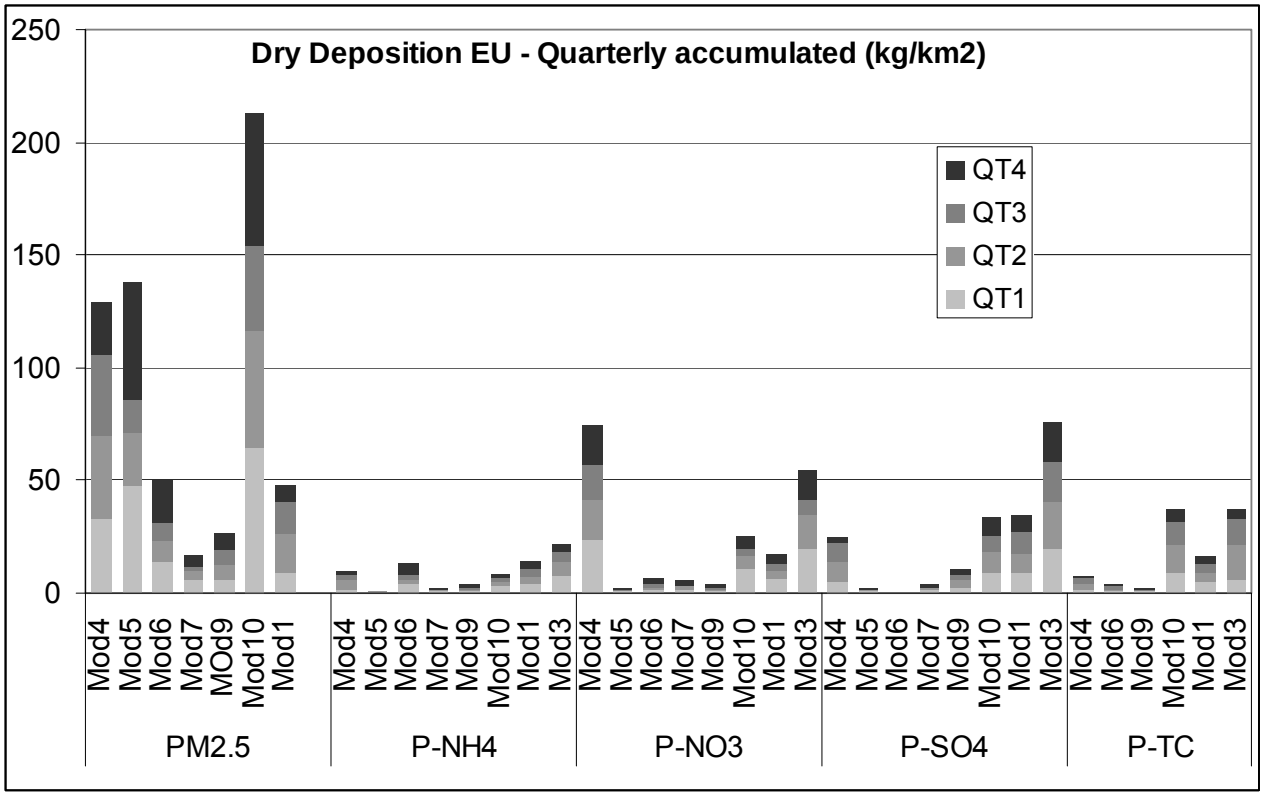

Figure 3a

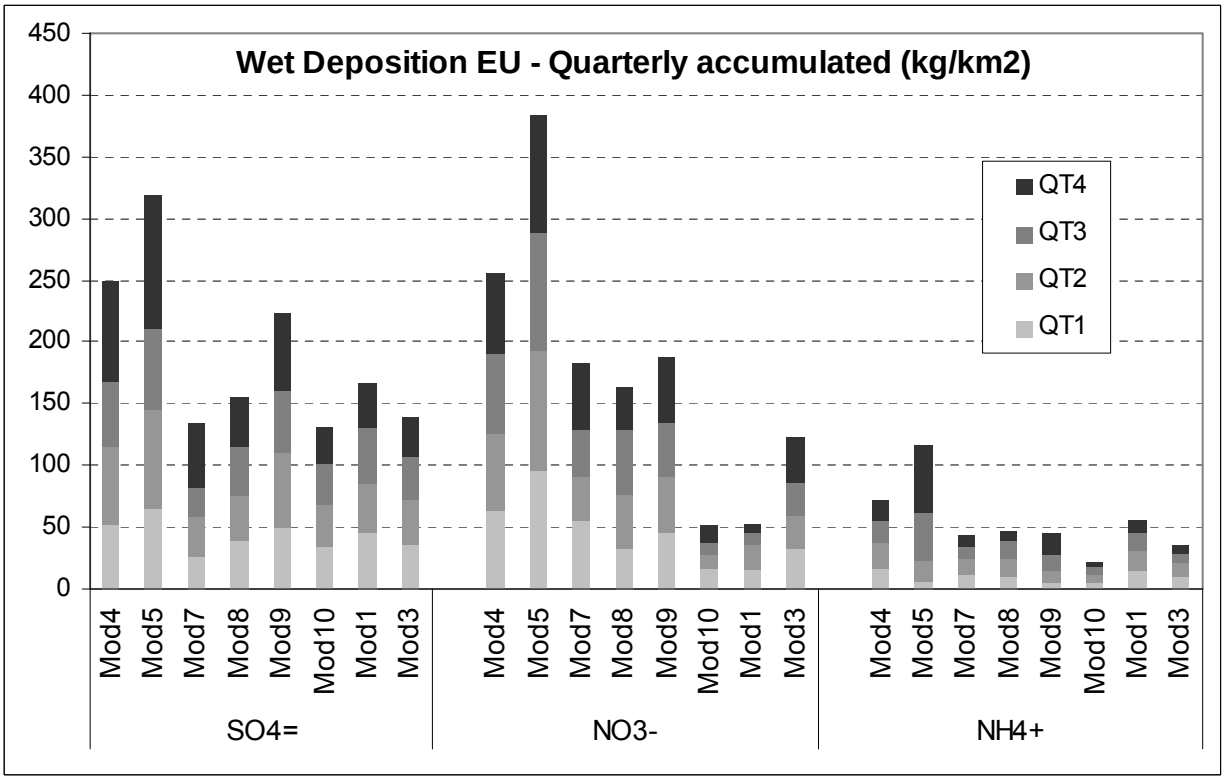

Figure 3b

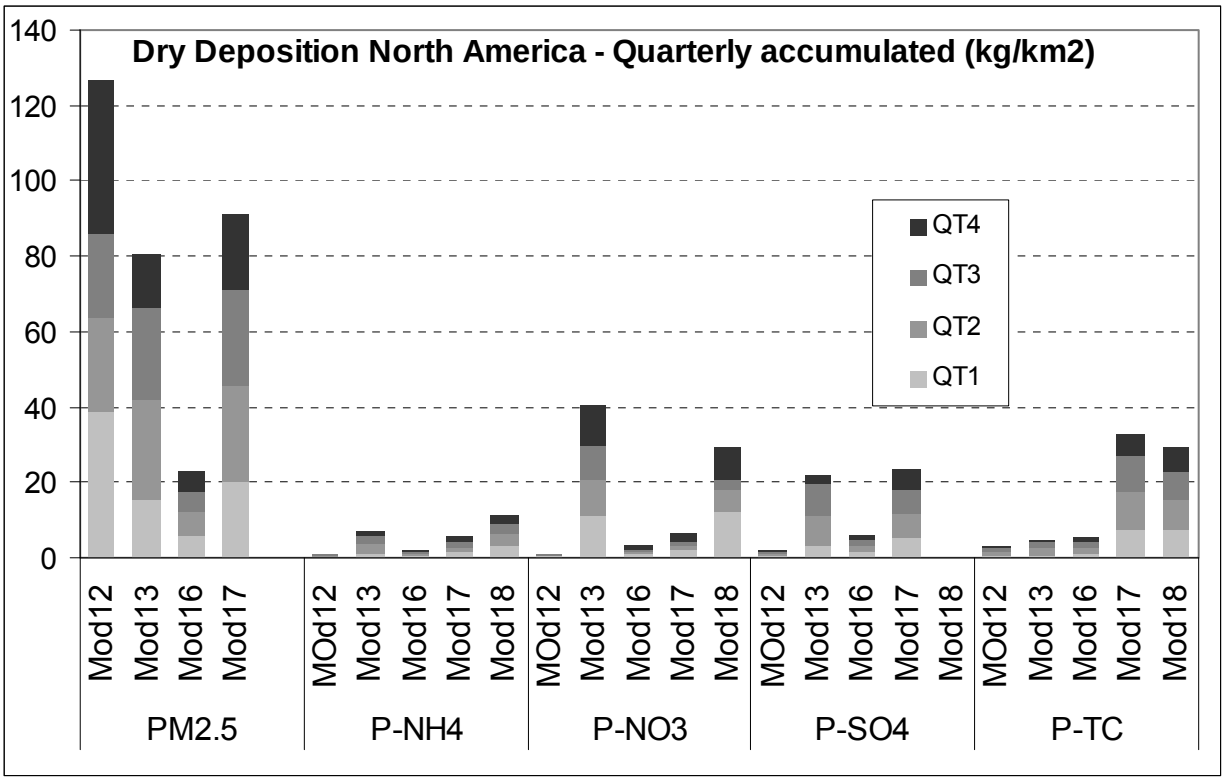

Figure 3c

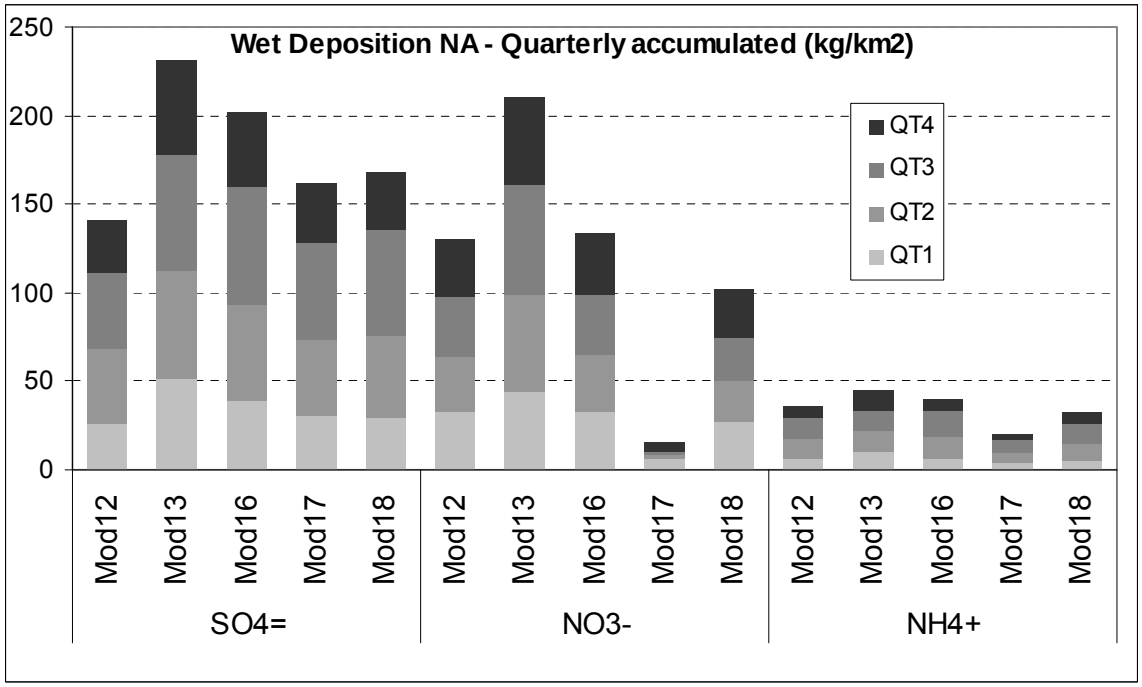

Figure 3d

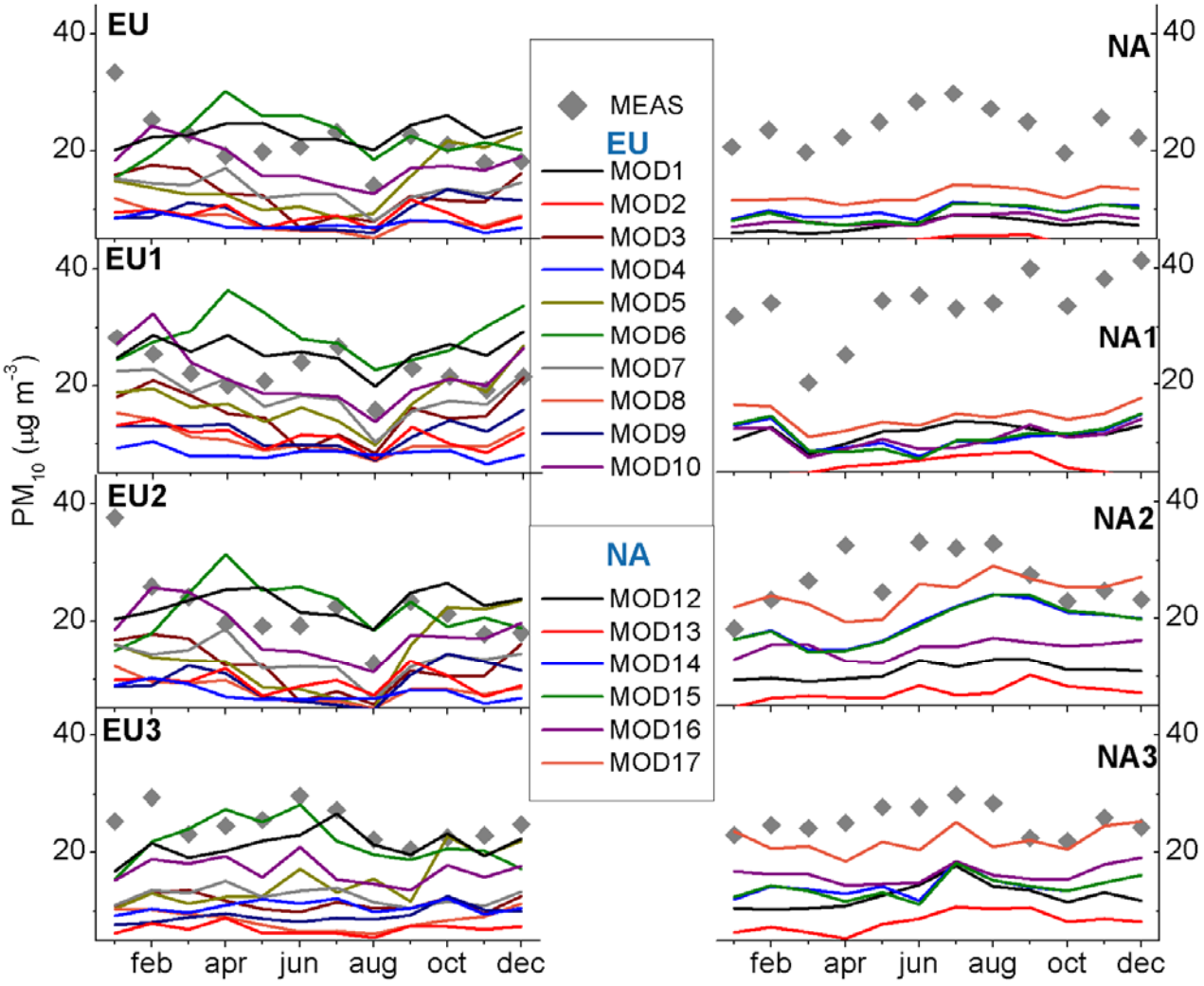
EU
EU1
EU2
EU3
NA
NA1
NA2
NA3
40
20
PM10 (µg m-3)
MEAS
EU
MOD1
MOD2
MOD3
MOD4
MOD5
MOD6
MOD7
MOD8
MOD9
MOD10
NA
MOD12
MOD13
MOD14
MOD15
MOD16
MOD17
feb
apr
jun
aug
oct
dec

Figure 4

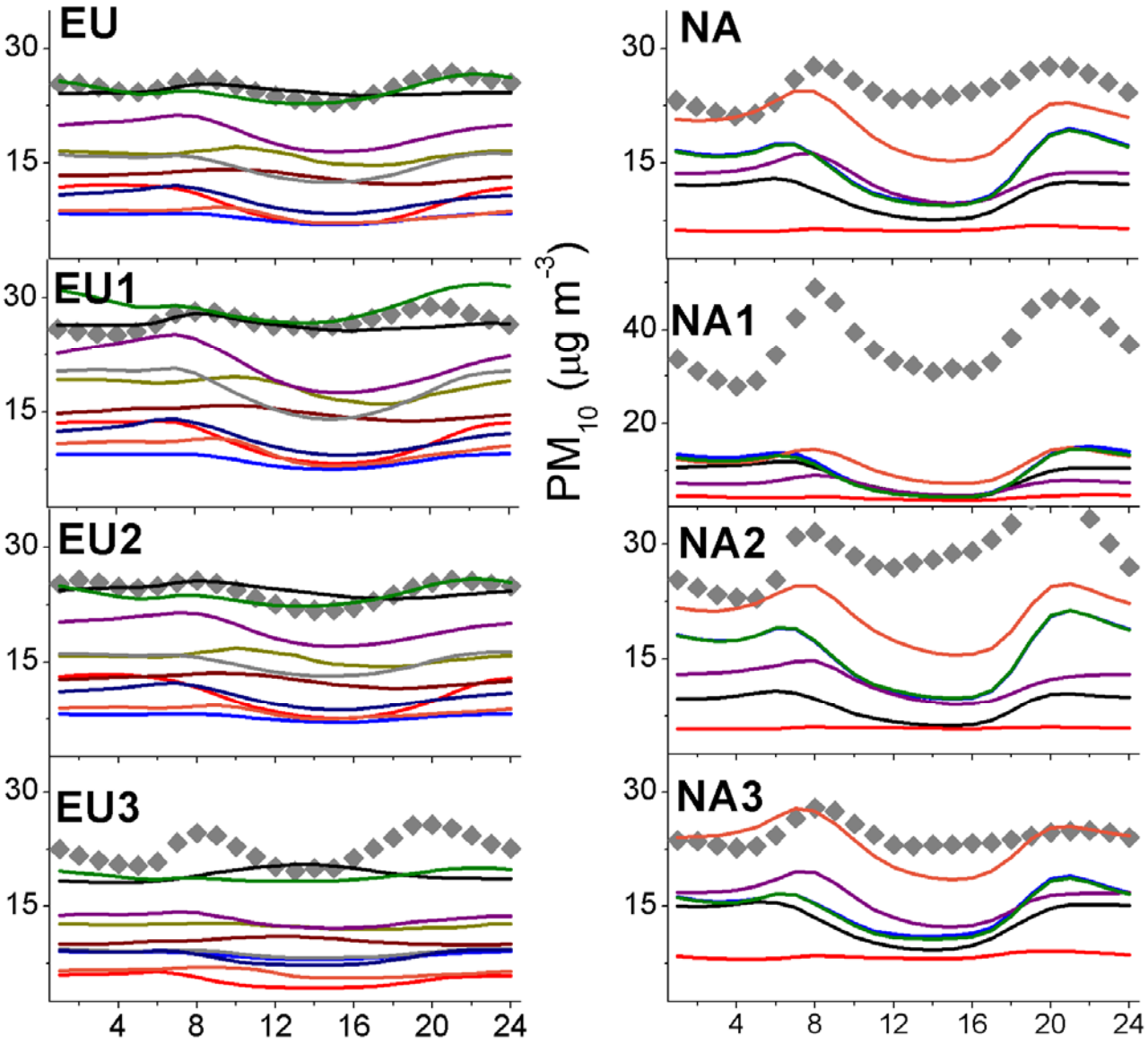

Figure 5

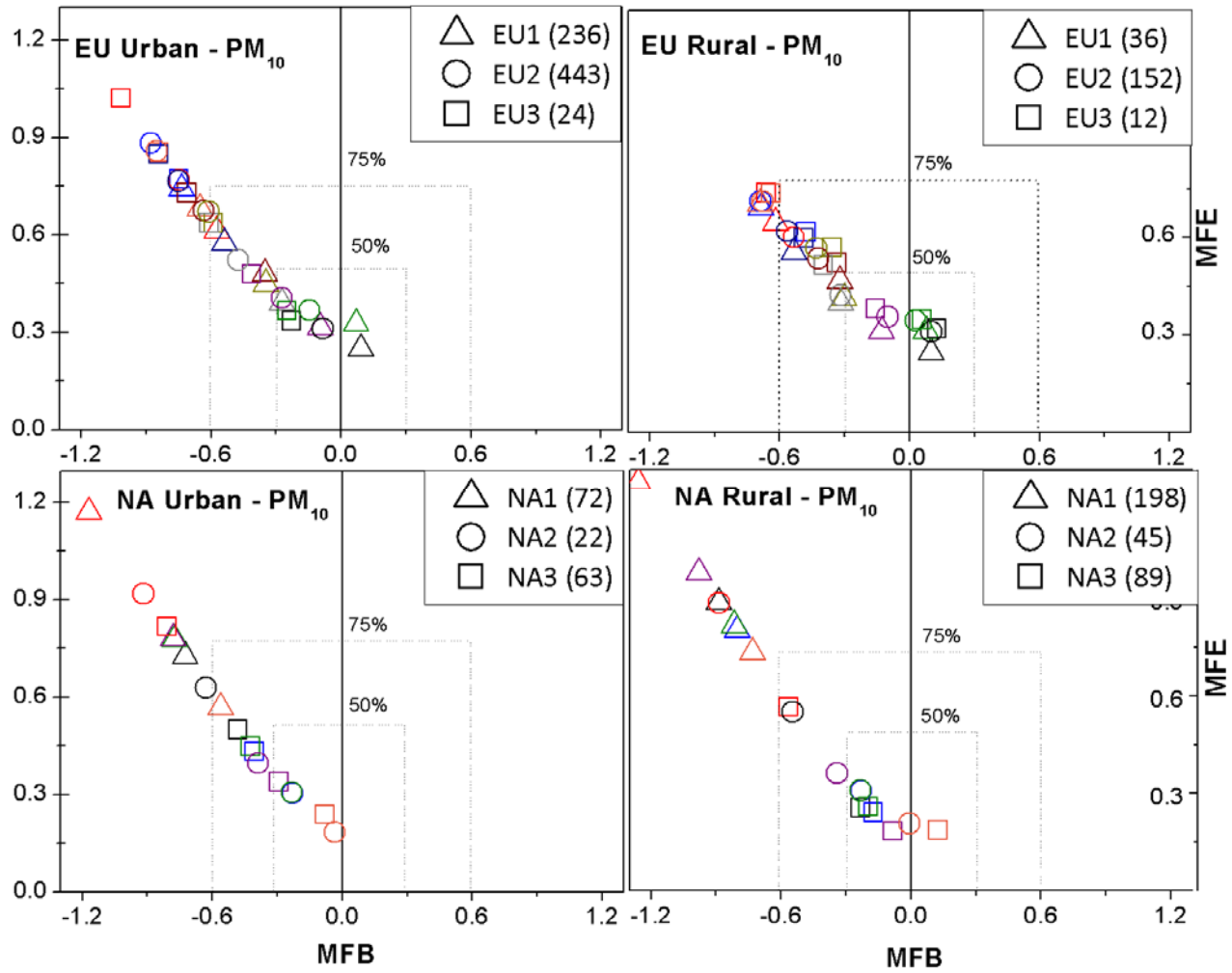
EU Urban - PM10
EU1 (236)
EU2 (443)
EU3 (24)
EU Rural - PM10
EU1 (36)
EU2 (152)
EU3 (12)
NA Urban - PM10
NA1 (72)
NA2 (22)
NA3 (63)
NA Rural - PM10
NA1 (198)
NA2 (45)
NA3 (89)
75%
50%
MFB
MFE

Figure 6

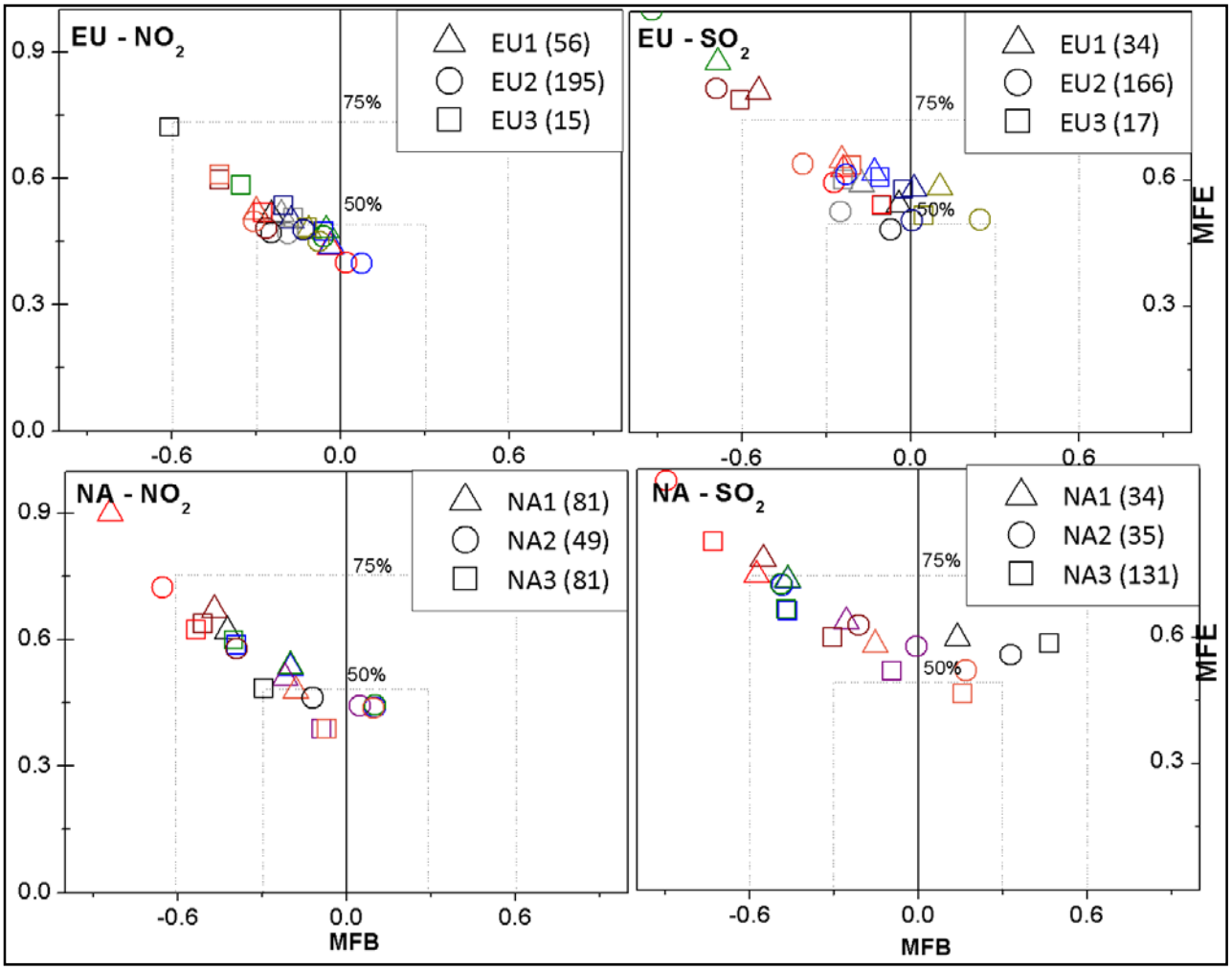
EU - NO2
EU1 (56)
EU2 (195)
EU3 (15)
EU - SO2
EU1 (34)
EU2 (166)
EU3 (17)
NA - NO2
NA1 (81)
NA2 (49)
NA3 (81)
NA - SO2
NA1 (34)
NA2 (35)
NA3 (131)
75%
50%
MFB
MFE
0.0
0.3
0.6
0.9
-0.6
0.0
0.6

Figure 7.

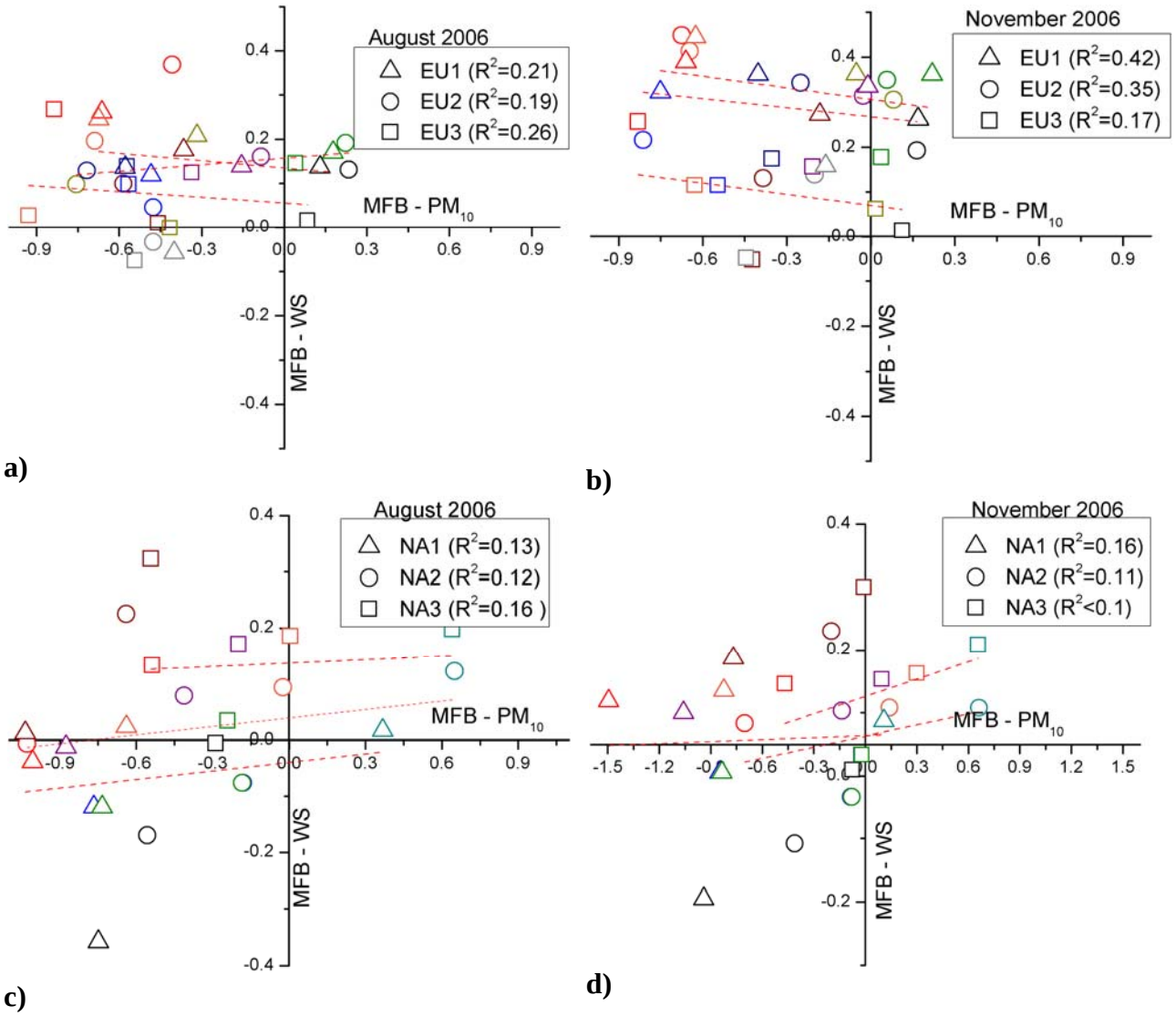
August 2006
EU1 (R^2=0.21)
EU2 (R^2=0.19)
EU3 (R^2=0.26)
MFB - PM_{10}
MFB - WS
a)
November 2006
EU1 (R^2=0.42)
EU2 (R^2=0.35)
EU3 (R^2=0.17)
MFB - PM_{10}
MFB - WS
b)
August 2006
NA1 (R^2=0.13)
NA2 (R^2=0.12)
NA3 (R^2=0.16)
MFB - PM_{10}
MFB - WS
c)
November 2006
NA1 (R^2=0.16)
NA2 (R^2=0.11)
NA3 (R^2<0.1)
MFB - PM_{10}
MFB - WS
d)

Figure 8

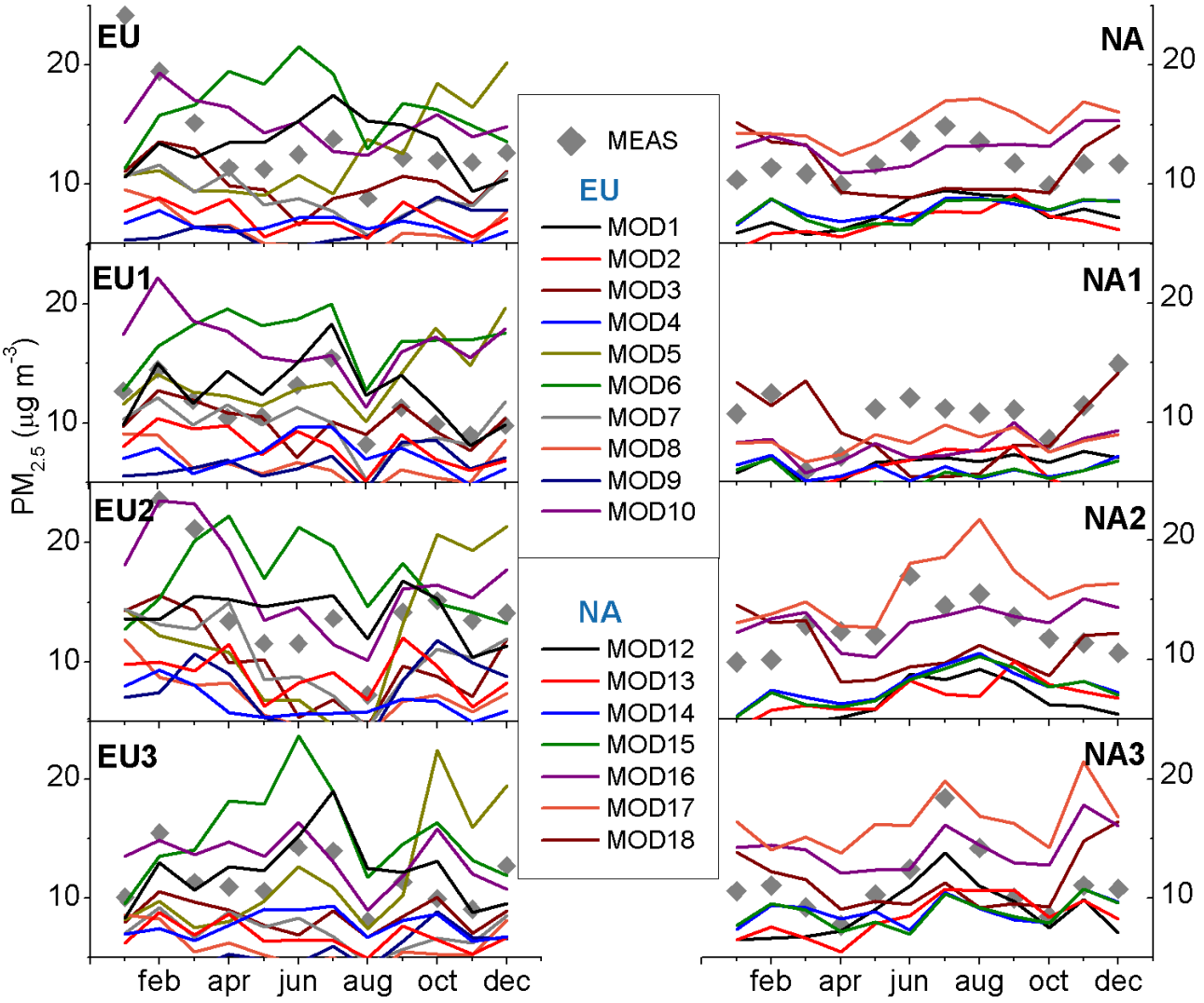
EU
EU1
EU2
EU3
NA
NA1
NA2
NA3
20
10
PM2.5 (µg m-3)
feb
apr
jun
aug
oct
dec
MEAS
EU
MOD1
MOD2
MOD3
MOD4
MOD5
MOD6
MOD7
MOD8
MOD9
MOD10
NA
MOD12
MOD13
MOD14
MOD15
MOD16
MOD17
MOD18

Figure 9.

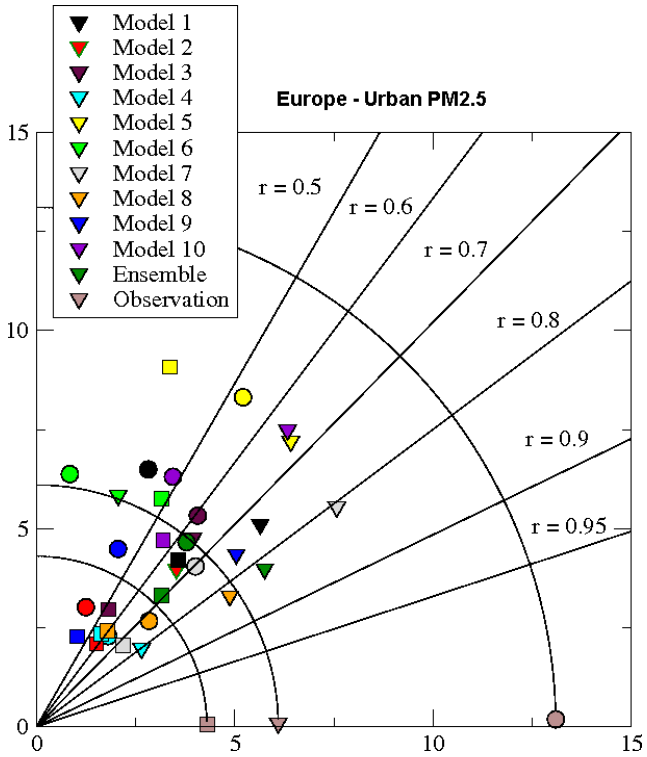

Figure 10a

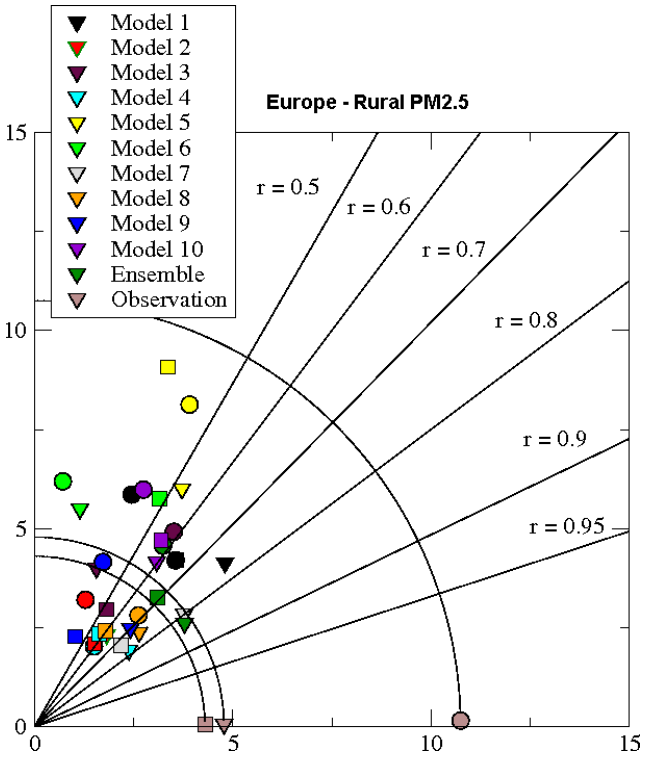

Figure 10b

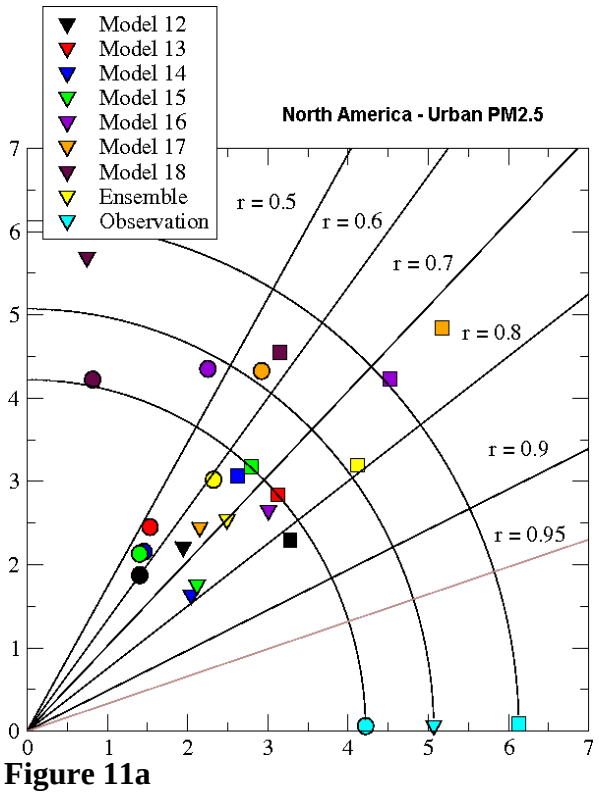

Figure 11a

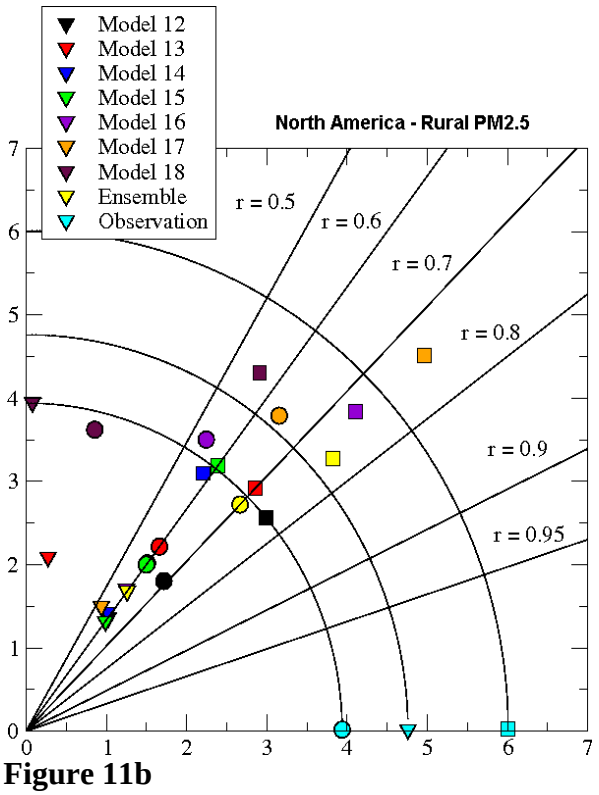

Figure 11b

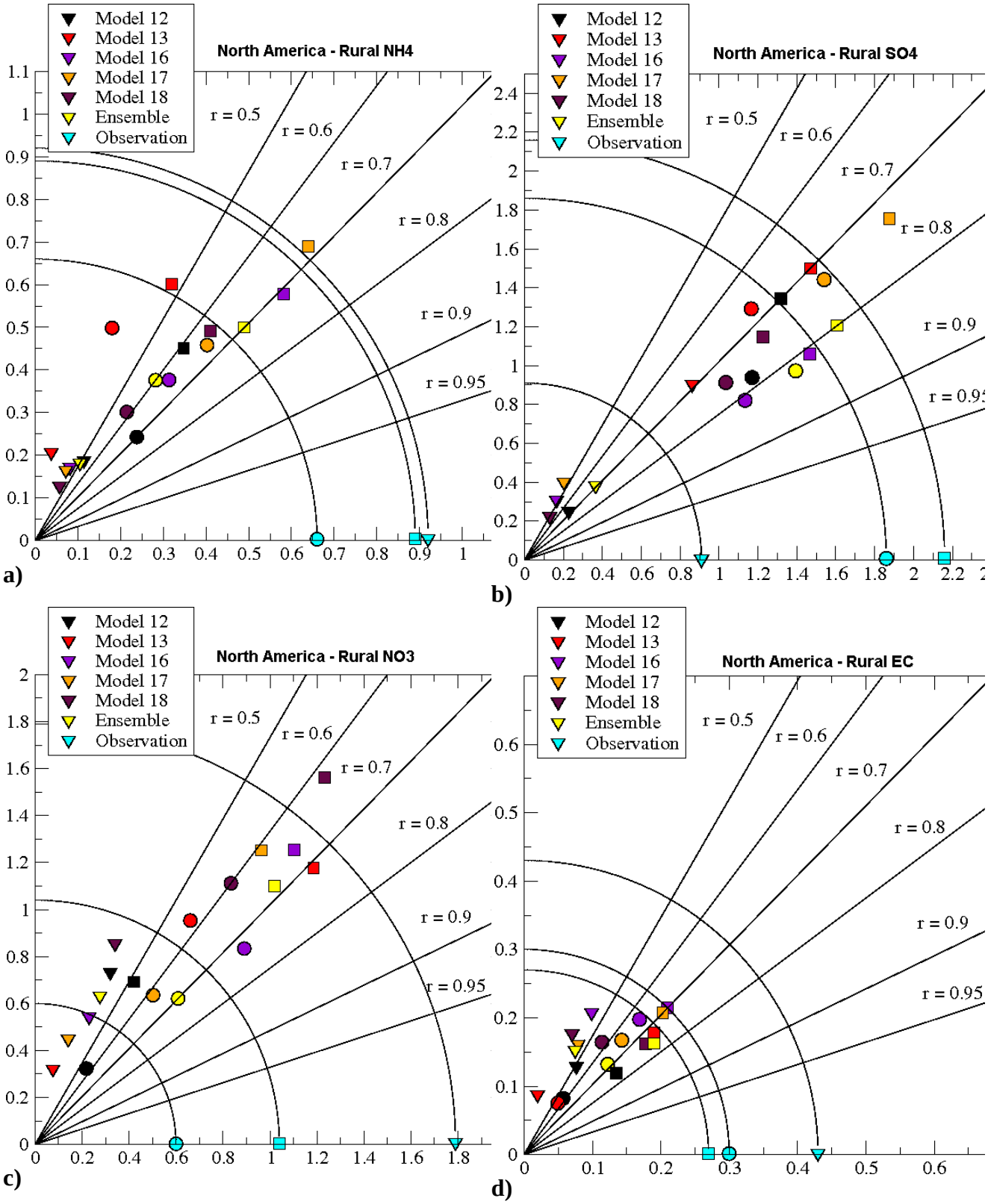
Model 12
Model 13
Model 16
Model 17
Model 18
Ensemble
Observation
North America - Rural NH4
r = 0.5
r = 0.6
r = 0.7
r = 0.8
r = 0.9
r = 0.95
a)
North America - Rural SO4
b)
North America - Rural NO3
c)
North America - Rural EC
d)

Figure 12

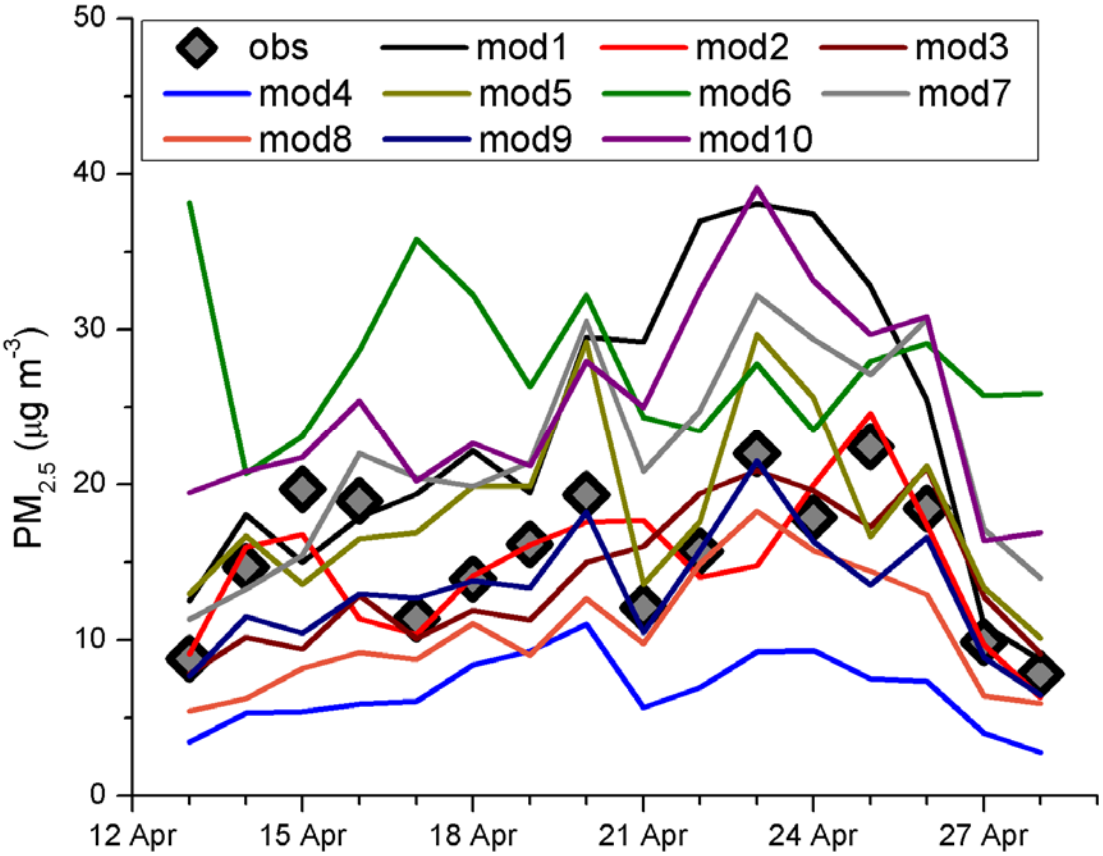
obs
mod1
mod2
mod3
mod4
mod5
mod6
mod7
mod8
mod9
mod10
PM2.5 (μg m-3)
0
10
20
30
40
50
12 Apr
15 Apr
18 Apr
21 Apr
24 Apr
27 Apr

Figure 13

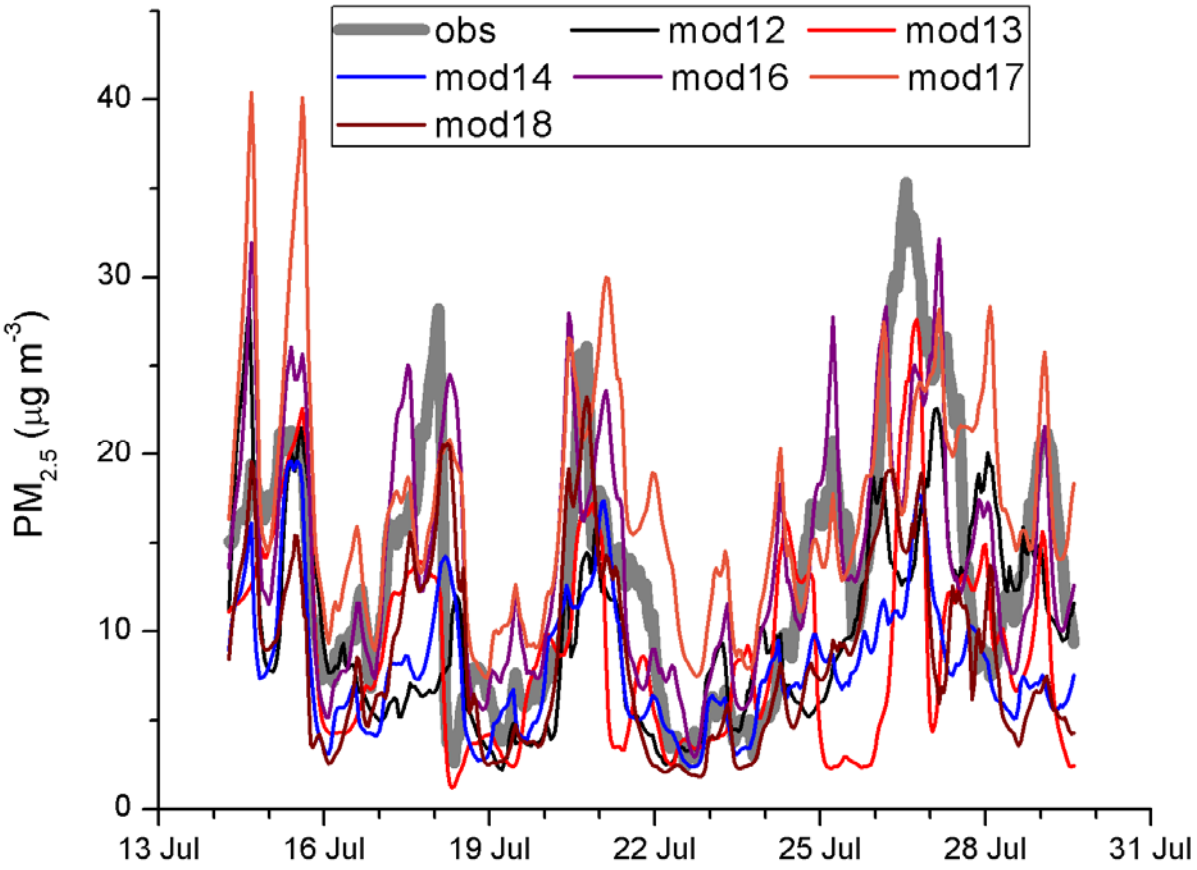
obs
mod12
mod13
mod14
mod16
mod17
mod18
PM2.5 (μg m-3)
0
10
20
30
40
13 Jul
16 Jul
19 Jul
22 Jul
25 Jul
28 Jul
31 Jul

Figure 14