Evaluation of the meteorological forcing used for the Air Quality Model Evaluation International Initiative (AQMEII) air quality simulations

Robert Vautard (1), Michael D. Moran (2), Efisio Solazzo (3), Robert C. Gilliam (4), Volker Matthias (5), Roberto Bianconi (6), Charles Chemel (7), Joana Ferreira (8), Beate Geyer (9), Ayoe B. Hansen (10), Amela Jericevic (11), Marje Prank (12), Arjo Segers (13), Jeremy D. Silver (10), Johannes Werhahn (14), Ralf Wolke (15), S.T. Rao (4), and Stefano Galmarini (3)

(1) Laboratoire des Sciences du Climat et de l'Environnement, IPSL Laboratoire CEA/CNRS/UVSQ, Orme des Merisiers, F-91191 Gif/Yvette CEDEX, France
(2) Air Quality Research Division, Science and Technology Branch, Environment Canada, 4905 Dufferin Street, Toronto, Ontario, M3H 5T4, Canada
(3) European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Italy
(4) Atmospheric Modeling and Analysis Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency Research, Triangle Park, NC 27711 USA
(5) Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Germany
(6) Enviroware srl, via Dante 142, I-20863 Concorezzo (MB), Italy
(7) National Centre for Atmospheric Science (NCAS), Centre for Atmospheric & Instrumentation Research (CAIR), University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, UK
(8) CESAM & Department of Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal
(9) Institute of Coastal Research Helmholtz-Zentrum Geesthacht
(10) Department of Atmospheric Environment, National Environmental Research Institute, Aarhus University, Denmark
(11) Meteorological and Hydrological Service in Croatia, Zagreb, Croatia
(12) Finnish Meteorological Institute, Helsinki, Finland
(13) Netherlands Organisation for applied scientific research (TNO), Earth Environmental and Life Sciences, Princetonlaan 6, NL-3584CB Utrecht, The Netherlands
(14) Institute for Meteorology and Climate Research Atmospheric Environmental Research, Garmisch Partenkirchen
(15) Leibniz Institute for Tropospheric Research, Permoserstr. 15, 04318 Leipzig, Germany

29 October 2011, revised version submitted to *Atmospheric Environment (Special issue AQMEII)*

Abstract

Accurate regional air pollution simulation relies strongly on the accuracy of the mesoscale meteorological simulation used to drive the air quality model. The framework of the Air Quality Model Evaluation International Initiative (AQMEII), which involved a large international community of modeling groups in Europe and North America, offered a unique opportunity to evaluate the skill of mesoscale meteorological models for two continents for the same period. More than 20 groups worldwide participated in AQMEII, using several meteorological and chemical transport models with different configurations. The evaluation has been performed over a full year (2006) for both continents. The focus for this particular evaluation was meteorological parameters relevant to air quality processes such as transport and mixing, chemistry, and surface fluxes. The unprecedented scale of the exercise (one year, two continents) allowed us to examine the general characteristics of meteorological models' skill and uncertainty. In particular, we found that there was a large variability between models or even model versions in predicting key parameters such as surface shortwave radiation. We also found several systematic model biases such as wind speed overestimations, particularly during stable conditions. We conclude that major challenges still remain in the simulation of meteorology, such as nighttime meteorology and cloud/radiation processes, for air quality simulation.

1. Introduction

Air quality (AQ) modeling has progressed significantly over the past decade. It has evolved from the investigation of limited case studies of a few days or weeks duration to operational use for decision makers. Models are now routinely used to produce operational AQ forecasts in several countries (Brandt et al., 2001; McHenry et al., 2004; McKeen et al., 2005; Otte et al., 2005; Tarasick et al., 2007; Honoré et al., 2008; Hogrefe et al., 2007; Balk et al., 2010; Menut et al., 2010; Kukkonen et al., 2011) and to provide a prospective evaluation of air pollutant emissions control scenarios for policy needs, as in the Clean Air For Europe program or the United States (U.S.) NOx State Implementation Plan Call (e.g., Amann et al., 2005; Gilliland et al., 2008; Gego et al., 2007). However, many uncertainties still remain and need to be reduced in order to improve the performance of such modeling systems so they would have high societal utility. Owing to the large number of interrelated processes in AQ models, biases in the representation of different processes are sometimes difficult to parse because of compensating errors, making it difficult to fully diagnose and attribute the different sources contributing to modeling uncertainty.

Uncertainties in AQ model simulations basically arise from three main classes of processes: (1) chemistry and aerosol physics; (2) fluxes (emissions, deposition, boundary fluxes); and (3) meteorological processes affecting transport and diffusion, chemistry, and surface fluxes (e.g., Pielke and Uliasz, 1998; Seaman, 2000). This paper looks at the influence of the last class of processes. More precisely, it will focus on the meteorological processes and parameters known to have a strong influence on air pollutant concentrations and their variability. The evaluation of such parameters in meteorological models is particularly important because the requirements of weather forecasts are different from those of air quality forecasts. For instance, an accurate prediction of the height of the boundary layer is crucial for air quality prediction while it is not for weather prediction although it does have an indirect impact on weather in terms of triggering convection.

The three-dimensional wind fields transport primary pollutants or, if chemical reactions occur en route, secondary pollutants from emissions sources to receptor areas. Wind speed overestimation typically result in the underestimation of primary pollutant concentrations through increased ventilation and dilution, but they can also increase the concentrations of secondary pollutants near certain sources. For example, in areas close to nitric oxide (NO) emissions sources, an overestimated wind speed may induce a change in the photochemical regime since over-dilution of NO concentration will reduce ozone titration by NO, thereby resulting in an overestimation of ozone in the near-field. Wind direction errors will affect the path of pollutant trajectories and, hence, the source-receptor relationships. The concentration of pollutants in the lower troposphere, especially at the ground level, is also strongly sensitive to the rate of pollutant mixing by atmospheric turbulence, the height of the planetary boundary layer (PBL), the amount of venting to the free troposphere and transport from the upper-troposphere to the PBL for ozone. Atmospheric turbulence is, in turn, controlled by the magnitude of vertical temperature gradient and wind shear.

Meteorological parameters driving chemical processing are numerous. Radiation and its variability due to the presence of clouds, water vapor, aerosols and temperature are strong chemistry and aerosol thermodynamics drivers. For example, excessive cloud formation predicted at any altitude leads to the underestimation of below-cloud secondary pollutant formation from gas-phase processes and an overestimate in aerosol scavenging, inducing a low bias in secondary organic aerosol concentration. Many chemical reaction rates are temperature-dependent. And aerosol activation and aqueous-phase chemistry can occur in fog and clouds. Finally, meteorological processes also drive surface fluxes (emissions, deposition). Temperature and shortwave radiation control the emission of biogenic volatile organic compounds by vegetation, and wind speed and soil moisture control wind-blown dust or pollen emissions. Dry deposition is influenced by radiation, wind speed/turbulence, temperature, and surface wetness, and wet deposition is influenced by

precipitation intensity, vertical distribution (washout, rainout) and form (e.g., drizzle, rain, snow) (Gilliam et al., 2011).

Seaman (2000) provided an extensive overview of the influence of meteorology in regional AQ modeling in which he gave a number of examples of the sensitivity of AQ predictions to different meteorological variables. Hanna et al. (2001) employed a Monte Carlo approach to investigate the impact of uncertainties of 128 input variables, including a number of meteorological parameters, on ozone predictions made by a regional photochemical grid model (UAM-V). They found that the UAM-V predictions were sensitive to wind speed and direction, relative humidity, and cloud cover. Zhang et al. (2007) followed a meteorological ensemble approach in which they considered small initial perturbations in wind and temperature on MM5 meteorological forecasts and their subsequent impact on ozone levels in Houston, Texas predicted by the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006) for an episode in summer 2000. For this particular episode, they found high uncertainties in predicted ozone. Urban-scale sensitivities of air quality predictions to different meteorological variables were also studied within the EU project FUMAPEX and COST Action 715.

A number of studies have considered the impact of supplying meteorological fields for the same case from two or more different mesoscale meteorological models to the same regional AQ model (Sistla et al., 1996; Biswas and Rao, 2001; Hogrefe et al., 2001; Smyth et al., 2006; Pirovano et al., 2007; de Meij et al., 2009; Appel et al., 2010). Biswas and Rao (2001) used two different prognostic meteorological models (MM5 and RAMS) with the UAM-V AQ model and found an uncertainty of about 20% in simulating episodic 1-h ozone maxima. Hogrefe et al. (2001) evaluated temperature, water vapor mixing ratio, and wind speed predictions from two different prognostic meteorological models (MM5 and RAMS3b) and found that model predictions were best for temperature and worst for wind speed and that neither model showed skill in predicting intra-day variability (i.e., periods less than 12 hours). Smyth et al. (2006) examined predictions of temperature, relative humidity, and

wind speed from two different prognostic meteorological models (GEM and MM5) and found that differences in these fields resulted in a range of differences in O_3, PM_{10}, $PM_{2.5}$, and speciated $PM_{2.5}$ fields predicted by the CMAQ AQ model. de Meij et al. (2009) used two different prognostic meteorological models (MM5 and WRF) with the CHIMERE AQ model for winter and summer simulations of air quality in the Po Valley of Italy and found differences of 60% in PM_{10} predictions, particularly in the wintertime when predictions of PBL height made by the two meteorological models were significantly different. Finally, Appel et al. (2010) compared predictions made by CMAQ driven by two prognostic meteorological models (MM5 and WRF) and attributed differences in predicted AQ fields to differences in predicted wind speed, PBL height, cloud cover, and friction velocity.

Weather Services and research groups of more than 20 European countries investigated the influence of mesoscale meteorological models on regional AQ simulations in the framework of COST 728 (www.cost728.org). Eleven different AQ modeling systems participated in an inter-comparison exercise. The task was to model concentrations of particulate matter (PM) during a complex high-pressure episode over Germany in winter 2003 (Stern et al., 2008). It was found that none of the chemical transport models (CTMs) was able to predict the observed high PM values in East Germany (Matthias et al., 2010). The largest meteorological influence on the simulated concentrations was connected with vertical mixing of the pollutants. However, it could not be concluded that the most accurate model results for meteorological quantities led to the most accurate CTM results since emission inventories that drive AQ models are uncertain. In some cases errors in the meteorological and AQ models cancelled out, resulting in reasonable pollutant concentration values. One of the conclusions of the COST 728 action was that extensive meteorological model testing on longer time scales is necessary to gain more insight into the meteorological effects that may cause errors in AQ modeling.

The framework of the Air Quality Model Evaluation International Initiative (AQMEII; Rao et al., 2011) offers a unique opportunity to evaluate AQ model strengths and weaknesses from a year-long AQ simulation for 2006 carried out by a large set of AQ models over two continents. This paper focuses on uncertainties associated with the meteorological inputs used by the AQMEII AQ modelers. It benefits immensely from the opportunity to inter-compare the performance of more than 10 meteorological models or model configurations for the same meteorological parameters on the same analysis grids for the same extended period for two continental-scale regions.

The AQMEII project has collected both meteorological observations of several meteorological parameters and asked participating modeling groups to extract equivalent model values in a format that it would allow direct comparison. However, the limited number of parameters routinely observed does not allow a full and comprehensive evaluation. Thus, we focus our analysis on a few issues. The main questions we address here concerning transport and mixing are:

- Are boundary-layer wind speed and PBL height accurately simulated?
- Are boundary-layer temperature, relative humidity profiles, and surface radiation influencing atmospheric chemistry accurately simulated?
- Are meteorological processes influencing surface fluxes (surface temperature, wind speed, shortwave radiation, and precipitation) accurately simulated?
- What are the spatial and seasonal distribution of the biases in both mean and variability of the studied parameters?
- Are there any systematic differences in the prevailing meteorology over the two continents?

It must be noted that the questions addressed here relate strictly to the ability of models to *simulate* in retrospect and not *forecast* the meteorology of the lower troposphere. Because data assimilation is used, it is assumed that atmospheric model simulations are “best attempts” to reconstruct the

state of the atmosphere retrospectively at a scale relevant to simulated air quality. This is generally done in two steps: an analysis or a reanalysis is carried out by a weather centre by blending cycling forecasts with new observations, followed by a simulation using a limited-area model with increased resolution and detailed surface and boundary-layer processes that may be combined with some form of data assimilation like analysis nudging. Our conclusions thus do not necessarily apply to weather forecasts, for which the additional uncertainty due to the forecast itself must be taken into account. However, they do help to quantify current uncertainties in a number of important meteorological parameters required by AQ simulation models. Finally, it should be noted that this study only provides investigation and evaluation of multi-model performance in general terms, and specific in-depth performance evaluations are also being carried out of individual models (e.g. Gilliam et al., 2011).

In addition to the evaluation and inter-comparison of the predictions of 2006 meteorology for North America (NA) and Europe (EU) made by the different meteorological models applied in the AQMEII study, this paper also reviews the weather conditions experienced during 2006 over both continents and the climatological representativeness of that year. After a description of the meteorological observations for 2006 in Section 2 and the AQMEII 2006 meteorological simulations in Section 3, a summary of 2006 weather is given in Section 4. Section 5 contains a quantitative multi-parameter evaluation of the set of meteorological simulations, and the paper concludes with a discussion of results and conclusions in Section 6.

2. Meteorological Observations

Surface-based observations for the evaluation of the annual Weather Research and Forecasting (WRF) model NA simulations were extracted from the Meterological Assimilation Data Ingest System (MADIS: http://madis.noaa.gov/) database. MADIS has both archived and real-time meteorological observations for North America including standard US and Canadian managed surface measurements

206 as well as mesonet, rawinsonde, wind profiler, aircraft, and satellite measurements. For the
207 European domain, the surface observations were extracted from the National Center for
208 Atmospheric Research (NCAR) global synoptic surface data archive
209 (http://dss.ucar.edu/datasets/ds464.0). The extracted observations for 10-m wind (speed and
210 direction), 2-m temperature, 2-m relative humidity and precipitation were ingested by the
211 ENSEMBLE system of the European Commission Joint Research Centre at Ispra, Italy (Galmarini et al.
212 2001, Bianconi et al. 2001, Galmarini et al. 2004), which allows matching in time and space with the
213 various model datasets in order to carry out model performance evaluations. Some technical
214 difficulties prevented the extraction of precipitation and relative humidity for the European domain,
215 so the evaluation of these parameters is only for the North American domain. Since a robust, high
216 resolution gridded precipitation dataset called the Parameter-elevation Regressions on Independent
217 Slopes Model (PRISM) was available, the only direct evaluation of model precipitation is focused on
218 the United States. The 4 km PRISM precipitation was aggregated up to the 12 km WRF grid so a direct
219 comparison of seasonal precipitation could be made.

220 For upper-air analysis, meteorological variables observed from ozone soundings were downloaded
221 from the WMO World Ozone and Ultraviolet Radiation Centre (www.woudc.org). Even though we do
222 not investigate ozone in this article, this choice was made in order to have collocation with ozone
223 measurements. Vertical profiles of pressure, temperature, relative humidity and wind speed were
224 obtained from these soundings. In this study, a set of six stations was selected for each continent to
225 serve as basis for model error statistics at the given altitudes of 0, 100, 250, 500, 750, 1000, 1500,
226 2000, 3000, 4000, 5000, 6000, 7500 and 8500 m above ground level. These stations were selected in
227 three ways:

228 • The data set should not be too small (i.e., it should contain 40 profiles or more)
229 • The station altitude should be close to the altitude of the respective model grid cell.
230 • The stations should cover different regions of the continent

3. Meteorological Models

AQMEII provided a 2006 meteorological reference simulation for each continent to all participants so as to encourage both maximum participation and model input harmonization, but the use of these simulations was not mandatory. The reference simulations for NA and EU were generated using the Weather Research and Forecasting (WRF) model version 3.1 (Skamarock et al., 2008) and MM5 (Dudhia, 1993), respectively. The choice of these two models was ad hoc as one group on each side of the Atlantic volunteered to share their meteorological simulations. For the study conducted in this paper, groups used five different meteorological models or model configurations to drive NA AQ simulations and 11 different meteorological models or model configurations to drive EU AQ simulations. In this article, we emphasize the two reference simulations, as more than one group made use of each of these simulations, but we also describe and evaluate the other meteorological simulations that were employed.

For NA, the Advanced Research WRF (ARW) core was employed, which is a fully-compressible, non-hydrostatic, mass-conserving numerical solver. The modeling domain has a horizontal grid scale of 12 km with 34 vertical levels extending from the surface to the 50 hPa pressure layer with 14 levels below 1 km and the first layer about 40 m thick. This 12-km domain aligns exactly with standard U.S. Environmental Protection Agency (EPA) modeling domains, including the 36-km modeling domain described in Otte (2008) and Gilliam et al. (2006) and the 12-km domain discussed in Gilliam and Pleim (2010) and Appel et al. (2010). The difference is that this AQMEII modeling domain was extended to the north and east in order to include some key emission sources in Canada. In addition to the domain used, most of the model physics and four-dimensional data assimilation (FDDA) techniques were adopted from previous U.S. EPA modeling research such as Otte (2008) and Gilliam and Pleim (2010), which provide guidance on using WRF and MM5 effectively for retrospective AQ

modeling applications although Gilliam et al. (2011) does suggests an updated technique that reduces transport errors in the lower troposphere.

Among the WRF physics options used were the Rapid Radiation Transfer Model Global (RRTMG) long- and short-wave radiation (Lacono et al., 2008), Morrison microphysics (Morrison et al., 2009), and the Kain-Fritsch 2 cumulus parameterization (Kain, 2004). For the land-surface model (LSM) and planetary boundary layer (PBL) model, the Pleim-Xiu LSM (Xiu and Pleim 2001; Pleim and Xiu 2003) and Asymmetric Convective Model version 2 (ACM2) (Pleim 2007a; Pleim 2007b) were used. These physics schemes, in particular, were developed explicitly for retrospective AQ modeling as the LSM employs an indirect soil moisture and temperature nudging scheme (Pleim and Gilliam, 2009). The soil nudging limits the error growth of critical near-surface fields such as temperature and moisture by adjusting surface energy fluxes to minimize the difference between the simulated 2-m temperature and moisture and that provided by an analysis. The ACM2 PBL scheme is also used in the CMAQ AQ model, so its use in WRF allows the mixing of pollutants to be consistent with the mixing of heat, moisture, and momentum within the PBL or other mixed layers in the atmosphere. Initialization and nudging follow the strategy described in Gilliam and Pleim (2010).

For EU, MM5 was run with lateral and surface (sea-surface temperature) boundary conditions obtained from the European Centre for Medium Range Weather Forecast (ECMWF) operational analyses, with a 6-hour sampling rate. Initial conditions (soil and atmospheric variables) were also taken from ECMWF analyses. The configuration used is Version 3.7, with most parameterizations as described in Chiriaco et al. (2006). Nudging to ECMWF analyses is applied with a relaxation time of about 3 hours for temperature and wind, and 15 hours for humidity. The 2006 simulation was split into twelve 1-month long simulations with new initializations 6 hours (spin-up time) before the first day of each month.

277 The vertical grid contains 32 sigma layers from surface to the top of the atmosphere, with 9 layers
278 below the first kilometer. The top of the first layer was taken at s=0.996 (about 40 m above the
279 surface, thus the middle of the first layer is 20 m). The horizontal grid is taken along a Mercator
280 projection, with grid spacing decreasing from south to north. It extends outside the chemical model
281 grid imposed by the AQMEII coordinates (15°W – 35°E; 35°N – 70°N). The exact domain boundaries
282 for MM5 are 18°W – 38°W and 33.3°N – 71.5°N. At 50°N, the grid size is about 20 km while it is
283 about 10 km at the northern boundary and 25 km at the southern boundary.

284 The planetary boundary layer (PBL) is described using the MRF PBL scheme (Hong and Pan, 1996).
285 The microphysics scheme is the Reisner2 scheme, which considers five states of water: vapor, rain,
286 cloud, ice, and graupel (Reisner et al., 1998). The cumulus scheme is taken from Grell and Devenyi
287 (2002). The NOAH LSM scheme is used (Ek et al., 2003), with the default four layer depths changed
288 to 7, 28, 100, and 289 cm to better match ECMWF model soil levels. The long-wave radiation scheme
289 used is the Rapid Radiation Transfer Model (RRTM; Mlawer et al., 1997).

290 In addition to these two meteorological reference simulations offered to AQMEII participants, some
291 of the groups performed their own meteorological simulations. A total of six different
292 meteorological models were used: COSMO, ECMWF, GEM, MM5, PARLAM-PS, and WRF. A summary
293 of some of the main characteristics of all of the models is given in Table 1. There is considerable
294 overlap between the models in terms of physical parameterizations and run strategies employed, but
295 five NA and 11 EU meteorological model configurations were distinct. The horizontal grid spacing
296 used ranged from 12 to 50 km, and the number of vertical levels ranged from 23 to 58. Data
297 assimilation techniques were employed by a minority of the models.

298 In Section 5, the five NA meteorological model configurations are denoted by the labels “M1NA” to
299 “M5NA” and the 11 EU configurations by the labels “M1EU” to “M11EU”. Three model
300 configurations were applied for both 2006 NA and EU simulations and have been assigned the labels

301 "M1NA" and "M1EU", "M2NA" and "M2EU", and "M3NA" and "M3EU". Note that the order in which
302 the labels have been assigned is different from the order of the model configuration descriptions in
303 Table 1 to keep anonymity.

304 4. 2006 Weather in North America and Europe

305 For a number of years the U.S. National Climatic Data Center (NCDC) has led an effort to characterize
306 the weather of recent years. Arguez et al. (2007) provides the summary of significant global weather
307 events and anomalies in 2006. The highlights specifically for North America and Europe are covered
308 here and will be used to provide context for the model evaluation where appropriate. One of the
309 most significant characteristics of 2006 was its rank as the 5th warmest (global) in the last century.
310 Regionally, parts of Europe (UK, Spain and the Netherlands) saw the warmest year on record and
311 parts of the U.S. and Canada experienced the second warmest year on record. Figure 1 provides the
312 850 hPa seasonal temperature anomaly, which clearly shows the warmer than normal weather. 500
313 hPa geopotential height anomalies (Figure 2) correlated well with the 850 hPa temperature
314 anomalies as the regions with warmer than normal temperatures almost always correlate to more
315 ridging aloft. The averaged temperature in January over U.S. was 3.9 K above normal, which is a full 1
316 K greater than the previous 100+ year record. Most of the central and western US was warmer than
317 normal in the summer as well. Areas in the central and southwestern U.S. that saw the higher
318 temperatures (Figure 1) and anomalous ridging aloft (Figure 2) also experienced very little rainfall
319 and, as a result, severe drought conditions. A record-breaking heat wave that reinforced the drought
320 conditions began in the northern Plains and upper Midwest in mid-July and spread to the western
321 U.S. in late July then back to the east, all the way to the East Coast for the first half of August (Arguez
322 et al., 2007).

323 The eastern U.S. and southern Ontario and Quebec experienced average to above average rainfall in
324 the spring and summer. The 500 hPa height anomalies in Figure 2 indicate the east coast of North
325 America did experience near to below normal 500 hPa height and temperature at 850 hPa (Figure 1)

326 in the spring and summer, which translates to above average rainfall. On the opposite side of the
327 Continent, areas of the Pacific Northwest U.S. and British Columbia saw heavy rainfall the last few
328 months of 2006 because of blocking ridge in the NW Pacific (Arguez et al., 2007).

329 For Europe, Arguez et al. (2007) showed annual near-surface temperature anomalies that were
330 generally greater than 0.5-1.0 K for most of Europe. An examination of the seasonal 850 hPa
331 temperature and 500 hPa geopotential height anomalies in Figure 1 and Figure 2, respectively, shows
332 cooler than normal temperature and lower 500 hPa heights for the first part of 2006 across much of
333 Europe. Arguez et al. (2007) identifies this large-scale weather pattern as common feature with the
334 negative phase of the North Atlantic Oscillation (NAO) that was in place for the first few months of
335 2006. Countries in the north and far western parts of Europe, like the British Isles and Scandinavia,
336 saw temperatures at or slightly above normal and normal precipitation in the winter of 2006.

337 The large-scale weather pattern made a transition from cooler and drier over much of Europe in
338 winter to warmer than normal, in general, for the rest of 2006. However, there was a substantial
339 month-to-month variability from spring to summer that the seasonal anomalies do not capture well.
340 For example, July 2006 was well above normal as an eastward extension of the Azores High
341 developed over central Europe leading to an extreme heat wave (Arguez et al., 2007). Many of the
342 central European countries, including Belgium, Netherlands, Germany, Czech Republic and Austria,
343 set all-time records in terms of mean July temperatures. This heat wave was also accompanied by a
344 large-scale pollution episode over Central Europe (Struzewska and Kaminski, 2008). In August
345 however, this warm pattern transitioned to a cooler than normal pattern. Precipitation was generally
346 lower than normal during the anomalously high temperatures and near or just above normal during
347 the cooler periods like what occurred in August.

348 Autumn was the most anomalous season of the year over Europe. It broke the record of seasonal
349 temperature by a large amount and was shown to have a temperature largely exceeding that

350 expected from analogue weather regimes in previous years (Yiou et al., 2007), presumably due to a
351 concurrence of a large Atlantic sea surface temperature anomaly and a persistent southerly flow
352 (Cattiaux et al., 2009). The 850 hPa temperature anomaly for autumn clearly shows that a large
353 anomaly that had been centered in the Northern Hemisphere (+3.0 K) was now centered over the
354 Denmark/Germany area and extended north to Scandinavia, west to the British Isles and south to
355 France as well as much of southern Europe that borders the Mediterranean Sea. The 500 hPa height
356 anomalies are in good agreement with the warm autumn temperatures as a persistent ridging is
357 centered over Germany and Poland. Precipitation amounts under and around this ridge, as one
358 would expect, were well below normal. Areas that did experience higher than normal autumn
359 precipitation are those countries to the west and southwest periphery of the 500 hPa ridge anomaly,
360 which includes Ireland, United Kingdom, western France and western Spain and Portugal. Much of
361 Europe that borders the Mediterranean was dry as the axis of the 500 hPa ridge anomaly extended
362 south into the Mediterranean Sea between Spain and Italy as shown by Figure 2 and describe in
363 detail by Arguez et al. (2007).

364 ## 5. Quantitative Evaluation

365 In this section, we quantitatively compare model simulations and observations of weather
366 parameters that are most relevant to air quality. For the sake of synthesis, we have focused on three
367 distinctive subregions on each continent that have qualitatively different climate and air quality
368 characteristics. These subregions are shown in Figure 3, together with the locations of
369 meteorological measurement sites.
370 For NA, subregion NA1, the southwestern U.S., was selected because of the combination of high
371 solar radiation, low relative humidity, large cities with poor air quality (Los Angeles, Phoenix), and
372 geographic location to the west of the Rocky Mountain barrier. Subregion NA2, the Texas area, was
373 selected for its hot, humid climate, large cities with poor air quality (Houston, Dallas), and location to

374 the east of the Rocky Mountain barrier. Subregion NA3, northeastern NA including parts of Canada,
375 has a marked seasonal cycle, three of the North American Great Lakes, the highest emissions areas in
376 NA, and large cities (New York City, Philadelphia, Toronto, Montreal). For EU, subregion EU1, the
377 British Isles and western France, was selected for its mid-latitude, mixed maritime-continental
378 climate and large cities (London, Paris). Subregion EU2, Central Europe, has a rather continental
379 climate with marked seasonality, many large cities, and large emissions areas. Subregion EU3, the Po
380 Valley of Italy, has a Mediterranean climate, poor air quality, and belongs to a separate air shed from
381 northern Europe due to the Alpine barrier.

382 **5.1 Transport and mixing**

383 The weather parameters that drive the transport and mixing of air pollutants are controlled by grid-
384 scale winds and subgrid-scale turbulence, including shallow and deep convection. We use here the
385 reduced set of available routine network observations described in Section 2. For resolving
386 transport, the analysis uses the 10-m wind observations and vertical wind profiles obtained from
387 ozonesonde launches.

388 Figure 4 shows the evolution of the 10-m wind speed averaged over all measurement station
389 locations in each subregion for each calendar month for each model and for the observations. In
390 general the seasonal cycle is well reproduced by all models in all subregions, but wind speed
391 amplitude spread is rather large and overestimated for EU. Model values differ by rather constant
392 multiplicative factors. This could be due to a combination of differences in the model resolution in
393 the lowest layers and differences in the methodology of diagnosing the 10 m wind amongst models.
394 A general overestimation is found in all regions but NA1 and NA2, and no obvious explanation was
395 found for this feature.

396 The amplitudes of the diurnal cycle of wind speed are underestimated (Figure 5). In the stable
397 nighttime boundary layer, wind speed is overestimated, probably as a result of the lack of vertical
398 resolution (i.e., layer height is approximately 40m) and overly strong vertical diffusion. For the NA

399 subregions, the intensification of wind speed due to the stronger vertical momentum fluxes that are
400 associated with the development of the convective boundary layer and associated increase in wind
401 speed is not marked enough and daytime wind speeds are generally underestimated. However,
402 biases are generally larger during the night, which indicates a general difficulty to simulate the stable
403 boundary layer. A particular situation occurs for EU3 (the Po Valley) where even the shape of the
404 diurnal cycle is not well simulated, probably due to the complex topography of the area and the land-
405 sea interface that induces complex mesoscale circulations.
406 The skill of the models in simulating the day-to-day variability of daily mean wind speed is
407 summarized in Figure 6, which shows Taylor diagrams (Taylor, 2001) for wind speeds in all subregions
408 studied. In all subregions, simulations have a correlation exceeding 0.5, and often reaching 0.9. For
409 NA, the amplitude of daily wind variability varies by a factor of two relative to the observed one, with
410 no systematic bias, while the variability is overestimated by all models in the EU case, which is
411 consistent with the general overestimation of wind speed. Over NA, there is a marked spread in
412 model skill. Correlation is generally higher for NA2 and NA3, where three models have a correlation
413 exceeding 0.9, than for NA1, where topographic and coastal effects dominate the meteorology. For
414 EU, models' skill is higher in maritime areas (EU1) and Central Europe (EU2), but is poor over the Po
415 Valley due to complex topography. The large spread in model skill leads to a skill of the ensemble
416 mean or median that is not higher than that of any model.
417 The spatial distribution of surface wind speed is fairly well simulated by the models (Figure 7). Over
418 NA, the differences between the windier mid-western areas and less windy eastern areas are
419 correctly reproduced, even though the observed winds are somewhat weaker than the simulated
420 winds. WRF also generally does well in simulating the strength of transport over the oceans. Over
421 EU, MM5 reproduces the northwest – southeast wind speed gradient. Regional discrepancies are
422 found, for example, in some mountainous areas (e.g., Scandinavia and Alpine regions), where poorly
423 resolved effects of topography probably explain the simulated wind overestimation.

Vertical profiles (based on ozone soundings) of wind speed are compared to the results of several models over NA and EU (Fig. 8). The statistical measures (bias, RMSE, correlation), were calculated for each of the stations and then averaged. Wind speed is well simulated along the profiles but markedly overestimated at lower altitudes for EU, confirming the results for 10-m wind speed. For NA, more scatter occurs among models for wind, but agreement between model and observations in terms of the mean wind speed is stronger in the lowest 500 m for three of the models. The RMSE is between 2 and 4 m s^{-1} (except for models M1NA and M2NA in North America) with slightly higher values in higher altitudes, which corresponds to higher wind speeds on average. The correlation is lowest close to ground, but may exhibit values exceeding 0.9 above 500 m in Europe and above 1500 m in North America. Two models (M1NA, M2NA) show poor correlation of 0 – 0.25 in North America. For simulating North America, the results from the European groups show clearly less agreement with the observations compared to Canadian and U.S. groups. However it must be taken into account that different run schemes and nudging techniques are used (see Table 1). If a model run is restarted every few days with initial conditions that stem from reanalysis data, the results will stay close to the observations because they are typically considered in the reanalysis. A continuous model run on the other hand, that is only nudged to the wind fields above the PBL has much more freedom to develop differently than the driving reanalysis fields. This should lead to a larger variability of the simulated quantities and therefore larger RMSE and lower correlation.

In order to evaluate the skill of the model in representing turbulent mixing, PBL heights calculated by the different meteorological models are compared to observations at Lindenberg, Germany (14.3°E, 52.1°N). The observational data has been derived from radiosondes using the bulk Richardson number method. The observational data has been derived from radiosondes using profiles of the bulk Richardson number Ri_b. The method is a standard and widely used approach to derive PBL height from the numerical weather prediction (NWP) models,

as well as from the radiosounding data (see e.g. the review by Seibert et al., 2000). Here, a critical Richardson number $Ri_c = 0.2$ was chosen. The top of the PBL is the altitude where $Ri_b > Ri_c$.

Each model has its own algorithm to diagnose the PBL height, many of them are based on similar approaches as the one applied to the observations. It was found that the models are able to simulate the PBL height at noon quite well (Fig. 9 and Table 2). This can be interpreted in a way that the PBL parameterizations are working reasonably well and the vertical mixing of pollutants under these conditions is likely represented adequately in the models. By contrast, particularly at 18 UTC and in the summer months, the modeled PBL height is much lower than observed (Fig. 9). This may be explained by the fact that this is a transition time to a stable PBL as static stability of the surface layer turns positive. In this transition phase the top of the PBL is not well defined and the models typically diagnose the top of the PBL to be one of the first few model layers while the radiosondes do not show this. Some of the models give very low PBL height around the top of the first model layer throughout the night which is clearly unrealistic, but default position of these non-TKE schemes. The morning ascent of the PBL, when strong mixing processes take place, could not, unfortunately, be investigated due to 6-hour observation sampling.

Table 2 gives the mean observed PBL heights at 0, 6, 12 and 18 UTC together with the bias, RMSE and the correlation of the model results when compared to the observations. Here, all observations including those when the PBL height was not well defined were taken into account. As mentioned above, the largest discrepancies between model results and observations occur at 18 UTC, at this time none of the models reproduce the observed values with reasonable accuracy. This is represented in poor correlation coefficients and a large negative bias. About 3-5 models show clear problems in representing the correct PBL height at all times except 12 UTC.

5.2 Chemistry drivers

Three of the meteorological parameters that drive atmospheric processing of emissions (chemistry and aerosol transformations, see Monks et al., 2009 for a full review) are evaluated here:

473 temperature, relative humidity and surface shortwave radiation. Biases of monthly means of 2-m
474 temperature are generally small (Figure 10). Over NA only one model has a moderate positive
475 temperature bias that occurs mainly in the winter season and is as large as 5 K. Otherwise, the
476 remaining ensemble members have little spread and agree well with the observed temperature in a
477 regional average sense. Likewise, in EU, biases remain small, with slightly more spread during winter
478 months, but the model ensemble envelopes the observations well.

479 The diurnal cycles of 2-m temperature are also fairly well reproduced by the models (Figure 11).
480 Unlike the 10-m wind speed, the amplitudes of the diurnal cycles for 2-m temperature are not
481 underestimated except for one model over NA, which also had the systematic positive wind speed
482 bias seen in Figure 5. Thus we expect that related temperature-dependent fields (clouds, longwave
483 radiation and sensible heat fluxes, see e.g. Liu et al., 2003) are fairly well accounted for in the
484 models.

485 The typical vertical temperature profile bias is between ±1 K (Figure 12). On average the temperature
486 is slightly underestimated by the models. The RMSE is between 1 and 2 K along the profile, best
487 agreement being achieved between 1000 and 6000 m altitude. The correlation is above 0.9, and at
488 many heights, even above 0.95.

489 For simulated ozone episodes to build up, it is essential that the highest diurnal temperatures are
490 well predicted by the models, other parameters also being important. In order to focus on this issue,
491 Figure 13 shows the 99.5th centiles of the models temperature distribution (hourly values) against the
492 corresponding observed 99.5th centiles. In most cases, considering both continents, the extreme
493 temperatures that were observed are greater than the model simulated temperatures. The
494 differences, however, remain moderate and do not exceed 3 K. This small bias should have the
495 effect of reducing gas-phase chemical reaction rates as well as slightly displacing the gas-particle
496 equilibrium for volatile species.

497 Relative humidity (RH) influences photochemistry through reactions between water vapor and the
498 oxygen radical, which forms the hydroxyl radical. Water vapor can be either an ozone sink or source,
499 depending on the availability of nitrogen oxides. Relative humidity at 2 m is not as well simulated as
500 temperature (Figure 14 vs. 10). Over NA, systematic biases are found for most models, and in
501 general RH is overestimated. The bias is particularly marked over the southwestern U.S. (subregion
502 NA1), the driest of the three NA subregions. This reveals model deficiencies in dry areas, with a
503 possible consequence of overestimation of soil moisture. However, the amplitude of the diurnal cycle
504 is simulated in a realistic manner (not shown).

505 Above the surface, relative humidity is overestimated by all models and in all regions (Figure 15), in
506 agreement with surface analysis for NA. Biases and RMSE both increase with height. This is not
507 surprising if one keeps in mind that the water vapor mixing ratio decreases rapidly with height and
508 therefore RH is sensible to small deviations of the mixing ratio. The overestimation of RH might be
509 connected with the underestimation of the temperature. The correlation of the time series, however,
510 is relatively large, with values between 0.6 and 0.8.

511 Model predictions of hourly gridded surface shortwave radiation (SSWR) were submitted to AQMEII
512 by most groups, but surface radiation components are not routinely measured at many stations in
513 either NA or EU. Since shortwave radiation plays an important role in photochemistry, the surface
514 energy budget, and biogenic emissions, it was still of interest to examine differences between
515 models, especially because SSWR will be modulated by cloud shading, which may vary considerably
516 between models due to the difficulties associated with predicting the presence and properties of
517 clouds.
518 The lefthand column of Figure 16 shows the monthly variation of mean mid-day SSWR at the centers
519 of the three NA subregions (see Fig. 3) predicted by four meteorological models. The highest
520 summer values for the three subregions are predicted over the southwestern U.S. (NA1), and the
521 largest differences between the models for this subregion occur in the spring (~400 Wm^{-2} or ~100%).

522 The lowest summer values are predicted for northeastern NA (NA3), and the largest differences
523 between models for this subregion occur in June (~400 Wm^{-2} or ~100%). These summertime
524 differences are surprisingly large and are likely due to differences in the predictions of clouds. The
525 righthand column of Figure 16 shows the same analysis for the center points of the three EU
526 subregions for nine meteorological models or model configurations. The largest actual difference
527 between models occurs in June for EU3 (~500 Wm^{-2} or ~125%) but relative differences are even
528 higher in the winter months. For the EU subregions the ranking between models is generally
529 constant between subregions and across seasons. These systematic differences in SSWR between
530 models may impact many other meteorological fields such as surface temperature and PBL height.
531 Figure 17 shows considerable variation in the model-simulated diurnal cycle of SSWR for the six
532 subregions. For NA there are systematic differences of 15% to 50% between the four models at local
533 noon and for EU there are differences of 30% to 60% between eight models (excluding one outlier).
534 As expected, the maximum daytime value tends to decrease with increasing latitude, but cloud cover
535 also plays a role; for example, the maximum daytime value is lower for subregion NA2 (31°N) than
536 for NA1 (36.5°N). For the EU subregions there is also a suggestion that local noon differs between
537 two clusters of models.
538 Figure 18 shows monthly variations in the standard deviation of mid-day hourly SSWR for each
539 month of 2006 for the same six locations. This quantity provides another measure of the impact of
540 differences in model predictions of hourly cloud fields on cloud shading. It is evident that there are
541 considerable differences between the models throughout the year, but these differences vary from
542 subregion to subregion. The differences are largest in spring and summer for the southwestern U.S.
543 (NA1) but fairly even throughout the year for northeast NA (NA3). For the Texas subregion (NA2) and
544 the three EU subregions, on the other hand, there is closer agreement between the models in the
545 cold season and less agreement in the warm season. One possible explanation is a higher frequency

546 of stratiform cloud in the winter, a higher frequency of convective cloud in the summer, and closer
547 agreement between model predictions of the former (see next section).

548 **5.3 Surface fluxes**

549 Biogenic emissions depend on a number of factors, including surface weather. Soil nitrogen oxides
550 (NOx) and vegetation volatile organic compound (VOC) emissions increase nonlinearly with
551 temperature, with sharp sensitivity at temperatures exceeding 30°C. The above analysis shows that
552 these emissions should be fairly well represented in most models, but an underestimation may be
553 expected due to moderate low temperature bias at highest temperatures. Biogenic VOC emissions
554 also depend on radiation, but the model skill for radiation could not be properly evaluated against
555 observations within this study.

556 A major driver of dry deposition fluxes is the stomatal resistance which also depends on temperature
557 and radiation. Dry deposition, particularly for ozone, is also driven by turbulent mixing near the
558 ground. Although we were not able to evaluate the model predictions of sensible heat fluxes, the
559 weak differences between simulated and observed 2-m temperatures indicates that aerodynamic
560 resistance should not undergo strong model biases.

561 For both aerosol particles and soluble gases, wet deposition fluxes depend on precipitation
562 frequency, duration, intensity, and type (e.g., Wang et al., 2010). Model predictions of hourly
563 precipitation for 2006 have been examined for the North American simulations. In terms of seasonal
564 accumulation, Table 3 lists mean winter (Dec.-Feb.) and summer (June-Aug.) precipitation amounts
565 for all measurement stations in each of the three NA analysis subregions and corresponding mean
566 model-predicted precipitation amounts for these three groups of stations. In 2006, the NA1
567 subregion received more precipitation in the winter than the summer while the opposite was true for
568 the NA2 and NA3 subregions. Most of the models reproduced this geographically-varying seasonal
569 cycle, but there is a wide variation in predicted amount and the models, including the ensemble
570 mean, tend to overpredict seasonal precipitation. This is particularly true in the summer when

571 convective precipitation typically dominates (e.g., Tremblay, 2005), since the simulation of
572 convective precipitation is challenging because of its small-scale and scattered nature.
573 Given that wintertime precipitation tends to be dominated by stratiform precipitation (Tremblay,
574 2005), and given that stratiform precipitation tends to be longer-lived with more wide spread
575 coverage than convective precipitation due to its synoptic forcing, it is useful to examine observed
576 and predicted hourly precipitation intensity. Figure 19 shows winter- and summer-season
577 histograms of observed and predicted occurrence frequencies for different hourly precipitation
578 amounts for the three NA analysis subregions. Both observations and models exhibit more high-
579 intensity precipitation events (i.e., a longer distributional "tail") in the summer than winter for
580 subregion NA3, about the same for subregion NA2, and fewer high-intensity events in subregion
581 NA1. In meso-β-scale models (i.e., horizontal grid spacing of 10-40 km) such as those considered
582 here, transport by convective precipitation systems will be associated with subgrid-scale circulations
583 and hence will not be resolvable. Figure 19, however, suggests that such high-intensity precipitation
584 occurs infrequently (note the semi-log scale). In terms of low-intensity precipitation forecasts, on the
585 other hand, most of the models underpredict non-precipitation events (i.e., the "< 0.5" bin includes
586 dry conditions and "trace" precipitation) but overpredict the occurrence of low-intensity
587 precipitation (i.e., 1-5 mm h^{-1}). There is also considerable variability amongst the models. Note that
588 it is likely that this difference between the measurements and models can be ascribed at least in part
589 to the comparison here of point measurements to grid-scale predictions, which introduces the
590 problem of representativeness error due to interpolation of model grid-cell values to station
591 locations (e.g., Tustison et al., 2001). Nevertheless, the combination of higher accumulation, longer
592 duration, and greater spatial coverage on average in the model predictions suggests that wet
593 removal may be overemphasized by the models in areas of more frequent convection, leading to a
594 tendency to underestimate ambient air concentrations of particles and water-soluble species such as
595 SO_2, HNO_3, and NH_3.

Finally, Figure 20 compares the spatial distribution of seasonally observed precipitation (PRISM) for two seasons with the corresponding spatial distribution predicted by the U.S. EPA WRF simulation. WRF agrees with PRISM quite well in winter when grid-scale stratiform precipitation is likely dominant, whereas in summer, when diurnally-forced convective precipitation is most common, the PRISM and WRF differ significantly in total summer precipitation.

6. Summary

This study was devoted to a collective operational evaluation of regional meteorological models that forced the air quality simulations carried out in the AQMEII regional AQ modeling system inter-comparison. It was the first time that a multi-model evaluation of this scale has been performed, with five participating meteorological models or model versions over North America (NA) and 11 models or model versions participating over Europe (EU). We emphasized model parameters that are major drivers of air quality variability. The focus was not to inter-compare the models and produce statistical metrics, but rather to discern general characteristics seen. This study produced a number of conclusions.

- There is considerable variability among model predictions, even for different configurations or post-processing of the same model. This is particularly clear for short wave radiation where noontime predicted values vary by a factor up to two. This scatter should contribute to variability in many other predicted fields, suggesting that prediction of the timing and location of clouds remains an ongoing challenge for both meteorological and AQ modeling.
- There are systematic positive model biases, particularly for EU, for surface and boundary-layer wind speed, which are confirmed both in 10-m wind and ozonesonde measurements. These biases should contribute to a tendency to underestimate surface concentrations of primary pollutants. The overestimation is particularly marked in stable wintertime or

nighttime conditions. The day-to-day variability of low-level wind speed is also systematically overestimated for EU.

- Developed planetary boundary layer (PBL) heights are, at one European site, well captured, but PBL height is poorly simulated at nighttime or transition times. Models generally underpredict PBL heights in these situations, which should lead to air pollutant concentration overestimation if this conclusion holds in other locations.
- Less clear conclusions hold for water vapor and precipitation, but we found large – albeit not systematic – differences for these parameters. These variables can significantly influence the predicted concentrations of fine particulate matter.
- The models have a tendency to underestimate the occurrence of non-precipitation conditions and extreme precipitation events but overpredict the occurrence of light to moderate precipitation conditions. This could lead to an overestimation of wet removal of particles and water-soluble gases.
- Not surprisingly, temperature is the best predicted of the variables that we analyzed in this study.

Our conclusions point to several systematic biases (e.g., overestimated wind speed, lack of long dry periods). These biases should induce significant and systematic concentration biases, in particular for primary pollutants. It is beyond the scope of this article to actually verify that concentrations undergo such biases. However, several of the conclusions of the AQMEII multi-model analysis of model skill in simulating particulate matter (PM) are consistent with our results (see Solazzo et al., 2011). In particular, model wind speed bias was found to be correlated with negative particulate matter biases. Overestimated rainfall frequency is also consistent with underestimated PM concentrations, but verification of this bias was not carried out.

Since the meteorological variables considered in this paper are known to have important influences on AQ predictions, the large variability in the predicted meteorological fields amongst the different

meteorological models and model versions will likely make an important contribution to the variability in the predicted AQ fields that has been quantified in companion AQMEII papers in this special issue. For primary pollutants and aerosols, dispersion (wind, boundary layer height) is the most important concentration driver. From our analysis, we conclude that model simulations of daytime meteorology have fewer deficiencies than simulations of nighttime meteorology. Nighttime concentrations undergo systematic overestimation of wind and underestimation of PBL height, which is a potential source of large error compensation for pollutant simulation. Therefore, nighttime meteorology remains a challenge for models. Finally, for photochemistry and secondary pollutants, shortwave radiation and its influence on cloud processes is probably the most critical process to improve as it is a major driver of ozone build up. We conclude that efforts must be made to reduce the uncertainty in the simulation of radiation and clouds.

Acknowledgements

The authors would like to acknowledge the following data providers. NOAA and Environment Canada provided North American surface meteorological data from national monitoring networks. The Data Support Section of the National Center for Atmospheric Research (NCAR) in Boulder, Colorado provided European surface meteorological data. The WMO World Ozone and Ultraviolet Data Centre (WOUDC) and its data-contributing agencies provided vertical meteorological profiles from both the North American and European ozonesonde networks. PRISM Climate Group, Oregon State University (http://www.prismclimate.org**)** provided the gridded observed precipitation. Boundary layer height data from radiosoundings at Lindenberg for 2006 have been kindly provided by Frank Beyrich from Meteorological Observatory Lindenberg - Richard-Aßmann-Observatory of the German Meteorological Service (DWD). The authors would also like to thank following individuals for their assistance: Junhua Zhang and Radenko Pavlovic of Environment Canada, Ana Isabel Miranda and Alexandra Monteiro of the University of Aveiro, Luksa Kraljevic from the meteorological service in Croatia, and Anna Carlin Benedictow. The present study is partially funded by the Center for

Energy, Environment and Health, financed by The Danish Strategic Research Program on Sustainable Energy under contract no 2104-06-0027. French teams acknowledge the LEFE/CHAT program of the Institut National de Sciences de l'Univers (INSU) through the support of the national AQMEII.fr project grant.

675 ## References

676 Amann, M., Bertok, I., Cofala, J., Gyarfas, F.., Heyes, C., Klimont, Z., Schöpp, W., Winiwater, W., 2005.
677 Clean Air for Europe (CAFE) program final report, Int. Inst. for Appl. Syst. Anal., Laxenburg, Austria.

678 Appel, K.W., Roselle, S.J., Gilliam, R.C., Pleim, J.E., 2010. Sensitivity of the Community Multiscale Air
679 Quality (CMAQ) model v4.7 results for the eastern United States to MM5 and WRF meteorological
680 drivers. Geoscientific Model Development 3, 169-188.

681 Arguez, A., Waple, A.M., Sanchez-Lugo, A.M., 2007. State of the climate in 2006: Executive summary.
682 Bulletin of the American Meteorological Society 88, 929-932 + Supplement.

683 Balk, T., J. Kukkonen, K. Karatzas, T. Bassoukos, V. Epitropou (2010) A European open access chemical
684 weather forecasting portal, Atmospheric Environment, doi:10.1016/j.atmosenv.2010.09.058

685 Belair, S., Crevier, L.-P., Mailhot, J., Bilodeau, B., Delage, Y., 2003. Operational implementation of
686 the ISBA land surface scheme in the Canadian regional weather forecast model. Part I: Warm season
687 results. Journal of Hydrometeorology 4, 352–370.

688 Betts, A. K., and M. J. Miller, 1993: The Betts-Miller scheme. The representation of cumulus
689 convection in numerical models, K. A. Emanuel and D. J. Raymond, Eds., Amer. Meteor. Soc., 246 pp.

690 Bianconi R., S. Galmarini and R. Bellasio, A WWW-based decision support system for the
691 management of accidental releases of radionuclides in the atmosphere. *Environmental Modelling &*
692 *Software*, **19**, 401–411, 2004.

693 Biswas, J., Rao, S.T., 2001. Uncertainties in episodic ozone modeling stemming from uncertainties in
694 the meteorological fields. Journal of Applied Meteorology 40, 117-136.

695 Brandt, J., Christensen, J.H., Frohn, L.M., Berkowicz, R., 2001. Operational air pollution forecasts
696 from regional scale to urban street scale. Part 2: Performance evaluation. Physics and Chemistry of
697 the Earth, Part B., 26, 825–830, doi:10.1016/S1464-1909(01)00092-2.

698 Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other
699 components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Applied
700 Mechanics Reviews, 59, 51-77.

701 Cattiaux, J., Vautard, R., Yiou, P., 2009. Origins of the extremely warm European fall of 2006.
702 Geophysical Research Letters 36, doi: 10.1029/2009GL037339, 5 pp.

703 Chen, F., and J. Dudhia, 2001: Coupling an advanced land-surface/hydrology model with the Penn
704 State/NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.,*
705 129, 569–585.

706 Chiriaco, M., Vautard, R**.**, Chepfer, H., Haeffelin, M., Wanherdrick, Y., Morille, Y., Protat, A., Dudhia,
707 J., 2006. The ability of MM5 to simulate ice clouds: systematic comparison between simulated and
708 measured fluxes and lidar/radar profiles at SIRTA atmospheric observatory. Monthly Weather
709 Review 134, 897-918.

710 Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A., 1998a. The
711 operational CMC/MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations
712 and formulation. Monthly Weather Review 126, 1373-1395.

713 Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., Staniforth, A., 1998b. The
714 operational CMC/MRB Global Environmental Multiscale (GEM) model. Part II: Results. Monthly
715 Weather Review 126, 1397-1418.

716 de Meij, A., Gzella, A., Cuvelier, C., Thunis, P., Bessagnet, B., Vinuesa, J.F., Menut, L., Kelder, H.M.,
717 2009. The impact of MM5 and WRF meteorology over complex terrain on CHIMERE model
718 calculations. Atmospheric Chemistry and Physics 9, 6611-6632.

719 Doms, G., Schättler, U., 2004. A description of the nonhydrostatic regional COSMO-Model. Part II:
720 Physical parametrizations. Deutscher Wetterdienst, Offenbach [available from http://www.cosmo-
721 model.org].

722 Doms, G., Forstner, J., Heise, E., Herzog, H.-J., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin,
723 R., Schulz, J.-P., Vogel, G., 2007. A description of the nonhydrostatic regional COSMO-Model. Part II:
724 Physical parametrizations. Deutscher Wetterdienst, Offenbach [available from http://www.cosmo-
725 model.org].

726 Dudhia, J., 1993. A nonhydrostatic version of the Penn State–NCAR mesoscale model: Validation
727 tests and simulation of an Atlantic cyclone and cold front. Monthly Weather Review 121, 1493–
728 1513.

729 Dudhia, J., 1996: A multi-layer soil temperature model for MM5. Preprints, The Sixth PSU/NCAR
730 Mesoscale Model Users' Workshop, 22-24 July 1996, Boulder, Colorado, 49-50. Available from
731 http://www.mmm.ucar.edu/mm5/mm5v2/whatisnewinv2.html.

732 Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003.
733 Implementation of Noah land surface model advances in the National Centers for Environmental
734 Prediction operational mesoscale Eta model. Journal of Geophysical Research 108, D22/8851.

735 Galmarini S., Bianconi R., Bellasio R., Graziani G., Forecasting consequences of accidental releases
736 from ensemble dispersion modelling, 2001, *Journal of Environmental Radioactivity* 57, 203-219.

737 Galmarini S., , R. Bianconi, W. Klug, T. Mikkelsen, R. Addis, S. Andronopoulos, P. Astrup, A. Baklanov,
738 J. Bartnicki, J. C. Bartzis, R. Bellasio, F. Bompay, R. Buckley, M. Bouzom, H. Champion, R. D'Amours, E.
739 Davakis, H. Eleveld, G. T. Geertsema, H. Glaab, M. Kollax, M. Ilvonen, A. Manning, U. Pechinger, C.
740 Persson, E. Polreich, S. Potempski, M. Prodanova, J. Saltbones, H. Slaper, M. A. Sofiev, D. Syrakov, J.
741 H. Sørensen, L. Van der Auwera, I. Valkama, R. Zelazny, 2004: Can the Confidence in Long Range
742 Atmospheric Transport Models be increased? The Pan European experince of ENSEMBLE*, Radiation*
743 *Protection Dosimetry* 109, 1-2, 19-24.

744 Gego, E., P.S Porter, A. Gilliland, S.T. Rao. Observation-based assessment of the impact of nitrogen
745 oxides emissions reductions on O3 air quality over the eastern United States. Journal of Applied
746 Meteorology and Climatology, 46, 994-1008, 2007.

747 Gilliam, R.C., Hogrefe, C., Rao, S.T., 2006. New methods for evaluating meteorological models used
748 in air quality applications. Atmospheric Environment 40, 5073-5086.

749 Gilliam, R.C., Pleim, J.E., 2010. Performance assessment of new land-surface and planetary boundary
750 layer physics in the WRF-ARW. Journal of Applied Meteorology and Climatology 49, 760-774.
751 Gilliam, R. C., J.M. Godowitch, W. Appel, and ST Rao, 2011: Improving the Characterization of Lower
752 Troposphere Transport with Four Dimensional Data Assimilation. J. Geophys. Res., submitted.

753 Gilliland, A.B., Hogrefe, C., Pinder, R.W., Godowitch, J.M., Foley, K.L., Rao, S.T., 2008.
754 Dynamic evaluation of regional air quality models: Assessing changes in O_3 stemming from changes
755 in emissions and meteorology. Atmospheric Environment 42, 5110-5123.

756 Grasselt, R., Schüttemeyer, D., Warrach-Sagi, K., Ament, F., Simmer, C., 2008. Validation of TERRA-ML
757 with discharge measurements. Meteorologische Zeitschrift, Vol. **17**, No. 6, 763-773.

758 Grell, G., Dudhia, J., Stauffer, D. (1994). A Description of Fifth-Generation Penn State/NCAR
759 Mesoscale Model (MM5), NCAR Tech. Note NCAR/TN-398+STR, Boulder, pp. 122.

760 Grell, G.A., Devenyi, D., 2002. A generalized approach to parameterizing convection combining
761 ensemble and data assimilation techniques. Geophysical Research Letters 29,
762 doi:10.1029/2002GL015311.

763 Hack, J. J., B. A. Boville, B. P. Briegleb, J. T. Kiehl, P. J. Rasch, and D. L. Williamson, 1993: Description
764 of the NCAR Community Climate Model (CCM2). NCAR Technical Note, NCAR/TN-382+STR

765 Hanna, S.R., Lu, Z., Frey, H.C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., Hansen, D.A.,
766 2001. Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V

767 photochemical grid model applied to the July 1995 OTAG domain. Atmospheric Environment
768 35, 891-903.

769 Hogrefe, C., Rao, S.T., Kasibhatla, P., Kallos, G., Tremback, C.J., Hao, W., Olerud, D., Xiu, A., McHenry,
770 J., Alapaty, K., 2001. Evaluating the performance of regional-scale photochemical modeling systems:
771 Part I - Meteorological predictions. Atmospheric Environment 35, 4159-4174.

772 Hogrefe, C. W. Hao, K. Civerolo, J.-Y. Ku, R.S. Gaza, L. Sedefian, G. Sistla, K. Schere, A. Gilliland, and R.
773 Mathur, Daily Photochemical Simulations of Ozone and Fine Particulates Over New York State:
774 Findings and Challenges, J. Appl. Meteor., 46, 961–979, 2007.

775 Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range
776 forecast model. Mon. Wea. Rev., 124, 2322-2339.

777 Hong, S.-Y., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of
778 entrainment processes. Monthly Weather Review 134, 2318-2341.

779 Honoré, C., Rouil, L., Vautard, R., Beekmann, M., Bessagnet, B., Dufour, A., Elichegaray, C., Flaud, J.-
780 M., Malherbe, L., Meleux, F., Menut, L., Martin, D., Peuch, A., Peuch, V.-H., Poisson, N., 2008.
781 Predictability of European air quality: The assessment of three years of operational forecasts and
782 analyses. J. Geophys. Res., 113, D04301, doi:10.1029/2007JD008761.

783 Lacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A., Collins, W.D., 2008.
784 Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer
785 models. Journal of Geophysical Research 113, D13103, doi:10.1029/2008JD009944.

786 IFS DOCUMENTATION - Cy31r1, Operational implementation 12 September 2006, PART IV: PHYSICAL
787 PROCESSES. ECMWF, Reading, U.K., 2007.
788 http://www.ecmwf.int/research/ifsdocs/CY31r1/PHYSICS/IFSPart4.pdf

789 Janjic, Zavisa I., 1990: The step-mountain coordinate: Physical package. Mon. Wea. Rev., 118, 1429-
790 1443.

791 Janjic, Zavisa I., 1994: The step-mountain eta coordinate model: Further development of the
792 convection, viscous sublayer, and turbulent closure schemes. Mon. Wea. Rev., 122, 927-945.

793 Kain, J.S., Fritsch, J.M., 1990. A one-dimensional entraining/detraining plume model and its
794 application in convective parameterization. Journal of the Atmospheric Sciences 47, 2784-2802.

795 Kain, J.S., Fritsch, J.M., 1993. Convective parameterization for mesoscale models: The Kain–Fritsch
796 scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteorological
797 Monograph No. 46, American Meteorological Society, Boston, 165–170.

798 Kain, J. S., 2004. The Kain-Fritsch convective parameterization: An update. Journal of Applied
799 Meteorology 43, 170–181.

800 Kukkonen, J., T. Balk, D. M. Schultz, A. Baklanov, T. Klein, A. I. Miranda, A. Monteiro M. Hirtl, V.
801 Tarvainen, M. Boy, V.-H. Peuch, A. Poupkou, I. Kioutsioukis, S. Finardi, M. Sofiev, R. Sokhi, K. Lehtinen,
802 K. Karatzas, R. San José, M. Astitha, G. Kallos, M. Schaap, E. Reimer, H. Jakobs, K. Eben, 2011.
803 Operational, regional-scale, chemical weather forecasting models in Europe. Atmos. Chem. and Phys.
804 Discussion, 11, 5985-6162.

805 Li, J., Barker, H.W., 2005. A radiation algorithm with correlated k-distribution. Part I: Local
806 thermodynamic equilibrium. Journal of the Atmospheric Sciences 62, 286-309.

807 Liu, G., Hogrefe, C. and S.T. Rao, 2003. Evaluating the performance of regional-scale meteorological

808 models:effect of clouds simulation on temperature prediction. Atmospheric Environment, 37, 1425-
809 1433.

810 Majewski, D., Liermann, D., Prohl, P., Ritter, B., Buchhold, M., Hanisch, T., Paul, G., Wergen, W.,
811 Baumgardner, J., 2002. The operational global icosahedral-hexagonal gridpoint model GME:
812 Description and high-resolution tests. Monthly Weather Review 130, 319–338.

813 Matthias, Volker, Builtjes, Peter, Chemel, Charles, Friese, Elmar, Galmarini, Stefano, Pérez, Juan L.,
814 Prank, Marje, Quante, Markus, Schlünzen, K. Heinke, San Jose, Roberto, Sofiev, Mikhail, Stern, Rainer,
815 Struzewska, Joanna, Vautard, Robert, Wolke, Ralf: Comparative analysis of chemistry transport
816 simulations of a particulate matter episode over Germany, International Aerosol Conference,
817 Helsinki, 29.8. – 3.9. 2010.

818 McHenry, J.N., Ryan, W.F., Seaman, N.L., Coats Jr., C.J., Pudykiewicz, J., Arunachalam, S., Vukovich,
819 J.M., 2004. A real-time Eulerian photochemical model forecast system. Bulletin of the American
820 Meteorological Society 85, 525-548.

821 McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E.-Y., Gong, W., Bouchet, V.,
822 Ménard, S., Moffet, R., McHenry, J., McQueen, J., Tang, Y., Carmichael, G.R., Pagowski, M., Chan, A.,
823 Dye, T., 2005. Assessment of an ensemble of seven real-time ozone forecasts over Eastern North
824 America during the summer of 2004. Journal of Geophysical Research 110, D21307,
825 doi:10.1029/2005JD005858, 16 pp.

826 Menut, L. and B. Bessagnet (2010) Atmospheric composition forecasting in Europe. Ann. Geophys.,
827 28, 61-74.

828 Mlawer, E.J., Taubman, S.J., Brown, P.D., Lacono, M.J., Clough, S.A., 1997. RRTM, a validated
829 correlated-k model for the longwave. Journal of Geophysical Research 102, 16663-16682.

830 Monks, P. S., C. Granier, S. Fuzzi, A. Stohl, M. Williams, H. Akimoto, M. Amman, A.Baklanov, U.
831 Baltensperger, I. Bey, N. Blake, R.S. Blake, K. Carslaw, O.R. Cooper, F. Dentener, E. Fragkou, G. Frost,
832 S. Generoso, P. Ginoux, V. Grewe, A. Guenther, H.C., Hansson, S. Henne, J. Hjorth, A. Hofzumahaus,
833 H. Huntrieser, M.E. Jenkin, J. Kaiser, M. Kanakidou, Z. Klimont, M. Kulmala, M.G. Lawrence, J.D. Lee,
834 C. Liousse, G. McFiggans, A. Metzger, A. Mieville, N. Moussiopoulos, J.J. Orlando, P.I. Palmer, D.
835 Parrish, A. Petzold, U. Platt, U. Poeschl, A.S.H. Prévôt, C.E. Reeves, S. Reiman, Y. Rudich, K. Sellegri, R.
836 Steinbrecher, D. Simpson, H. ten Brink, J. Theloke, G. van der Werf, R. Vautard, V. Vestreng, Ch.

837 Vlachokostas, R. vonGlasow, 2009, Atmospheric composition change – global and regional air quality,
838 Atmospheric Environment, 43, 5268-5350.

839 Morrison, H., Thompson, G., Tatarskii, V., 2009. Impact of cloud microphysics on the development of
840 trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment
841 schemes. Monthly Weather Review 137, 991-1007.

842 Noilhan, J., Planton, S., 1989. A simple parameterization of land surface processes for meteorological
843 models. Monthly Weather Review 117, 536–549.

844 Otte, T.L., Pouliot, G., Pleim, J.E., Young, J.O., Schere, K.L., Wong, D.C., Lee, P.C.S., Tsidulko, M.,
845 McQueen, J.T., Davidson, P., Mathur, R., Chuang, H.-Y., DiMego, G., Seaman, N.L., 2005. Linking the
846 Eta Model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air
847 quality forecasting system. Weather and Forecasting 20, 367-384.

848 Otte, T.L., 2008. The impact of nudging in the meteorological model for retrospective air quality
849 simulations. Part I: Evaluation against national observation networks. Journal of Applied
850 Meteorology and Climatology 47, 1853–1867.

851 Pielke, R.A., Uliasz, M., 1998. Use of meteorological models as input to regional and mesoscale air
852 quality models – limitations and strengths. Atmospheric Environment 32, 1455-1466.

853 Pirovano, G., I. Coll, M. Bedogni, S. Alessandrini, M.P. Costa, V. Gabusi, F. Lasry, L. Menut, R. Vautard,
854 2007: On the influence of meteorological input on photochemical modelling of a severe episode over
855 a coastal area. Atmospheric Environment, 41, 6445-6464.

856 Pleim, J.E., Xiu, A., 2003: Development of a land-surface model. Part II: Data assimilation. Journal of
857 Applied Meteorology, 42, 1811–1822.

858 Pleim, J.E., 2007a. A combined local and nonlocal closure model for the atmospheric boundary layer.
859 Part I: model description and testing. Journal of Applied Meteorology and Climatology 46, 1383–
860 1395.

Pleim, J.E., 2007b. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: application and evaluation in a mesoscale meteorological model. Journal of Applied Meteorology and Climatology 46, 1396–1409.

Pleim J.E., Gilliam, R., 2009. An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land surface model. Journal of Applied Meteorology and Climatology 48, 1362-1376.

Pudykiewicz, J., Benoit, R., Mailhot, J., 1992. Inclusion and verification of a predictive cloud water scheme in a regional weather prediction model. Monthly Weather Review 120, 612–626.

Rao, S.T., Galmarini, S., Puckett, K., 2011. Air Quality Model Evaluation International Initiative (AQMEII): Advancing the state of the science in regional photochemical modeling and its applications. Bulletin of the American Meteorological Society 92, 23-30.

Reisner, J., Rasmussen, R.M., Bruintjes, R.T., 1998. Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. Quarterly Journal of the Royal Meteorological Society 124 B, 1071-1107.

Ritter, B., Geleyn, J.-F., 1992. A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. Monthly Weather Review 120, 303-325.

Rockel, B; Will, A; Hense, A, 2008. The Regional Climate Model COSMO-CLM(CCLM), Meteorologische Zeitschrift, 17, 4, 347-348.

Sass, B.H., Rontu, L., Raisanen, P., 1994. HIRLAM-2 radiation scheme: Documentation and tests. HIRLAM Technical Report No. 16. [Available from: http://hirlam.org/publications/TechReports/index.html].

Schättler, U., Doms, G., Schraff, C., 2009. A description of the nonhydrostatic regional COSMO-Model. Part VII: User's guide. Deutscher Wetterdienst, Offenbach, 2009 [available from http://www.cosmo-model.org].

884 Schrodin, R., Heise, E., 2001. The Multi-Layer Version of the DWD Soil Model TERRA-LM, COSMO
885 Technical Report No. 2, Offenbach.

886 Seaman, N.L., 2000. Meteorological modeling for air-quality assessments. Atmospheric Environment
887 34, 2231-2259.

888 Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, Ph.: Review and
889 intercomparison of operational methods for the determination of the mixing height, Atmos. Environ.,
890 34, 1001-1027, doi:10.1016/S1352-2310(99)00349-0, 2000.

891 Seifert A. and K.D. Beheng, 2001. A double-moment parameterization for simulating autoconversion,
892 accretion and selfcollection. Atmospheric Research 59–60, 265–281.

893 Seifert, A., and S. Crewell, 2008: A revised cloud microphysical parameterization for operational
894 numerical weather prediction using the COSMO model. Proc. 15th International Conference on Clouds
895 and Precipitation. Cancun, Mexico.

896 Sistla, G., Zhou, N., Hao, W., Ku, J.-Y., Rao, S.T., Bornstein, R., Freedman, F., Thunis, P., 1996.
897 Effects of uncertainties in meteorological inputs on Urban Airshed Model predictions and ozone
898 control strategies. Atmospheric Environment 30, 2011-2025.

899 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang,
900 W., Powers, J. G., 2008. *A description of the Advanced Research WRF Version 3*. Technical Note
901 NCAR/TN-475+STR, National Center for Atmospheric Research, Boulder, Colorado, 113 pp. [Available
902 from www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf]

903 Smyth, S., Yin, D., Roth, H., Jiang, W., Moran, M.D., Crevier, L.-P., 2006. The impact of GEM and MM5
904 meteorology on CMAQ air quality modeling results in eastern Canada and the northeastern United
905 States. Journal of Applied Meteorology 45, 1525–1541, DOI: 10.1175/JAM2420.1.

Steppeler, J., Doms, G., Schättler, U., Bitzer, H.W., Gassmann, A., Damrath, U., Gregoric, G., 2003. Meso-gamma scale forecasts using the nonhydrostatic model LM. Meteorology and Atmospheric Physics 107, 4576, doi:10.1007/s00703-001-0592-9, 75–96.

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

von Storch, H; Langenberg, H; Feser, F, 2000. A spectral nudging technique for dynamical downscaling purposes. Monthly Weather Review 128, 10, 3664-3673.

Struzewska, J., Kaminski, J.W., 2008. Formation and transport of photooxidants over Europe during the July 2006 heat wave - Observations and GEM-AQ model simulations. Atmospheric Chemistry and Physics 8, 721-736.

Tarasick, D.W., Moran, M.D., Thompson, A.M., Carey-Smith, T., Rochon, Y., Bouchet, V.S., Gong, W., Makar, P.A., Stroud, C., Ménard, S., Crevier, L.-P., Cousineau, S., Pudykiewicz, J.A., Kallaur, A., Moffet, R., Ménard, R., Robichaud, A., Cooper, O.R., Oltmans, S.J., Witte, J.C., Forbes, G., Johnson, B.J., Merrill, J,. Moody, J.L., Morris, G., Newchurch, M.J., Schmidlin, F.J., Joseph, E., 2007. Comparison of Canadian air quality forecast models with tropospheric ozone profile measurements above mid-latitude North America during the IONS/ICARTT campaign: Evidence for stratospheric input. Journal of Geophysical Research 112, D12S22, doi: 10.1029/2006JD007782.

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

Tiedtke, M., 1989. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon. Wea. Rev., 117, 1779-1799.

Tremblay, A., 2005. The stratiform and convective components of surface precipitation. Journal of the Atmospheric Sciences 62, 1513-1528.

930 Tsyro, S., Støren, E., 1999. New meteorological model for air pollution transport models.
931 EMEP/MSC-W Note 3/99, July 1999.
932 Tustison, B., Harris, D., Foufoula-Georgiou, E., 2001. Scale issues in verification of precipitation
933 forecasts. Journal of Geophysical Research 106, 11775-11784.

934 Wang, X., Zhang, L., Moran, M.D., 2010. Uncertainty assessment of current size-resolved
935 parameterizations for below-cloud particle scavenging by rain. Atmospheric Chemistry and Physics
936 10, 5685-5705, doi:10.5194/acp-10-5685-2010.

937 Xiu, A., Pleim, J.E., 2001. Development of a land-surface model. Part I: Application in a mesoscale
938 meteorological model. Journal of Applied Meteorology 40, 192-209.

939 Yiou, P., Vautard, R., Naveau, P., Cassou, C., 2007. Inconsistency between atmospheric dynamics and
940 temperatures during the exceptional 2006/2007 fall/winter and recent warming in Europe.
941 Geophysical Research Letters 34, L21808, doi:10.1029/2007GL031981.

942 Zhang, F., Bei, N., Nielsen-Gammon, J.W., Li, G., Zhang, R., Stuart, A., Aksoy, A., 2007. Impacts of
943 meteorological uncertainties on ozone pollution predictability estimated through meteorological and
944 photochemical ensemble forecasts. Journal of Geophysical Research 112, art. no. D04304,

945

Figure Captions

Figure 1: Seasonal 850 hPa temperature anomalies (K) for 2006.

Figure 2: Seasonal 500 hPa geopotential height anomalies (dam) for 2006.

Figure 3: Six subregions selected for model and observation comparisons: (left panel) North America (24°N-54°N, 130°W-60°W); (right panel) Europe (30°N-70°N;15°W-30°E). The exact subregion boundaries are the following: (1) NA1, 31°N-42°N, 125°W-112°W; (2) NA2, 25°N-37°N, 104°W-90°W; (3) NA3, 36.5°N-48.5°N, 85°W-69°W; (4) EU1, 42°N-60°N, 10°W-5°E; (5) EU2, 46°N-56°N, 5°E-25°E; and (6) EU3, 43°N-46°N, 7°E-15°E. Dots indicate the location of the observation stations considered in this study. Sites used for profile calculations, where ozone soundings are launched are marked with a "+" sign.

Figure 4: Monthly averages of subregional mean wind speeds as observed (thick solid black lines) and as simulated from the various meteorological models used in AQMEII.

Figure 5: Mean annual diurnal cycle of wind speed by subregion as observed (thick solid black lines) and as simulated from the various meteorological models used in AQMEII.

Figure 6: Taylor plots for the simulation of daily wind speed over each continent (left panel: NA; right panel: EU). Each symbol type stands for a subregion. The amplitude of variability is the radial distance to origin. The amplitude of observation for a given subregion is shown by the symbol on the x axis. Larger symbols indicate the skill of the ensemble mean (open symbol) and the ensemble median (solid symbol).

Figure 7: Spatial distribution of the mean annual wind speed at 10 m as observed at measurement sites and simulated over the two continents by WRF for North America and MM5 for Europe.

Figure 8: Comparison of vertical profiles of wind speed for NA and EU soundings. The observations are based on irregular ozone soundings at six stations for EU and six stations for NA. The statistical

parameters bias, root mean square error and correlation were derived for time series in given altitudes.

Figure 9: Annual times series of PBL heights at Lindenberg, Germany, derived from radiosondes (obs) and from two selected models at 12UTC and 18 UTC.

Figure 10: Simulated and observed monthly mean 2-m temperature values for the six subregions.

Figure 11: Same as Figure 5 for the mean diurnal cycle of 2-m temperature.

Figure 12: same as Figure 8 but for temperature profiles

Figure 13: Simulated vs. observed 99.5^{th} centiles of area-average hourly temperature distributions for each continental subregion of Figure 3. Each point represents a model and each color a different subregion. The area names are indicated on the figure.

Figure 14: Left panels: Seasonal cycle of relative humidity (%) at 2 m as averaged over observations (thick black line) or model simulations (other lines) for three NA subregions; Right panels: As in left panels for hourly precipitation rate (in mm).

Figure 15: Same as Figure 8 but for Relative humidity

Figure 16: Left panels: Mean monthly mid-day (hours 10-14 local time) surface shortwave radiation (W m^{-2}) predicted by four meteorological models at center points of three NA subregions [NA1: 36.5°N, 118.5°W; NA2: 31°N, 97°W; NA3: 42.5°N, 77°W]; Right panels: Same plots for nine models and center points of three EU subregions [EU1: 51°N, 2.5°W; EU2: 51°N, 15°E; EU3: 44.5°N, 11°E].

Figure 17: Left panels**:** Mean annual diurnal cycle (UTC) of surface shortwave radiation (W m^{-2}) predicted by four meteorological models at center points of three NA subregions [NA1: 36.5°N, 118.5°W; NA2: 31°N, 97°W; NA3: 42.5°N, 77°W]; Right panels: Same but for nine models and center points of three EU subregions [EU1: 51°N, 2.5°W; EU2: 51°N, 15°E; EU3: 44.5°N, 11°E].

991 **Figure 18:** Same as Fig. 16 but for mean monthly standard deviation of hourly surface shortwave
992 radiation (W m^{-2}) for mid-day period (hours 10-14 LT).

993 **Figure 19**: Histograms of percentage occurrence of observed and predicted hourly precipitation
994 amount (mm h^{-1}) for the (a) winter and (b) summer season for the NA1 subregion, (c) winter and (d)
995 summer season for the NA2 subregion, and (e) winter and (f) summer season for the NA3 subregion.

996 **Figure 20**: Spatial distribution of seasonal accumulated precipitation (mm) for the US1 WRF
997 simulation and observations, which are represented by the Parameter-elevation Regressions on
998 Independent Slopes Model (PRISM). Left panels represent winter (DJF) and right summer (JJA).

999

1000 **Tables**

Research Group that operated simulations and processing	Model	Appx. horiz. resol. (km)	# of vertical levels; # of levels < 1 km; model top;	Key parameterizations LSM = Land Surface Model; PBL = Planetary Boundary Layer Scheme; MP = Microphysics Scheme; CuP = Cumulus Parameterization; LWR = Long-Wave Radiation Scheme	Analysis and initialization AI) , integration (IN), boundary conditions (BC), data assimilation (DA)
North America					
Environment Canada (CA)	GEM (Côté et al., 1998a,b)	15 (0.1375°)	58 8 10 hPa	LSM: ISBA (Noilhan and Planton, 1989; Belair et al., 2003) PBL: TKE scheme (Belair et al., 2003) MP: Sundqvist (Pudykiewicz et al., 1992) CuP: KFC (Kain and Fritsch, 1990, 1993) LWR: Li and Barker (2005)	AI: Global 0.33° analysis every 6 h IN: 1.25 d segments with 0.25 d overlap BC: None (global variable grid) DA: None
Helmholtz-Zentrum Geesthacht (DE)	COSMO-CLM (Steppeler et al., 2003; Schättler et al., 2009, Rockel et al., 2008)	24	40 11 20 hPa	LSM: multi-layer model TERRA-LM (Grasselt et al., 2008) PBL: TKE closure, Doms and Schaettler, 2004,Doms et al., 2008 MP: Seifert and Beheng, 2001 CuP: Tiedtke, 1989 LWR: Ritter and Geleyn, 1992	AI: 1.875° NCEP1 reanalysis IN: continuous run, 1 month spin up time BC: same as AI DA: Spectral nudging of wind in higher altitudes (von Storch et al., 2000)
Univ. Aarhus (DK)	MM5	50	29 11 100 hPa	LSM: NOAH (Ek et al., 2003) PBL: Eta MY (Janjic, 1990, 1994) MP: mixed phase Reisner 1 (Reisner et al. 1998) CuP: BM (Betts and Miller, 1993) LWR: CCM2 (Hack et al., 1993)	AI/IN/DA: One continuous simulation with grid nudging FDDA using 1º NCEP-FNL global analysis every 6 h. Relaxation/inflow-outflow lateral BCs.
Univ. Aveiro (PT)	MM5 Dudhia, 1993; Grell et al., 1994)	27	23 15 100 hPa	LSM: Five-Layer Soil model (Dudhia, 1996) PBL: MRF (Hong and Pan, 1996) MP: Reisner 2 (Reisner et al., 1998) CuP: Grell and Devenyi (2002) LWR: RRTM (Lacono et al 2008)	AI: 1° NCEP-FNL global analysis every 6 h IN: 5.25 d segments with 0.25 d overlap BC: same as AI DA: Not used
Environmental Protection Agency (US)	WRF (Skamarock et al., 2008)	12	34 14 50 hPa	LSM: PX LSM (Xiu and Pleim, 2001; Pleim and Xiu, 2003) PBL: ACM2 (Pleim, 2007a,b) MP: Morrison et al. (2009) CuS: Kain-Fritsch2 (Kain, 2004)	AI: 12-km NAM analysis + radiosondes every 6 h IN: 5.5 d segments with 0.5 d overlap BC: same as AI DA: V, T, q nudging in atmosphere; T, q

				LWR: RRTMG (Lacono et al 2008)	nudging in soil
Europe					
IFT (DE)	COSMO (Steppeler et al., 2003; Schättler et al., 2009)	24	40 total 9 below 1 km	LSM: multi-layer model TERRA-ML (Grasselt et al. 2008) PBL: prognostic TKE, 2.5 closure scheme (Doms et al. 2008) MP: Kessler type bulk scheme, ice phase, prognostic precipitation (Doms et al. 2007; Seifert and Crewell, 2008) CuP: mass flux scheme of Tiedke (1989) LWR: ∂-two-stream (Ritter and Geleyn 1992)	Initialization and boundary conditions from the GME system (Majewski et al. 2002)
IMK-IFU (DE)	WRF/Chem	22.5	36 total 13 below 1 km	LSM: NOAH (Chen and Dudhia 2001, Ek et al, 2003) LWR: RRTM (Mlawer et al., 1997) CuP: Grell (Grell and Devenyi 2002) PBL : Hong et al. (2006)	Initialization and nudging from NCEP GFS 1° analyses. Nudging above PBL detailed in Gilliam and Pleim (2010). Note: the run was done with aerosol radiation effects (direct and indirect) and also included some aqueous chemical reactions (see Forkel et al., in preparation in this issue).
Helmholtz-Zentrum Geesthacht (DE)	COSMO-CLM (Steppeler et al., 2003; Doms and Schättler, 2004, Rockel et al., 2008)	24	40 11 20 hPa	LSM: Multi-layer model TERRA-LM (Grasselt et al., 2008) PBL: TKE closure, Doms and Schaettler, 2004 MP: Seifert and Beheng, 2001 CuP: Tiedtke, 1989 LWR: Ritter and Geleyn, 1992	AI: 1.875° NCEP1 reanalysis IN: continuous multidecadal run BC: same as AI DA: Spectral nudging of wind in higher altitudes (von Storch et al., 2000)
Univ. Aarhus (DK)	MM5	50	29 11 100 hPa	LSM: NOAH (Ek et al., 2003) PBL: Eta MY (Janjic, 1990, 1994) MP: mixed phase Reisner 1 (Reisner et al. 1998) CuP: BM (Betts and Miller, 1993) LWR: CCM2 (Hack et al., 1993)	AI/IN/DA: One continuous simulation with grid nudging FDDA using 1º NCEP-FNL global analysis every 6 h. Relaxation/inflow-outflow lateral BCs.
FMI	ECMWF IFS	25	4 2 3.5 km	Physics from the IFS forecasting / assimilation system, interpolated to the grid (IFS, 2007)	ECMWF operational global forecasts
TNO	ECMWF IFS	25 (0.5^{o}x0.25^{o})	4 2 3.5 km	Physics from the IFS forecasting / assimilation system, interpolated to the grid (IFS, 2007)	ECMWF operational global forecasts
IPSL	MM5 Dudhia, 1993; Grell et al., 1994)	20	32 9 100 hPa	LSM: NOAH (Ek et al., 2003) CuP: Grell and Devenyi (2002) LWR: RRTM (Mlawer et al., 1997) PBL: MRF PBL scheme	BC, initial conditions and nudging from ECMWF analyses

				MP: Reisner 2 (Reisner et al., 1998)	
Univ. Aveiro	MM5 Dudhia, 1993; Grell et al., 1994)	27	23 14 100 hPa	LSM: Five-Layer Soil model (Dudhia, 1996). CuP: Grell and Devenyi (2002) LWR: RRTM (Mlawer et al., 1997) PBL: MRF PBL scheme MP: Reisner 2 (Reisner et al., 1998)	
Univ. Hertfordshire (UK)	WRF (Skamarock et al., 2008)	18	**52** total **11** below 1 km	LSM: NOAH (Ek et al., 2003) PBL: Hong et al (2006) Microphysics Morrison et al (2009) CuP Grell and Devenyi (2002) LWR: RRTMG (Lacono et al 2008)	BC, initial conditions and nudging from ECMWF analyses
MSC (HR)	PARLAM-PS Tsyro and Støren 1999	50	**20** total **2** below 1 km	Most parameterizations from the HIRLAM model, see description in Sass et al (1994)	BC from ECMWF analyses, then forecasts 4x a day
NOAA	WRF/Chem	22.5	**36** total **13** below 1 km	LWR: RRTM (Mlawer et al., 1997) CuP: Grell (Grell and Devenyi 2002) LSM: NOAH (Chen and Dudhia 2001, Ek et al, 2003) PBL : Hong et al. (2006)	Initialization and nudging from NCEP GFS 1° analyses. Nudging above PBL detailed in Gilliam and Pleim (2010)

Table 1: Summary of some key characteristics of the meteorological models or model configurations participating in AQMEII.

Hour (UTC)	Mean Obs	M1EU	M2EU	M3EU	M4EU	M5EU	M7EU	M8EU	M9EU	M10EU
					Bias (m)					
All hours	628	30	39	-476	-210	-3	-139	-361	-288	-55
0	363	11	34	-326	-246	22	-113	-317	-293	54
6	366	8	76	-326	-132	32	-126	-313	-276	29
12	1078	167	237	-612	72	33	-264	-280	-150	-170
18	705	-68	-193	-638	-538	-99	-52	-532	-434	-134
					RMSE (m)					
All hours		542	464	645	481	358	386	538	471	433
0		443	223	403	310	208	215	394	343	281
6		547	253	412	271	206	209	400	343	253
12		550	589	797	410	402	504	527	478	536
18		615	631	836	768	514	503	750	650	565
					Correlation					
All hours		0,66	0,70	0,56	0,70	0,78	0,77	0,69	0,76	0,62
0		0,41	0,78	0,03	0,59	0,85	0,73	0,09	0,73	0,56
6		0,27	0,69	-0,02	0,45	0,80	0,76	0,02	0,61	0,57
12		0,72	0,73	0,53	0,81	0,79	0,72	0,75	0,78	0,58
18		0,51	0,25	0,19	0,25	0,49	0,67	0,39	0,50	0,35

Table 2: Comparison of simulated PBL heights with observations at Lindenberg, at 0, 6, 12 and 18 UTC, and for all hours. On total, 1457 values were taken into account.

Region	Season	N	Obs	M1NA	M2NA	M3NA	M4NA	M5NA	Ensemble
NA1	DJF	115	93	130	66	158	135	206	139
NA1	JJA	115	16	9	2	37	22	30	20
NA2	DJF	203	106	137	143	123	132	195	146
NA2	JJA	203	125	283	99	274	290	258	241
NA3	DJF	291	152	186	207	194	184	235	201
NA3	JJA	291	208	431	418	314	351	440	391

Table 3: Observed and model-predicted 2006 mean seasonal precipitation accumulations at available measurement stations (mm) in three North American analysis subregions. The "Ensemble" column corresponds to mean of model values.

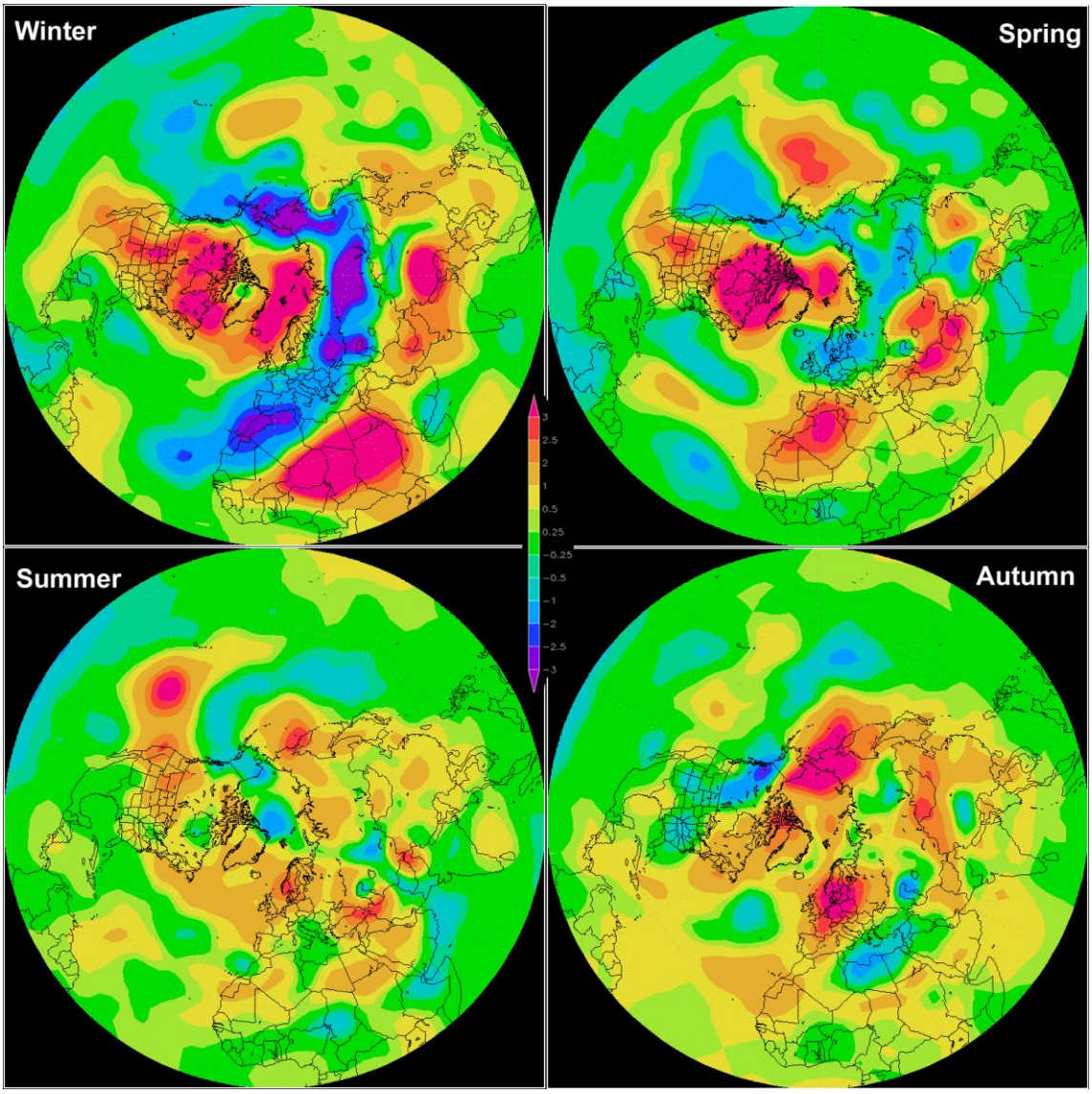

Figure 1

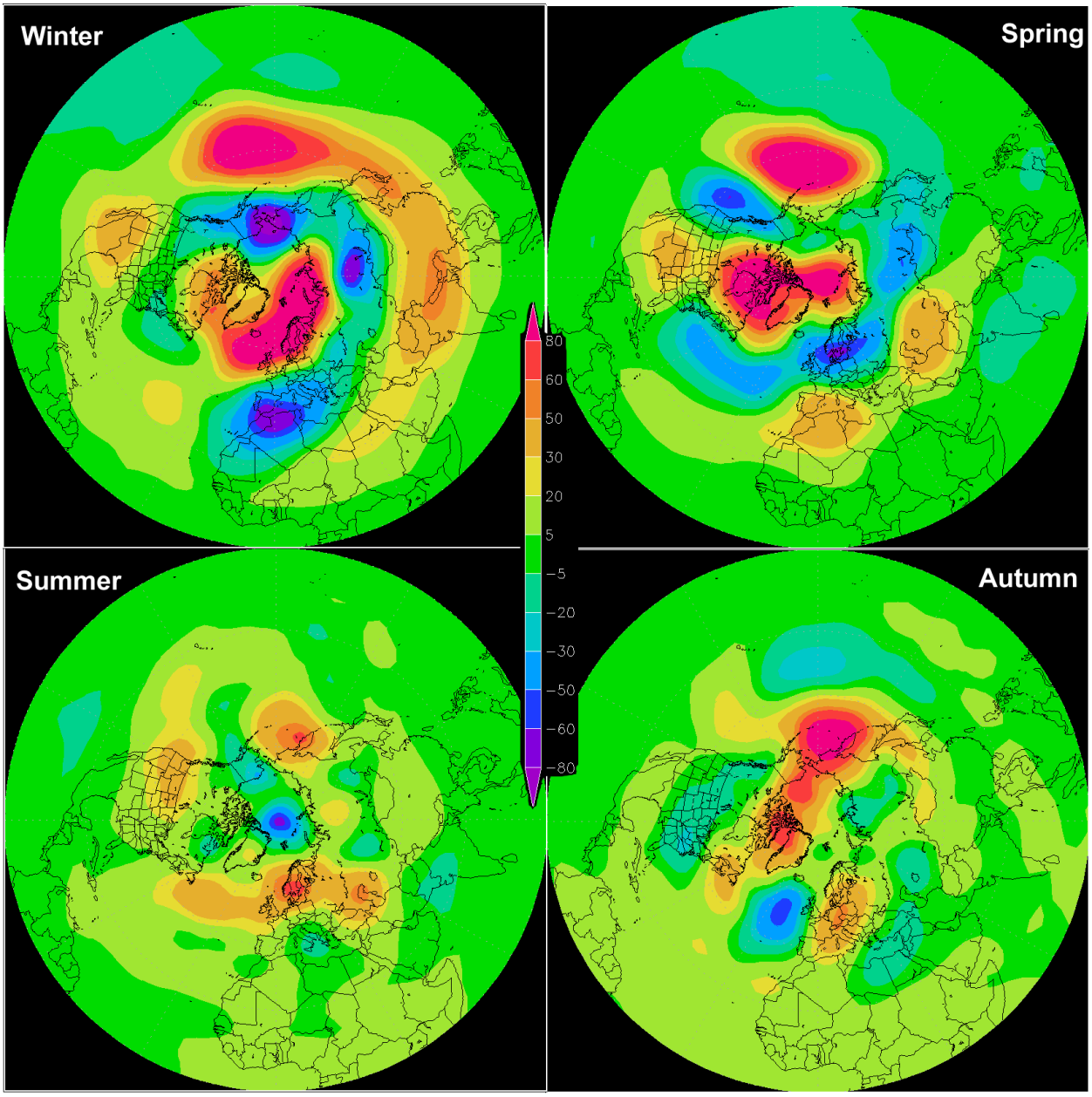

Figure 2

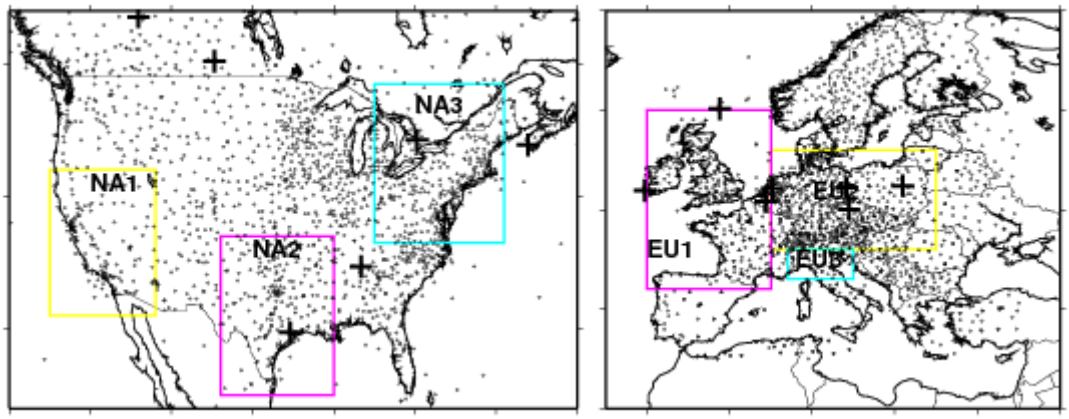

Figure 3

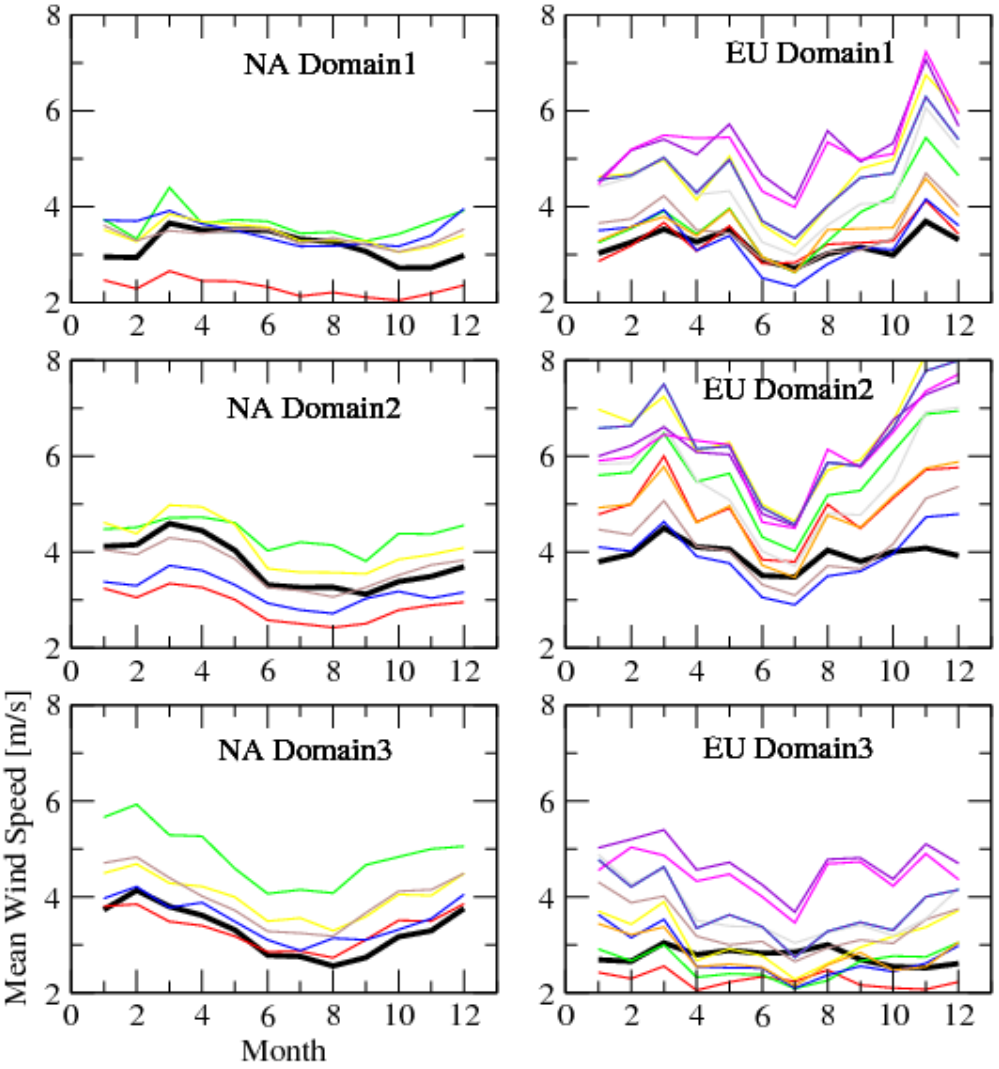
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
Mean Wind Speed [m/s]
Month

Figure 4

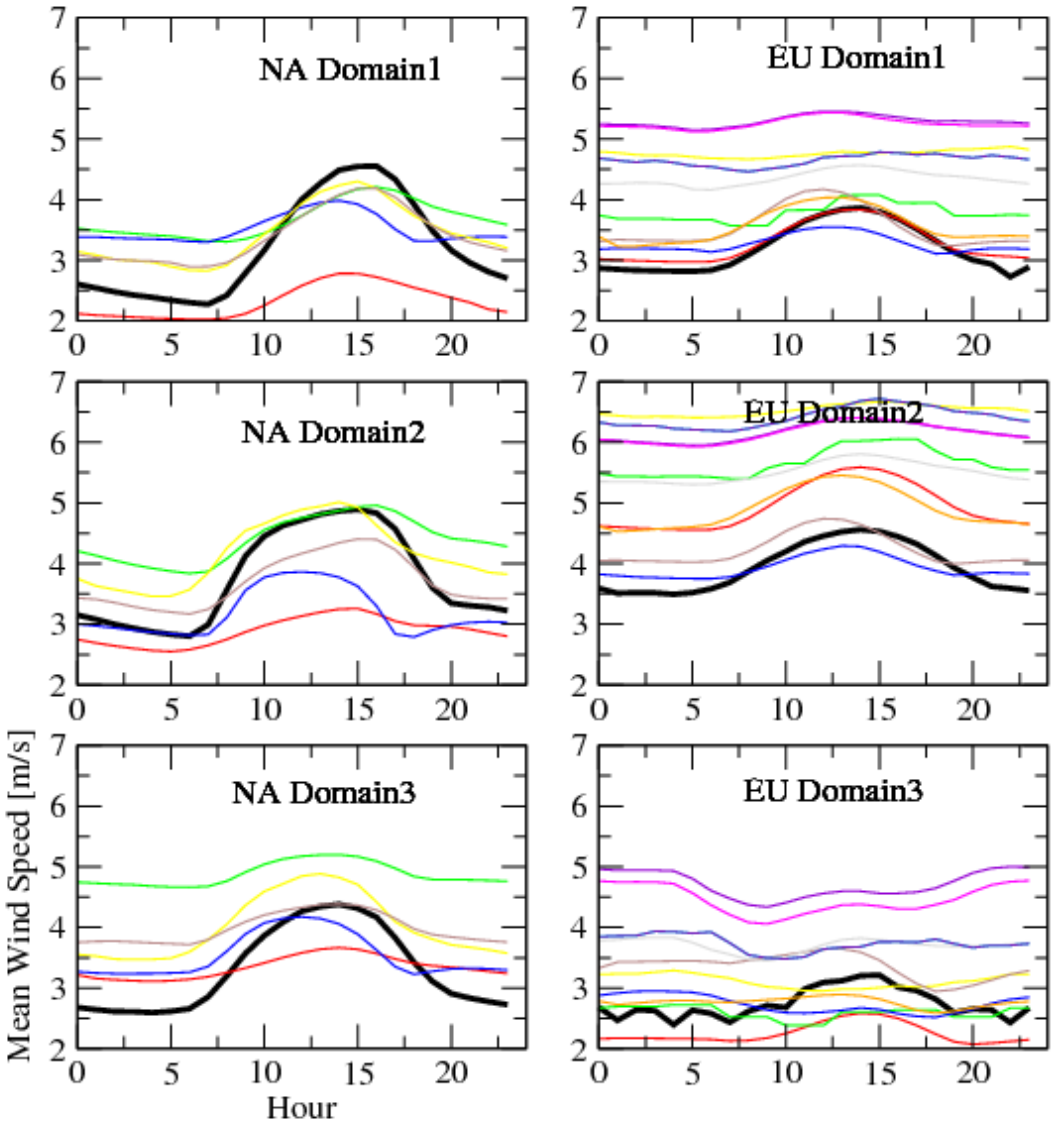
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
Mean Wind Speed [m/s]
Hour

Figure 5

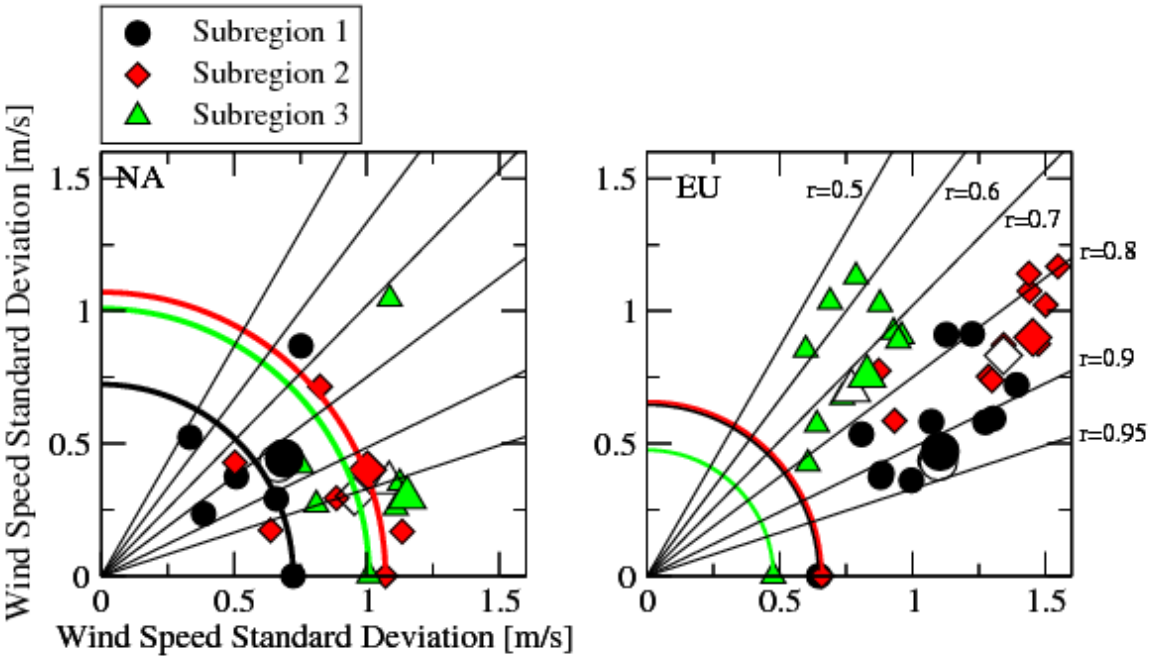

Figure 6

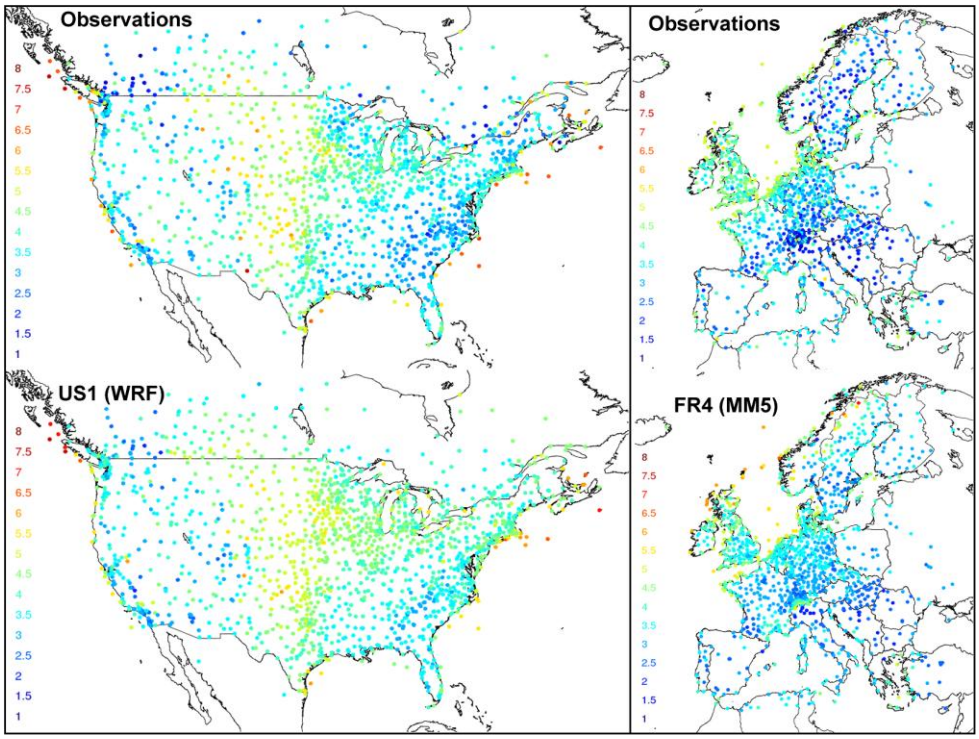

Figure 7

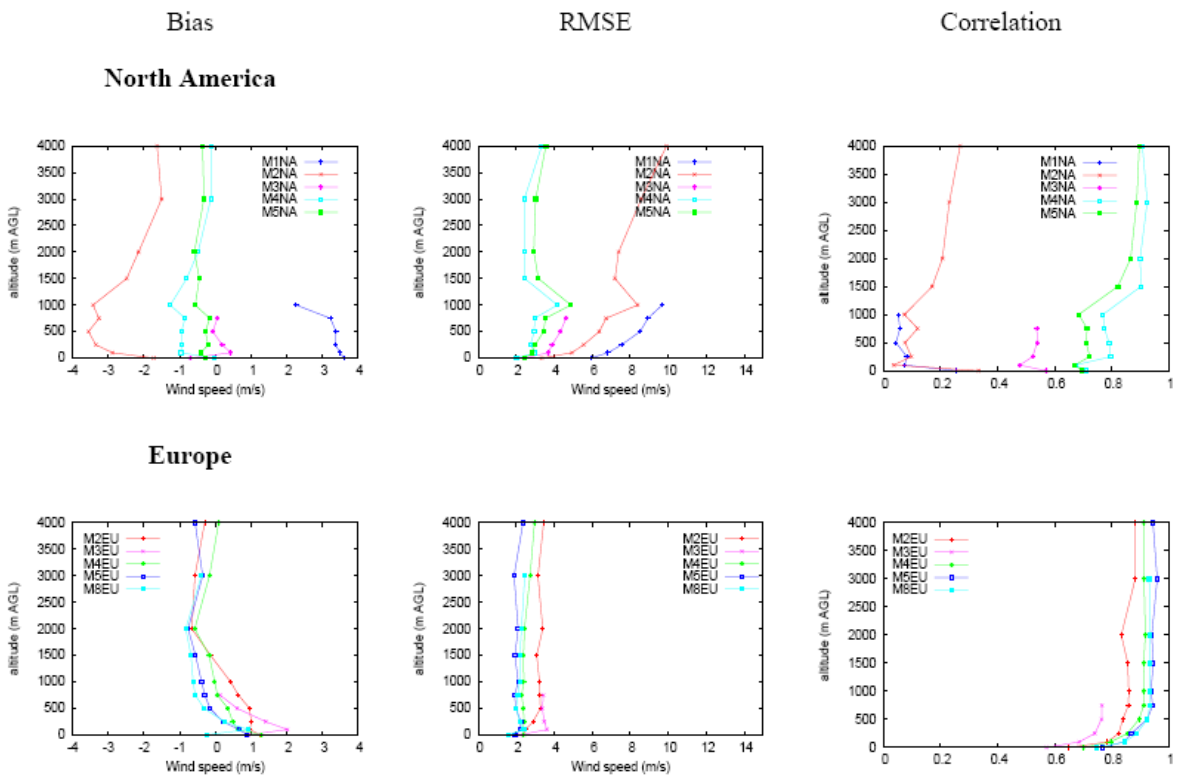

Figure 8

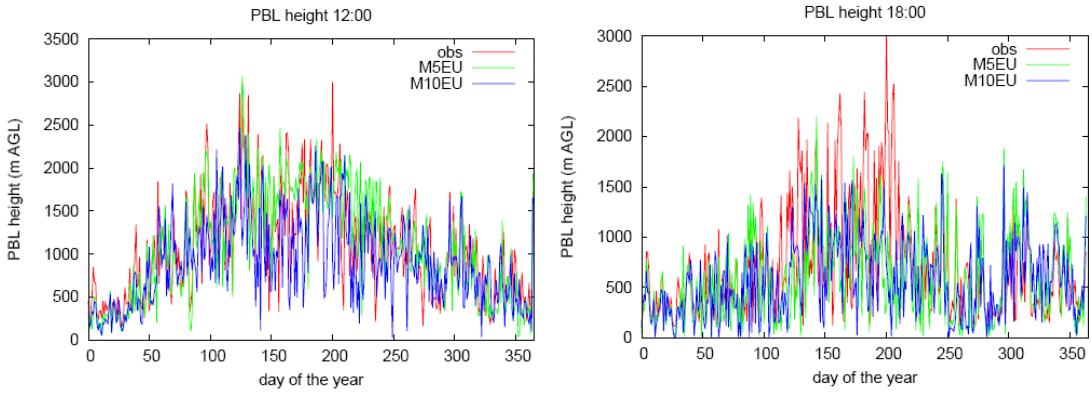

Figure 9

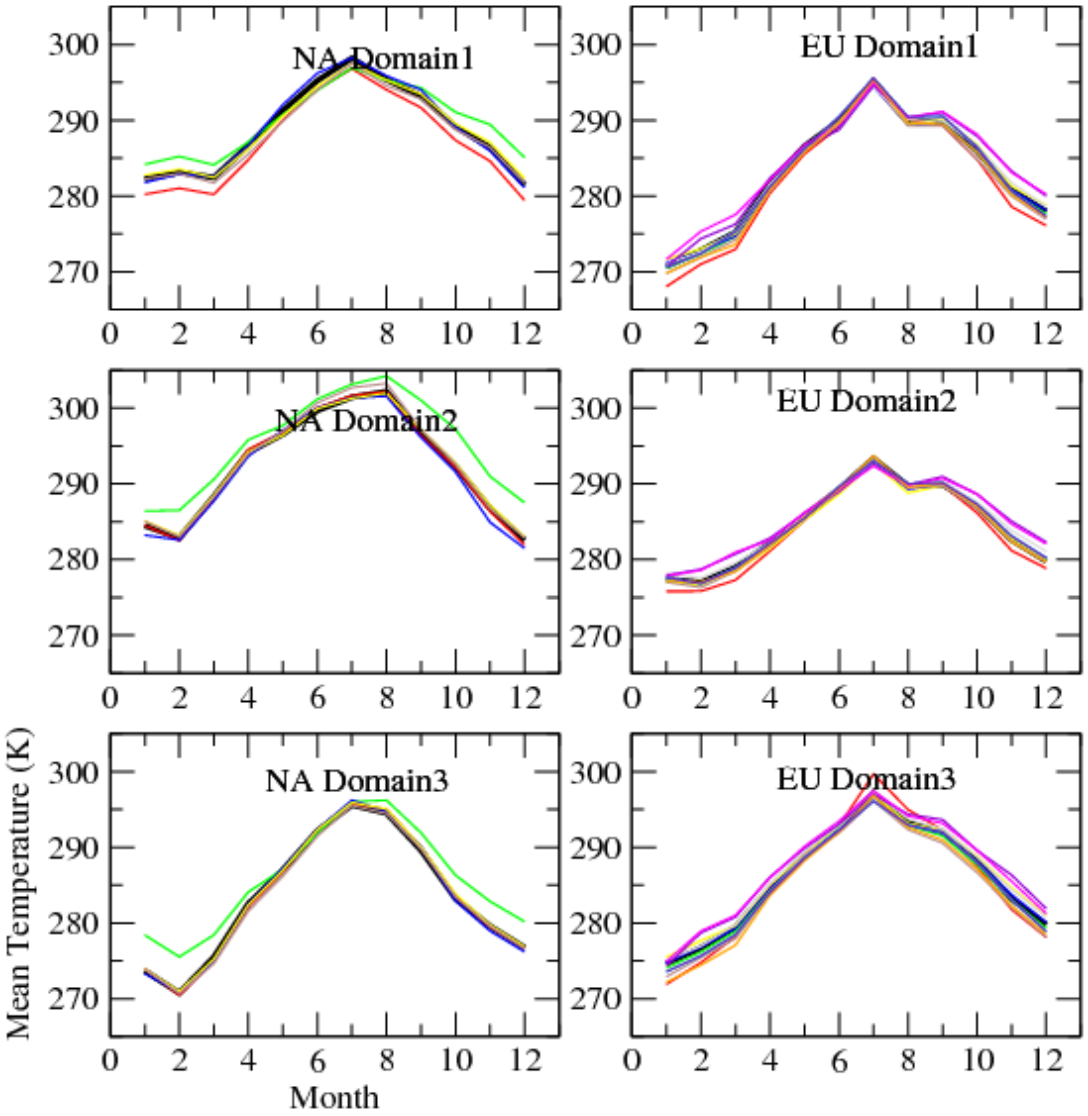
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
Mean Temperature (K)
Month
300
290
280
270
0
2
4
6
8
10
12

Figure 10

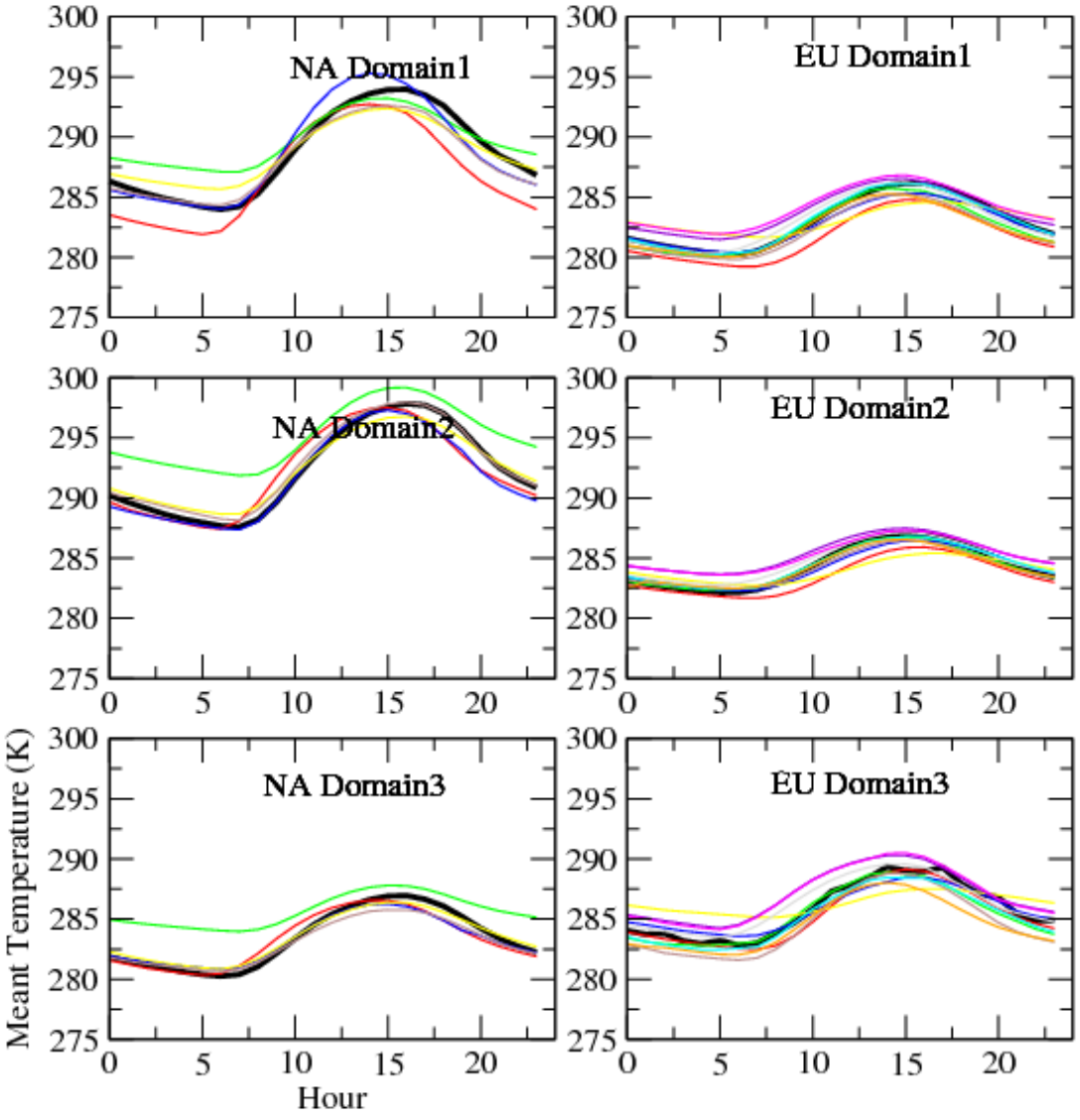
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
Meant Temperature (K)
Hour
300
295
290
285
280
275
0
5
10
15
20

Figure 11

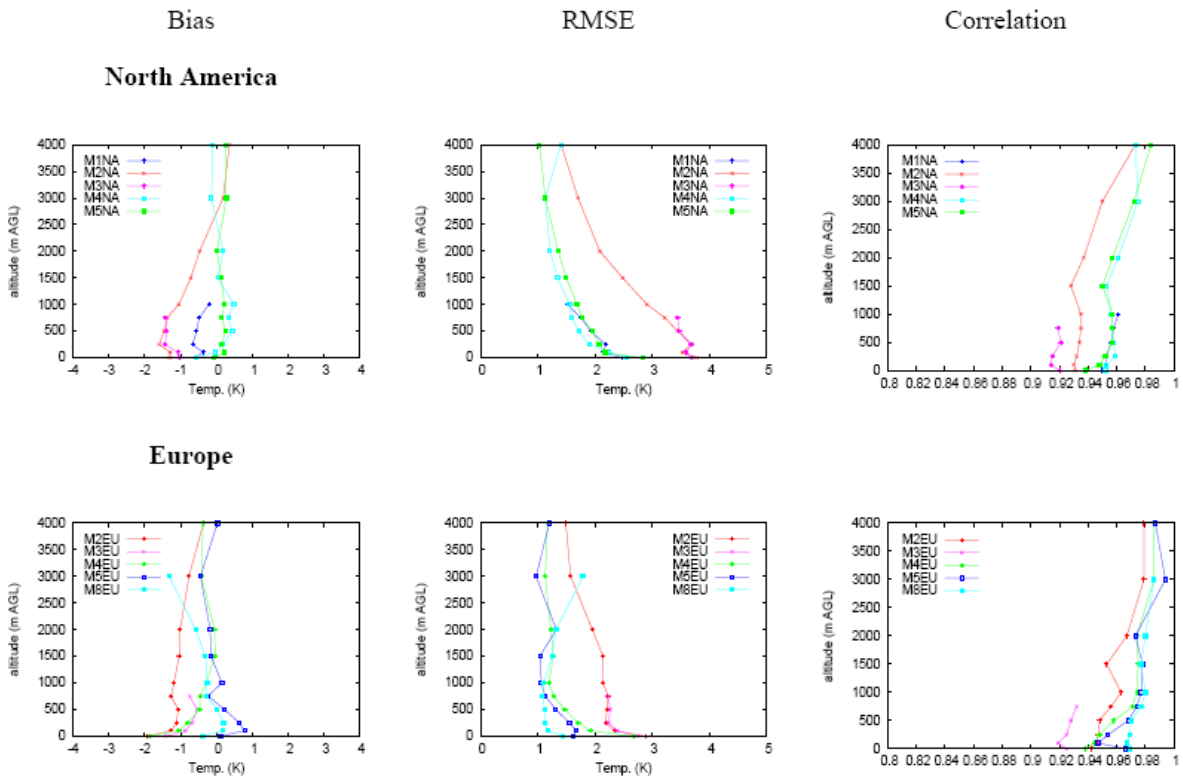
Bias
RMSE
Correlation
North America
Europe
M1NA
M2NA
M3NA
M4NA
M5NA
M2EU
M3EU
M4EU
M5EU
M8EU
altitude (m AGL)
Temp. (K)

Figure 12

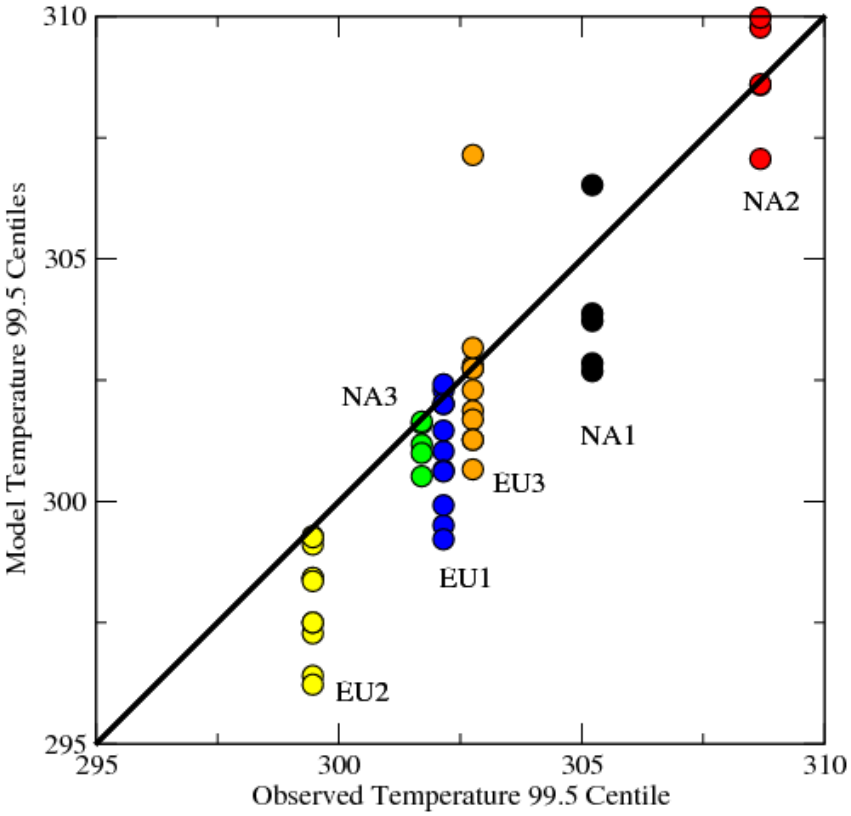
Model Temperature 99.5 Centiles
Observed Temperature 99.5 Centile
310
305
300
295
295
300
305
310
NA2
NA3
NA1
EU3
EU1
EU2

Figure 13

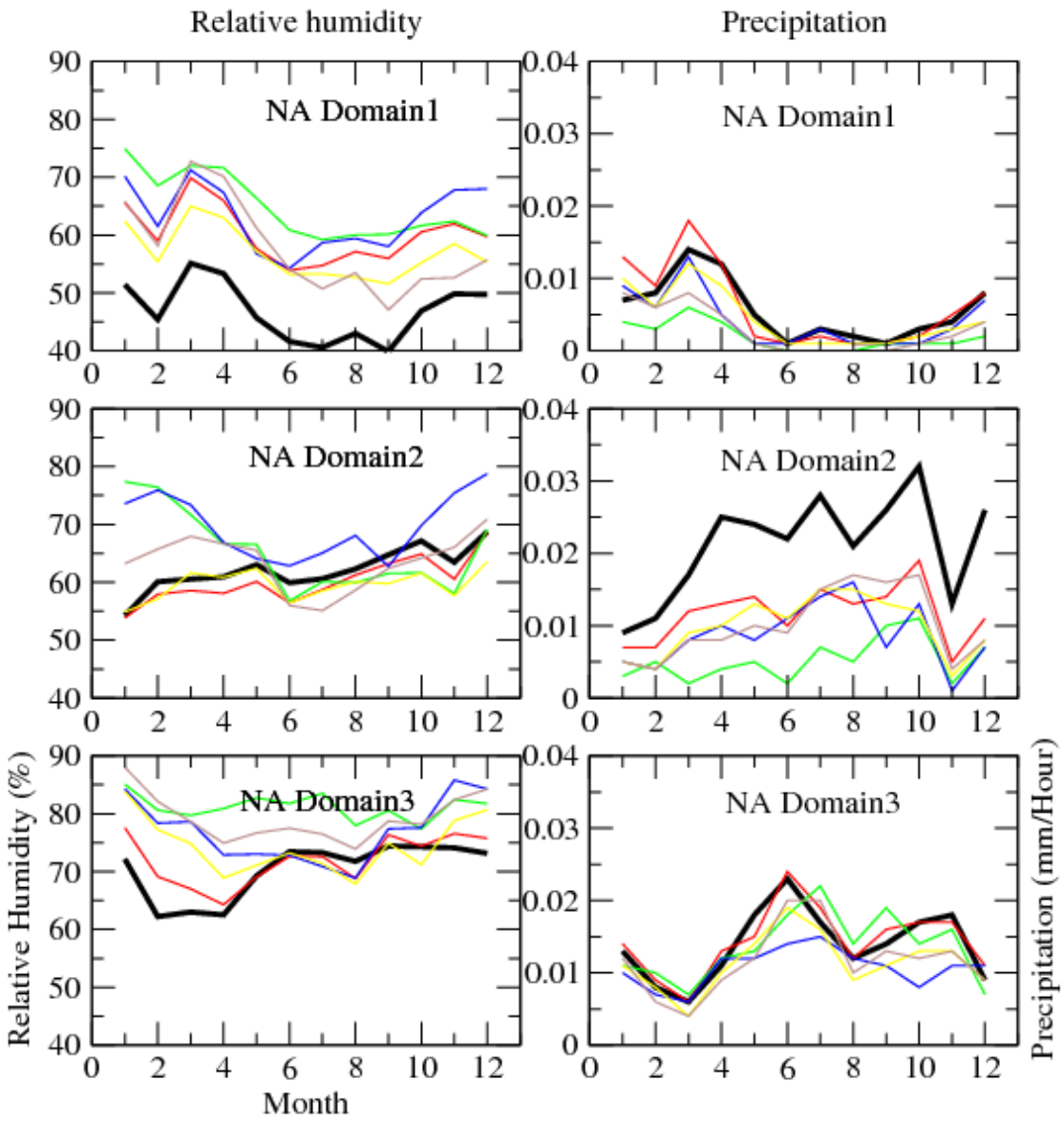
Relative humidity
Precipitation
NA Domain1
NA Domain2
NA Domain3
Relative Humidity (%)
Precipitation (mm/Hour)
Month

Figure 14

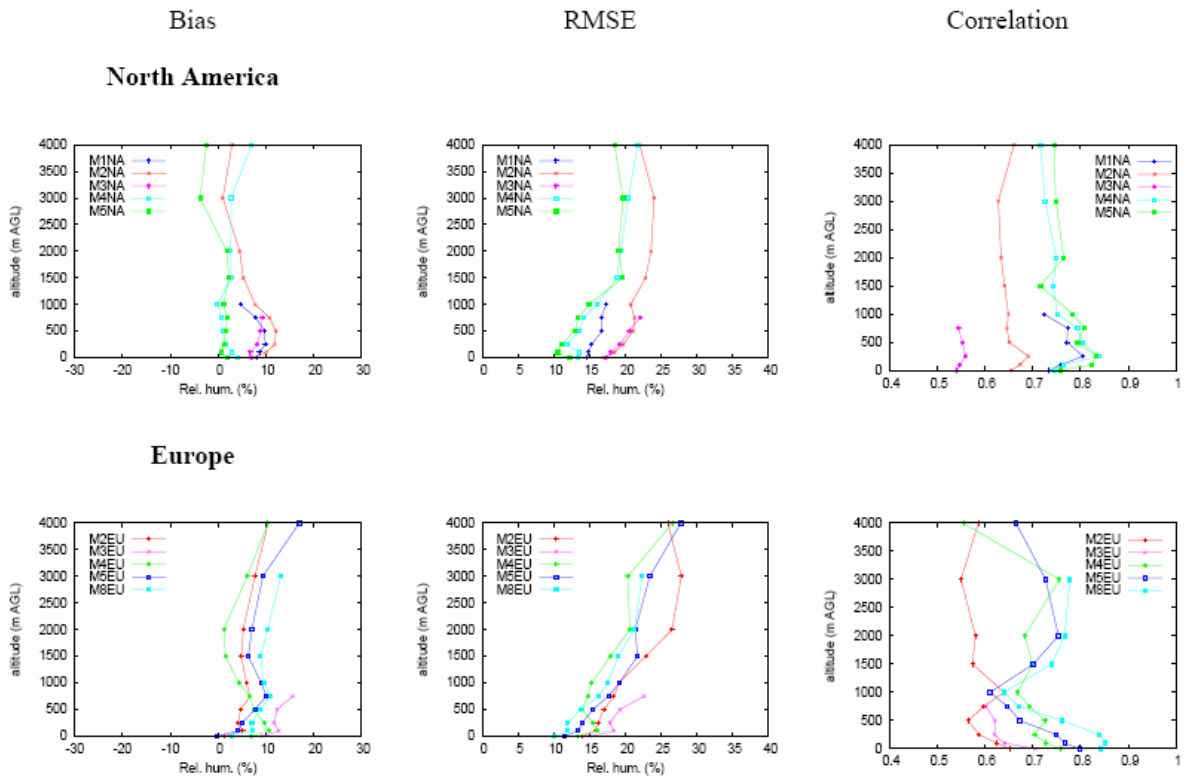
Bias
RMSE
Correlation
North America
Europe
M1NA
M2NA
M3NA
M4NA
M5NA
M2EU
M3EU
M4EU
M5EU
M8EU
altitude (m AGL)
Rel. hum. (%)

Figure 15

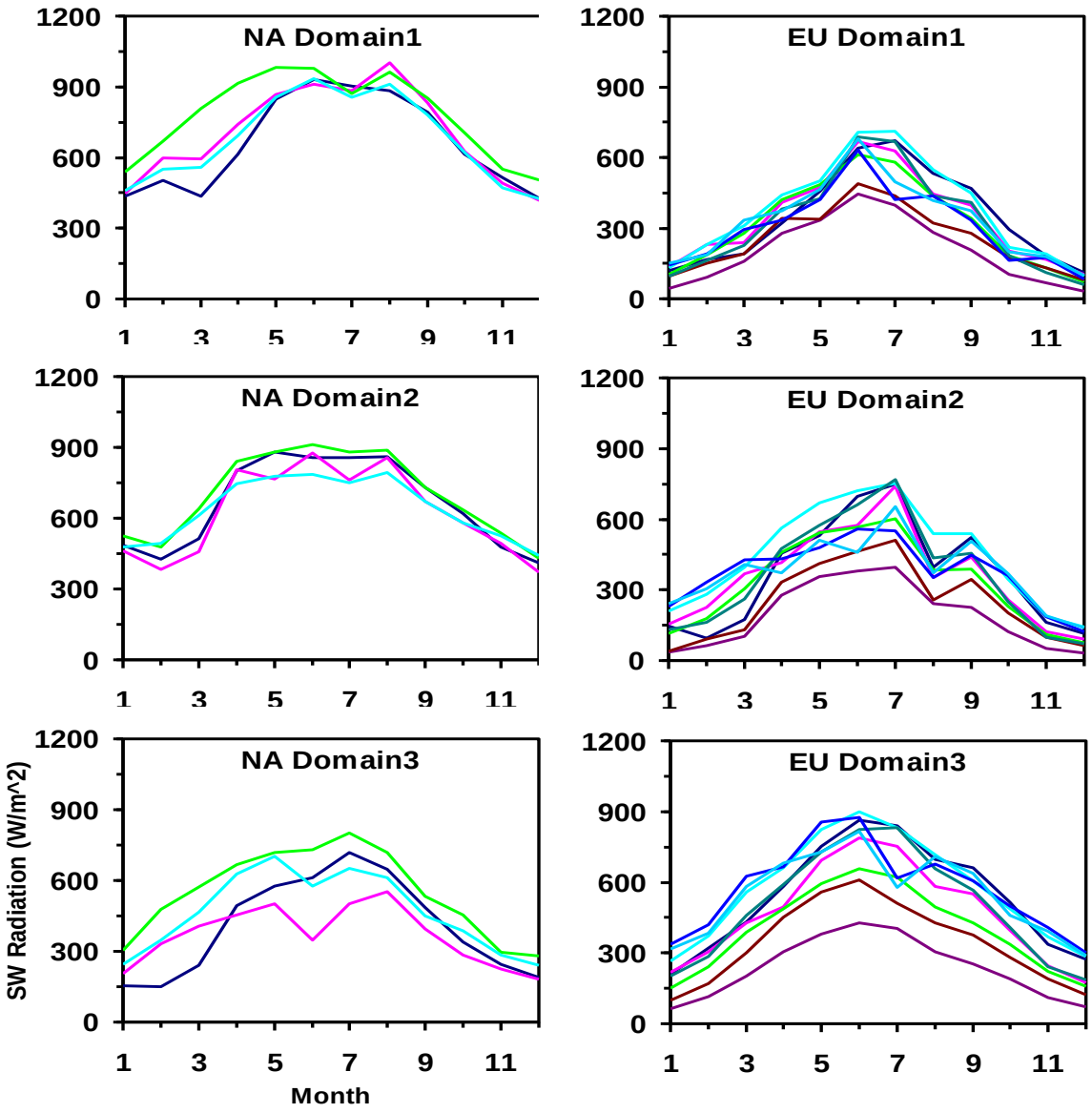
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
SW Radiation (W/m^2)
Month
0
300
600
900
1200
1
3
5
7
9
11

Figure 16

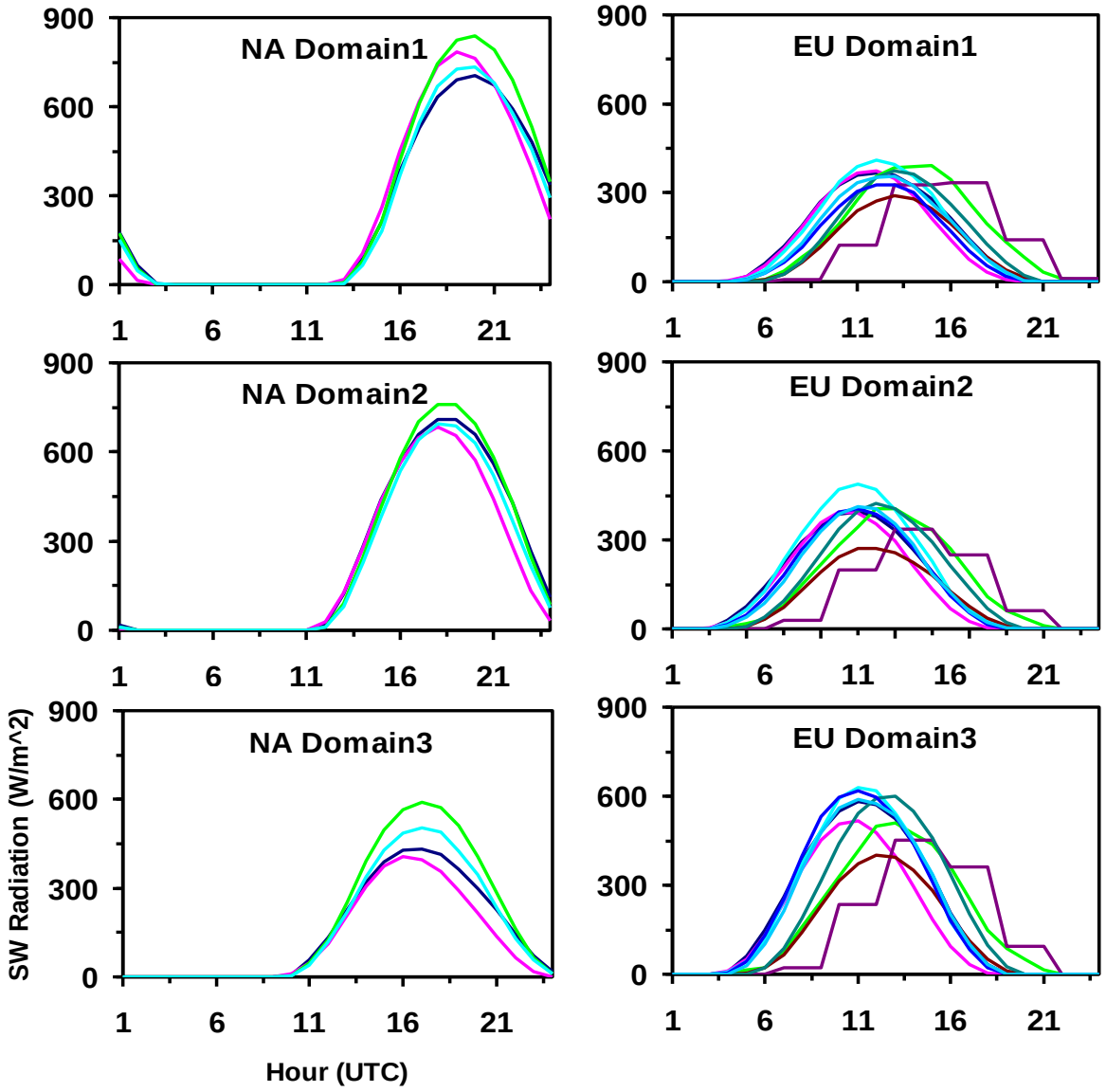
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
900
600
300
0
1
6
11
16
21
SW Radiation (W/m^2)
Hour (UTC)

Figure 17

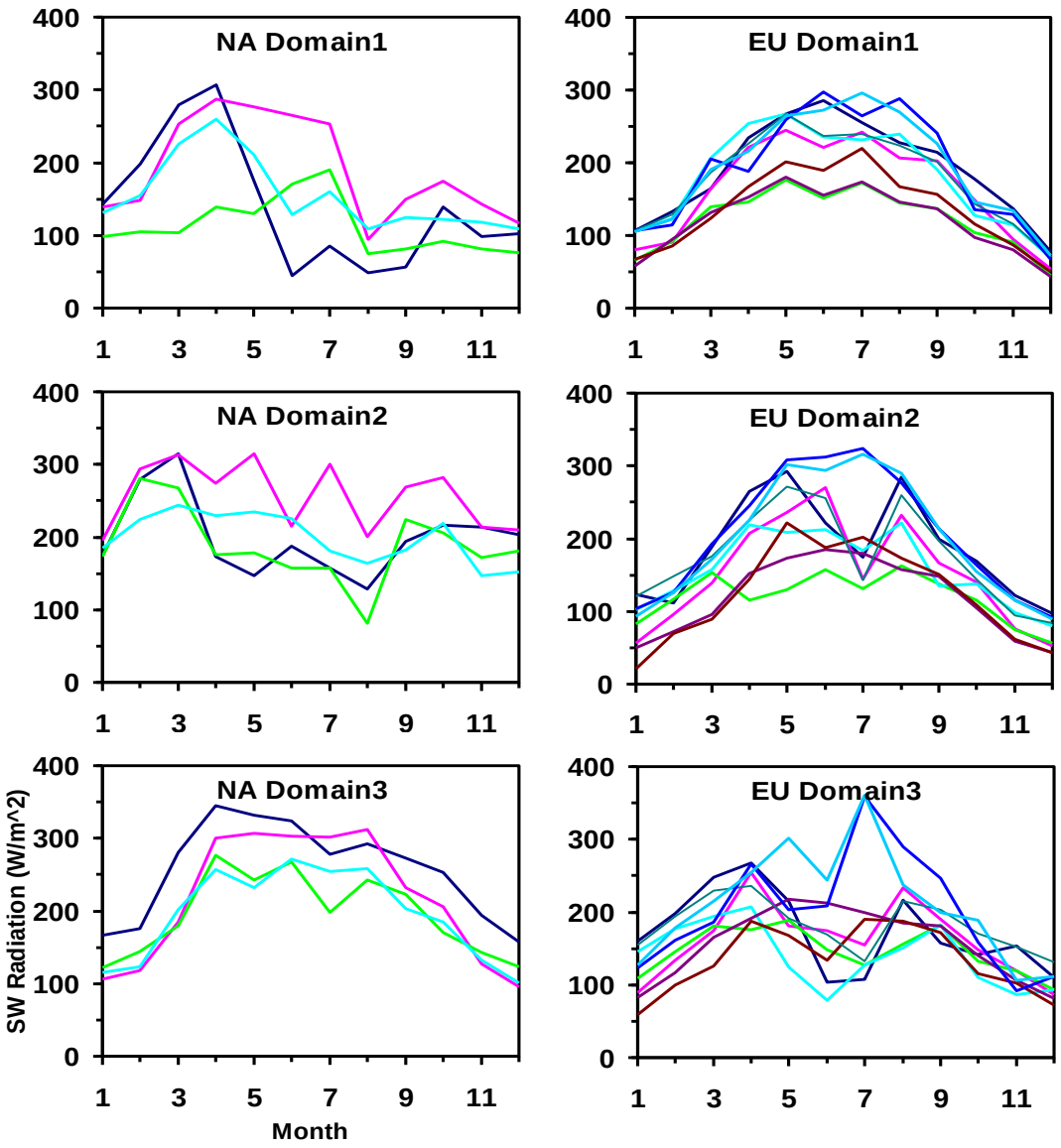
NA Domain1
EU Domain1
NA Domain2
EU Domain2
NA Domain3
EU Domain3
SW Radiation (W/m^2)
Month
0
100
200
300
400
1
3
5
7
9
11

Figure 18

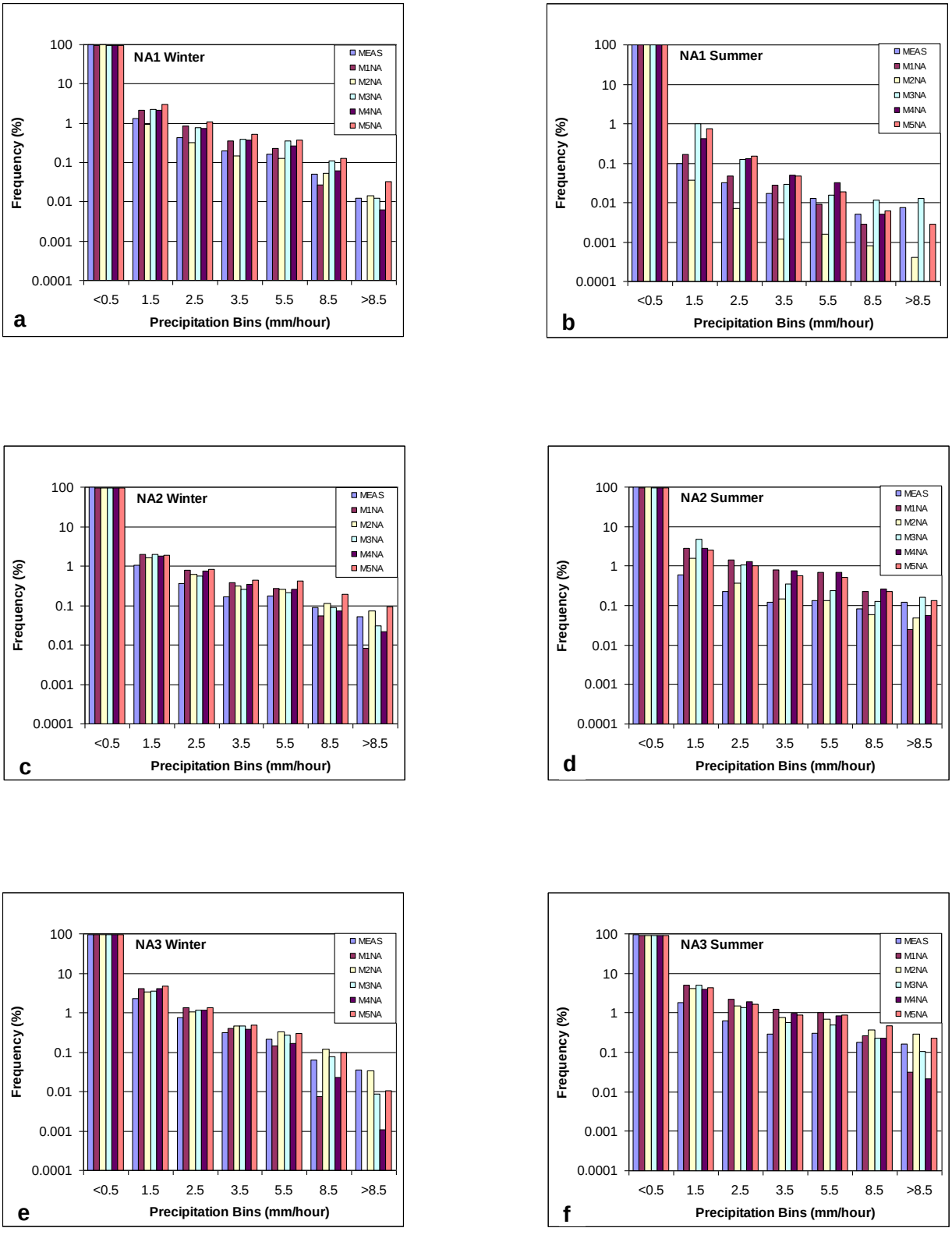
NA1 Winter
NA1 Summer
NA2 Winter
NA2 Summer
NA3 Winter
NA3 Summer
MEAS
M1NA
M2NA
M3NA
M4NA
M5NA
Frequency (%)
100
10
1
0.1
0.01
0.001
0.0001
<0.5
1.5
2.5
3.5
5.5
8.5
>8.5
Precipitation Bins (mm/hour)
a
b
c
d
e
f

Figure 19

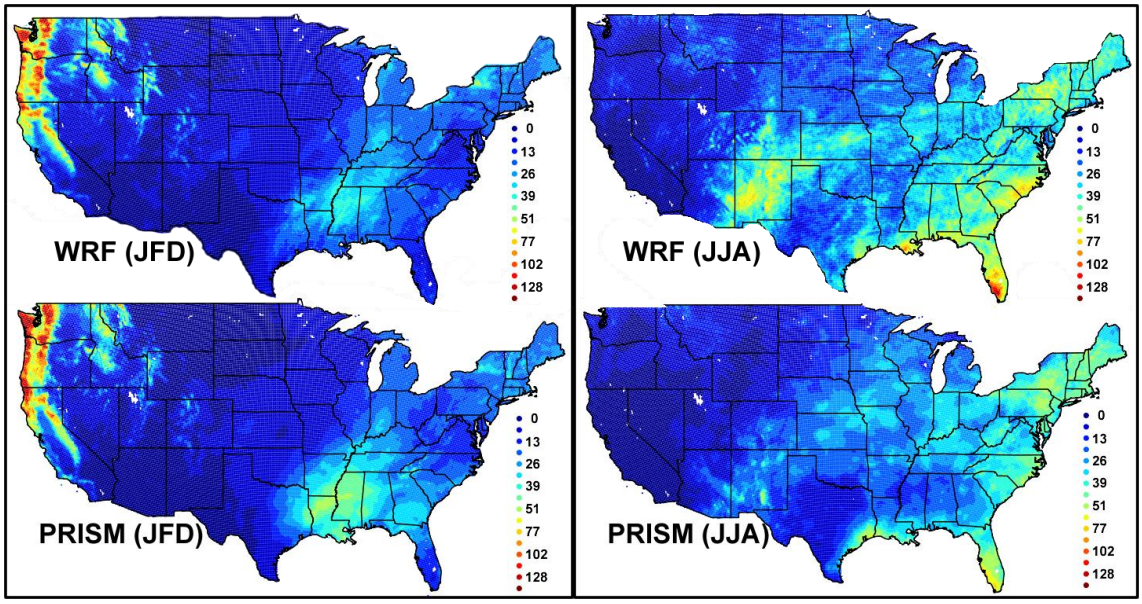

Figure 20