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

Ulas Im, Roberto Bianconi, Efisio Solazzo, Ioannis Koutsisoukis, Alba Badia, Alessandra Balzarini, Rocío Baró, Roberto Bellasio, Dominik Brunner, Charles Chemel, Gabriele Curci, Hugo Denier van der Gon, Johannes Flemming, Renate Forkel, Lea Giordano, Pedro Jiménez-Guerrero, Marcus Hirtl, Alma Hodzic, Luka Honzak, Oriol Jorba, Christoph Knot, Paul A. Makar, Astrid Manders-Groot, Lucy Neal, Juan L. Pérez, Guido Pirovano, George Pouliot, Roberto San Jose, Nicholas Savage, Wolfram Schroder, Ranjeet S. Sokhig, Alfieida Torian, Paolo Tuccella, Kai Wang, Johannes Werhahn, Ralf Wolke, Rahela Zabkar, Yang Zhang, Junhua Zhang, Christian Hogrefe, Stefano Galmarini.

* Corresponding author.
E-mail addresses: ulasim77@gmail.com (U. Im), Stefano.galmarini@jrc.europa.eu (S. Galmarini).

Contents lists available at ScienceDirect
Atmospheric Environment
journal homepage: www.elsevier.com/locate/atmosenv

http://dx.doi.org/10.1016/j.atmosenv.2014.08.072
© 2014 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Im, U., et al., Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate matter, Atmospheric Environment (2014), http://dx.doi.org/10.1016/j.atmosenv.2014.08.072
1. Introduction

Particulate matter (PM) is related to respiratory and cardiovascular diseases as well as to mortality (Schwartz et al., 1996; Bernard et al., 2001). PM has direct and indirect effects on climate (IPCC, 2007) and in turn, climate may have a significant impact on PM levels and composition (Jacob and Winner, 2009). PM has both anthropogenic and natural sources and is emitted as primary aerosols or is chemically formed from gaseous precursors in the atmosphere. PM levels are still a concern, particularly in the urban areas and its adverse effects on climate and health are expected to persist (Winker et al., 2013). Due to the greater potential of PM$_{2.5}$ (PM with an aerodynamic diameter smaller than 2.5 $\mu$m) to cause adverse effects on public health compared to PM$_{10}$ (PM with an aerodynamic diameter below 10 $\mu$m), PM$_{2.5}$ attracted more scientific attention that led to air quality model (AQM) development to focus more on this size of PM and its composition. PM can lead to reductions in radiation reaching the earth and therefore impact the temperature, wind speed and humidity, and it can also modify cloud droplet size and number (Baklanov et al., 2014; Brunner et al., 2014). On-line coupled AQMs can simulate the aerosol feedbacks on meteorology that can be important on a wide range of temporal and spatial scales (Zhang, 2008; Grell and Baklanov, 2011).

The Air Quality Model Evaluation International Initiative (AQMEII) is designed to promote policy-relevant research on regional air quality model evaluation across the atmospheric modeling communities in Europe (EU) and North America (NA) through the exchange of information on current practices and the identification of research priorities (Galmarni and Rao, 2011). Standardized observations and model outputs were made available through the ENSEMBLE web-based system (http://ensemble2.jrc.ec.europa.eu/public/) that is hosted at the Joint Research Centre (JRC; Bianconi et al., 2004; Galmarni et al., 2012). The first phase of AQMEII focused on the evaluation of off-line atmospheric modeling systems against large sets of monitoring observations over Europe and North America for the year 2006 (Solazzo et al., 2012a,b; 2013; Vautard et al., 2012; Hogrefe et al., 2014). The results from this first phase demonstrated a large underestimation by all models throughout the year and a large variability among models in representing emissions, deposition and concentrations of PM and their composition (Solazzo et al., 2012b).

The second phase of AQMEII extends this model assessment to on-line air quality models. In this study, we analyze PM$_{10}$ and PM$_{2.5}$ mass concentrations simulated by eight on-line-coupled models, which have been run by seventeen independent groups from Europe and North America (a companion study is devoted to the analyses of ozone, Im et al., 2014). The surface PM levels simulated by the individual models as well as their ensemble mean and median are compared with the observational data provided by the ENSEMBLE system. As multi-model ensemble analyses is not the scope of this paper, further analyses have been performed by Kioutsiouki et al. (2014) for the EU case using the multi-model data presented in the present paper. The aim of the study is to evaluate the performances of widely used operational on-line coupled models in EU and NA in simulating PM and its chemical components on a sub-regional and seasonal basis employing an experimental set up with common anthropogenic emission and boundary conditions and thus, to identify areas of model improvements and the links to policy applications.

2. Materials and methods

2.1. Models

In the context of AQMEII2, thirteen modeling groups from EU and four modeling groups from NA have submitted PM simulations for the year 2010 (Table 1). One European group (BG2) employed an off-line coupled model while the rest of the groups performed their simulations using their operational on-line models. Nine groups used WRF/Chem model (Grell et al., 2005) and its variant (e.g. Wang et al., 2014), having different gas-phase mechanisms (see Table 1 in Im et al., 2014) but similar aerosol modules that employ different size distributions approaches (modal/bin) and inorganic/organic aerosol treatments as seen in Table 1. The IT2 simulation is
performed with an experimental version of WRF/Chem v. 3.4, where the new secondary organic aerosol scheme VBS was coupled to the aerosol indirect effects modules. Therefore, the bias of IT2 run should not be regarded as the bias of the general WRF/Chem modeling system, but only of this particular version under development. The simulations were conducted for continental-scale domains of EU and NA covering continental U.S., southern Canada and northern Mexico (Fig. 1). To facilitate the cross-comparison between models, the participating groups interpolated their model output to a common grid with 0.25° resolution for both continents. Model values at observation locations were extracted from the original model output files for comparison to observations (described below).

2.2. Emissions and boundary conditions

Standard anthropogenic emissions were provided by the TNO (Netherlands Organization for Applied Scientific Research) for EU (Kuenen et al., 2014) and by U.S. EPA (United States Environmental Protection Agency) and Environment Canada for NA (Pouliot et al., 2014). The NA emissions were processed by the US EPA for all models except for GEM-MACH, where a different grid projection required separate processing by Environment Canada. Different assumptions were used for snow reduction of fugitive dust emissions in these two efforts. More information on the implementation of these emissions is provided in Im et al. (2014). The spatial distribution of annually-integrated anthropogenic PM2.5 emissions for EU and NA domains are depicted in Fig. 1. Anthropogenic PM10 emissions per km² in NA (76 ktons km⁻² yr⁻¹) are larger than those in EU (69 ktons km⁻² yr⁻¹) while EU is characterized by larger PM2.5 emissions density (49 ktons km⁻² yr⁻¹) compared to NA (29 ktons km⁻² yr⁻¹). EU also has more than a factor of two larger NOx, NMVOC and NH3 emission densities compared to NA (Im et al., 2014). Note that the emissions over the oceans represent those originating only from the maritime sector (Kuenen et al., 2014; Schwede et al., 2005; Zaveri et al., 2008; Ackermann et al., 1998; Scheff et al., 2001; Shaw et al., 2008; Jones and Creighton, 2011; Appel et al., 2008; Riener et al., 2003; Ahmadov et al., 2012; Fountoukis and Nenes, 2007; Appel et al., 2013; Savale et al., 2013; Makar et al., 2014a,b). Sauer et al., 2012).
3-D daily chemical boundary conditions were provided by the ECMWF IFS-MOZART model (referred as MACC hereafter) run in the context of the MACC-II project (Monitoring Atmospheric Composition and Climate – Interim Implementation) on 3-hourly and 1.125° spatial resolution (Inness et al., 2013). The aerosol chemical species available in the reanalysis included sea-salt, dust, organic matter, black carbon and sulfate. However, following the AQMEII Phase 1 experience described in Schere et al. (2012), MACC-II sea-salt concentrations were not used as chemical boundary conditions for the NA domain.

2.3. Observations

Observations of hourly and daily rural and urban surface PM$_{10}$ and PM$_{2.5}$ mass concentrations with a data availability of at least 75% from different measurement networks in EU (EMEP (European Monitoring and Evaluation Programme; http://www.emep-int/) and AirBase (European AQ database; http://acm.eionet.europa.eu/databases/airbase/)) and NA (the Canadian National Atmospheric Chemistry (NATChem) Database and Analysis Facility operated by Environment Canada (http://www.ec.gc.ca/natchem/) that contains measurements from the Canadian National Air Pollution Surveillance Network (http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx), the Canadian Air and Precipitation Monitoring Network (http://www.ec.gc.ca/natchem/), the U.S. Clean Air Status and Trends Network (http://java.epa.gov/castnet/clearsession.do), the U.S. Interagency Monitoring of Protected Visual Environments Network (http://views.cira.colostate.edu/web/DataWizard/), and the U.S. Environmental Protection Agency’s Air Quality System database for U.S. air quality data (http://www.epa.gov/ttn/airs/airsaqsdetaildata/downlaodadsadata.htm)) have been used in order to evaluate the model performances in simulating the surface PM concentrations in the two continents (Fig. 1). Daily averages were calculated using the hourly observations from the station where daily measurements were not available and the analyses were performed on the daily averaged PM concentrations. Daily observations from 1525 stations (439 rural and 1076 urban) in EU and 469 stations (158 rural and 311 urban) in NA were used for PM$_{10}$ comparisons. For PM$_{2.5}$, data from 517 stations in EU (139 rural and 378 urban) and 659 stations in NA (311 rural and 348 urban) were used. A geographical breakdown into three sub-regions for each continent that is similar to that in Solazzo et al., 2012a,b was applied, which is based on emission and climatological characteristics (Fig. 1). The European sub-region EU1 can be characterized by north-western European sources with a transition climate between marine and continental and hosts 618 stations for PM$_{10}$ (216 rural and 402 urban) and 255 stations for PM$_{2.5}$ (66 rural and 189 urban). EU2 covers the north-eastern and central Europe sources as well as Germany with 433 stations for PM$_{10}$ (124 rural and 309 urban) and 255 stations for PM$_{2.5}$ (66 rural and 189 urban). EU3 is characterized by the Mediterranean type climate and sources covering 375 stations for PM$_{10}$ (92 rural and 283 urban) and 94 stations for PM$_{2.5}$ (44 rural and 50 urban). Sub-region NA1 consists of the arid southwestern part of the U.S. with the western slope of the Rocky
mountains on the east and hosts 113 stations for PM$_{10}$ (44 rural and 69 urban) and 70 stations for PM$_{2.5}$ (37 rural and 33 urban). NA2 covers the more humid southeastern U.S. with 45 stations for PM$_{10}$ (17 rural and 28 urban) and 117 stations for PM$_{2.5}$ (52 rural and 65 urban). NA3 consists of the northeastern NA that is characterized by its highest emission sources in NA covering 64 stations for PM$_{10}$ (11 rural and 53 urban) and 188 stations for PM$_{2.5}$ (78 rural and 110 urban).

2.4. Statistical analyses

A number of statistical parameters, including Pearson's correlation coefficient (PCC), root mean square error (RMSE); normalized mean standard error (NMSE) and normalized mean bias (NMB) are calculated (Im et al., 2014) in order to compare the individual model performances as well as the ensemble mean and median. The comparisons are performed individually for the two domains and their sub-regions for the whole simulation period and on a seasonal basis, in order to identify which regions and/or seasons have systematic errors.

3. Results and discussion

3.1. PM$_{10}$

3.1.1. Seasonal and regional surface levels over Europe

Comparisons of observed and simulated annual and domain-averaged PM$_{10}$ and PM$_{2.5}$ concentrations over the rural and urban monitoring stations in EU and NA are presented in Table 2. The temporal variation of the rural PM$_{10}$ levels over EU are moderate-to-well-reproduced by the models (PCC = 0.18–0.86), while the variations at urban sites were reproduced with slightly lower agreement (PCC = 0.06–0.82). For both station types, the lowest correlations are calculated for DE4, ES1 and UK4 (PCC<0.25) while BG2 and UK5 well-captured the variation of PM$_{10}$ with PCC values between 0.4 and 0.9 and lowest in the sub-region EU2 (PCC = 0.2–0.9). Similar to the continental scale (EU0), in all sub-regions, the smallest biases are calculated for the UK4 model while the largest are calculated for the IT2 model. For the urban PM$_{10}$ levels, EU2 and EU3 have the largest biases (up to 81%). UK4 model has the lowest MBV values while IT2 model is again associated with has the largest biases. The temporal variation was best reproduced by the UK5 model for all sub-regions except for EU3 where highest PCC is calculated for IT1 model.

<table>
<thead>
<tr>
<th>Models</th>
<th>PM$_{10}$</th>
<th>Rural</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>NMSE</td>
<td>NMB</td>
</tr>
<tr>
<td>AT1</td>
<td>0.40</td>
<td>55.34</td>
<td>-43.55</td>
</tr>
<tr>
<td>BG2</td>
<td>0.74</td>
<td>55.30</td>
<td>-46.86</td>
</tr>
<tr>
<td>CH1</td>
<td>0.42</td>
<td>29.94</td>
<td>-28.52</td>
</tr>
<tr>
<td>DE3</td>
<td>0.63</td>
<td>45.54</td>
<td>-41.88</td>
</tr>
<tr>
<td>DE4</td>
<td>0.18</td>
<td>59.13</td>
<td>-43.64</td>
</tr>
<tr>
<td>ES1</td>
<td>0.22</td>
<td>74.83</td>
<td>-49.19</td>
</tr>
<tr>
<td>ES3</td>
<td>0.35</td>
<td>77.96</td>
<td>-50.74</td>
</tr>
<tr>
<td>IT1</td>
<td>0.57</td>
<td>21.70</td>
<td>-25.12</td>
</tr>
<tr>
<td>IT2</td>
<td>0.26</td>
<td>168.38</td>
<td>-66.10</td>
</tr>
<tr>
<td>NE2</td>
<td>0.61</td>
<td>34.54</td>
<td>-35.68</td>
</tr>
<tr>
<td>SI1</td>
<td>0.62</td>
<td>17.63</td>
<td>-21.52</td>
</tr>
<tr>
<td>UK4</td>
<td>0.25</td>
<td>31.91</td>
<td>-23.29</td>
</tr>
<tr>
<td>UK5</td>
<td>0.86</td>
<td>50.34</td>
<td>-46.32</td>
</tr>
<tr>
<td>EU mean</td>
<td>0.64</td>
<td>43.49</td>
<td>-40.29</td>
</tr>
<tr>
<td>EU median</td>
<td>0.68</td>
<td>50.52</td>
<td>-43.50</td>
</tr>
<tr>
<td>CA2f</td>
<td>-0.10</td>
<td>49.37</td>
<td>-19.79</td>
</tr>
<tr>
<td>ES1</td>
<td>0.41</td>
<td>344.08</td>
<td>-76.91</td>
</tr>
<tr>
<td>US6</td>
<td>0.21</td>
<td>63.65</td>
<td>-38.22</td>
</tr>
<tr>
<td>US7</td>
<td>0.30</td>
<td>34.17</td>
<td>-17.21</td>
</tr>
<tr>
<td>US8</td>
<td>0.31</td>
<td>438.30</td>
<td>-80.09</td>
</tr>
<tr>
<td>NA median</td>
<td>0.24</td>
<td>83.01</td>
<td>-46.45</td>
</tr>
<tr>
<td>NA median</td>
<td>0.18</td>
<td>115.82</td>
<td>-54.21</td>
</tr>
</tbody>
</table>

Table 2

Statistical comparisons of observed and simulated annual and domain-mean surface PM$_{10}$ and PM$_{2.5}$ over EU and NA.
Fig. 2. Observed and simulated monthly continental and sub-regional rural PM10 concentrations over EU (upper panel) and NA (lower panel). Note scale differences.
Fig. 3. Observed and simulated monthly continental and sub-regional urban PM$_{10}$ concentrations over EU (upper panel) and NA (lower panel). Note scale differences.
Fig. 4. Box-and-whisker plots for observed and simulated PM_{10} (upper panel) and PM_{2.5} (lower panel) concentrations over rural and urban stations in North America.
Fig. 5. Soccer plots for simulated seasonal and regional rural PM10 levels over Europe (upper panel) and North America (lower panel) for winter (a,e), spring (b,f), summer (c,g) and autumn (d,h).

U. Im et al. / Atmospheric Environment xxx (2014) 1–21

Please cite this article in press as: Im, U., et al., Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate matter, Atmospheric Environment (2014), http://dx.doi.org/10.1016/j.atmosenv.2014.08.072
Fig. 6. Soccer plots for simulated seasonal and regional PM10 levels over Europe (upper panel) and North America (lower panel) for winter (a,e), spring (b,f), summer (c,g) and autumn (d,h).
The seasonal and regional model evaluations are conducted through scatter plots presented in Figs. 5 and 6, summarizing the performance in both domains for the rural and urban sites, respectively. The observed and modeled surface rural PM$_{10}$ levels over EU are compared in Fig. 5a–d (upper panel). The results show a systematic underestimation for all models in almost all seasons and sub-regions. The largest underestimations for the rural PM$_{10}$ are calculated for the EU3 sub-region (Mediterranean), particularly during winter (Fig. 5a). In sub-region EU1, underestimations of 2% (in summer by SI1) to 74% (in winter by IT2) are calculated. In EU1, surface PM$_{10}$ levels in autumn were overestimated by 1% and 4% by IT1 and SI1, respectively. In sub-region EU2, the highest underestimation (85%) was calculated for IT2 model again for the winter period (Fig. 5a) while SI1 model had the smallest underestimations with values from 23% to 57%. UK4 model had the lowest underestimations for the spring and summer levels (Fig. 5a,d) by 14% and 11%, respectively. Overall, the largest biases were calculated for the winter period (by up to 85%). Similar results were calculated for the urban surface PM$_{10}$ levels in EU with slightly lower biases (Fig. 6a–d).

### 3.1.2. Seasonal vs regional surface levels over North America

Over NA, the temporal variation of rural PM$_{10}$ levels is poorly reproduced by majority of the models with PCC of 0.22–0.38 (Table 2). CA2f model fails to reproduce the temporal variation (PCC = −0.05). The low values for this last model may be due to the lack of snow reduction factors in the reprocessing of emissions of fugitive dust for this model in this experiment (see Pouliot et al., 2014). On the other hand, the temporal variation at the urban sites are slightly better captured by the models (PCC = 0.18–0.54). The NMB values do not differ much between the rural and urban stations on the continental scale (NA0) as seen in Table 2. Over both station types, ES1 and US8 models have the largest biases (>70%) while other models have much lower biases (<40%). The monthly variations in NA0 (NA0) are better captured compared to the daily variability as seen in Figs. 2 and 3. In sub-region NA1, particularly over the rural stations, the majority of the models fail to reproduce both the temporal variation and the magnitudes. In sub-regions NA2 and NA3, the temporal variability is relatively better captured by the models. The variability in the observed PM$_{10}$ concentrations are relatively well represented by CA2f and US7 with low biases (<20%) as seen in Fig. 4 (upper panel), but also by US6 with a larger bias over the rural (~39%) and urban (~34%) stations (Table 2). Similar to the EU domain, the MACC model largely underpredicts the observed variability.

The temporal variability of rural PM$_{10}$ levels over the NA1 sub-region was poorly reproduced by all models with PCC values ranging from 0.03 (CA2f) to 0.52 (US6). In NA2, PCC values were also low: −0.16 (ES1) to 0.56 (US7). Temporal variations over NA3, however, were reproduced reasonably well by most models (PCC = 0.69–0.74) except for the ES1 model (PCC = 0.28). There is a general underestimation by all models in all sub-regions. As can be seen in Fig. 2, the largest underestimation occurs in NA1 (MNB = −57% to −84%) with the exception of US7 overestimating by 15%. Over NA2 and NA3, underestimations from 20% to 88% are calculated. The largest underestimations are calculated for ES1 (MNB > 80%) while US7 had the smallest biases (≤25%). Urban PM$_{10}$ levels over NA are best reproduced in NA3 with PCC over 0.60 except for ES1 (PCC = 0.33). PCC values range from 0.11 to 0.55 over NA1 and from −0.15 to 0.72 over NA2. There are generally underestimations by up to 87% in the sub-regions while CA2f and US7 overestimate the urban PM$_{10}$ levels over NA1 by 11% and 20%, respectively. The largest biases are calculated for the ES1 model in all sub-regions (MNB = 80%–87%).

Soccer plots for the seasonal and geographical model performance for the rural and urban surface PM$_{10}$ levels over NA are presented in Figs. 5 and 6 (lower panels). Over NA, there are no systematic seasonal trends in model performance except for the ES1 and US8 models having the largest biases for rural PM$_{10}$ levels in all seasons and sub-regions (Fig. 5e–h). ES1 model follows US8 with slightly lower biases. The largest underestimations were calculated for the spring and summer periods in all sub-regions by up to 90% and 93%, respectively. There is a general underestimation in all seasons and sub-regions, with the exception of overestimations calculated for US7 model by 3%–67% over NA1. On a continental scale, US7 model slightly overestimates the rural PM$_{10}$ levels by 3%. The model performances for the urban PM$_{10}$ levels over NA (Fig. 6e–h) are similar to those for the rural levels, with slightly lower biases.

The large differences in PM$_{10}$ predictions among those models and their performances at rural and urban sites can be attributed mainly to the use of different online dust emission modules. For example, US7 and US8 use two different dust emission modules available in WRF/Chem version 3.4.1, i.e., the MOSAIC/GOCART dust module of Zhao et al. (2010) and AER/AFWA dust module of Jones and Creighton (2011). The simulated coarse dust concentrations by the two dust emission modules used by US7 and US8 are significantly different in terms of locations and magnitudes (Fig. S1). While both simulate dust emissions from the Mojave desert in southeastern California and the Sonoran Deserts in southern Arizona, the MOSAIC/GOCART dust module gives much higher coarse dust emissions than the AER/AFWA dust module in these areas with a much broader areal coverage and also predict dust emissions in many other areas in the continental U.S. and northern Mexico. As reported by Raman and Arellano (2013), the AER/AFWA dust emission module in WRF/Chem v. 3.4.1 significantly underpredicted dust emissions over Phoenix area in Arizona, U.S., resulting in significant underpredictions of PM$_{10}$ (~50 mg m$^{-3}$) compared to the observed concentration of 1800 mg m$^{-3}$. While differences in the dust emission modules explain most differences in coarse dust, another reason for much lower dust concentrations by US8 is the use of a simplified surface drag parameterization of Mass and Ovens (2011). While this parameterization helps reduce the overpredictions of wind speeds (Wang et al., 2014; Yahya et al., 2014a,b), it reduces dust emissions which depend strongly on wind speeds. The sensitivity simulation without the parameterization of Mass and Ovens (2011) gives dust concentrations that are higher by about a factor of two than the one with this parameterization. The substantial differences in coarse dust concentrations contribute to large differences in coarse PM between the two model simulations. Differences in sea-salt emissions predicted by US7 and US8 also contribute to differences in coarse PM concentrations, although their contributions to differences in PM$_{10}$ performance at rural and urban locations are negligible (in particular, for sites located inland). Although US7 and US8 use the same sea-salt emission module of Gong et al. (1997), US8 gives lower sea-salt emissions (thus lower sea-salt concentrations) over oceanic areas because of the use of a simplified surface drag parameterization of Mass and Ovens (2011) that gives lower wind speeds.

### 3.2. PM$_{2.5}$

#### 3.2.1. Seasonal and regional surface levels over Europe

All models show a very similar behavior for simulated continental surface rural and urban PM$_{2.5}$ levels compared to the simulated PM$_{10}$ levels, with lower biases, as seen in the box-and-whisker plots presented in the lower panel of Fig. 4. PCC values calculated for the simulated PM$_{2.5}$ levels are very similar in general to those calculated for the PM$_{10}$ levels (Table 2). Over the rural
Fig. 7. Observed and simulated monthly continental and sub-regional rural PM$_{2.5}$ concentrations over EU (upper panel) and NA (lower panel). Note scale differences.
Fig. 8. Observed and simulated monthly continental and sub-regional PM$_{2.5}$ concentrations over EU (upper panel) and NA (lower panel). Note scale differences.
stations, the underestimations range from 2% (CH1) to 60%, with the highest bias calculated for the IT2 model similar to PM$_{10}$. For the urban stations, the largest bias was again calculated for the IT2 model (MNB = 68%). UK4 model overestimated the rural PM$_{10}$ concentrations by 20% (Table 2) as can also be seen in Fig. 7. The sub-regional analyses show that these overestimations are mostly due to the large overestimations particularly during summer in the Mediterranean region (EU3) by up to 72%. Further analyses have shown that these overestimates for UK4 are due to excessive model PM from wildfire emissions on the Iberian Peninsula where the vast majority of PM observations are located (Fig. 8). The UK4 model has not previously been run for a domain with large sources of wildfires and it seems likely that the implementation of these sources needs further improvement in this model configuration. The MACC model underestimates the continental and annual mean levels as shown in Fig. 4, as well as in all sub-regions and seasons, suggesting that these overestimations are not due to the boundary conditions, but may be due to the emissions or deposition. Dry deposition of PM$_{2.5}$ calculated by the models (Fig. 5a) show that IT2 and SI1 models simulate significantly larger deposition compared to the other models. This can explain the systematic largest underestimations associated with the IT2 model compared to the other models.

The soccer plots presented in Figs. 10a and 11a show that winter levels are underestimated by all models in all sub-regions, in general by more than 50%, particularly over the urban stations. In other seasons, the underestimations are lower. CH1 and UK4 models overestimate in spring and in particular during summer. IT1 and SI1 overestimate rural EU3 PM$_{2.5}$ levels by 4% and 5%, respectively (Fig. 10b). Similar overestimations hold for UK4 over the urban stations (Fig. 11b). In summer, there is general underestimation by the majority of the models by up to 49% and 59% (by IT2 in EU2) over the rural and urban stations, respectively (Figs. 10c and 11c). Autumn levels are underestimated by up to 72% over the rural (Fig. 10d) and by up to 77% over the urban stations (Fig. 11d) depending on the region with the maximum bias calculated for EU2 by the IT2 model.

### 3.2.2. Seasonal vs regional surface levels over North America

The temporal variations for the domain-averaged surface PM$_{2.5}$ concentrations over both rural and urban stations are much better captured by the majority of the models compared to the PM$_{10}$ levels (Table 2). PCC values for the urban stations (0.31–0.78) are higher than those for the rural values (0.05–0.61) for all models, as can also be seen from the monthly time series plots in Figs. 7 and 8. ES1 model had the lowest correlations while US7 had the highest values. ES1 model also had the largest biases (MNB = –68% and –71% for rural and urban stations, respectively) while US8 simulated the surface PM$_{2.5}$ levels with the lowest bias (MNB = –26% and –17%, respectively). The large underestimation calculated for the ES1 model can be attributed to the significantly larger dry deposition compared to the other models as can be seen in Fig. 9b. As discussed in Section 3.1.2, the underestimation in the PM$_{10}$ levels for the US8 model suggests that the dust particles in both coarse and fine modes are significantly underestimated by this model. US7 model overestimated the domain-averaged PM$_{2.5}$ levels over both station types by ~48%, likely due to an overprediction in dust and sea-salt concentrations in PM$_{2.5}$ size sections. PM$_{2.5}$ concentrations predicted by US7 are much higher than those from US8 (Fig. S1). Such differences can be attributed to several factors. First, US7 and US8 use different dust emission modules, which give very different concentrations of dust in the PM$_{2.5}$ size sections/modes. Second, US7 and US8 use different splitting fractions between coarse and fine dust emissions. US7 allocates 9% and 68% of the total dust emission to PM$_{2.5}$ and coarse PM, respectively. Since MOSAIC only describes aerosols up to 10 µm, the emissions for particles with diameter greater than 10 µm are neglected (which is ~23% of the total emissions). For comparison, US8 allocates 3% of dust emissions in the accumulation mode and the rest of 97% in the coarse mode. Third, US7 and US8 give different predictions of primary and secondary organic aerosols (POA and SOA), due possibly

![Fig. 9. Calculated annual dry deposition of fine inorganic aerosols (SO$_4$, NO$_3$ and NH$_4$), total organic carbon (TOC) PM$_{2.5}$, crustal material (CM) and sea-salt (SS) over a, b) EU and c, d) NA.](image-url)
Fig. 10. Soccer plots for simulated seasonal and regional rural PM$_{2.5}$ levels over Europe (upper panel) and North America (lower panel) for winter (a,c,e), spring (b,d), summer (c,g), and autumn (d,h).
Fig. 11. Scatter plots for simulated seasonal and regional urban PM$_{2.5}$ levels over Europe (upper panel) and North America (lower panel) for winter (a,e), spring (b,f), summer (c,g) and autumn (d,h).

Please cite this article in press as: Im, U., et al., Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate matter, Atmospheric Environment (2014), http://dx.doi.org/10.1016/j.atmosenv.2014.08.072
to the use of different SOA modules and different conversion factors between primary organic carbon emissions and the POA simulated in the model. As seen in Fig. 4, the models have similar profiles for both rural and urban stations while the MACC model overestimates the rural and underestimates the urban PM$_{2.5}$ concentrations, implying that the simulated levels were due to local contributions rather than regional transport.

US7 model overestimates both the rural and urban PM$_{2.5}$ concentrations in all seasons and sub-regions (Figs. 10 and 11e–h). The overestimations simulated by US7 model are smallest during winter from 16% to 96% over the rural and 51%–82% over the urban stations. The figures also show that ES1 model underestimates in all seasons and sub-regions. With the exception of ES1 model, all models fall into the 75% error range in all seasons and sub-regions, while excluding US7, the error decreases to the 50% range (Figs. 10 and 11e–h). Compared to the PM$_{10}$ levels, the figures show that majority of the models are grouped around the zero line of the soccer plots. The differences in all seasons are highest in sub-region NA1 over both rural (MNB up to 143%) and urban stations (MNB up to 95%).

3.2.3. PM$_{2.5}$ speciated components

Simulated surface sulfate (SO$_4^{2-}$), nitrate (NO$_3^{-}$) and ammonium (NH$_4^{+}$) components of PM$_{2.5}$ aerosols are compared with observations from five, six, and five rural stations in EU, respectively, and 250, 148 and 149 station in NA, respectively. The results are presented in Fig. 12 in the soccer plots for the continental and sub-regional levels in 2010 over EU and NA. Over EU, the continental SO$_4^{2-}$ levels are underestimated by a majority of the models (AT1, DE4, ES1, ES3, IT1, IT2 and UK5) by 22%–61% (Fig. 12a) while few groups (BG2, CH1, NL2, SI1 and UK4) overestimated the SO$_4^{2-}$ levels by 7%–52%. The results show that the underestimating models were all WRF/CHEM models, with the exception of SI1 that overestimates. The largest underestimation of SO$_4^{2-}$ by IT2 can be attributed to the large SO$_4^{2-}$ dry deposition calculated by this model (Fig. 9a). SO$_4^{2-}$ underestimation can also be attributed to absence of SO$_2$ oxidation in cloud water in the heterogeneous phase (e.g. the IT1 model: Balzarini et al., 2014). As seen in Fig. 12b and c, simulated NO$_3^{-}$ and NH$_4^{+}$ are higher than the observed levels. NO$_3^{-}$ levels are overestimated by majority of the models in all regions by more than 75%, particularly in EU2 and EU3 (Fig. 12b). NH$_4^{+}$ levels are also underestimated largely in EU3. In other sub-regions, the differences for simulated NH$_4^{+}$ levels are lower (50%–75%). The results suggest ammonium nitrate (NH$_4$NO$_3$) formation dominating over the ammonium sulfate ((NH$_4$)$_2$SO$_4$) formation over EU as well as possible underestimations in heterogeneous (cloud) SO$_4$ formation and generation of fine sea-salt emissions.

The picture is completely opposite over the NA domain as seen in Fig. 12d–f. SO$_4^{2-}$ levels are particularly overestimated over NA1 as well as over the continent. Particularly CA2f model largely overestimates SO$_4^{2-}$ levels in all sub-regions. NA2 and NA3 are characterized by underestimated SO$_4^{2-}$ levels by the majority of the models. The differences from the observations are in general below 75% except for the CA2f model that has much larger bias. CA2f model has the smallest differences for both NO$_3^{-}$ and NH$_4^{+}$ while ES1 model has the largest underestimations by more than a factor of 2.

3.3. Aerosol optical depth (AOD)

The reconstructed AOD at 555 nm (AOD555) are compared with observations from 35 Aerosol Robotic Network (AERONET: http://aeronet.gsfc.nasa.gov/new_web/index.html) stations from each domain. Soccer plots and the diurnal profiles for the model performances in 2010 for the continental and sub-regional AOD555 levels are presented in Figs. 13a, c. Over EU (Fig. 12a), the majority of the model performed within the 50% error range. The DE3 model had the largest underestimations (MNB = 60%) in all regions (Fig. 13c) while the BG2 model had the largest overestimations (MNB up to 70%). The large underestimation by the DE3 model can be attributed to the approach in estimating the AOD555. While the majority of the models consider SO$_4$, NO$_3$, NH$_4$, primary and secondary organic aerosols (POA/SOA), elemental carbon (EC), dust

![Fig. 12. Soccer plots for simulated regional rural fine SO$_4$ (a,d), NO$_3$ (b,e) and NH$_4$ (c,f) levels over Europe (upper panel) and North America (lower panel).](Image)
and sea-salt (Curci et al., 2014) in their AOD estimations, the DE3 model does not consider EC, POA/SOA and sea-salt. The smallest bias was calculated for SI1 (MNB = +7%) and for AT1 (−12%). In general, models BG2, CH1, NL2 and UK5 overestimated the observed AOD555 levels while other models underestimate. The observed hourly diurnal variation over the continent was moderately captured by the models with a maximum and minimum PCC of 0.65 (AT1) and 0.25 (DE3), respectively. WRF/CHEM models were associated with very similar temporal variations (PCC = ~0.6). Over NA (Fig. 13b,d), CA2f model failed to reproduce both the temporal variation (PCC = 0.23) and the magnitude of the continental AOD555 with an overestimation of 29%. US6 model reproduced the temporal variation better than the other models (PCC = 0.73), but with the largest bias (MNB = −32%). US7 also overestimated the continental AOD555 by 25% and captured the temporal variability (PCC = 0.70) while US8 underestimated the observations by 17% with a temporal agreement of 0.65. Further discussion on model uncertainty on AOD calculation may be found in Curci et al. (2014).

4. Summary and conclusions

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

Results show that over EU, particularly in winter, the monthly temporal variations were not captured by any of the models while the majority of the models produced spring and autumn peaks, particularly for the rural stations while these are not observed in the measurements or the MACC model, suggesting that the anthropogenic emissions or the online-simulated natural dust emissions can be responsible for these peaks. Over EU, the rural PM10 concentrations are underestimated by all models by up to 66% while the underestimations are much larger for the urban PM10 concentrations (up to 75%), suggesting that the urban emissions were not able to represent the actual emissions. The results show a systematic underestimation for all models in almost all seasons and sub-regions, with the largest underestimations for the Mediterranean region. The results also show overestimations in PM2.5 levels suggesting the large underestimations in the PM10 levels can be attributed to the natural emissions. Over NA, there are no systematic seasonal trends in model performances except for the ES1 and US8 models having the largest biases for rural PM10 levels in all seasons and sub-regions. There is a general underestimation in all...
seasons and sub-regions, with the exception of overestimations calculated for US7 model by 3%–67% over western US. The highest underestimations were calculated for the spring and summer periods in all sub-regions by up to -90%. In general, majority of the models simulating the NA case have smaller biases compared to those simulating the EU case, in particular regarding PM$_{2.5}$, which suggests a better representation of the anthropogenic emissions in NA.

SO$_4$ levels over EU are underestimated by majority of the models by up to 61% while few groups overestimated the SO$_4$ levels by 7%–52%. NO$_3$ levels are overestimated by majority of the models in all regions by more than 75%, particularly in east and south Europe while NH$_4$ levels are also underestimated largely in south Europe. SO$_4$ levels over NA are particularly overestimated over western US that is characterized by large anthropogenic emissions. Eastern US is characterized by underestimated SO$_4$ levels by the majority of the models. Regarding the AOD555, the majority of the model performed within the 50% error range over EU. Differences in models can be attributed to differences in approaches in estimating the AOD such as the aerosol components considered in these estimations. The observed hourly diurnal variation over the continent was moderately captured by the models while WRF/Chem models were associated with very similar temporal variations. Over NA, the CA2F and US7 models overestimate the observed AOD555 levels by up to 29% while the US6 and US8 models underestimate by up to 32%. Results show that the simulated dry deposition simulated can lead to substantial differences among the models.

Overall, the results show that representation of dust and sea-salt emissions can largely impact the simulated PM concentrations and that there are still major challenges and uncertainties in simulating the PM levels and identifying the source of the bias in the models. It should be noted that as the results presented in this paper are temporally and spatially averaged over the seasons and sub-regions, cases where feedback mechanisms are of importance must be further studied and evaluated in order to better evaluate the skills of these models in simulating the feedback mechanisms and their impact on the surface PM levels.

Acknowledgments

We gratefully acknowledge the contribution of various groups to the second air Quality Model Evaluation international Initiative (AQMEII) activity: U.S. EPA, Environment Canada, Mexican Secretariat of the Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales-SEMARNAT) and National Institute of Ecology (Instituto Nacional de Ecológia-INE) (North American national emissions inventories); U.S. EPA (North American emissions processing); TNO (European emissions processing); ECMWF/MACC project & Météo-France/CNRM-GAME (Chemical boundary conditions). Ambient North American concentration measurements were extracted from Environment Canada’s National Atmospheric Chemistry Database (NATChem) PM database and provided by several U.S. and Canadian agencies (AQ5, CAPMoN, CASTNet, IMPROVE, NAPS, SEARCH and STN networks); North American precipitation-chemistry measurements were extracted from NATChem’s precipitation-chemistry database and were provided by several U.S. and Canadian agencies (CAPMoN, NADP, NBPMN, NSPNSN, and REPF networks); the WMO World Ozone and Ultraviolet Data Centre (WOUDC) and its data-contributing agencies provided North American and European AOD measurements; the MOZART Data Centre and its contributing airlines provided North American and European aircraft takeoff and landing vertical profiles; for European air quality data the following data centers were used: EMEP European Environment Agency/European Topic Center on Air and Climate Change/AirBase provided European air- and precipitation-chemistry data. The Finish Meteorological Institute is acknowledged for providing biomass burning emission data for Europe. Data from meteorological station monitoring networks were provided by NOAA and Environment Canada (for the US and Canadian meteorological network data) and the National Center for Atmospheric Research (NCAR) data support section. Joint Research Center Ispra/Institute for Environment and Sustainability provided its ENSEMBLE system for model output harmonization and analyses and evaluation. The co-ordination and support of the European contribution through COST Action ES1004 UleNetChem is gratefully acknowledged. The views expressed here are those of the authors and do not necessarily reflect the views and policies of the U.S. Environmental Protection Agency (EPA) or any other organization participating in the AQMEII project. This paper has been subjected to EPA review and approved for publication. C. Knote was supported by the DOE grant DE-SC0006711. The UPM authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the Centro de Supercomputación y Visualización de Madrid (CSEVIMA) and the Spanish Supercomputing Network (JSC). G. Curci and P. Tuccella were supported by the Italian Space Agency (ASI) in the frame of PAMS project (contract n.1/017/11/0). The Center of Excellence for Space Sciences and Technologies SPACE-SI is an operation partly financed by the European Union, European Regional Development Fund and Republic of Slovenia, Ministry of Higher Education, Science, Sport and Culture. Y. Zhang acknowledges funding support from the NSF Earth System Program (AGS-1049200) and high-performance computing support from Yellowstone by NCAR’s Computational and Information Systems Laboratory, sponsored by the National Science Foundation and Stampede, provided as an Extreme Science and Engineering Discovery Environment (XSEDE) digital service by the Texas Advanced Computing Center (TACC). The technical assistance of Bert van Ulf (KNMI) and Arjo Segers (TNO) in producing the results of the RACMO2-LOTOS-EUROS system is gratefully acknowledged. L. Giordano was supported by the Swiss SERI COST project C11.0144. UH-CAIR acknowledges support from the TRANSFORM (FP7) project which provided the basis for their modelling approaches.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atmosenv.2014.08.072.

References


