Atmospheric Environment xxx (2012) 1–17

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

# Investigating impacts of chemistry and transport model formulation on model performance at European scale

G. Pirovano [a,b,*], A. Balzarini [a], B. Bessagnet [b], C. Emery [c], G. Kallos [d], F. Meleux [b], C. Mitsakou [d], U. Nopmongcol [c], G.M. Riva [a], G. Yarwood [c]

[a] RSE S.p.A., via Rubattino 54, 20134 Milano, Italy
[b] INERIS, Parc Technologique Alata BP2, 60550 Verneuil-en-Halatte, France
[c] ENVIRON International Corporation, 773 San Marin Drive, Suite 2115, Novato, CA 94998, USA
[d] University of Athens, School of Physics, Division of Environmental Physics-Meteorology, Bldg PHYS-V, 15784 Athens, Greece

ARTICLE INFO

*Article history:*
Received 27 May 2011
Received in revised form
16 November 2011
Accepted 21 December 2011

*Keywords:*
Model performance evaluation
Wilcoxon ranked test
Model intercomparison
Ozone
Particulate matter
CAMx
CHIMERE

ABSTRACT

The CAMx and CHIMERE chemistry and transport models were applied over Europe for the year 2006 in the framework of the AQMEII inter-comparison exercise. Model simulations used the same input data set thus allowing model performance evaluation to focus on differences related to model chemistry and physics. Model performance was investigated according to different conditions, such as monitoring station classification and geographical features. An improved evaluation methodology, based on the Wilcoxon statistical test, was implemented to provide objectivity in the comparison of model performance.

The models demonstrated similar geographical variations in model performance with just a few exceptions. Both models displayed great performance variability from region to region and within the same region for $NO_2$ and $PM_{10}$. Station type is relevant mainly for pollutants directly influenced by low level emission sources, such as $NO_2$ and $PM_{10}$.

The analysis of the differences between CAMx and CHIMERE results revealed that both physical and chemical processes influenced the model performance. Particularly, differences in vertical diffusion coefficients (Kz) and 1st layer wind speed can affect the surface concentration of primary compounds, especially for stable conditions. Differently, differences in the vertical profiles of Kz strongly influenced the impact of aloft sources on ground level concentrations of both primary pollutants such as $SO_2$ as well as $PM_{10}$ compounds. CAMx showed stronger photochemistry than CHIMERE giving rise to higher ozone concentrations that agreed better with observations. Nonetheless, in some areas the more effective photochemistry showed by CAMx actually compensated for an underestimation in the background concentration.

Finally, $PM_{10}$ performance was rather poor for both models in most regions. CAMx performed always better than CHIMERE in terms of bias, while CHIMERE score for correlation was always higher than CAMx. As already mentioned, vertical mixing is one cause of such discrepancies in correlation. Differently, the stronger underestimation experienced by CHIMERE was mainly influenced by temporal smoothing of the boundary conditions, underestimation of low level emissions (mainly related to fires) and more intense depletion by wet deposition.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

Chemistry and transport models (CTMs) are essential tools to investigate the atmospheric fate of pollutants as well as to design and evaluate effective emission reduction strategies. CTMs include descriptions of the main chemical and physical processes driving air concentration of primary and secondary pollutants, such as sulphur and nitrogen oxides, ozone (Jacobson et al., 1996; Russell and Dennis, 2000) and particulate matter (Jacobson, 1997; Seigneur, 2001; Vautard et al., 2009).

For regional simulations, present data availability allows computed results to be compared against tens to hundreds of measuring sites in Europe and North America (Tesche et al., 2006; Morris et al., 2006; Van Loon et al., 2007) requiring the

* Corresponding author. RSE S.p.A., via Rubattino 54, 20134 Milano, Italy.
*E-mail addresses:* guido.pirovano@rse-web.it (G. Pirovano), bertrand.bessagnet@ineris.fr (B. Bessagnet).

1352-2310/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2011.12.052

**Table 1**
Comparison of CAMx and CHIMERE domain-wide emissions, also split between surface and aloft sources (ton year$^{-1}$). $PM_{10}$ emissions account for anthropogenic sources and wild fires, but not for sea salts. Aloft sources account for the total amount of emissions injected from the 2nd layer up to the domain top.

| | Surface | | Aloft | | Total | |
|---|---|---|---|---|---|---|
| | CAMx | CHIMERE | CAMx | CHIMERE | CAMx | CHIMERE |
| CO | 51 899 170 | 35 819 210 | 29 982 570 | 27 793 180 | 81 881 740 | 63 612 390 |
| NOx | 12 986 190 | 12 377 892 | 5 282 967 | 5 115 994 | 18 269 157 | 17 493 886 |
| $NH_3$ | 5 350 228 | 5 151 898 | 492 906 | 160 960 | 5 843 134 | 5 312 858 |
| $SO_2$ | 4 396 024 | 4 318 913 | 9 657 167 | 9 291 435 | 14 053 191 | 13 610 348 |
| $PM_{10}$ | 5 497 295 | 2 707 733 | 4 378 300 | 4 803 718 | 9 875 595 | 7 511 451 |
| ISOP | 4 920 623 | 5 817 524 | 139 | – | 4 920 762 | 5 817 524 |
| TERP | 2 491 000 | 2 741 965 | 433 | – | 2 491 433 | 2 741 965 |
| FORM | 239 430 | 67 583 | 217 245 | 165 578 | 456 675 | 233 161 |
| ETH | 1 008 042 | 357 950 | 9456 | 25 251 | 1 017 498 | 383 200 |
| TOL | 686 099 | 609 417 | 14 561 | 7977 | 700 660 | 617 394 |

development of suitable methodologies, which enable robust findings and conclusions about model performance. In the last decades, several efforts were made to develop a systematic framework for model performance evaluation (MPE, Weil et al., 1992; Chang and Hanna, 2004). More recently Dennis et al. (2010) proposed a rather complete approach identifying four main components including: operational, diagnostic, dynamical and probabilistic model evaluation.

The Air Quality Modelling International Initiative (AQMEII) was launched as a joint effort between the North American and European modelling communities (Rao et al., 2011). The first phase of the project and hence the content of this paper deal with the first component of the Dennis et al. approach, the operational model evaluation. The CAMx and CHIMERE models were applied and compared over the European domain for calendar year 2006. The models were driven by the same inputs (meteorology, emissions, boundary conditions) provided by the AQMEII organizers. In contrast to many previous inter-comparison studies (Cuvelier et al., 2007; Van Loon et al., 2007), the present application focused on analysis of differences in model chemistry and physics, allowing reliable conclusions on the influence of model formulation on the modelled concentrations. To this aim, a thorough evaluation and comparison of the model results were performed. Model performance was investigated by sub-dividing the observational data set according to different criteria, such as station classification and geographical features. This effort was made to assess possible differences in model performance within the larger regional domain. In order to objectively evaluate differences between CAMx and CHIMERE, the Wilcoxon test was adopted (Wilks, 2006).

The following section describes the CAMx and CHIMERE modelling systems and the main features of the observed data set. Section 3 describes the methodology implemented to evaluate and compare model performance. In Section 4, a detailed analysis of the modelling results is presented. Finally, Section 5 summarises the main findings and conclusions.

## 2. Models and observations

CAMx is a widely used three-dimensional photochemical Eulerian model that simulates the atmospheric fate of ozone and PM (ENVIRON, 2010). This study used CAMx version 5.21 with Carbon Bond 2005 (CB05) gas phase chemistry (Yarwood et al., 2005). The CAMx modelling domain was defined in latitude and longitude with 207 by 287 grid cells of resolution of 0.25° longitude by 0.125° latitude and 23 vertical layers. The CAMx surface layer exactly matched the MM5 surface layer and was about 30 m deep. Further details on the CAMx set up can be found in Nopmongcol et al. (in this issue).

In this study, the CHIMERE model (Bessagnet et al., 2004; Vautard et al., 2005) was used in a configuration similar to that presented in Bessagnet et al. (2010) with MELCHIOR gas phase chemistry (Latuatti, 1997). In AQMEII, CHIMERE was applied over a domain covering part of the Europe continent (from 15°W to 35.25°E in longitude and from 35°N to 70.25 °N in latitude), with a constant horizontal resolution of 0.25° × 0.25°. The vertical grid contained 9 layers expressed in a hybrid-sigma pressure coordinate system, from the surface to 500 hPa. The first ground layer height was 20 m. The model documentation is available at http://euler.lmd.polytechnique.fr/chimere. For both ozone and $PM_{10}$ and its components, the model has undergone extensive intercomparisons with other CTMs at European and urban scales (Bessagnet et al., 2004, 2010; Vautard et al., 2007; Van Loon et al., 2007).

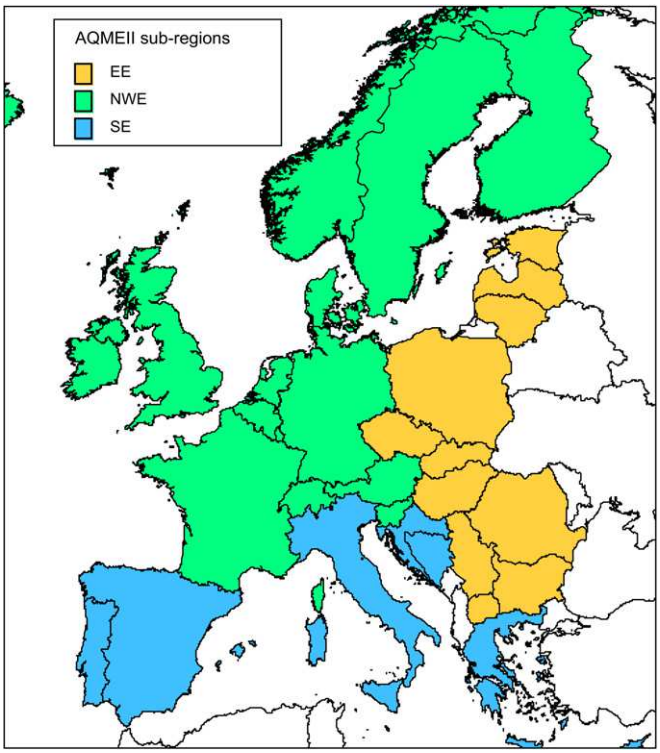


**Fig. 1.** Sub-regions identified within the computational domain: Southern Europe (SE), North-Western Europe (NWE), Eastern Europe (EE). Countries without available observations are in white.

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

### 2.1. Input data

AQMEII participants were provided with a meteorological simulation for the year 2006, generated with MM5 model (Dudhia, 1993) for the European domain with resolution of 0.25° × 0.25°. The MM5CAMx preprocessor for CAMx was used to collapse the 32 vertical layers used by MM5 to 23 layers in CAMx and convert from the Mercator grid used by MM5 to a latitude–longitude grid. Both models used the planetary boundary layer (PBL) heights derived from MM5, apart from cloudy days during which the CHIMERE model considers the development of neutral conditions up to the cloud base (Bessagnet et al., 2009). The models adopt different parameterisations of vertical diffusion below the PBL height which, as discussed below, influenced pollutant dispersion under stable conditions (e.g. night-time).

The AQMEII emissions were prepared by TNO (*Netherlands Organization for Applied Scientific Research*), which provided a gridded emissions database for the year 2005 and 2006 (Pouliot et al., in this issue). The dataset consists in European anthropogenic emissions for the 10 SNAP sectors and international shipping with resolution of 0.125° longitude by 0.0625° latitude. A fire emissions inventory was provided by the Finnish Meteorological Institute (FMI).

The models shared the same emission inventories, but the model-ready input files were prepared independently for each model giving rise to some discrepancies (Table 1). Particularly: 1) $NO_X$, $SO_2$ and $NH_3$ emissions are slightly different because the CAMx computational domain is slightly larger than the CHIMERE domain; 2) the emission vertical distribution was defined from vertical profiles with less detail than the vertical structure of the two models giving rise to discrepancies in the fraction assigned to the surface layer; 3) the models adopted different assumptions to vertically distribute fire emissions which explains why the main differences occur for CO and $PM_{10}$ emissions (Table 1). Biogenic VOC emissions were computed by both models by applying the MEGAN emission model (Guenther et al., 2006). Sea salt emissions were computed separately using published algorithms (Monahan, 1986 for CHIMERE; de Leeuw et al., 2000; Gong, 2003 for CAMx) driven by MM5 meteorological fields. Boundary conditions for both models were derived from GEMS data (Schere et al., in this issue) provided by the European Centre for Medium-Range Weather Forecasts (ECMWF).

### 2.2. Observations

Observed concentrations for calendar year 2006 were provided by the European database of national operational networks (AirBase). Data are available on the AirBase web site for all countries of European Union (http://air-climate.eionet.europa.eu/databases/airbase/). Observations of CO, $NO_2$, $NO_X$, $SO_2$, $O_3$, $PM_{10}$ and $PM_{2.5}$ were selected. Stations were chosen with data availability of 75% and higher. Stations showing outliers in yearly statistics were rejected. Only background stations (rural, suburban and urban) were chosen. A set of 1410 stations were found to fulfil the selection criteria with a total number of 29 countries represented. The station density of the selected dataset was adequate for $NO_2$, $SO_2$ $O_3$ and

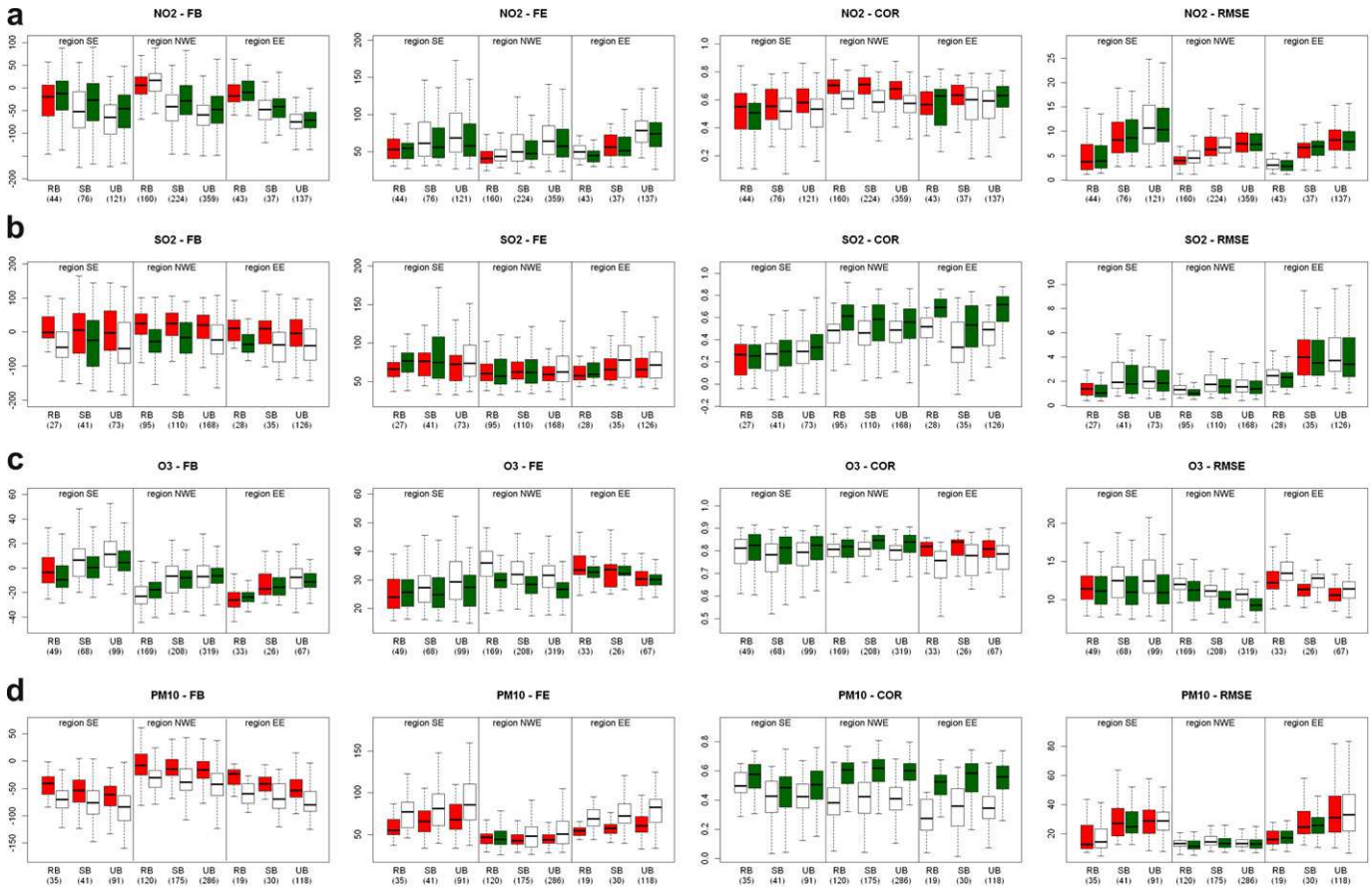


**Fig. 2.** Box-whisker plots of the distribution of the different metrics computed for CAMx (red) and CHIMERE (green) in each sub-region and for each station type. The number of stations included in each subset is reported in brackets. If the performance is significantly different, the plot is unfilled for the worst model (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

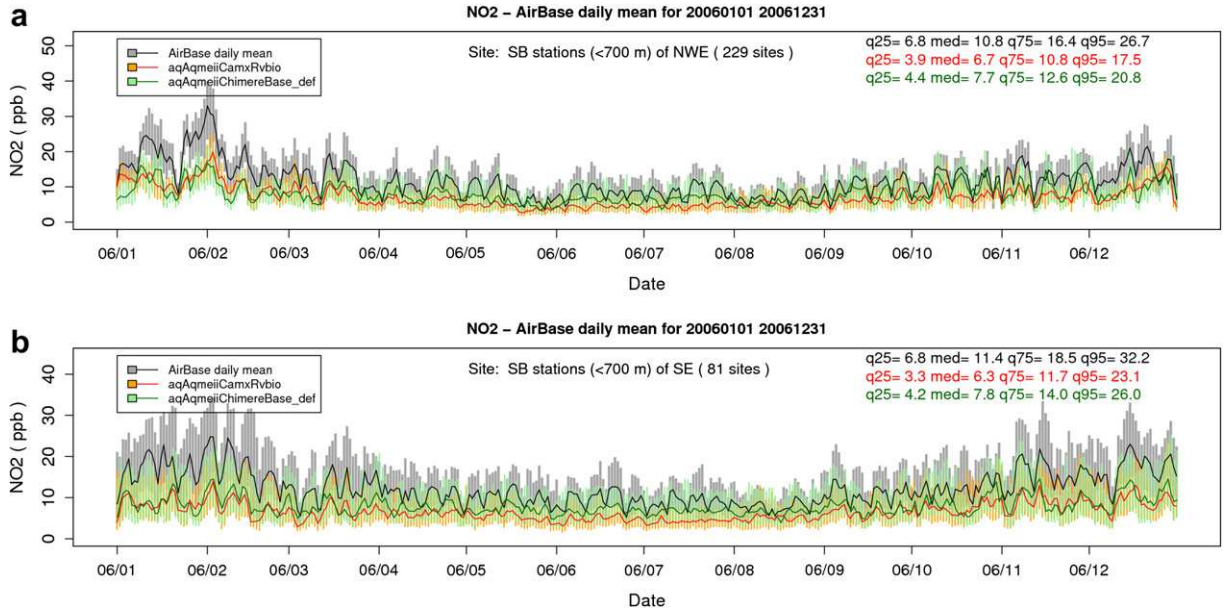


**Fig. 3.** Time series of daily Box-Whisker plots showing the distribution of the observed and computed $NO_2$ concentration at Suburban Background sites of NWE (a) and SE (b) regions. Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green. Bars show the 25th–75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole yearly series are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$PM_{10}$ in Western Europe, while fewer stations were available for Eastern Europe (Table S.1). For $NO_X$, CO and $PM_{2.5}$, monitoring stations were available only for a few countries.

Observations for PM in 2006 from the EMEP (European Monitoring and Evaluation Programme) database were used too. The observational data were available on the EMEP Chemical Coordinating Centre (EMEP/CCC) web site at http://www.emep.int/. The $PM_{10}$ measurements were available from 16 countries mostly on daily basis. The $PM_{2.5}$ measurements were available from 11 countries also on daily basis. The PM measurements were performed using high volume samplers, Whatman quartz fibre filters or Tapered Element Oscillating Microbalance (TEOM). The measured quantities were analysed mostly with the gravimetric method, whereas in some countries the micro balance technique was implemented. Sulphate, nitrate and ammonium daily data were available from 24–36 EMEP stations

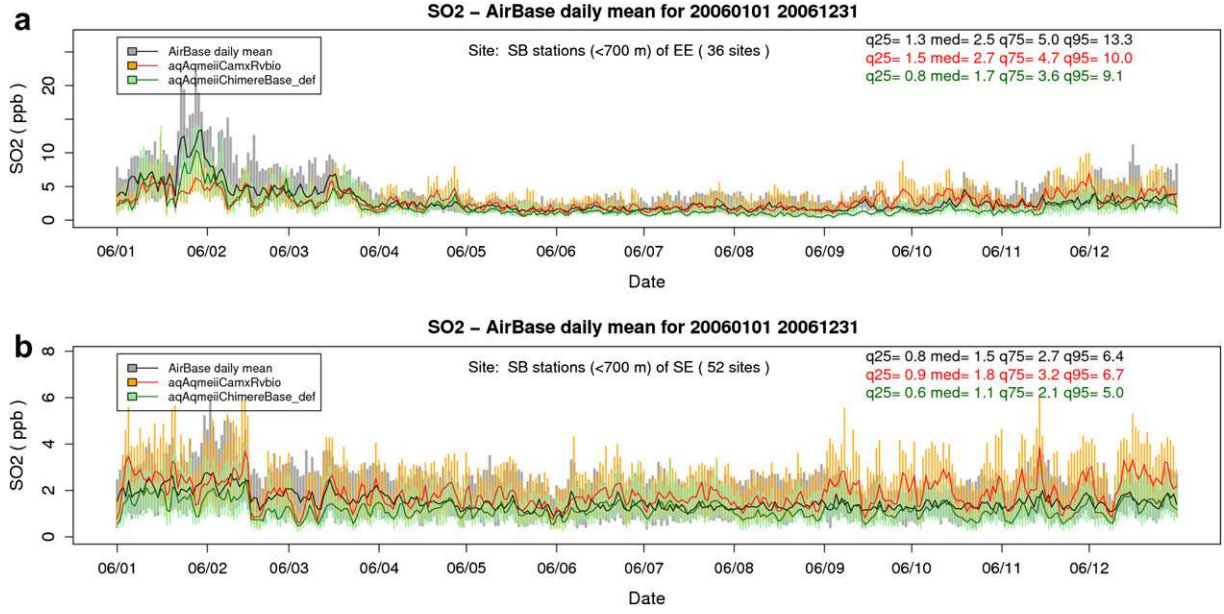


**Fig. 4.** Time series of daily Box-Whisker plots showing the distribution of the observed and computed $SO_2$ concentration at Suburban Background sites of EE (a) and SE (b) regions. Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green. Bars show the 25th–75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole yearly series are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

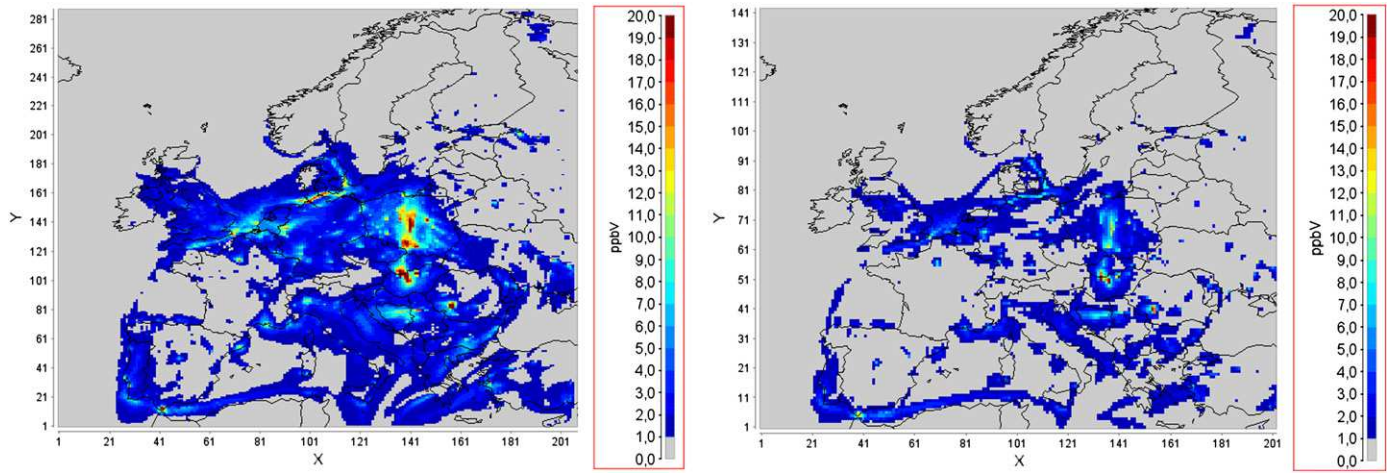


**Fig. 5.** $SO_2$ daily mean concentration computed by CAMx (left) and CHIMERE (right) for October 10th, 2006.

Four species, namely $NO_2$, $SO_2$, $O_3$ and $PM_{10}$, were selected for use in model evaluation because they provided a rather homogenous spatial coverage, in contrast to $NO_X$, CO and $PM_{2.5}$. Where necessary, $PM_{10}$ bulk observations data were integrated with $PM_{2.5}$ data as well as with aerosol composition data for nitrate, sulphate and ammonium. $NO_2$, $SO_2$ and $PM_{10}$ concentrations were expressed as daily means whereas the daily maximum of the 8-hour running average was chosen for $O_3$. The selected metrics for $PM_{10}$ and $O_3$ are used to establish air quality standard in the European legislation (EU, 2008).

## 3. Methods

Several concentration statistics and evaluation metrics can be selected to assess model performances (Boylan and Russel, 2006; Schluenzen and Sokhi 2008; Dennis et al., 2010; Denby, 2011) and to compare results produced by different models (Potempski et al., 2008; Vautard et al., 2009; Thunis et al., 2011). To provide a comprehensive evaluation we selected 7 metrics whose mathematical expression is reported in the appendix: Normalised Mean Bias (NMB), Normalised Mean Error (NME), Mean Fractional Bias

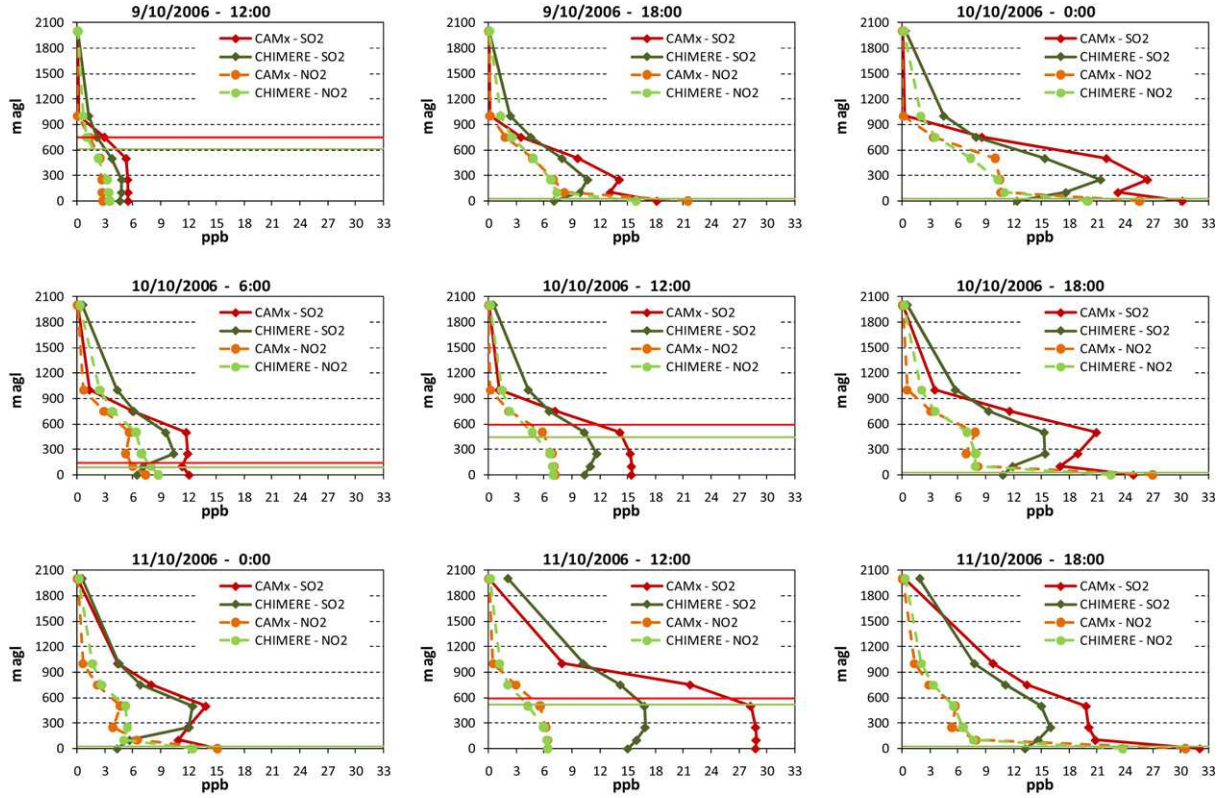


**Fig. 6.** Selection of hourly vertical profiles of $NO_2$ and $SO_2$ computed by CHIMERE (light and dark green) and CAMx (orange and red) at a site belonging to the industrial area of Katowice (Poland), between October 9th and 11th, 2006. Plots also display PBL height adopted by CAMx (red) and CHIMERE (green) at the same site (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

(FB), Mean Fractional Error (FE), correlation (r), Index of Agreement (IA), Root Mean Square Error (RMSE). A preliminary analysis of model performance (not shown) revealed that some metrics provided very similar responses; for this reason detailed analysis was limited to a subset of 4 metrics: FB, FE, *r* and RMSE. One goal of this paper was to investigate the influence of geographical features on model skill. Following the approach of Putaud et al. (2010), the computational domain was split into 3 sub-regions (Fig. 1): Southern Europe (SE), Northwestern Europe (NWE) and Eastern Europe (EE). The SE sub-region is characterised by complex circulation conditions due to coastal areas and complex terrain, it experiences hot summers enhancing photochemical activity (Millán et al., 2000; Gangoiti et al., 2001) and it can subject to dust episodes more frequently than the rest of Europe (Kallos et al., 2007; Mitsakou et al., 2008). The NWE sub-region is characterised by more homogenous circulation conditions than SE and comparison of $PM_{10}$ composition reveals higher fractions of sea salt and, to a lesser extent, nitrate than other sub-regions (Putaud et al., 2010). The EE sub-region is characterised by a higher $PM_{10}$ fraction of total carbon (Putaud et al., 2010) that could be related emission characteristics that still distinguish Eastern European countries. Observation sites were also categorised according to station type, following the official classification proposed by the European Environment Agency (EU, 1997): rural background stations (RB), suburban background stations (SB) and urban background stations (UB).

The Wilcoxon matched-pairs rank test (WMP, Wilks, 2006) was applied to perform the comparison between CAMx and CHIMERE

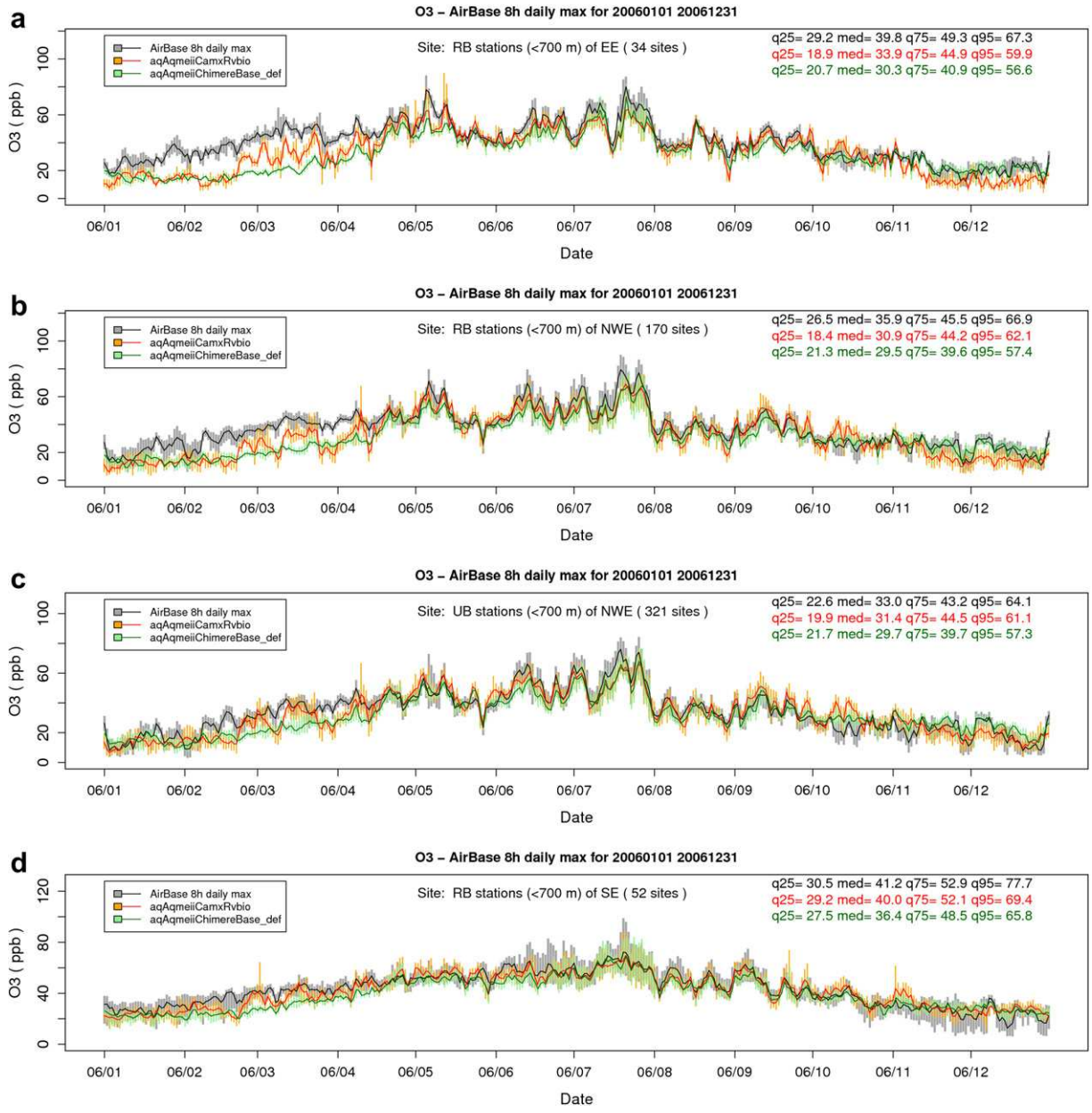


**Fig. 7.** Time series of daily Box-Whisker plots showing the distribution of the observed and computed $O_3$ concentration at RB sites of EE (a), NWE (b) SE (d) regions and at UB sites of NWE region (c). Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green. Bars show the 25th–75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole yearly series are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

skill. The WMP test is the non-parametric counterpart of the matched-pairs Student *t*-test. Being non-parametric, the test relaxes the constraint on normality of the underlying distributions (Gego et al., 2006). Firstly observation sites were categorised in subsets according to sub-region and stations type. For each subset, the pairs of metrics computed by CAMx and CHIMERE were submitted to the WMP to investigate whether the null hypothesis (i.e. the two series of metrics are not different) could be rejected or not. The probability level ($p$) of rejecting the hypothesis was set to 5%. In case of rejection (i.e. $p < 5\%$), model performance could be considered significantly different and a better performing model was identified.

The Wilcoxon–Mann–Witney test for unpaired series (WMU, Wilks, 2006) was applied to investigate differences in model performance within subsets of either region or station type. Subsets of metrics were compared by using an index computed as follows: The WMU test was applied to each possible combination of subsets (e.g. NWE versus SE; NWE versus EE and SE versus EE, in case of regional comparison) of the 7 metrics previously described and then summing the total number of "non-negative scores"(NNS) of each subset. A non-negative score takes place when a subset performs not worse than the other one. The score ranges between 0 (the subset is always worse than the others for each metric) and 14 (the subset is always better or not significantly different than the others). NNS were computed for each model separately; for each station type in case of regional comparison and, conversely, for each region in case of station type comparison.

## 4. Results and discussion

### 4.1. Nitrogen dioxide

Fig. 2a provides a concise comparison of model performance for $NO_2$ for each sub-region and station type. CHIMERE and CAMx showed a rather coherent behaviour, meaning that in most cases they provided their best or worst performance in the same region or for the same kind of station. Best performance usually occurred at rural stations in the NWE sub-region. Model estimates show FB very close to 0, absolute errors (FE) lower than 50% on average and a small spread of the distribution for all metrics, suggesting that the level of performance is fairly homogenous in the whole region. In contrast, $NO_2$ performance in the SE sub-region was systematically worse than for other sub-regions (see also Fig. S.1a) due to circulation conditions that are strongly influenced by local scale features, such as sea-land interface and complex terrain, often associated with low wind speeds and stable conditions. In all 3 sub-regions, both models showed a worsening in performance moving from rural to urban stations (see also Fig. S.1a), driven by the growing influence of local scale emissions. RMSE was very sensitive to station type because it grows according to the square of the difference between observed and computed concentrations. On the other hand, correlation was less sensitive to station type and, quite surprisingly, displayed better performances at urban or suburban stations than rural ones for SE and EE regions. This is probably due to the stronger variability at urban sites, between winter and summer observed concentrations, slightly enhancing correlation score.

The WMP test was used to discriminate when CAMx and CHIMERE performance can be considered significantly different ($p < 5\%$). As illustrated in Fig. 2a, CAMx performed significantly better than CHIMERE in terms of correlation, while CHIMERE performed better than CAMx, when assessed by FB and FE.RMSE failed to distinguish model skill, hence it does not provide useful insights about the models behaviour. This is because the RMSE formulation is sensitive to both bias and temporal agreement (Murphy, 1988) which can mutually compensate for $NO_2$.

The WMP test also allowed detection of differences in model performance that are not obvious from box-whisker plots. As an example, RMSE distribution at SB and UB sites of SE region seems to show comparable performance for CAMx and CHIMERE at both station types. However, the WMP test reveals statistically significant difference between the two models. This result stems from the WMP approach that takes into account the number of times that one model performs better than the other one. In this case, this means that CHIMERE is slightly but systematically more skilful than CAMx at UB stations.

Fig. 3 shows the daily Box-whisker plots of the distribution of the observed and computed $NO_2$ concentration at SB sites of NWE and SE region. CHIMERE concentrations are almost always higher than CAMx, thus explaining the better score in FB and FE. Conversely, CAMx seems to better reproduce the weekly cycle of $NO_2$ concentrations, giving rise to a higher correlation

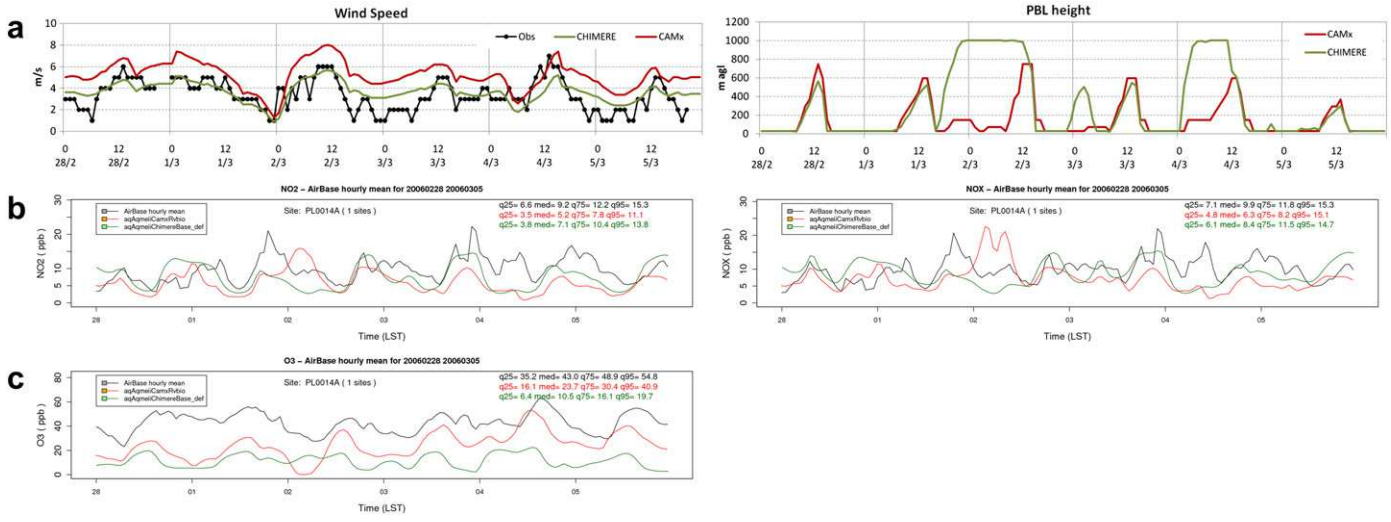


**Fig. 8.** Hourly time series of observed and computed fields at PL0014A site from 2/28/2006 to 3/5/2006: wind speed and PBL height (a); NOx and $NO_2$ concentrations (b); ozone concentration (c). Observations are in black; CAMx in red and CHIMERE in green. As for chemical species, the 25th, 50th, 75th and 95th quantile of the hourly series are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

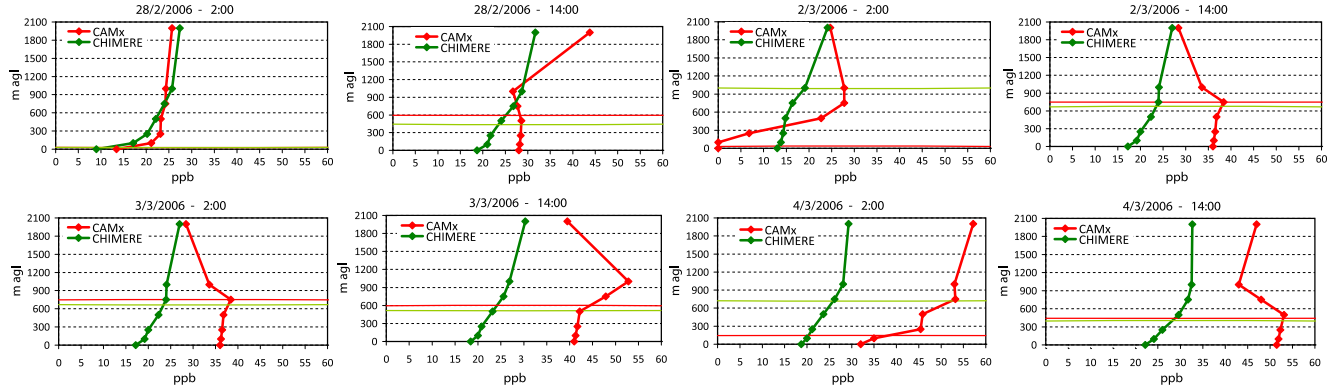


**Fig. 9.** Selection of hourly ozone vertical profiles computed by CHIMERE (green) and CAMx (red) at site PL0014A, between February 28th and March 4th. Plots also display PBL height adopted by CAMx (red) and CHIMERE (green) at the same site (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

skill. As discussed below (Section 4.3), such differences are related to the different assumptions underlying the reconstruction of the vertical diffusion and the first layer wind speed in the two models.

### 4.2. Sulphur dioxide

Fig. 2b compares CAMx and CHIMERE performance in simulating $SO_2$ concentrations. Concerning FB, CAMx performs better

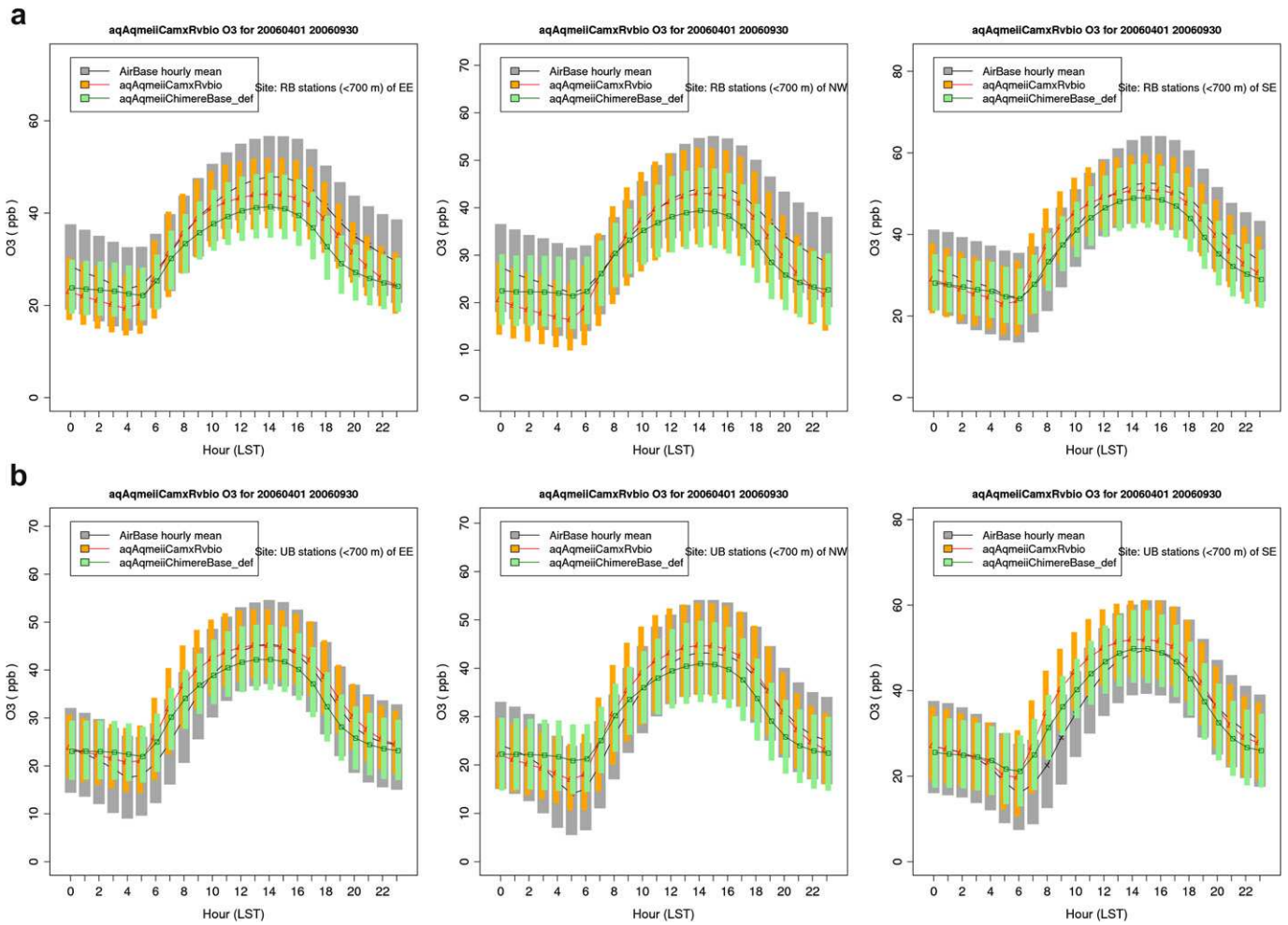


**Fig. 10.** Ozone mean day concentration at RB (a) and UB (b) stations of EE, NWE and SE regions. Each bar represents 25th–75th quantile interval of the distribution of the concentrations at all stations for the same hour. Lines display the median of the distribution. Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

than CHIMERE showing a slight overestimation, whereas CHIMERE concentrations are underestimated. In contrast, CHIMERE shows better skill than CAMx in capturing temporal variability of observed concentrations, as shown by the correlation values. There is a clear worsening in model performance in the SE sub-region as discussed above for $NO_2$. Overall, geographical region is less important for $SO_2$ in CAMx (left panel in Fig. S.1b) but CHIMERE performance is clearly worst in the SE sub-region. Station type is less influential for $SO_2$ model skill (right panel in Fig. S1.b) because $SO_2$ emissions mainly come from aloft sources (Table 1) which disperse emissions widely. An exception is presented by RMSE for UB and SB stations in the EE region that show a clear worsening, increasing from 2 to 4 ppb, on average (Fig. 2b). This happens because surface level sources of $SO_2$ are still relevant in the EE region and they influence the observed concentrations at UB and SB sites (Hjellbrekke and Fjæraa, 2008).

Fig. 4 shows the box-whisker time-series of $SO_2$ daily concentrations at SB stations for the EE and SE regions. CAMx computes higher concentrations that result in better performance for FB but without reproducing the time series variability. Indeed, it can be noted that over EE stations, CAMx overestimates the lowest quantiles of the yearly series, while underestimating the highest ones. Moreover, the model tends to underestimate January–March concentrations, whereas the October–December period is overestimated. Similar conclusions can be drawn for SE stations where the spread of the observed distribution is well reproduced by both models, but not the temporal variability. Moreover, the seasonal cycle at SE stations is very smooth, causing further worsening in correlation estimates. Because point source stack parameters were lacking from the emission inventory both models were forced to assume static vertical profiles to distribute point source emission rather than calculating time-varying plume rise based on meteorological conditions. Fig. 5 compares the daily mean $SO_2$ concentration computed by CAMx and CHIMERE for October 10th, 2006. CAMx ground level concentrations are always higher than CHIMERE, above all urban and industrialised areas. To investigate the differences between the two models, Fig. 6 compares the $NO_2$ and $SO_2$ vertical profiles computed by CAMx and CHIMERE from October 9th to October 11th at an industrial area close to Katowice (Poland), where $SO_2$ maximum concentrations are found. $NO_2$ concentrations display a rather typical hourly profile driven by emissions and the Planetary Boundary layer (PBL) evolution and the models show similar behaviour, although CAMx concentrations are higher than CHIMERE. Maximum ground level $NO_2$ concentrations are observed late in the evening due to ozone-NO titration combined with low PBL height. The latter also creates a sharp vertical gradient in $NO_2$ concentrations dropping from 20–30 ppb for CHIMERE and 25–35 ppb for CAMx at ground level to less than 12 ppb at 100 m above ground level (agl). Conversely, minimum values (around 6 ppb) are observed during daytime, due to chemical removal of $NO_2$ and a deeper PBL. $SO_2$ shows a relatively different profile, especially on October 10th and 11th when the highest concentrations are observed for both models. Both models

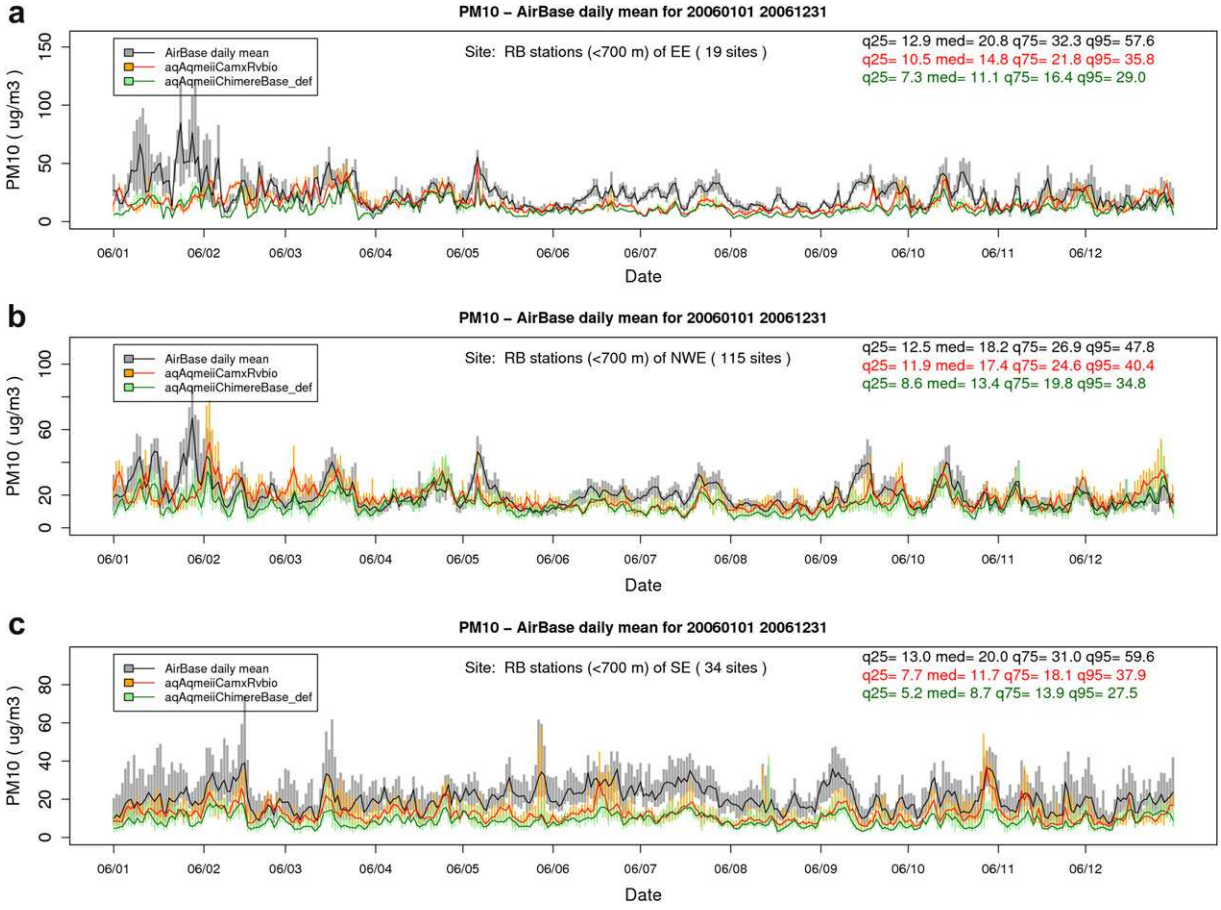


**Fig. 11.** Time series of daily Box-Whisker plots showing the distribution of the observed and computed $PM_{10}$ concentration at RB sites of EE (a), NWE (b) SE (c) regions. Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green. Bars show the 25th–75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole yearly series are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

show comparable profiles at noon on October 9th, but when the PBL collapses the models behave very differently: Both models display a concentration peak between 250 and 500 m agl, confirming the importance of aloft sources for $SO_2$ but CAMx has higher concentrations in the surface layer than aloft whereas CHIMERE has lower concentrations in the surface layer than aloft. Similar behaviour, even enhanced, is shown on October 11th. Differences in the models results can be explained noting that: a) CAMx displays a sharp gradient close to ground level during stable conditions, while CHIMERE maxima take place mostly at higher altitude; b) CAMx concentrations are usually higher than CHIMERE inside the PBL, while at higher altitude (over 1000 m agl) CHIMERE values can be greater than CAMx. Considering that the models: a) shared the same emission inventory b) adopted the same vertical distribution for point source emissions c) showed similar dry deposition fields (Fig. S.2), the differences showed by the models can be ascribed to different assumptions in the description of the PBL processes for stable conditions.

### 4.3. Ozone

CAMx and CHIMERE also present coherent behaviour for secondary pollutants. The best performance for ozone (Fig. 2c) takes place in the SE sub-region, with FB values close to 0, whereas NWE and EE are characterised by a negative bias ranging from 10 to 30%. Rather surprisingly, model performance clearly improves moving from rural to urban stations, where FB is close to 0. By examining the FE distribution conclusions similar to FB can be extracted, with values ranging, on average, from 20 to 30% in SE region and being greater than 30% in Eastern Europe. In contrast to $NO_2$, CHIMERE and CAMx did not show any statistically significant difference from region to region (Fig. S.1c). For both metrics, CHIMERE skills are significantly better than CAMx for most subsets. Both models present a noticeable skill in terms of correlation. CHIMERE performed very well in SE and NWE regions showing correlation values higher than 0.8 at more than 50% of the selected sites and being significantly better than CAMx. Conversely, CAMx performed better at EE sites, whereas CHIMERE correlation drops to values lower than 0.8 at most sites.

A clear worsening in CAMx performance is presented only for UB stations for the SE sub-region. This happens because $O_3$ concentrations at urban stations are rather influenced by local scale effects (e.g. titration) that are not well captured in the SE sub-region. CHIMERE displays a very different behaviour in the EE region where performance is always the worst. Such a discrepancy is driven by correlation at EE sites, which is clearly lower than the other regions. Comparing $O_3$ model estimates for different station types (right panel in Fig. S.1c) showed a rather surprising outcome, where rural stations are usually worse than others. Such unexpected behaviour could be driven by an error compensation or it could indicate that station classification is not correctly identified. This result possibly denotes that only selecting rural monitoring stations for the evaluation of ozone simulations could provide misleading conclusions, while including other station types (urban, suburban) produced more representative results.

Fig. 7 presents the box-whisker plots for daily maximum 8-hour $O_3$ for a few station subsets. Generally speaking, both models are able to follow the seasonal cycle of daily maximum ozone also reproducing most of the episodes taking place in the summer season. Simulated concentrations are slightly underestimated by

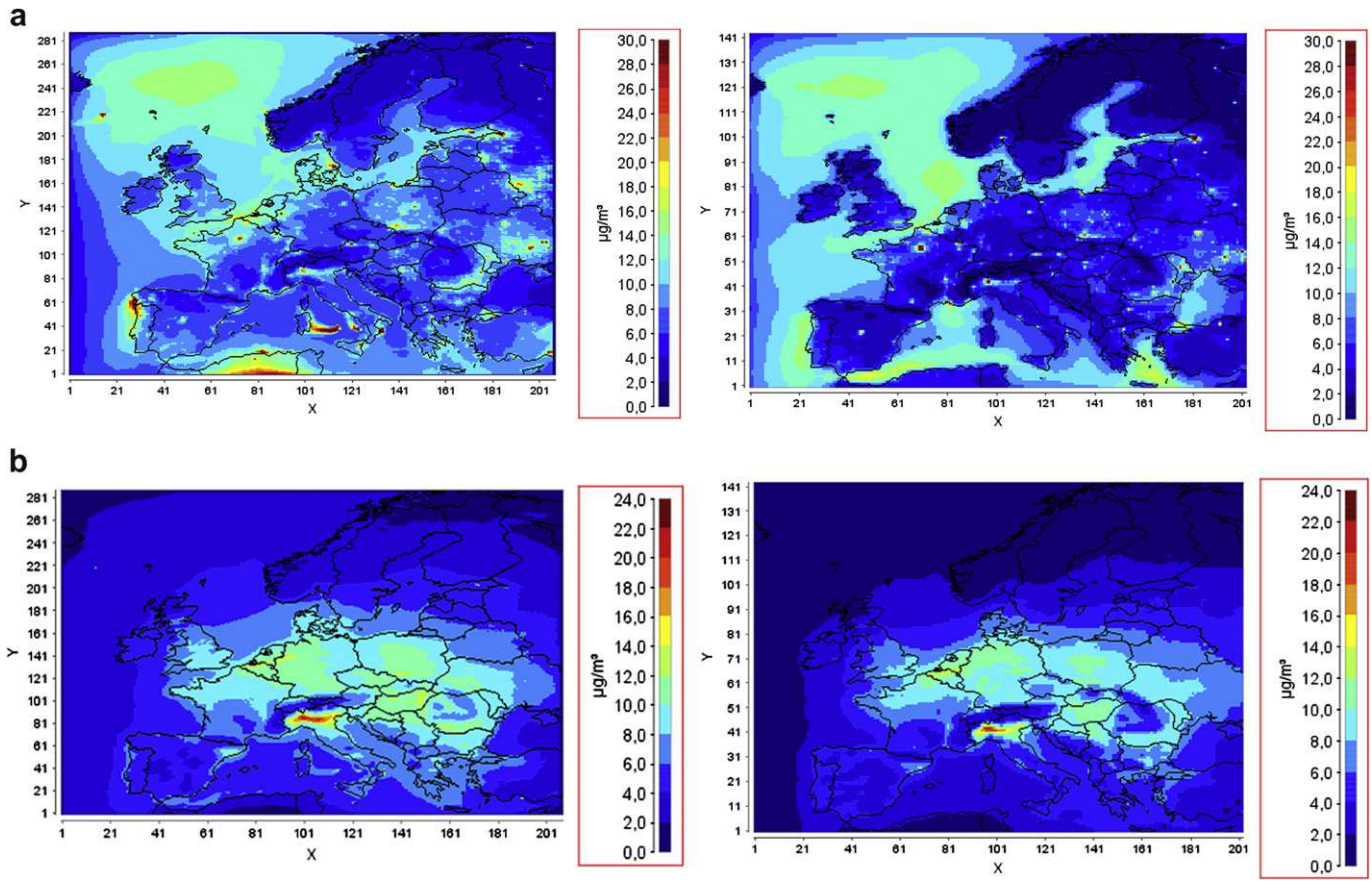


**Fig. 12.** Primary (a) and secondary (b) $PM_{10}$ yearly mean concentration computed by CAMx (left) and CHIMERE (right).

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

both models, as shown by comparing the quantiles. CHIMERE performed better in reproducing the low concentrations, while CAMx better reproduced the median and high quantiles. The analysis of the temporal evolution shows that model bias is mainly driven by a strong underestimation taking place during the first part of the year in the NWE and EE regions when both models show the strongest discrepancy. This feature is very clear at EE stations, thus explaining the significant differences in correlation. This underestimation of ozone during the first part of the year results from a lack of ozone in the northern boundary conditions, as explained through sensitivity simulations by Nopmongcol et al. (in this issue).

To investigate differences in model behaviour, Fig. 8 displays the temporal evolution of selected variables at a rural site in the EE region (PL0014A). The 5 day period, from February 28th to March 3rd, is characterised by the development of a spring ozone episode, with observed concentrations reaching 60 ppb. Observed wind speed ranges between 1 and 6 m $s^{-1}$. The models are able to capture the hourly evolution of wind speed but CHIMERE has lower wind speeds than CAMx because CHIMERE has a shallower surface layer (20 m) than MM5 (30 m) and therefore adjusted down the MM5 wind speeds, whereas CAMx has the same surface layer depth as MM5 and used the MM5 winds directly. CAMx and CHIMERE PBL heights are both derived from MM5, but the influence of the CHIMERE modification to PBL height during cloudy days is clearly evident for example on $NO_X$ and $NO_2$ concentrations at night on March 2nd. Conversely, when the PBL is very low in both models, the $NO_X$ and $NO_2$ concentrations simulated by CHIMERE are often higher than CAMx (e.g. evening hours of March 2–4). These differences result from the wind speed and the minimum value of the vertical dispersion coefficient (Kz) adopted by the models. The influence of these differences in meteorological fields is rather systematic as it can be inferred from the computed quantiles, better reproduced by CHIMERE than CAMx (Fig. 9b). The differences in the reconstruction of wind and vertical diffusion aim in explaining the resulting differences in the night-time ozone concentrations. However comparing ozone time series, it can be noted that models differ in the reconstruction of the daytime build-up too, being stronger in CAMx than CHIMERE. Higher ozone concentrations can be also observed along the vertical profile, as shown in Fig. 10, which illustrates the increase of CAMx concentrations from February 28th to March 4th. This difference could be attributed to the chemical mechanism, suggesting that CB05 is more effective than MELCHIOR in producing ozone. This behaviour is clearly displayed by the increasing discrepancy between CAMx and CHIMERE daytime vertical profiles, inside the PBL. Moreover, ozone produced during daytime is accumulated in upper layers, enhancing the differences between the two models along the development of the episode. The development of higher ozone concentrations is confirmed also by the comparison of other species, such as $H_2O_2$ (not shown) where CAMx daily maximum is twice the values produced by CHIMERE. As a final result, CAMx performed better

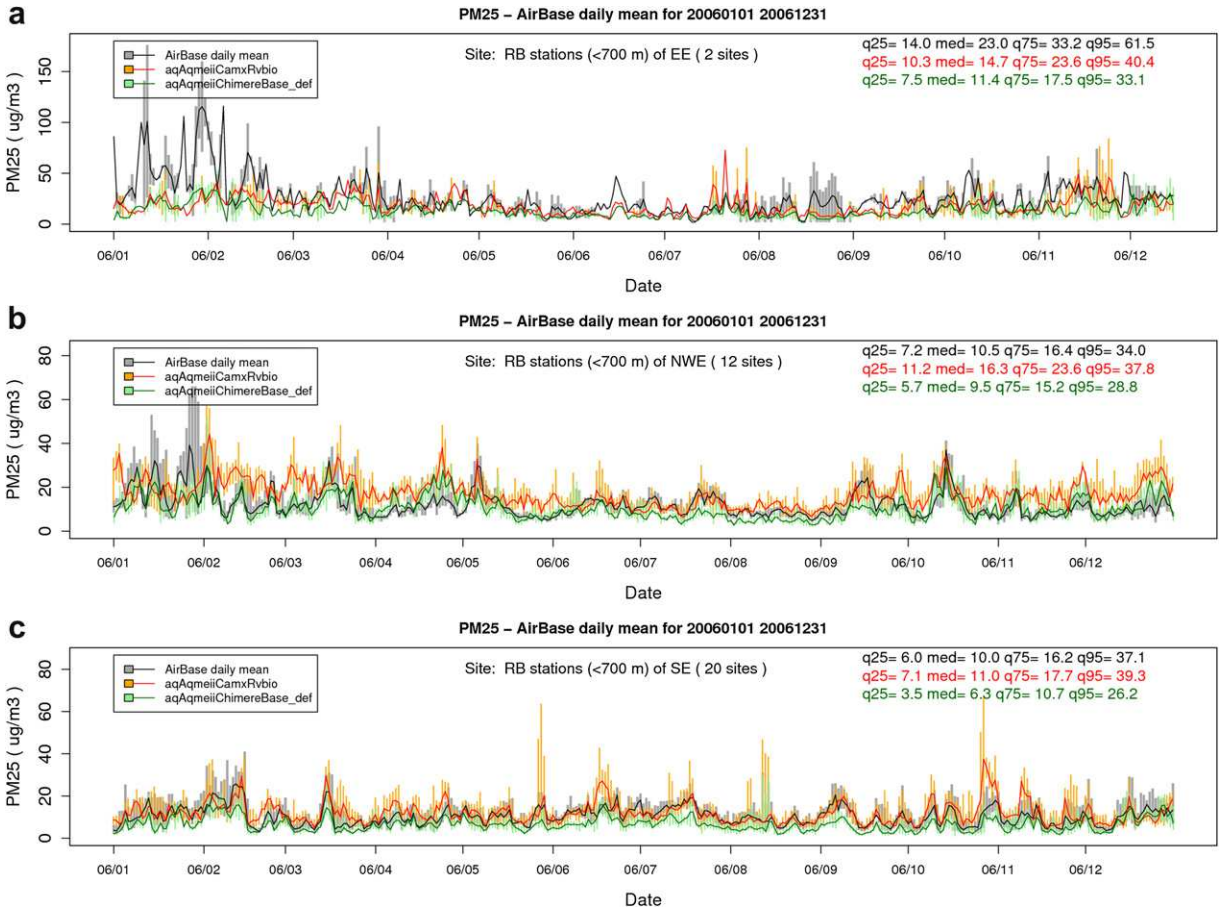


**Fig. 13.** Time series of daily Box-Whisker plots showing the distribution of the observed and computed $PM_{10}$ concentration at RB sites of EE (a), NWE (b) SE (c) regions. Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green. Bars show the 25th–75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole yearly series are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

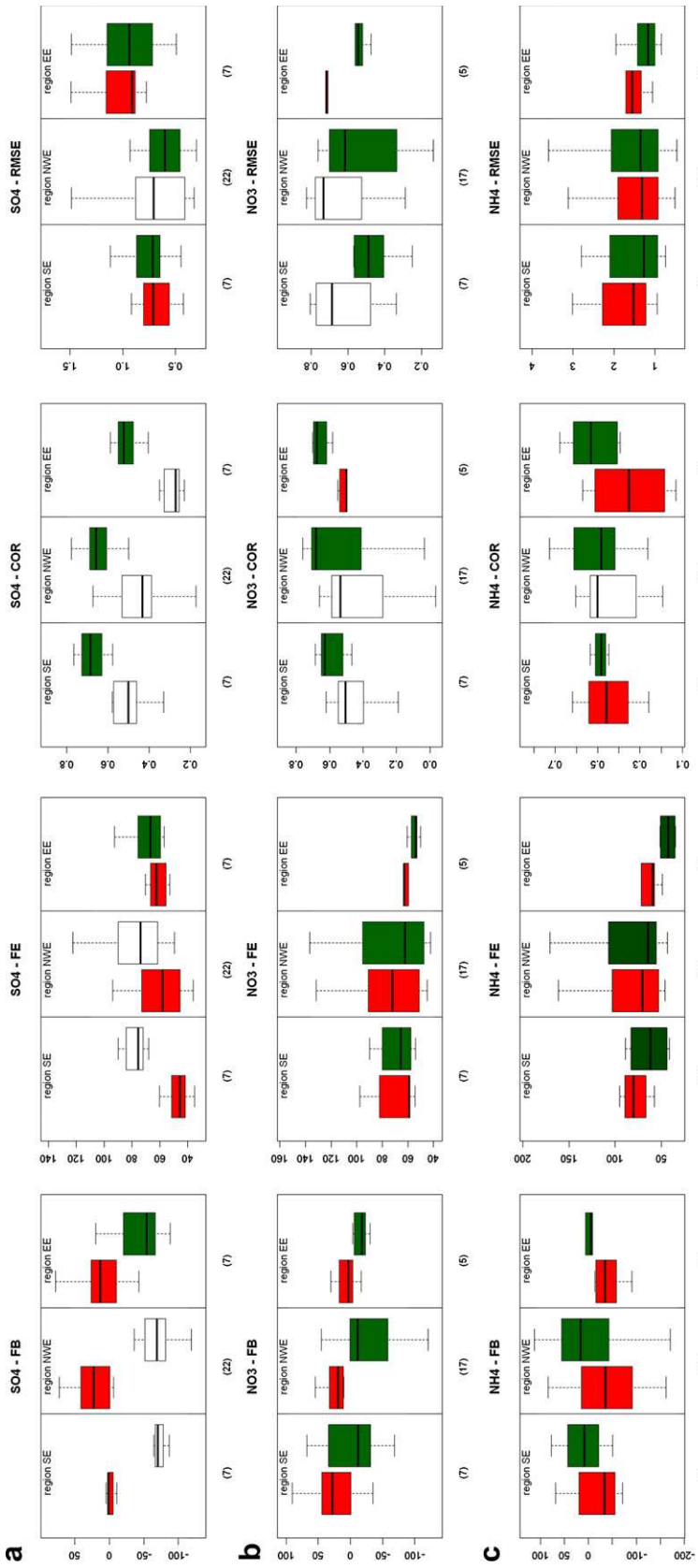


**Fig. 14.** Box-whisker plots of the distribution of the different metrics computed for CAMx (red) and CHIMERE (green) in each region and for EMEP RB stations. The number of stations included in each subset is reported in brackets. If the performances are significantly different, the plot is unfilled for the worst model (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

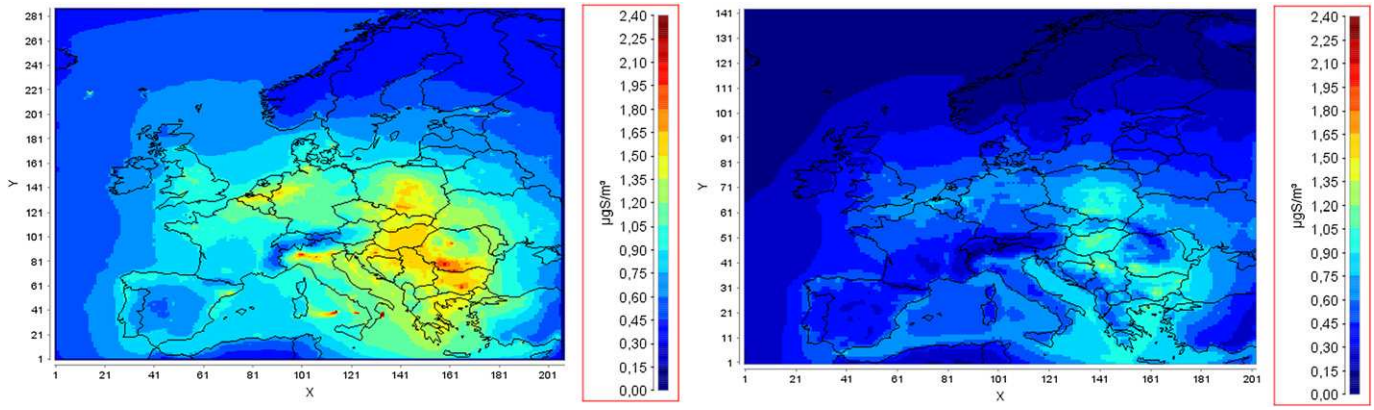


**Fig. 15.** Sulphate yearly mean concentration computed by CAMx (left) and CHIMERE (right).

than CHIMERE, because the stronger ozone production in CAMx compensated for underestimation in the background ozone caused by the boundary conditions.

To better investigate possible differences in photochemistry, the ozone mean day concentrations from April to September are compared in Fig. 11. The seasonal analysis confirms that the increase in the daytime concentration is systematically higher in CAMx than CHIMERE. Discrepancies are stronger during the first part of the daytime period, supporting the hypothesis that CB05 produces more ozone than MELCHIOR. This finding seems to be confirmed also by the spread of the computed concentrations that is higher in CAMx than CHIMERE. CAMx skills show better in terms of hourly ozone peak, but not over the whole daytime period. This result could provide an explanation for the more accurate CHIMERE performance with respect to the daily maximum 8-hour ozone.

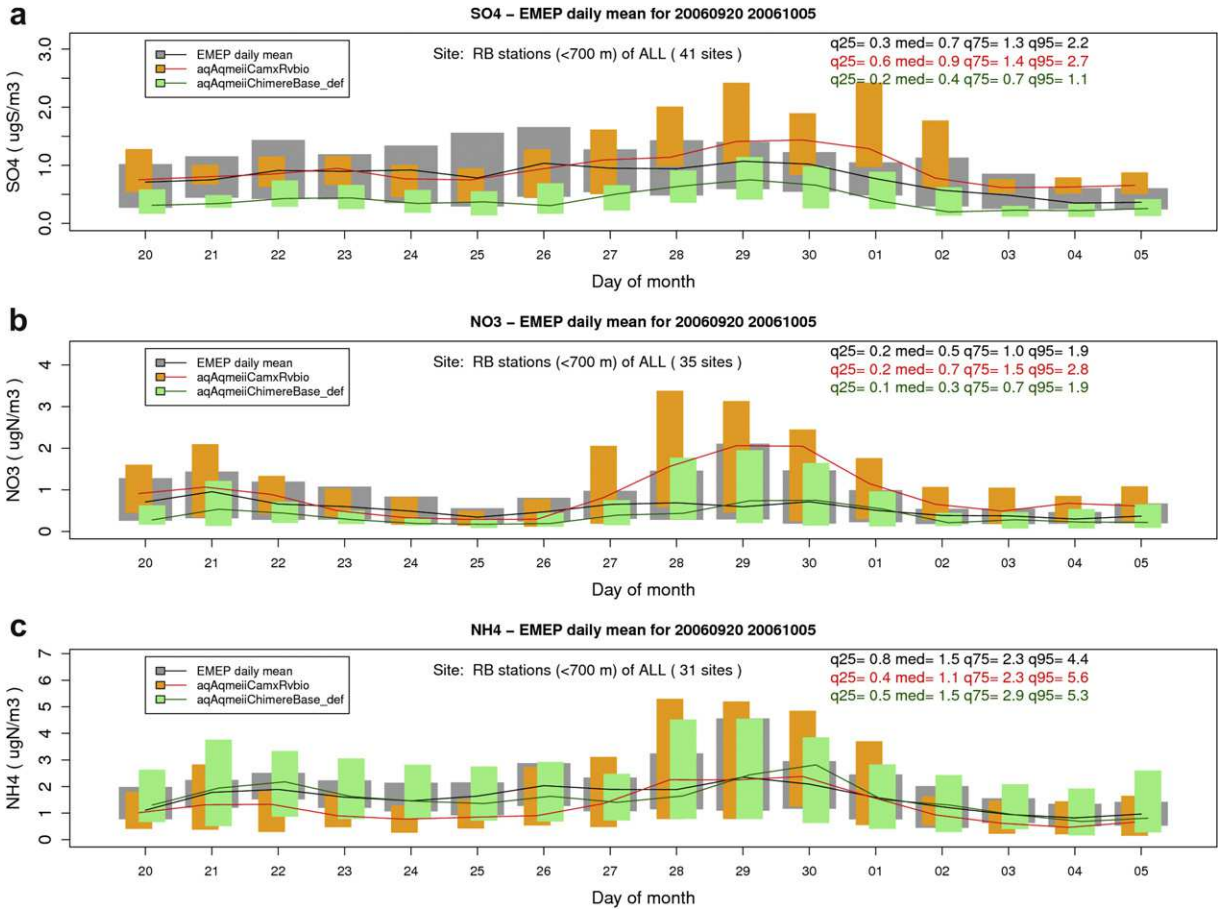


**Fig. 16.** Time series of daily Box-Whisker plots showing the distribution of the observed and computed concentration of sulpahte (a), Nitrate (b) and ammonium (c) at EMEP sites. Observations are in black/grey; CAMx in red/orange and CHIMERE in dark green/light green. Bars show the 25th–75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole period are reported too (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

### 4.4. Particulate matter ($PM_{10}$ and $PM_{2.5}$)

In analysing $PM_{10}$ performance, shown in Fig. 2d, it appears once again that the models provided a coherent answer both comparing different regions as well as considering different types of station. Similar to $NO_2$, the best $PM_{10}$ performance take place in NWE region and for RB stations. The WMP test shows that CAMx systematically provided better FB and FE scores than CHIMERE. CAMx bias is close to 0 in NWE region, whereas SE and EE stations display a negative bias. CHIMERE presents a similar pattern but characterised by a stronger negative bias. Similar findings can be derived by FE scores. Analysing the WMP results for correlation displays a very different pattern, with CHIMERE performing always better than CAMx. The RMSE values display a less clear pattern, where differences between CAMx and CHIMERE are reduced, due to the influence of both bias and temporal variability, which mutually compensate.

Differences among regions depend neither on the model nor the station type, confirming the strong influence of geographical features on $PM_{10}$ simulations. As shown in Fig. S.1d, differences in region to region skill are mainly driven by bias, which ranges around 0 in the NWE area, dropping down to −100% in the SE region. Discrepancies in $PM_{10}$ FB and FE are stronger than $NO_2$ (Fig. 2a), indicating that there is inaccuracy in the reconstruction of

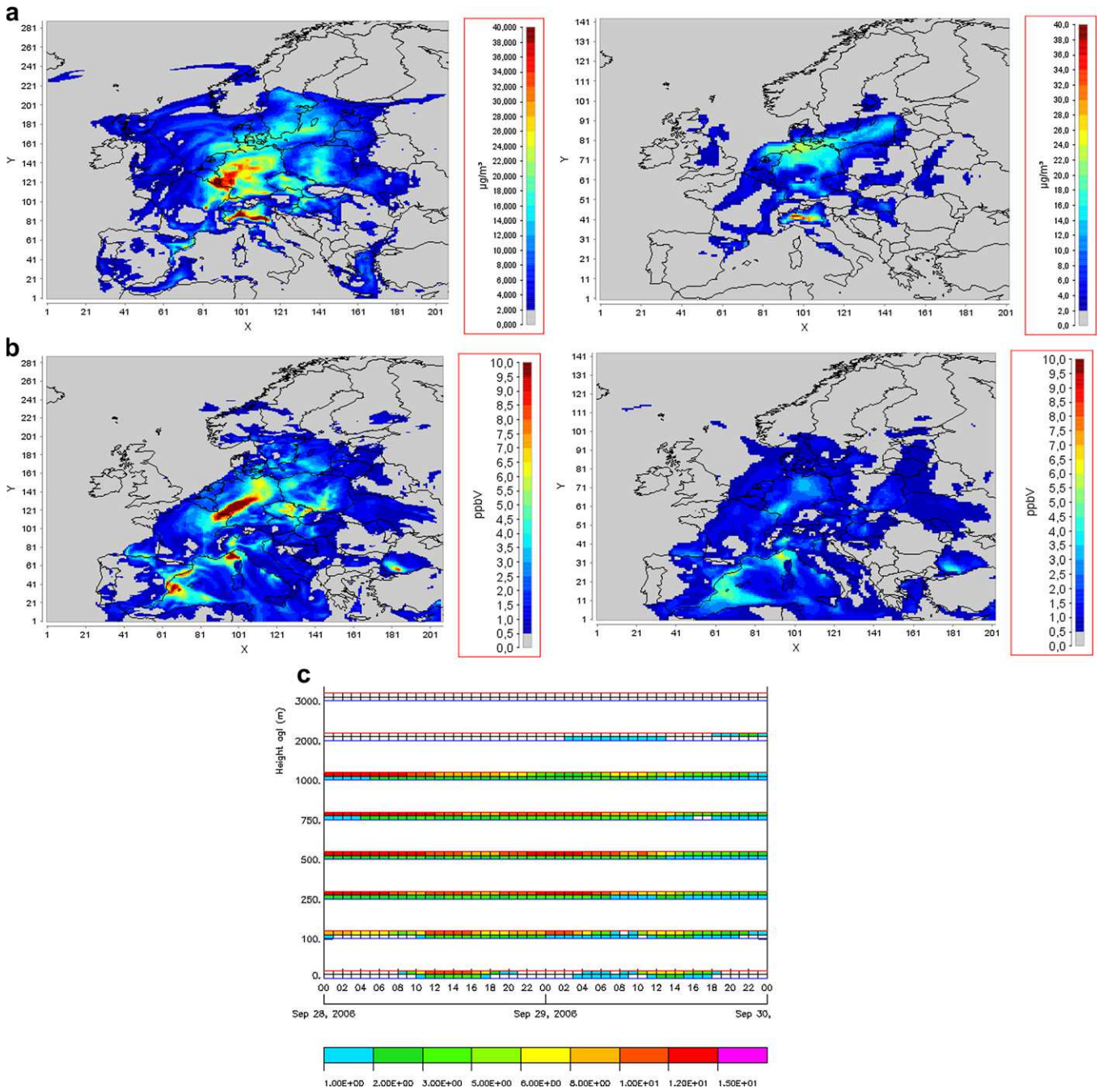


**Fig. 17.** (a) Nitrate hourly mean concentration computed by CAMx (left) and CHIMERE (right) on September 29th at 07:00; (b) Nitric acid hourly mean concentration computed by CAMx (left) and CHIMERE (right) on September 28th at 12:00; (c) Hourly time series of nitric acid vertical profiles computed by CHIMERE (lower band) and CAMx (upper band) at site (8.00E, 50.00N) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

either emission sources or aerosol processes or both, influencing $PM_{10}$ concentration in the EE and SE regions.

To investigate the differences in PM model performances, Fig. 12 compares the yearly mean concentration of primary and secondary $PM_{10}$ computed by both models. The spatial patterns look similar, but CAMx concentrations are generally higher than CHIMERE, mainly for primary $PM_{10}$, thus helping to explain the stronger bias exhibited by CHIMERE. The discrepancies between the two models can be ascribed to: a lower contribution of dust at boundaries in CHIMERE due to a smoothing filter applied to peak events; a lower emission of $PM_{10}$ at ground level (see Table 1)); more efficient wet deposition scavenging in CHIMERE (see Fig. S.3).

As a further step in evaluating PM performance, $PM_{2.5}$ concentrations at RB sites were compared. Due to the lower number of $PM_{2.5}$ stations, the results shown in Fig. 13 cannot be strictly compared to Fig. 11. However, available stations show that $PM_{2.5}$ modelled concentrations are closer to observations than $PM_{10}$. A clear improvement in model performance can be observed for CHIMERE at NWE sites and for CAMx at SE sites. Comparing $PM_{2.5}$ and $PM_{10}$ shows that, as expected, observed concentrations clearly decrease when just the fine PM fraction is considered. However, the quantiles of the modelled concentration distributions are rather constant. This result suggests that emissions of coarse PM are missing from both models.

Finally, it is worth noting that CAMx overestimated the $PM_{2.5}$ concentration at NWE sites. This behaviour was investigated further. Fig. 14 provides an overview of CAMx and CHIMERE performance in reproducing the three main inorganic aerosol compounds at RB sites, the only type of stations available in the EMEP dataset. Comparing sulphate performance shows that CAMx provided better FB and FE scores, whereas the model was clearly worse than CHIMERE for correlation. CAMx concentrations were higher than CHIMERE (see Fig. 15), due to the corresponding higher availability of $SO_2$, as discussed previously. Both models presented comparable FB and FE scores for both nitrate and ammonium. Similar to $PM_{10}$, CHIMERE correlation estimates are generally better than CAMx, especially for sulphate and nitrate. Also in this case, the worsening in CAMx performances is due an overestimation of the variability of computed concentrations, both in space and time. Fig. 16 provides an example of such behaviour. Differences between the models can be clearly detected by analysing the episode that occurred between September 27th and October 2nd, where CAMx overestimates both sulphate and nitrate concentrations. Fig. 17 helps in explaining model discrepancies. Panels (a) show the nitrate hourly concentrations on September 29th at 07:00, when CAMx concentrations are higher than CHIMERE, above all in a large area across France and Germany. Differences in nitrate concentrations can be related to a higher availability of nitric acid, whose concentration is higher in CAMx, starting from the day before (as shown in Fig. 17b). The increase in $HNO_3$ concentration that takes place during daytime hours is caused by the development of the PBL that favour the vertical mixing of pollutants produced by high level sources. Between 250 and 500 m agl, CAMx exhibits $HNO_3$ night-time concentrations higher than 12 ppb, while CHIMERE is lower than 3 ppb. As soon as the PBL starts growing a downward mixing takes place, giving rise to a strong increase in ground level concentrations. This result suggests that, similar to $SO_2$, the discrepancies between the two models are driven by different assumptions in simulating PBL processes.

## 5. Conclusions

Two CTMs were evaluated and compared over Europe for calendar year 2006 in the framework of the AQMEII project. The analysis sub-setted the observational sites according to geographical region and station type. Performance statistics were compared objectively by application of a non-parametric statistical test of matched pairs rank.

The models demonstrated similar geographical variations in model performance with just a few exceptions: for $SO_2$ in the SE sub-region and $O_3$ in the EE sub-region. Both models displayed great performance variability from region to region and within the same region for $NO_2$ and $PM_{10}$. Station type is relevant mainly for pollutants directly influenced by low level emission sources, such as $NO_2$ and $PM_{10}$, while station type is not influential for region to region comparisons. The analysis demonstrated that selecting RB stations is not necessarily a good "a-priori strategy" for model evaluation. For some pollutants, like $SO_2$ including urban and suburban stations could enrich the database, giving further chances to investigate model performance leading to more reliable findings.

Investigation of model performance differences showed that FB and FE metrics together with correlation index (or index of agreement) often highlighted significant differences in model scores, usefully guiding model users to further analysis of model behaviour. Conversely, in several cases, RMSE proved uninformative, being unable to identify significant differences in model performance. The WMP test provided a clear and robust answer about the significance of the differences between the models, also allowing systematic differences in model performance to be distinguished, even at low concentrations.

A more detailed analysis of the likely causes of the differences between CAMx and CHIMERE results revealed that:

- Differences in the reconstruction of vertical diffusion coefficients (Kz) and wind speed in the first model layers can affect the surface concentration of primary compounds, especially for stable conditions. Lower threshold for minimum Kz could enhance $NO_2$ peaks in CHIMERE, improving FB. Also, taking into account the influence of clouds on PBL height can modify the reconstruction of the daily variability yielding different correlation values.
- Differences in the vertical profiles of Kz strongly influenced the impact of aloft sources on ground level concentrations of both primary pollutants such as $SO_2$ as well as $PM_{10}$ compounds such as sulphate and nitrate. CAMx vertical mixing proved to be more efficient than CHIMERE (note that since CAMx vertical mixing is determined by input Kv profiles this finding may be specific to this application). As a consequence, CAMx often performed better in terms of bias, while CHIMERE was better than CAMx for correlation. This happened because the stronger mixing produced a general increase of ground level concentrations, but also caused the overestimation of several episodes.
- CAMx showed stronger photochemistry than CHIMERE giving rise, on average, to higher ozone concentrations that agreed better with observations, as shown by analysis of the diurnal variation during the summer season. However, CHIMERE performance on daily basis was better than CAMx because the greater variability of the CAMx concentrations yielded worsening bias and correlation. The only exception was for the EE region where CAMx estimates were more accurate. Nonetheless, this result seems to be due to an error compensation, where the more effective photochemistry showed by CAMx compensated for an underestimation in the background concentration. Concerning this last point, it is worth noting that, the best model performance was observed in the SE region, the only area not influenced by a strong underestimation taking place in the early spring.

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052

- $PM_{10}$ performance was rather poor for both models, except for the NWE region. Model results were sensitive to geographical features and station type similar to $NO_2$. However, differences in model performance between the NWE region and the other two areas were stronger than for $NO_2$, suggesting that either further emission sources, or processes, or both are missing for $PM_{10}$ in the SE and EE regions. Moreover, $PM_{10}$ performance was very different between regions, while secondary inorganic aerosol scores were relatively homogenous. This suggests that $PM_{10}$ underestimation has to be ascribed to other compounds (e.g. PM coarse, Particulate Organic Matter and dust). This finding has been confirmed by comparing $PM_{2.5}$ stations, which exhibited a lower bias than $PM_{10}$ sites. This result proved that coarse PM sources are still missing from emission inventories. Beside these shared features, comparing the two models displayed a rather unexpected result, with CAMx performing always better than CHIMERE in terms of bias, while CHIMERE score for correlation was always higher than CAMx. As already mentioned, vertical mixing is one cause of such discrepancies. The previous analysis also suggested that the stronger underestimation experienced by CHIMERE was mainly influenced by temporal smoothing of the boundary conditions, underestimation of low level emissions (mainly related to fires) and more intense depletion by wet deposition.

## Acknowledgements

RSE contribution to this work has been partially financed by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE (formerly known as ERSE) and the Ministry of Economic Development - General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency, stipulated on July 29th, 2009 in compliance with the Decree of March 19th, 2009. The CAMx model application was supported by the Coordinating Research Council Atmospheric Impacts Committee (CRC Project A-75).

## Appendix A

The statistical indicators selected to evaluate the model performances have been defined as follows:

Normalised Mean Bias (NMB):

$$\mathrm{NMB} = \frac{\frac{1}{N}\sum_{t=1}^{N}(C_{\mathrm{mod}}(x,t) - C_{\mathrm{obs}}(x,t))}{\frac{1}{N}\sum_{t=1}^{N} C_{\mathrm{obs}}(x,t)} \quad (A.1)$$

Normalised Mean Error (NME):

$$\mathrm{NME} = \frac{\frac{1}{N}\sum_{t=1}^{N}|C_{\mathrm{mod}}(x,t) - C_{\mathrm{obs}}(x,t)|}{\frac{1}{N}\sum_{t=1}^{N} C_{\mathrm{obs}}(x,t)} \quad (A.2)$$

Mean Fractional Bias (FB)

$$\mathrm{FB} = \frac{1}{N}\sum_{t=1}^{N}\frac{C_{\mathrm{mod}}(x,t) - C_{\mathrm{obs}}(x,t)}{(C_{\mathrm{obs}}(x,t) + C_{\mathrm{mod}}(x,t))/2} \quad (A.3)$$

Mean Fractional Error (FE)

$$\mathrm{FE} = \frac{1}{N}\sum_{t=1}^{N}\frac{|C_{\mathrm{mod}}(x,t) - C_{\mathrm{obs}}(x,t)|}{(C_{\mathrm{obs}}(x,t) + C_{\mathrm{mod}}(x,t))/2} \quad (A.4)$$

Correlation Index ($r$)

$$r = \frac{\sum_{t=1}^{N}(C_{\mathrm{mod}}(x,t) - \overline{C}_{\mathrm{mod}}(x))\cdot(C_{\mathrm{obs}}(x,t) - \overline{C}_{\mathrm{obs}}(x))}{\sqrt{\sum_{t=1}^{N}(C_{\mathrm{mod}}(x,t) - \overline{C}_{\mathrm{mod}}(x))^2}\cdot\sqrt{\sum_{t=1}^{N}(C_{\mathrm{obs}}(x,t) - \overline{C}_{\mathrm{obs}}(x))^2}} \quad (A.5)$$

Index of Agreement (IA)

$$\mathrm{IA} = 1 - \frac{\sum_{t=1}^{N}(C_{\mathrm{mod}}(x,t) - C_{\mathrm{obs}}(x,t))^2}{\sum_{t=1}^{N}(|C_{\mathrm{mod}}(x,t) - \overline{C}_{\mathrm{obs}}(x)| + |C_{\mathrm{obs}}(x,t) - \overline{C}_{\mathrm{obs}}(x)|)^2} \quad (A.6)$$

Root Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(C_{\mathrm{mod}}(x,t) - C_{\mathrm{obs}}(x,t))^2} \quad (A.7)$$

$C_{\mathrm{mod}}(x,t)$ – computed concentration; $C_{\mathrm{obs}}(x,t)$ – observed concentration; $n$ – number of pairs.

A cut-off threshold has been applied to the observed concentrations to avoid numerical problems due to unrealistic observations. Thresholds have been defined as follows:

$NO_2 = 0.5$ ppb; $O_3 = 5$ ppb; $PM_{10} = 1\ \mu g\ m^{-3}$; $SO_2 = 0.2\ \mu g\ m^{-3}$; $SO_4 = 0.01\ \mu g\ m^{-3}$; $NO_3 = 0.01\ \mu g\ m^{-3}$; $NH_4 = 0.01\ \mu g\ m^{-3}$.

## Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.atmosenv.2011.12.052.

## References

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

Bessagnet, B., Khvorostyanov, D., Menut, L., Monge, J.L., Vautard, R., 2009. Documentaion of the chemistry transport model CHIMERE. http://www.lmd.polytechnique.fr/chimere/docs/CHIMEREdoc2008.pdf.

Bessagnet, B., Seigneur, C., Menut, L., 2010. Impact of dry deposition of semi-volatile organic compounds on secondary organic aerosols. Atmos. Environ.. ISSN: 1352-2310 44 (14). ISSN: 1352-2310, 1781–1787. doi:10.1016/j.atmosenv.2010.01.027.

Boylan, J., Russel, A., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmos. Environ. 40, 4946–4959.

Chang, J.C., Hanna, S.R., 2004. Air quality model performance. Meteorol. Atmos. Phys. 87, 167–196.

Cuvelier, C., Thunis, P., Vautard, R., Amann, M., Bessagnet, B., Bedogni, M., Berkowicz, R., Brandt, J., Brocheton, F., Builtjes, P., Carnevale, C., Copalle, A., Denby, B., Douros, J., Graf, A., Hellmuth, O., Honoré, C., Hodzic, A., Jonson, J., Kerschbaumer, A., de Leeuw, F., Minguzzi, E., Moussiopoulos, N., Pertot, C., Peuch, V.H., Pirovano, G., Rouil, L., Sauter, F., Schaap, M., Stern, R., Tarrason, L., Vignati, E., Volta, M., White, L., Wind, P., Zuber, A., 2007. CityDelta: A model intercomparison study to explore the impact of emission reductions in European cities in 2010. Atmos. Environ. 41 (1), 189–207.

de Leeuw, G., Neele, F.P., Hill, M., Smith, M.H., Vignati, E., 2000. Production of sea spray aerosol in the surf zone. J. Geophys. Res. 105, 29,397–29,409.

Denby, B. (Ed.), 2011. Guidance on the Use of Models for the European Air Quality Directive (A FAIRMODE Working Document). http://fairmode.ew.eea.europa.eu/ETC/ACC report, ver 6.2.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S.T., Scheffe, R., Schere, K., Steyn, D., Venkatram, A., 2010. A framework for

evaluating regional-scale numerical photochemical modeling systems. Environ. Fluid Mech.. doi:10.1007/s10652-009-9163-2.

Dudhia, J., 1993. A nonhydrostatic version of the Penn State/NCAR mesoscale model: validation tests and simulation of an Atlantic cyclone and cold front. Month. Weath. Rev. 121, 1493–1513.

ENVIRON, 2010. User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx). Version 5.20. Report prepared by ENVIRON International Corporation Novato, CA.

EU, 05/02/1997. Council Decision of 27 January 1997 establishing a reciprocal exchange of information and data from networks and individual stations measuring ambient air pollution within the Member States (Exchange of Information (97/101/EC)). Off. J. L 035, 14–22. See. http://ec.europa.eu/environment/air/quality/legislation/reporting.htm.

EU, 11.6.2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Off. J. L 152, 1–44. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF.

Gangoiti, G., Millán, M., Salvador, R., Mantilla, E., 2001. Long-range transport and recirculation of pollutants in the western Mediterranean during the project Regional Cycles of Air Pollution in the West-Central Mediterranean Area. Atmos. Environ. 35, 6267–6276.

Gego, E., Porter, P.S., Hogrefe, C., Irwin, J.S., 2006. An objective comparison of CMAQ and REMSAD performances. Atmos. Environ. 40 (26), 4920–4934.

Gong, S.L., 2003. A parameterization of sea-salt aerosol source function for sub- and super-micron particles. Glob. Biogeochem. Cycles 17, 1097–1104.

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

Hjellbrekke, A.G., Fjæraa, A.M., 2008. Data Report 2006. Acidifying and eutrophying compounds. EMEP/CCC-Report 1/2008, Norwegian Institute for Air Research.

Jacobson, M.Z., Lu, R., Turco, R.P., Toon, O.B., 1996. Development and application of a new air pollution model system – Part I: Gas-phase simulations. Atmos. Environ. 30, 1939–1963.

Jacobson, M.Z., 1997. Development and application of a new air pollution modeling system. Part III: Aerosol-phase simulations. Atmos. Environ. 31A, 587–608.

Kallos, G., Astitha, M., Katsafados, P., Spyrou, C., 2007. Long-range transport of anthropogenically and naturally produced particulate matter in the Mediterranean and North Atlantic: current status of knowledge. J. Appl. Meteorol. Climatol. 46, 1230–1251.

Latuatti, M., 1997. Contribution `a l'étude du bilan de l'ozone troposphérique `a l'interface de l'Europe et de l'Atlantique Nord: Modélisation lagrangienne et mesures en altitide, PhD thesis, Université de Paris 6, Paris.

Millán, M., Mantilla, E., Salvador, R., Carratalá, A., Sanz, M.J., Alonso, L., Gangoiti, G., Navazo, M., 2000. Ozone cycles in the Western Mediterranean Basin: interpretation of monitoring data in complex Coastal Terrain. J. Appl. Meteorol. 39, 487–508.

Mitsakou, C., Kallos, G., Papantoniou, N., Spyrou, C., Solomos, S., Astitha, M., Housiadas, C., 2008. Saharan dust levels in Greece and received inhalation doses. Atmos. Chem. Phys. 8, 7181–7192.

Monahan, E.C., 1986. The Ocean as a Source for Atmospheric Particles. The Role of Air–Sea Exchange in Geochemical Cycling. In: Buat-Menard, P. (Ed.). D. Reidel, pp. 129–163.

Morris, R.E., Yarwood, G., Emery, C.A., Wilson, G., Koo, B., 2006. Regional modeling using one-atmospheric models to address regional haze, 8-hour ozone and PM2.5 air quality. Presented at the 99th A&WMA Annual Conference, New Orleans, LA, (Paper # 06-A-503-AWMA), June.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Month. Weath. Rev. 116, 2417–2424. doi:10.1175/1520-0493(1988)116. <2417:SSBOTM> 2.0.CO.

Nopmongcol, U., Koo, B., Tai, R., Jung, J., Piyachaturawat, P., Emery, C., Yarwood, G., Pirovano, G., Mitsakou, C., Kallos, G. Modeling Europe with CAMx for the Air Quality Model Evaluation International Initiative (AQMEII). Atmos. Environ., in this issue.

Potempski, S., Galmarini, S., Addis, R., Astrup, P., Bader, S., Bellasio, R., Bianconi, R., Bonnardo, T.F., Buckley, R., D'Amours, R., van Dijk, A., Geertsema, G., Jones, A., Kaufmann, P., Pechinger, U., Persson, C., Polreich, E., Prodanova, M., Robertson, L., Sorensen, J., Syrakov, D., 2008. Multi-model ensemble analysis of the ETEX-2 experiment. Atmos. Environ. 42, 7250–7265. doi:10.1016/j.atmosenv.2008.07.027.

Pouliot, G., Pierce, T., Denier van der Gon, H., Schaap, M., Moran, M., Nopmongcol, U. Comparing emission inventories and model-ready emission datasets between Europe and North America for the AQMEII Project. Atmos. Environ., in this issue.

Putaud, J.P., Van Dingenen, R., Alastuey, A., Bauer, H., Birmili, W., Cyrys, J., Flentje, H., Fuzzi, S., Gehrig, R., Hansson, H.C., Harrison, R.M., Herrmann, H., Hitzenberger, R., Hüglin, C., Jones, A.M., Kasper-Giebl, A., Kiss, G., Kousa, A., Kuhlbusch, T.A.J., Löschau, G., Maenhaut, W., Molnar, A., Moreno, T., Pekkanen, J., Perrino, C., Pitz, M., Puxbaum, H., Querol, X., Rodriguez, S., Salma, I., Schwarz, J., Smolik, J., Schneider, J., Spindler, G., ten Brink, H., Tursic, J., Viana, M., Wiedensohler, A., Raes, F., 2010. A European aerosol phenomenology – 3: physical and chemical characteristics of particulate matter from 60 rural, urban, and kerbside sites across Europe. Atmos. Environ. 44, 1352–2310. doi:10.1016/j.atmosenv.2009.12.011.

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. BAMS 92 (1), 23–30.

Russell, A., Dennis, R., 2000. NARSTO critical review of photochemical models and modeling. Atmos. Environ. 34, 2283–2324.

Schere, K., Flemming, J., Vautard, R., Chemel, C., Colette, A., Hogrefe, C., Bessagnet, B., Meleux, F., Mathur, R., Roselle, E., Hu, R.M., Sokhi, R.S., Rao, S.T., Galmarini, S. Trace gas/aerosol boundary concentrations and their impacts on continental-scale AQMEII modeling domains. Atmosp. Environ., in this issue.

Schluenzen, K.H., Sokhi, R.S. (Eds.), 2008. Overview of Tools and Methods for Meteorological and Air Pollution Meso-scale Model Evaluation and User Training, ISBN 978-1-905313-59-4 Joint Report of COST Action 728 and GURME, WMO/TD-No.1457.

Seigneur, C., 2001. Current status of air quality models for particulate matter. J. Air Waste Manage. Assoc. 51 (11), 1508–1521.

Tesche, T.W., Morris, R., Tonnesen, G., McNally, D., Boylan, J., Brewer, P., 2006. CMAQ/CAMx annual 2002 performance evaluation over the eastern US. Atmos. Environ. 40, 4906–4919. doi:10.1016/j.atmosenv.2005.08.046.

Thunis, P., Georgieva, E., Galmarini, S., 2011. A procedure for air quality models benchmarking. FAIRMODE document, URL: http://fairmode.ew.eea.europa.eu/models-benchmarking-sg4.

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

Vautard, R., Bessagnet, B., Chin, M., Menut, L., 2005. On the contribution of natural aeolian sources to particulate matter concentrations in Europe: testing hypotheses with a modelling approach. Atmos. Environ 39, 3291–3303.

Vautard, R., Builties, P.H.J., Thunis, P., Cuvelier, C., Bedogni, M., Bessagnet, B., Honore, C., Moussiopoulos, N., Pirovano, G., Schaap, M., Stern, R., Tarrason, L., Wind, P., 2007. Evaluation and intercomparison of ozone and PM10 simulations by several chemistry transport models over four European cities within the CityDelta project. Atmos. Environ. 41, 173–188.

Vautard, R., Schaap, M., Bergstrom, R., Bessagnet, B., Brandt, J., Builtjes, P.J.H., Christensen, J.H., Cuvelier, C., Foltescu, V., Graff, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouil, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., Wind, P., 2009. Skill and uncertainty of a regional air quality model ensemble. Atmos. Environ. 43, 4822–4832. doi:10.1016/j.atmosenv.2008.09.083.

Weil, J.C., Sykes, R.I., Venkatram, A., 1992. Evaluating air quality models: review and outlook. J. Appl. Meteorol. 31, 1121–1145.

Wilks, D.S., 2006. Statistical methods in the atmospheric sciences. In: International Geophysics Series, second ed., vol. 59. Academic Press, 627 pp.

Yarwood, G., Rao, S., Yocke, M., Whitten, G., 2005. Updates to the Carbon Bond Chemical mechanism: CB05, report, Rpt. RT-0400675, US EPA, Res. Tri. Park.

Please cite this article in press as: Pirovano, G., et al., Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment (2012), doi:10.1016/j.atmosenv.2011.12.052