Presentation of the EURODELTA III inter-comparison exercise - Evaluation of the chemistry transport models performance on criteria pollutants and joint analysis with meteorology

B. Bessagnet[1], G. Pirovano[2], M. Mircea[3], C. Cuvelier[4], A. Aulinger[5], G. Calori[6], G. Ciarelli[7], A. Manders[8], R. Stern[9], S. Tsyro[10], M. García Vivanco[11], P. Thunis[12], M.-T. Pay[13], A. Colette[1], F. Couvidat[1], F. Meleux[1], L. Rouïl[1], A. Ung[1], S. Aksoyoglu[7], J. M. Baldasano[13], J. Bieser[5], G. Briganti[3], A. Cappelletti[3], M. D'Isidoro[3], S. Finardi[6], R. Kranenburg[8], C. Silibello[6], C. Carnevale[14], W. Aas[15], J.-C. Dupont[16], H. Fagerli[10], L. Gonzalez[17], L. Menut[18], A. S. H. Prévôt[7], P. Roberts[17], L. White[19]

[1]INERIS, National Institute for Industrial Environment and Risks, Parc Technologique ALATA, F-60550 Verneuil-en-Halatte, France
[2]RSE S.p.A., via Rubattino 54, 20134 Milano, Italy
[3]ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Via Martiri di Monte Sole 4, 40129 Bologna, Italy
[4]Ex European Commission, Joint Research Centre JRC Institute for Environment and Sustainability I-21020 Ispra (Va), Italy
[5]HZG, Helmholtz-Zentrum Geesthacht, Institute for Coastal Research, Max-Planck-Straße 1, 21502 Geesthacht, Germany
[6]ARIANET Srl, Via Gilino n.9 20128, Milano, Italy
[7]PSI, LAC, Paul Scherrer Institute, 5232 Villigen PSI, Switzerland
[8]TNO, Dept. Climate, Air and Sustainability, P.O. Box 80015, 3508 TA Utrecht, The Netherlands
[9]Freie Universität Berlin, Institut für Meteorologie Troposphärische Umweltforschung Carl-Heinrich-Becker Weg 6-10, D-12165 Berlin, Germany
[10]Climate Modelling and Air Pollution Division, Research and Development Department, Norwegian Meteorological Institute (MET Norway) P.O. Box 43, Blindern, N-0313 Oslo, Norway
[11]CIEMAT, Atmospheric Pollution Unit, Avda. Complutense, 22, 28040 Madrid, Spain
[12]European Commission, Joint Research Centre JRC Institute for Environment and Sustainability I-21020 Ispra (Va), Italy
[13]BSC, Barcelona Supercomputing Center, Centro Nacional de Supercomputación, Nexus II Building, Jordi Girona, 29, 08034 Barcelona, Spain
[14]Department of Electronics for the Automation, University of Brescia, via Branze 38, I 25123 Brescia, Italy
[15]Norwegian Institute for Air Research (NILU), Box 100, 2027 Kjeller, Norway
[16]Institut Pierre-Simon Laplace, CNRS-Ecole Polytechnique, 91128 Palaiseau, Paris, France
[17]CONCAWE, Boulevard du Souverain 165, B-1160 Brussels, Belgium
[18]Laboratoire de Météorologie Dynamique, École Polytechnique, ENS, UPMC, CNRS, Institut Pierre-Simon Laplace, 91128 Palaiseau, France
[19]AERIS EUROPE Ltd., Strouds Church Lane West Sussex RH17 7AY, United Kingdom

Correspondence to: B. Bessagnet (bertrand.bessagnet@ineris.fr)

Abstract.

The EURODELTA III exercise has facilitated a comprehensive inter-comparison and evaluation of chemistry transport model performances. Participating models performed calculations for four one-month periods in different seasons in the years 2006 to 2009, allowing the influence of different meteorological conditions on model performances to be evaluated. The exercise was performed with strict requirements for the input data, with few exceptions. As a consequence, most of differences in the outputs will be attributed to the differences in model formulations of chemical and physical processes. The

models were evaluated mainly for background rural stations in Europe. The performance was assessed in terms of bias, root mean square error and correlation with respect to the concentrations of air pollutants (NO_2, O_3, SO_2, PM10 and PM2.5), as well as key meteorological variables. Though most of meteorological parameters were prescribed, some variables like the planetary boundary layer (PBL) height and the vertical diffusion coefficient were derived in the model pre-processors and can partly explain the spread in model results. In general, the daytime PBL height is underestimated by all models. The largest variability of predicted PBL is observed over the ocean and seas. For ozone, this study shows the importance of proper boundary conditions for accurate model calculations and then on the regime of the gas and particle chemistry. The models show similar and quite good performance for nitrogen dioxide, whereas they struggle to accurately reproduce measured sulphur dioxide concentrations (for which the agreement with observations is the poorest). In general, the models provide a close-to-observations map of particulate matter (PM2.5 and PM10) concentrations over Europe with rather correlations in the range 0.4 – 0.7 and a systematic underestimation reaching -10 µg m^{-3} for PM10. The highest concentrations are much more underestimated particularly in wintertime. Further evaluation of the mean diurnal cycles of PM reveals a general model tendency to overestimate the effect of the PBL height rise on PM levels in the morning, while the intensity of afternoon chemistry leading to formation of secondary species to be underestimated. This results in larger modelled PM diurnal variations than the observations and this is so for all seasons. The models tend to be too sensitive to the daily variation of the PBL. All in all, in most cases model performances are more influenced by the model set-up than the season. The good representation of temporal evolution of wind speed is most responsible for models' skillfulness in reproducing the daily variability of pollutant concentrations (e.g. the development of peak episodes), while the reconstruction of the PBL diurnal cycle seems to play a larger role in driving the corresponding pollutant diurnal cycle and hence determine the presence of systematic positive and negative biases detectable on daily basis.

1 Introduction

The ongoing project EURODELTA has very successfully extended the European Air Quality Modelling capability by providing a forum in which modelling teams could share experiences in simulating technically interesting and policy relevant problems. The joint exercises contribute to further improve modelling techniques as well as to quantify and understand the sources of model uncertainties related to the parameterization of processes and the quality of input data. EURODELTA is now an activity contributing to the scientific work of the UNECE (United Nations Economic Commission for Europe) Task Force on Measurement and Modelling (TFMM) under the Convention on Long-range Transboundary Air Pollution (CLRTAP). The TFMM was established in 2000 to provide a forum to the Parties, the EMEP (European Monitoring and Evaluation Programme) centres and other international organizations for scientific discussions to evaluate measurements and modelling and to further develop working methods and tools. These are used for policy studies in support of the Gothenburg Protocol signed in 1999 which is a multi-pollutant protocol of the Convention designed to reduce

acidification, eutrophication and ground-level ozone by setting emissions ceilings for sulphur dioxide, nitrogen oxides, volatile organic compounds, fine particulate matter and ammonia.

In 2004, EURODELTA I (van Loon *et al*., 2007) examined the common performance of the chemistry transport models (CTM) in predicting recent (2000) and future (2020) air quality in Europe using the concept of a model ensemble to measure robustness of predictions. The spread of model predictions about the *ensemble* mean gave a measure of uncertainty for each predicted value. In a 2020 world the effect of making emission reductions for key pollutants in specific geographic areas was investigated. The pollutants were NOx (nitrogen dioxide), SO_2 (sulphur dioxide), VOC (Volatile Organic Compound), PM (particulate matter as PM10 and PM2.5 for particle diameters below 10 µm and 2.5 µm respectively) and NH_3 (ammonia). The countries were France, Germany and Italy. The effect of reducing NOx and SOx in sea areas was also investigated. Source-receptor relationships used in integrated assessment (IA) modelling were derived for all the models and compared to assess how model choice might affect this key input. EURODELTA II (Thunis *et al.*, 2008) built on this project by taking a closer look at how the different models represent the effect on pollutant impacts on a European scale by applying emission reductions to individual emission sectors.

In the recent literature, several inter-comparison and evaluation exercises of regional-scale chemistry transport models for PM have been reported: McKeen *et al.* (2007), van Loon *et al.*, (2007), Vautard *et al.* (2007), Hayami *et al.* (2008), Stern *et al.* (2008), Smyth *et al.* (2009), Vautard *et al.* (2009), Solazzo *et al.* (2012), Pernigotti, *et al.* (2013) and Prank *et al.* (2016).. In one of the most recent exercises, AQMEII (Solazzo *et al.*, 2012), models clearly tend to underestimate PM10 background concentrations in US and EU regions. Model results for PM2.5 concentrations showed better performances but large uncertainty remained certainly due to the simulation of secondary organic aerosols. Prank *et al*. (2016) stressed the problems of emission underestimates to explain the model discrepancies.

The new EURODELTA III (ED-III) exercise was designed to exploit and interpret intensive measurement campaigns carried out by EMEP (Aas *et al*., 2012). As far as possible the models have been run in ED-III with the same input data (emissions, meteorology, boundary conditions) and over the same domain (domain extension and resolution). This distinguishes the study from other model inter-comparisons. ED-III focussed on four EMEP intensive measurement periods:

- 1 Jun - 30 Jun 2006
- 8 Jan - 4 Feb 2007
- 17 Sep - 15 Oct 2008
- 25 Feb - 26 Mar 2009

The four different periods, within a rather limited number of years, allowed the influence of different meteorological conditions on model performance to be evaluated.. The list of modelling teams participating in the ED-III is reported in Table 1. FUB ran two of the four periods The ED-III framework (emissions, model configurations) was also used to assess the impact of the horizontal resolution on the performance of air quality models (Schaap *et al.*, 2015).

The ED-III exercise allowed a very comprehensive inter-comparison and evaluation of chemistry transport model performance with a joint analysis of some meteorological variables to be made. A first evaluation on the 2009 campaign with an interim version of models was published in Bessagnet *et al.* (2014). Moreover, the selected periods coincide with EMEP

intensive measurement periods so that an extended set of observational data were available. Therefore, in addition to EMEP operational monitoring data, size disaggregated (in PM2.5 and PM10) aerosol data and hourly measurements for studying diurnal cycles have been used. Additional AirBase data (Mol and de Leeuw, 2005) were used to evaluate the impact of meteorology on air pollutant concentrations. Finally, the exercise was performed under strict requirements (with some exceptions) concerning the input data. As a consequence, most of differences in the outputs will be attributed to the simulation of chemical and physical processes. The objective of this paper is twofold, (i) to present the exercise, the input data and the participating models, and (ii) to analyse the behaviour of models in the four campaigns focussing on the criteria pollutants PM10, PM2.5, O_3, NO_2 and SO_2 as defined in the EU directive on air quality 2008/50/EC (EC, 2008), and relevant meteorological variables. Complementary analyses of deposition fluxes and PM composition data at high temporal resolution will be discussed in companion papers in order to better understand the behaviour of models.

2 Description of models

2.1 Overall description of models

The models are synthetically described in Table 2 and Table 3. All the models were run on the same domain at 0.25°x0.25° resolution in longitude and latitude except CMAQ. CMAQ simulations were performed on a Lambert-conformal conic projection with the standard parallels at 30 and 60 degrees and a grid of 112 by 106 cells of size 24km x 24km. The results of the CMAQ simulations were interpolated to the prescribed EURODELTA grid.

Participants delivered both air concentrations and meteorological parameters. Most of variables were delivered on an hourly basis, while dry and wet deposition fluxes were provided on a daily basis. The output species include, among others: O_3, NO_2 and SO_2, total PM mass concentrations both in 2.5 and 10 µm fractions (PM10 and PM2.5). Secondary inorganic aerosols such as ammonium (NH_4^+), sulphate (SO_4^{2-}) and nitrate (NO_3^-) and other PM components relevant for the analysis as well as wet deposition of sulphur and nitrogen compounds were also collected and will be used in companion papers. The delivered air concentrations should approximately correspond to the standard measurement height (typically 3 m) and were directly derived from the first model layer, except for LOTOS-EUROS and EMEP that corrected the concentrations from the first layer to be representative of 3-m concentrations. The PM2.5 and PM10 concentrations are calculated as follows in each model:

$$PM_{xx} = PPM_{xx} + |SO_4^{2-}|_{xx} + |NO_3^-|_{xx} + |NH_4^+|_{xx} + |SOA|_{xx} + |Dust|_{xx} + |Sea\ Salts|_{xx}$$

where *xx*=2.5 µm or 10 µm, PPM stands for Primary Particulate Matter and includes Elemental carbon, Primary organic aerosol and primary non-carbonaceous aerosol, SOA represents Secondary Organic Aerosol, *Sea Salt* and *Dust* represent the contribution of the corresponding natural processes mainly controlled by the wind speed.

The participating models differ in the availability of PM components and formation routes. For instance, EMEP, LOTOS-EUROS and RCG contain coarse mode nitrate formation (produced by reaction of nitric acid with sea salts and dust),

whereas the others do not. In CMAQ additional anthropogenic dust is calculated as 90% of unspecified PM coarse emissions and attributed to fugitive dust (Binkowsky and Roselle, 2003). CAMx did not activate the parameterisation of sea salts in this exercise.

Based on the set-up of models and completeness of datasets, an "ENSEMBLE" called **ENS** has been built based on mean values of model outputs. To compare the behaviour of models for all pollutants and campaigns, only CHIMERE, MINNI, LOTOS-EUROS and EMEP constitute the "ENSEMBLE". CAMx, CMAQ and RCG were not included in the ensemble for three reasons: (i) CAMx did not account for sea salts leading to very different PM patterns over the oceans and seas, (ii) CMAQ used a different meteorology and (iii) RCG did not cover the four campaigns.

2.2 PBL height and mixing in models

CAMx

In ED-III the Planetary Boundary Layer was directly taken from the IFS-ECMWF data (Integrated Forecast System of the European Centre for Medium-Range Weather Forecasts). The PBL height was then used by CAMx pre-processor to derive K_z profiles. For ED-III the O'Brien scheme (1970) has been used to derive K_z profiles as Eq.1:

$$K_z = K_A + \frac{(z-z_A)^2}{(z_A-z_B)^2}\left\{K_B - K_A + (z-z_B)\left(K'_B + 2\frac{K_B-K_A}{z_A-z_B}\right)\right\} \quad \text{(Eq. 1)}$$

Where K_z is a value of K_A at the height of the atmospheric boundary layer, z_A, and K_B at the height of the surface layer z_B, the so-called constant-flux layer. Minimum K_z values have been set to 1 $m^2\ s^{-1}$. Any values of K_z calculated below, will be set to this value. By default, CAMx employs a standard "K-theory" approach for vertical diffusion to account for sub-grid scale mixing layer-to-layer.

CHIMERE

In this study, the Planetary Boundary Layer is directly taken from the IFS ECMWF data. Horizontal turbulent fluxes were not considered. Vertical turbulent mixing takes place only in the boundary layer. The formulation uses K-diffusion following the parameterization of (Troen and Mahrt, 1986), without a counter-gradient term. In each model column, diffusivity K_z is calculated as Eq. 2:

$$K_z = \mathrm{k}\,\mathrm{w_s}\,\mathrm{z}\left(1-\frac{\mathrm{z}}{\mathrm{h}}\right)^{1/3} \quad \text{(Eq. 2)}$$

where w_s is a vertical velocity scale given by similarity formulae.

- In the stable case (surface sensible heat flux < 0): $w_s = u_*/(1 + 4.7\,z/L)$
- In the unstable case: $w_s = ({u_*}^3 + 2.8e{w_*}^3)^{1/3}$

where e = *max(0.1,z/h)*, L is the Monin-Obukhov Length, w_* is the convective velocity scale, u_* the friction velocity and h the boundary layer height. The minimum value of K_z is assumed to be 0.01 $m^2\ s^{-1}$.

K_z and the wind speed were corrected in urban zones according Terrenoire *et al.* (2015) by applying a correction factor to limit the diffusion within the urban canopy, but this correction has very little effect at this resolution.

CMAQ

The boundary layer height in COSMO is calculated with the turbulent kinetic energy (TKE) method (Doms *et al.* 2011). CMAQ directly used the PBL fields from COSMO.

In CMAQ the vertical turbulent mixing is estimated using the Asymmetric Convective Model scheme version 2 (ACM2, Pleim, 2007a,b). The ACM2 replaces the simple eddy viscosity (K-theory) scheme. ACM2 scheme allows the non-local mixing, which means upward turbulent mixing from the surface across non-adjacent layers through the convective boundary layer. Pleim (2006) compared the eddy viscosity and the ACM2 schemes in CMAQ, finding that the ACM2 schemes tends to predict larger concentrations of secondary pollutants and smaller concentrations of primary pollutants at the surface, and has a more well-mixed profile in the PBL than the eddy viscosity scheme.

CMAQv5 has also an improved version of the minimum allowable vertical eddy diffusivity scheme. The new version interpolates between urban and nonurban land cover, allowing a larger minimum vertical diffusivity value for grid cells that are primarily urban. Moreover, the minimum eddy diffusivity values were reduced from 0.5 m^2 s^{-1} to 0.01 m^2 s^{-1}, and from 2.0 m^2 s^{-1} to 1.0 m^2 s^{-1} for urban areas.

EMEP

The mixing height is calculated using a slightly modified Richardson number (Ri_B) following Jeričevič *et al.* (2010) and defined as the lowest height at which the Ri_B>0.25. Finally, the PBL is smoothed with a second order Shapiro filter in space. The PBL height is not allowed to be less than 100 m or exceed 3000 m.

The initial calculation of the vertical exchange coefficients is done using the Ri number and wind speed vertical gradient for the whole domain. Then, K_z values within the PBL are recalculated based on Jeričevič *et al.* (2010) for stable and neutral conditions. For unstable situations K_z is calculated based on the similarity theory of Monin-Obukhov for the surface layer, whereas K_z profiles from O'Brian (1970) are used for the PBL above the surface layer. For more detail see Simpson *et al.* (2012).

LOTOS-EUROS

The first model layer is by definition the mixing layer, with height equal to the boundary layer height as given by ECMWF. Horizontal diffusion is not used, but for vertical mixing the vertical diffusion coefficient is calculated according to Eq. 3:

$$K_z = \frac{\kappa u^*}{\Phi(z/L)} \qquad \text{(Eq. 3)}$$

where κ is the von Karman constant, u^* the friction velocity, Φ the functions proposed by Businger (1971) for stable, neutral or unstable atmosphere, z the height and L the Monin-Obhukov length. The friction velocity is calculated depending on the

wind at reference height (10 m), the Businger functions and the roughness length per land use class. The vertical structure of LOTOS-EUROS is determined by the mixing layer height, with a shallow surface layer (25 m) to avoid too fast mixing of near-surface emissions and a second layer equal to the mixing layer as given by ECMWF.

MINNI

In MINNI, the friction velocity and Monin-Obukhov length are determined by using the Holtslag and van Ulden (1983) iterative scheme for unstable conditions and the Venkatram (1980) iterative method for stable conditions. Micro-meteorological parameters over water are derived with the profile method, using air-sea temperature difference (Hanna *et al.*, 1985), with the needed roughness length, depending on wind speed, supplied by the Hosker (1974) parameterization.

During daytime both convective and mechanical heights are determined, keeping then the maximum value between the two parameters. The convective height is calculated following the Maul (1980) version of Carson (1973) algorithm, essentially based on heat conservation equation. The mechanical mixing height is estimated by using the Venkatram (1980) algorithm. During nighttimes, the Bulk Richardson number method is applied (Sorensen, 1998), in which the height of the boundary layer is given by the smallest height at which the bulk Richardson number reaches the critical value fixed to 0.25.

RCG

The mixing layer depth in the model is the height of the layer closest to the input boundary layer height taken from the IFS ECMWF data. Vertical diffusion parameters for stable and unstable conditions are derived using the Monin-Obukhov similarity theory for the diabatic surface layer. The friction velocity and Monin-Obukhov-length are calculated iteratively depending on the 10 m wind, the stability correction factors and the roughness length determined from land use.

3 Input data

3.1 Anthropogenic emissions

The first step in the emission preparation was to calculate the spatial pattern of emissions for the reference year 2007, that was selected because it was a key year for the TNO-MACC inventory (Kuenen *et al.*, 2011). The anthropogenic emission input was harmonized following the methodology described in Terrenoire *et al.* (2015). The total emissions per sector and country were then scaled to the corresponding year of the campaigns: 2006, 2007, 2008 and 2009. Emission categories are broken down into 11 classes called SNAP (Selected Nomenclature for Air Pollutants): (1) Public Power stations, (2) Residential and Comm./inst. Combustion, (3) Industrial combustion, (4) Production processes, (5) Extraction and distribution fossil fuel, (6) Solvents use, (7) Road traffic, (8) Other mobile sources (trains, shipping, aircraft, ...), (9) Waste treatment, and (10) Agriculture. Natural emissions (11) were calculated by the models as set out in section 3.2.

The gridded distribution of anthropogenic emissions was provided by INERIS and it was based on a merging of different databases from:

- TNO 0.125°×0.0625° emissions for 2007 from MACC (Kuenen *et al.*, 2011)
- EMEP 0.5°×0.5° emission inventory for 2009 (Vestreng *et al.*, 2007)
- Emission data from the GAINS database (http://gains.iiasa.ac.at/gains).

Emission re-gridding was based on INERIS expertise and performed by means of various proxies:

- population data coming from the EEA database merged with global data (from the Socioeconomic Data and Applications Center http://sedac.ciesin.columbia.edu) to fill gaps in Europe.;
- the US Geophysical Survey land use at 1 km resolution (http://www.usgs.gov/).
- French bottom-up emission data for wood combustion to derive a proxy based on population density;
- EPER data for industries; the EPER Decision is based on Article 15(3) of Council Directive 96/61/EC (EC, 1996) concerning integrated pollution prevention and control. EPER is a web-based register, which enables the public to view data on emissions to water and air of 50 key pollutants from large and medium-sized industrial point sources in the European Union.

The TNO-MACC dataset provides two distinct datasets (i) large point sources (LPS) with the coordinates of stacks and (ii) surface emissions on a fine grid (0.125°×0.0625°). In the gridding process, the first step consisted in summing up LPS emissions from the TNO-MACC emissions inventory for 2007 with surface emissions to obtain total emissions as in the EMEP inventory. LPS were aggregated with surface emissions because no data were available to calculate plume rise heights for point sources emissions. For the various SNAP sectors the processing steps were the following:

- SNAP 2: The country emissions were re-gridded with coefficients based on population density and French bottom-up data, the methodology (Terrenoire *et al*., 2015) was extrapolated to the whole Europe. For PM2.5 emissions, the annual EMEP national totals were kept except for the countries: Czech Republic, Bosnia and Herzegovina, Belgium, Belarus, Spain, France, Croatia, Ireland, Lithuania, Luxemburg, Moldavia, Republic of Macedonia, Netherland, Turkey. For these countries, PM2.5 emissions from GAINS were used as this database provides higher numbers and certainly more realistic since wood burning is known to be underestimated in the EMEP database (Denier van der Gon *et al*., 2015). Additional factors were applied on two Polish regions for both PM2.5 and PM10 emissions. As a preliminary solution, domestic combustion emissions from provinces with active coal mines were multiplied by a factor of 8, while those in neighbouring provinces were adjusted by a factor of 4 (Kiesewetter *et al*., 2015). The former activity in coal mine regions still leads to high emissions of PM due to domestic uses of coal.
- SNAP 3,7,8,9,10: TNO-MACC emission spatial distribution was used as proxy to regrid EMEP 0.5°x0.5° annual totals into the finer modelling grid.
- SNAP 1,4,5,6: EMEP 0.5°x0.5° emissions were regridded by using "artificial area" (or built-up area), except for industries where EPER data were used.

For countries where TNO-MACC emissions were not available, the EMEP 0.5°×0.5° emissions were used (Iceland, Liechtenstein, Malta and Asian countries) and regridded with adequate proxies ("artificial land use", EPER data for industries).

The following emitted species were used in the models: methane (this species comes from the TNO-MACC inventory), carbon monoxide, ammonia, sulphur oxides, non-methane volatile organic compounds (NMVOC), nitrogen oxides, primary particulate matter.

Residential emissions of particulate matter are dominant in wintertime. In most countries, they come from wood burning or coal uses. Germany, Sweden, Spain clearly have the lowest levels of PM2.5 emissions for this activity sector. Romania, Poland and France have the highest levels of annual total emissions per country (Terrenoire *et al.*, 2015). For this activity sector, the PM2.5 emissions by components are provided in supplementary material S8.

The time profiles are those used in Thunis *et al.* (2008). Three types of profiles were provided:

- Seasonal factors : one value per species, month, activity sector and country
- Weekly factors : one value per species, day type (Monday – Sunday), activity sector and country
- Hourly factors : one value per hour (local time), species and activity sector

The vertical injection profile in CTMs was prescribed according to Bieser *et al.* (2011) where industrial sectors and residential heating were assigned in lower levels compared to the lower vertical levels than other literature default profiles (Mailler *et al*., 2013).

Since only PM2.5 and coarse PM emissions were provided by EMEP, a PM speciation profile provided by IIASA (based on Klimont *et al*., 2013) was used to estimate the fraction of non-carbonaceous species, Elemental Carbon and Organic Matter per activity sectors and country. Models used their own split for NOx, SOx and NMVOC emissions. This emission inventories did not account to recent changes in the way to account for Semi Volatile Organic Compounds from wood burning emissions as discussed in Denier van der Gon *et al.* (2015).

The full emission dataset is available on request to INERIS.

3.2 Natural emissions

Biogenic VOC emissions from vegetation

CHIMERE and MINNI used the version 2.04 of the MEGAN model while CAMx used the 2.1 version (Guenther *et al.*, 2006, 2012). The Model of Emissions of Gases and Aerosols from Nature (MEGAN) is a modelling framework for estimating fluxes of biogenic compounds between terrestrial ecosystems and the atmosphere using simple mechanistic algorithms to account for the major known processes controlling biogenic emissions. It is available as an offline code and has also been coupled into land surface and atmospheric chemistry models.

EMEP, LOTOS-EUROS and RCG used parameterizations derived from Simpson *et al.* (1999) for the temporal variations according to temperature and light, with maps of tree species from Koeble and Seufert (2001).

CMAQ used the BEIS (Biogenic Emission Inventory System: Vukovich and Pierce, 2002) module developed by the US EPA. BEIS estimates volatile organic compound (VOC) emissions from vegetation and nitric oxide (NO) and carbon monoxide (CO) emissions from soils. Because of resource limitations, recent BEIS development has been incorporated into

the Sparse Matrix Operational Kernel Emissions (SMOKE) system (available at https://www.cmascenter.org/smoke), so that the native version of BEIS is built within the SMOKE architecture.

Soil Nitrogen Monoxide (NO) emissions

CHIMERE and MINNI used the version 2.04 and CAMX used version 2.1 of the MEGAN model to calculate the NO emissions. RCG used a parameterization of NO emissions described in Simpson *et al.* (1999). LOTOS-EUROS did not include NO emissions in this simulation. CMAQ used the BEIS (Biogenic Emission Inventory System) module developed by the US EPA. The soil NO emission parameterization for EMEP is described in Simpson *et al.* (2012)

Sea salt emissions

All models host very different schemes based on Monahan (1986) with some updates from Martensson *et al.* (2003) for LOTOS-EUROS, and Gong *et al.* (1997) for RCG. CMAQ and MINNI used the Zhang *et al.* (2005) parameterization and CAMx had no sea salts for this exercise due to too high uncertainty in sea salt parameterization. EMEP used parameterisation from Monahan (1986) for larger sizes of sea spray and Martensson *et al.* (2003) for smaller sizes.

CMAQ emits also sea salts sulphate using a fraction of 7.76% of emitted sea salts split into the accumulation and coarse modes.

NO emissions from lightning

The only model to describe NO emissions from lightning is the EMEP model, following Köhler *et al.* (1995).

Wildfire emissions

Fire emissions were provided by the GFASv1.0 database (Kaiser *et al.*, 2012) only for the 2006 campaign. The Global Fire Assimilation System (GFASv1.0) calculates biomass burning emissions by assimilating Fire Radiative Power (FRP) observations from the MODIS instruments onboard the Terra and Aqua satellites. It corrects for gaps in the observations, which are mostly due to cloud cover, and filters spurious FRP observations of volcanoes, gas flares and other industrial activities. For all models the wildfire emissions were assigned in the whole PBL layer. Only the following species were taken into account: CO, CH_4, NOx, SO_2, PM2.5, TPM (Total Primary Matter), OC (Organic Carbon) and EC (Elemental Carbon).

Dust emissions

For CAMx, CHIMERE and CMAQ, no natural dust module is activated for this exercise. For these three models, natural dust only comes from the boundary conditions. For EMEP, windblown dust parameterisation is documented in Simpson *et al.* (2012), road dust calculations are included in the calculations from Denier van der Gon *et al.* (2009). LOTOS-EUROS contains emission parameterizations for several sources of mineral dust (Schaap *et al* 2009). Only wind-blown dust, resulting

from wind erosion of bare soils, was taken into account here, together with dust from boundary conditions. Other sources (agricultural activities, road dust resuspension) were not activated in ED-III. In MINNI, dust emissions from local erosion and particle resuspension (Vautard *et al.*, 2005) with attenuation in the presence of vegetation from Zender *et al.* (2003) is activated in this exercise. RCG considers resuspension of mineral aerosol as a function of friction velocity and the nature of soils. Two mechanisms are treated: direct release of small dust particles by the wind (Loosmore and Hunt, 2000), and indirect release by collision with bigger soil grains that are lifted by the wind but return to the surface by sedimentation (saltation process from Claiborn *et al.*, 1998).

3.3 Meteorology

All models except CMAQ and RCG share the same meteorological dataset at 0.2° resolution based on ECMWF IFS (Integrated Forecast System) calculations.

Because of its importance for applications (*e.g.* in air pollution modelling), the boundary layer height, as diagnosed in the IFS-ECMWF model, was made available. The parameterization of the mixed layer (and entrainment) uses a boundary layer height from an entraining parcel model. But, in order to get a continuous field, in neutral and stable situations the bulk Richardson method proposed by Troen and Mahrt (1986) is used as a diagnostic, independently of the turbulence parameterization. Boundary layer height is defined as the level where the bulk Richardson number, based on the difference between quantities of energy at that level and the lowest model level, reaches the critical value $Ri_{cr} = 0.25$.

For RCG, a different meteorological data set was used. The 3D-data for wind, temperature, humidity and density were produced employing a diagnostic meteorological analysis system developed at Freie Universität (Berlin, Germany) and based on an optimum interpolation procedure on isentropic surfaces. The system takes into account all available observed synoptic surface and upper air data as well as topographical and land use information (Reimer and Scherer, 1992). Rain data, cloud data and boundary layer heights were retrieved from the IFS data set. Boundary layer parameters as friction velocity and Monin-Obukhov length were calculated on-the-fly by applying standard boundary layer theory.

The CMAQ model used meteorological variables calculated with the COSMO model in CLimate Mode (COSMO-CLM) version 4.8 clm 11. The COSMO model is the non-hydrostatic operational weather prediction model applied and further developed by the national weather services joined in the COnsortium for SMall scale MOdeling (COSMO) described in Bettems *et al.* (2015).

3.4 Boundary conditions

In this study, the MACC reanalysis was used as input data for the boundary conditions (Inness *et al.*, 2013; Benedetti *et al.*, 2009). The MACC II project (Modelling Atmospheric Composition and Climate) established the core global and regional atmospheric environmental service delivered as a component of the COPERNICUS initiative (http://copernicus.eu/). The reanalysis production stream provides analyses and 1-day forecasts of global fields of O_3, CO, NO_2, SO_2, HCHO, CO_2, CH_4, and aerosols. Other reactive gases are available from the coupled chemistry transport model. The reanalysis covers the period 2003 – 2011 with a one-month spin-up. It runs at approximately 78 km by 78 km horizontal resolution over 60 levels. The coupled chemistry transport model has the same 60 vertical levels and a horizontal resolution of 1.125° x 1.125°. For aerosols only elemental carbon, organic carbon, dust and sulphate were used.

Stratospheric ozone fields from the MACC reanalysis agree with ozone sondes and ACE-FTS (Atmospheric Chemistry Experiment Fourier Transform Spectrometer) data within ±10% in most seasons and regions. In the troposphere, the reanalysis shows biases of −5% to +10% with respect to ozone sondes and aircraft data in the extratropics, while larger negative biases are shown in the tropics. Area-averaged total column ozone agrees with ozone fields from a multi-sensor reanalysis data set within a few percent. For aerosols, the observed Aerosol Optical Depth (AOD) is assimilated in the model with a feedback on individual PM species (sea salts, dust, elemental carbon, organic carbon and sulphate). When available, the MACC reanalysis is compared with observations, the model acronym in the supporting material is MACCA.

4 Observation dataset and statistics

4.1 Air pollutant concentrations

The evaluation was carried out with the available EMEP standard monitoring (Tørseth *et al.*, 2012) and intensive period observations for 2006, 2007, 2008 and 2009 (Aas *et al.*, 2012) on hourly and daily bases (see supplementary material S8 for the description of background sites). Elevated sites above 1500 m in altitude have been excluded from the analysis. The measurements were downloaded from the EBAS database (http://ebas.nilu.no/). Additional AirBase data (Mol and de Leeuw, 2005) were used to evaluate the impact of meteorology on air pollutant concentrations in section 7.2.

It is important to note that daily measurements for a day, N, is the averaged value between day N HH:00 and day N+1 HH:00, with HH usually varying in the range [00, 09] in GMT. For most of the species, measurements on daily and hourly bases are not necessarily performed for the same set of stations. Deposition and the PM composition are also available; the dataset will be detailed in the companion papers.

4.2 Meteorology

Temperature and wind speed

The temperature, wind speed and precipitation measurements come from 2016 synoptic stations in Europe reported by the European meteorological centres. The data are provided on an hourly basis. The temperature is measured at 2 m and the wind speed at 10 m. Some meteorological data are also reported at some EMEP sites. At EMEP sites, daily accumulated measurements (*e.g.* precipitation) for a day N represent the integral between day N HH:00 to day N+1 HH:00, with HH usually varying in the range [00, 09] in GMT.

Planetary Boundary Layer (PBL) height

The soundings data were extracted from the University of Wyoming database (http://weather.uwyo.edu/). For each site and for each day, two soundings are available at 00:00 and 12:00 GMT. The provided meteorological parameters are: pressure (hPa), the corresponding height above ground level (m), dew point temperature (°C), relative humidity (%), mixing ratio (g kg^{-1}), wind direction (degrees) and wind speed (expressed in knot and converted to m s^{-1} by applying the conversion factor 0.514), potential and virtual potential temperature (K). For the present study, data were extracted over 77 stations in Europe. The boundary layer height is estimated using the calculation of the Bulk Richardson number profile and searching for the altitude where the critical value of $Ri_{cr}=0.25$ is reached. The analysis was limited to the first 25 vertical points, roughly corresponding to an altitude of 5000m above ground level. Since the boundary layer height is a concept valid only for convective periods, only the soundings of 12:00 GMT were analyzed and used for the model evaluation.

In addition to the previous PBL data, hourly heights of the atmospheric boundary layer were calculated from LIDAR measurements in a background site near Paris (SIRTA in Palaiseau, France). A new objective method for the determination of the atmospheric boundary layer depths using routine LIDAR measurements has been used (Pal *et al.*, 2013).

4.3 Error statistics for the evaluation of model performances

The errors statistics considered in this report are presented in Table 4. In supplementary material S0-S1 the performances of all models for the four campaigns are reported. For a given pollutant or meteorological variable, model performance is computed for a common set of stations (over the same common geographic area). All maps of pollutant concentrations and meteorological variables concerning individual models and ensemble are provided in supplementary material (S2-S6).

For the analysis of the "ensemble" a coefficient of variation *VAR* is defined as follows in Eq. 4:

$$VAR = \frac{1}{C_{ENS}} \sqrt{\frac{1}{M} \sum_m (C_m - C_{ENS})^2} \qquad \text{(Eq. 4)}$$

With C_m the concentration of individual model *m* included in the *ensemble* (CHIMERE, LOTOS-EUROS, MINNI and EMEP), M is the number of models, and C_{ENS} is the *ensemble* mean concentration.

5 Evaluation of the meteorology

Some general features for each campaign can be provided, they are taken from the NOAA (National Oceanic and Atmospheric Administration) global analysis (https://www.ncdc.noaa.gov/sotc/global/).

June 2006 temperatures were above average everywhere in Europe with low precipitation except in Balkan countries and Spain compared to the 1961-1990 base period.

January 2007 was characterized by windy conditions in Europe with temperatures above the average everywhere except in Spain where temperatures were close to the average values. In the beginning of February temperatures were particularly low in Scandinavia. Precipitations were low over the Mediterranean basin but above the climate average, compared to the 1961-1990 period in the rest of Europe.

In September- October 2008, no clear general characteristics were recorded; this transition period was characterized by slight negative temperature anomalies over the western part of Europe, mainly France, United Kingdom and north of Spain.

After some cold spells in the end of February, March 2009 turned milder with on average warmer temperatures compare to the 1961-1990 base period. Precipitation was below average in the west part of Europe and above average in the central and east part of Europe.

2-m temperature

As summarized in supplementary material S0, the models using ECMWF data show comparable high temporal correlation coefficients based on hourly values over the whole domain ($0.88 < R < 0.94$), with highest correlations values in northern Germany and France when looking on a daily basis. Correlations are lower for all models over north of Italy and Austria. On average for the considered period, the bias is negative for all models in the range [-0.3 K, -0.7 K] for CAMx, CHIMERE, EMEP and LOTOS-EUROS. The negative bias for this group of models is more important for the two wintertime campaigns, however in Switzerland and Austria this bias exceeds -2K for all campaigns. Since this group of models shares the same meteorology, the error statistics are very similar; the discrepancies are due to the different interpolation methods used to regrid the 3D and 2D ECMWF variables to the final CTM grid.

RCG displays a very low absolute bias close to zero for the 2009 campaign, and CMAQ displays the lowest negative bias up to -2K for the 2009 campaign. CMAQ has lower correlation coefficient particularly in Germany and Poland for the 2008 and 2009 campaigns.

As displayed in Fig. 1, the negative bias is driven by afternoon temperatures that are underestimated by all models, this statement is valid for all campaigns. The night-time temperatures are more in line with the observations. The RCG diurnal cycle is rather different with a flatter profile but for the other models using ECMWF or COSMO data, the general pattern is well captured.

10-m wind speed

All the models using ECMWF data overestimate the wind speed from +0.1 to +0.9 m s^{-1}, while CMAQ, driven by COSMO, showed on average the lowest absolute bias. The biases are the highest for the two winter (2009) and fall (2008) campaigns, while for the summer campaign (2006) the biases are lower. It is worth noting that the 2007 campaign was the most windy period, showing a mean observed wind speed of 4.77 (m/s).

The bias is generally higher in eastern and northern Europe than in western and Mediterranean areas. In Europe, the spatial pattern of biases shows high positive bias in several coastal areas and negative bias in mountainous areas (Alps). This clearly points out a problem in some regions for the calculation of some emissions directly relying on IFS U10 fields. According to Ingleby *et al.* (2013) ECMWF 10 m wind speeds are slightly overestimated especially at night. In the IFS only 10m winds from ocean going ships are used in the data assimilation due to problems with station representativeness for inland sites. Moreover, errors on wind speed measurements are higher for low winds. For the lowest winds, the comparison of the predicted diurnal cycle with observations shows a larger positive bias at night than during the afternoon (Fig. 1), this behaviour could lead to an overestimation of the advection process in the chemistry transport models.

Time correlations are better for models using ECMWF data but all models exhibit low correlations over the Alpine region (North of Italy, South East of France, Switzerland and Austria). The RCG model shows higher correlation coefficients over northern Europe (Finland and Sweden) for the 2009 campaign.

Planetary boundary layer (PBL) and mixing

As explained in section 4.2, the observed PBL height was calculated at 12:00 because of methodology hypotheses, except at the SIRTA site where hourly measurements are available for 2008 and 2009. All models have a negative bias, the lowest RMSE are shown for CAMx and CHIMERE which use the ECMWF PBL, the biases are in the range -237 m and -100 m for these two models. It is worth noting that CAMx and CHIMERE exhibit exactly the same performance, while LOTOS-EUROS and EMEP that adopted IFS PBL too, show partially different performances., Some differences are attributed to different interpolation schemes and the use of minimum PBL values during night-times as for EMEP. The largest underestimation of the PBL height is usually found for MINNI particularly for the 2006 campaign (up to -616 m) and EMEP (up to -451 m) and the correlation coefficients for these models are lower compared to the others. CMAQ has the lowest bias for most of campaigns. Models using IFS PBL data showed the best performance for temporal correlation (see supplementary material S0), the main discrepancies are observed for the 2006 campaign with several sites in Europe with negative correlations. The largest negative biases are observed in the south of the domain, in these regions CMAQ performs better. In some regions over the Mediterranean basin, particularly in coastal areas, the MINNI's PBL is sometimes strongly biased up to -1000m. The obtained results suggest that either the Carlson algorithm or the micro-meteorological parameterization implemented by MINNI tends to underestimate the intensity of convection.

The spatial representation of the PBL for the 2009 campaign shows higher differences between the models mainly over the ocean and seas where the coefficient of variation reaches 40% in some areas (Fig. 2). While LOTOS-EUROS, CHIMERE, RCG and CAMx use the PBL from IFS with some differences on spatial and time interpolations, the other models use their own parameterizations discussed in section 2.2. The diurnal cycles displayed in Fig. 2 show that MINNI simulates a higher PBL at night and a lower PBL during daytime compared to ECMWF. The difference in the afternoon PBL is quite important over countries influenced by the ocean like the Great Britain. CMAQ and EMEP simulate over France and Great Britain the

highest PBL at night. The hourly times series at the SIRTA site confirm the underestimation of the ECMWF PBL but at this station, the negative bias of MINNI is of the same order of magnitude as those of the other models. The correlations based on hourly values are somewhat lower for CMAQ, EMEP, MINNI (below 0.50) compared to the models using ECMWF data.

The differences in treatment of advection and mixing as reported in section 2.2 lead to differences in the reconstruction of pollutant dispersion. Fig. 3 shows the mean coefficient of variation of CO concentrations predicted by the models sharing the same raw meteorology (IFS) for the 2006 campaign. This pollutant can be considered as a tracer with low influences of deposition and chemistry processes, most of the differences on concentrations are related to transport and mixing. The figure clearly shows that mixing in emission areas, such as big cities, produces the highest differences exceeding 20% of variations. The next highest coefficients of variation are observed over the seas and ocean which are related to the differences of PBL predicted by the models (Fig. 2), elsewhere this coefficient remains below 10%.

6 Overall model performance evaluation on criteria pollutants

6.1 Ozone

The model performances (supplementary material S1) are very different from campaign to campaign. Most of the models overestimate ozone concentrations in 2006, 2007 and 2008 (Fig. 4). Only the 2009 campaign show a systematic underestimation of observed ozone concentrations from -5 to -16 µg m^{-3}. The large positive bias in 2007 and negative in 2009 are largely explained by the boundary conditions that are biased respectively of +8 and -20 µg m^{-3} (Supplementary material S1). For the positive bias in 2007 the boundary conditions cannot be the sole reason, chemical processes play an important role. Correlations are similar for all models in the range 0.5-0.6, only CMAQ has lower correlations on average. For the summertime campaign 2006 CHIMERE and CMAQ display the lowest correlation for daily averaged concentrations but CHIMERE has the lowest bias with EMEP. The low correlation for CMAQ and CHIMERE is due to the difficulties to reproduce both spatial patterns and day to day variations. For this campaign most models underestimate concentrations in the mountainous regions in Spain and over the Alps (Fig. 5). The models tend to over predict ozone concentrations on background stations influenced by large urban areas like GR01 station in Greece and IT01 close to Rome. All models simulate high ozone concentrations over the Mediterranean Sea, most of them behave satisfactorily in Malta and Cyprus stations in agreement with the ozone concentrations pattern over the seas for the "ensemble" shown in Fig. 5 and particularly in Malta (Nolle *et al.,* 2002). The diurnal cycles in Fig. 6 reflect the overall performances depicted previously. All models fairly simulate the timing of the daily peak. For campaign 2007, except MINNI the models overshoot during nighttime and daytime. For campaign 2008, the very good shape of the LOTOS-EUROS diurnal cycle is remarkable. For the summertime campaign 2006, CHIMERE and EMEP provide on average the best diurnal cycles. Focussing on 2006 and 2008 campaigns, the two campaigns which are not biased by the boundary conditions, LOTOS-EUROS show the best performances regarding the bias. For these two campaigns, CAMx has a strong positive bias particularly at night. CAMx and CHIMERE use exactly

the same PBL height of, but night-time performances of the two models are rather different. In Fig. 5, the right side is the gridded coefficient of variation that is a standardized measure of the dispersion of model results. It is defined as the ratio of the standard deviation to the mean of models. This coefficient is very low for the 2006 campaign, below 10%, the models have different responses along the ship tracks. The coefficients of variation are the highest for the 2007 campaign (supplementary material S2) associated with low performances of the "ensemble" (high normalized root mean square errors). France, Spain and Norway show the lowest coefficient of variation indicating a more coherent behaviour among the models, but not necessarily corresponding to better model performance than other areas.

At Mace Head (IE31) located on the west part of the domain the time series of model results versus ozone observations show flat shape for the two winter campaigns with very low time correlations in 2009 (Fig. 7). The best correlation coefficients are observed for 2006 and 2008, the models are able to capture the peaks. At this station the negative bias mentioned in 2009 is roughly the same for LOTOS-EUROS, MINNI and RCG and comparable to the MACC analysis (-20 µg m^{-3}), the other models CAMx, EMEP, CHIMERE and CMAQ have a lower absolute bias (about -10 µg m^{-3}). This behaviour shows that concentrations close to boundary conditions are quickly modified certainly because the regional models restore their own chemical equilibrium in relation with dynamical processes like deposition and vertical dispersion.

6.2 Nitrogen dioxide

For NO_2, all models perform similarly in terms of correlation with value in the range 0.6-0.7 (Fig. 4 and supplementary material S1). The spatial correlation is much higher in the range 0.7-0.9 for all models. Only CMAQ strongly overestimates the mean concentrations and CAMx underestimates the concentrations for all campaigns. Bessagnet et al. (2014) showed rather low concentrations of elemental carbon compared to other models, this inert species is particularly sensitive to vertical mixing and CAMx presents the highest minimum diffusion coefficient that is of major importance during stable conditions and partly explaining the lower NO_2 concentrations. For CAMx, the enhanced mixing influences also O_3 concentrations that are higher than other models.

The spatial pattern of the "ensemble" shown for 2009 (Fig. 8) displays high concentrations over the Benelux, North Italy, the biggest cities and over the shipping tracks. The bias of the "ensemble" is rather good except for one station in Serbia (RS05) with high observed values, probably due to local sources. The gridded coefficients of variation provided in Fig. 8 show that most of differences between models are observed over remote areas far from emission regions even if errors are expected to occur more frequently for low values. As shown for a less reactive species like CO, the differences of mixing in models over emission areas lead to large differences in modelled concentrations, this effect can be clearly seen over the East Mediterranean for maritime emissions where the PBL is different from model to model. Over land the NO_2 chemistry and the different biogenic NO emission modules in the models are believed to explain a large part of the differences on NO_2 concentrations far from urban areas. As shown in Fig. 8, the root mean square errors of the models are the highest for the stations close to the emission areas. The diurnal cycles in Fig. 9 show a general underestimation during the afternoon. It

should be pointed out that the observed NO_2 concentrations can be slightly overestimated. For some types of analyzers, NO_2 is catalytically converted to NO on a heated molybdenum surface and subsequently measured by chemiluminescence after reaction with ozone. The drawback of this technique is that other oxidized nitrogen compounds such as peroxyacetyl nitrate and nitric acid are also partly converted to NO (Steinbacher et al., 2007). In the observations, the presence of two peaks on NO_2 concentrations is related to the traffic emissions peaks occurring in the morning and the evening. The timing of the peak occurrences is also modulated by the meteorology, for the 2006 and 2008 campaigns performed with identical summer time shift we clearly see a time shift of +1 and -1 hour respectively for the morning and evening peaks corresponding to a later rise and earlier fall of the PBL. Thus, as expected, the narrowest time lag between the two peaks is observed for the 2007 campaign. Most of the models predict the first peak too early, particularly CHIMERE and CMAQ for the 2006 campaign, and the second peak generally occurs too late.

CMAQ shows the strongest night-time bias, that contributes to explain the overall overestimation shown by the model in all campaigns. CMAQ was driven by a different meteorology that was characterized by very good performance with respect to both wind speed and PBL height mean bias. Conversely IFS-driven models overestimated night-time wind speed. As night-time vertical mixing is mainly driven by mechanical forces, the model results suggest that models tend to underestimate mixing during stable conditions and, as a consequence, that IFS-driven models show better results suggesting compensation processes.

6.3 Sulphur dioxide

The correlations are rather low for all models in the range 0.2-0.4 for the 2006 campaign to 0.5-0.6 for the 2007 campaign (Fig. 4 and supplementary material S1 for all statistics). Two groups of models are identified CAMx, MINNI and RCG that largely overestimate the concentrations and CHIMERE, CMAQ, EMEP and LOTOS-EUROS which are closer to the observations on average with the best performances on the RMSE. The overestimation in the MINNI model could be partially explained by the low model PBL height. For CAMx, the possible reasons such as the vertical distribution of SO2 emissions near the harbours and coastal areas, insufficient conversion to sulphate and too low deposition were discussed in Ciarelli et al. (2016). This leads to a positive bias of the "ensemble" as shown in Fig. 10 (supplementary material S4) particularly in Western Europe; the normalized RMSE is frequently above 100% in most part of Europe. The main hot spots are located in the Eastern Europe in addition with high concentrations along the shipping routes. The coefficient of variation is the lowest over emission areas but very high in remote areas like over the oceans far from shipping tracks and over mountain areas. This behaviour, very different from a primary species like CO, is a first indication of the very different way to simulate the SO_2 chemistry and deposition processes in the models.

The diurnal cycles presented in Fig. 11 show a peak at about 10:00 – 12:00. This peak is coherent with the hourly emission profiles of the industrial sector showing an emission peak at the same hours; however, most of models predict a larger decrease in the afternoon. Only CMAQ for the 2007 campaign captures satisfactorily the diurnal profile.

6.4 PM10

Concerning the RMSE, on average the performances of the models are similar except CMAQ which has the highest values driven by low correlations and high negative biases particularly for the 2006 campaign (Fig. 4). All models underestimate the concentrations generally in the range -3 to -10 µg m^{-3}. Except CMAQ, the correlations are in the range 0.4 – 0.6, but CHIMERE and EMEP reach 0.7 for the 2006 campaign. MINNI has the lowest absolute biases for the 2007, 2008 and 2009 campaigns. The "ensemble" provides a good picture of the PM10 concentrations in Europe (Fig. 12 and supplementary material S5) except for two stations IT01 in Italy and CY02 in Cyprus with high recorded values. For CY02, high PM10 concentrations are linked to high calcium concentrations (Bessagnet et al., 2014) due to dust events issued from North Africa. This dust event can be clearly observed for EMEP in Fig. 14. The spatial patterns show low concentrations below 5 µg m^{-3} in remote Scandinavia and three hot spots in the Po valley, Benelux and South Poland. The coefficient of variation of model results is rather high over the seas and arid areas as well as over areas influenced by biogenic emissions as in Scandinavia,. This coefficient is generally the lowest over the Western Europe. The best RMSE of the "ensemble" are observed for the summer campaign 2006 with values below 50% of the observations data.

EMEP has higher concentrations over North Africa because the model generates dust in this part of the domain and sea salt concentrations are generally higher over the seas. EMEP and CHIMERE perform well for the spatial correlations (Table 5), EMEP captures better the high concentrations in the south of the domains whereas CHIMERE performs better over the Benelux (supplementary material S5). In 2008, RCG has particularly good spatial correlation compared to the other models. The missing sea salt emission for CAMx is clearly observed over the ocean with very low PM10 concentrations impairing the spatial correlations.

As shown in supplementary material S5, most of models underestimate the highest PM10 concentrations observed in 2008 and 2009 by a factor of 2. For the 10% highest PM10 concentrations, MINNI has the lowest underestimations for these two campaigns whereas EMEP behaves rather well for the 2006 campaign regarding the bias and the correlation. As shown in Bessagnet *et al.* (2014) the large underestimation in 2009 is driven by the underestimation of organic species.

The observed diurnal cycles of PM10 are very flat for all campaigns with a small peak in the evening (Fig. 13). The systematic underestimation of PM10 can be clearly observed but the shape of cycle is not very well captured, the evening peak is not reproduced. The models simulate low concentrations in the afternoon mainly driven by the elevation of the PBL. For the 2009 campaign, MINNI reproduces very well the diurnal cycle until 16:00. As shown in Fig. 14, dust concentrations are higher for MINNI in the centre of the domain. MINNI uses a parameterisation for wind blown dust very productive over any land cover types (Vautard *et al.*, 2005). EMEP mainly produces dust by traffic resuspension and a little over arable land. This higher production of dust by MINNI in Europe certainly improves the negative bias for PM usually observed in chemistry transport models, particularly in the afternoon when the wind speed is higher and the soil moisture content lower.

Most of the underestimation of PM10 by the models is driven by too low daytime PM10 concentrations. It is noteworthy that MINNI calculate the lowest PBL that could explain its relatively higher PM10 concentrations. For the summer campaign

2006, the PM10 observations show an increase of concentrations in the afternoon while all other models tend to predict a decrease, indicating that all models are too sensitive to dynamical processes (meteorology) and not sufficiently to the chemical formation.

6.5 PM2.5

Performances on PM2.5 concentrations are rather different compared to PM10 (Fig. 4). MINNI generally shows a slight positive bias while all models underestimate the averaged concentrations, with CMAQ showing the highest negative bias. The performances of CHIMERE on the correlation are very good for all campaigns, its RMSE being the lowest for three campaigns. As for PM10, the "ensemble" captures rather well the spatial patterns of PM2.5. The concentrations in the south of Europe (Fig. 15 and supplementary material S6) are not specifically underestimated except in Cyprus where dust events also contribute to increase the PM2.5 concentrations. For all campaigns the coefficient of variation for PM2.5 is the lowest in Spain but the RMSE of the "ensemble" is not particularly low in this region. The coefficient of variation is generally high over the north east part of the domain. For all campaigns the models simulate a hot spot over the north of Italy. As shown in the supplementary material S6, CMAQ captures the PM2.5 concentrations in Ispra (IT04) for 2007 and 2008 campaigns better than the other models. This station located at the border of the Po valley hot spot is usually underestimated by the models due to the very stably stratified meteorological conditions in this region. The spatial correlations are usually better for PM2.5 for all models except for the summer campaign (Table 5).

As for the PM10 concentrations, the diurnal cycle of PM2.5 is rather flat with very small morning and evening peaks (Fig. 16). The models have a different behaviour; they simulate a sharp decrease of concentrations in the afternoon consistent with PM10 diurnal cycles. This confirms the lack of secondary production during daytime. The chemical schemes for the production of organic matter are still incomplete for one main reason. As suggested by Jathar *et al.* (2014) a large part of the "unspeciated" fraction of organic species react and produces secondary organic matter and gasoline vehicles could be an important contributor, as well as wood burning emissions according Denier van der Gon *et al.* (2015). This unspeciated fraction is not included in our emission inventory explaining a part of the negative bias of models observed either in winter and summer campaigns particularly during the afternoon. This suggests that models with negative biases on PM2.5 concentrations are consistent with the level of the completeness of our inventory and the state-of-the-art of knowledge on SOA modelling.

7 Impact of meteorology on pollutant concentrations

7.1 Impact of the PBL parameterization with MINNI results for the 2009 campaign

As shown in the previous section, MINNI underestimates the PBL heights calculated at 12:00 from measurements but it is in a better agreement with hourly data available at SIRTA (Fig. 2). In order to test the effect of PBL heights on air quality predictions, the MINNI model has been run using the PBL from IFS instead of its own parameterization for PBL heights. As Shown by Curci *et al.* (2015), processes in the PBL can greatly affect the PM2.5 ground concentrations, for instance temperature and relative humidity can favour the production of ammonium nitrate in the upper PBL.

Fig. 17 shows the average PBL heights and the average concentrations of O_3, NO_2 and PM10 using MINNI's parameterizations (left graphs) and the percentage difference between the average concentrations calculated with PBL heights given by IFS (PBL_{IFS}) and by MINNI's parameterizations (PBL_{MINNI}) (right graphs) using the following formula: $(PBL_{IFS}-PBL_{MINNI})/PBL_{MINNI}$.

It can be seen that over the seas, on average, PBL heights calculated with MINNI's parameterizations (PBL_{MINNI}) are lower than PBL heights given by IFS (PBL_{IFS}) but over the lands PBL_{MINNI} is higher than PBL_{IFS} in coastal areas, North Africa, Scandinavian mountains and middle of Russian plains, and lower over the rest. Over the sea, PBL_{IFS} are higher than PBL_{MINNI} more than 50% while over the land the differences are between -30 and +30%.

Fig. 17 also shows that the O_3 concentrations increase in correspondence of the increase of PBL heights up to 10% and more, and decrease where the PBL heights decrease. This behaviour is explained by the fact that with a higher PBL more O_3 is entrained from high altitudes where O_3 concentrations are higher than at surface. Since the NO_2 sources are mainly at surface, the NO_2 concentrations generally decrease with the increase of PBL heights and increase with the decrease of PBL heights as a consequence of more or, respectively, less effective dilution. Over most of Europe, the NO_2 concentrations decrease up to 8% when PBL_{IFS} heights are used. The PM10 concentrations respond to PBL heights variation in the same way as NO_2. The use of PBL_{IFS} heights produces a 4 % decrease of PM10 concentrations in most parts of Europe but an increase of 6-8% in coastal areas and Russian plains.

In terms of statistics, the use of the PBL from IFS in MINNI slightly improves the correlations mainly driven by an improvement of time correlations. PM10, PM2.5 and NO_2 concentrations are decreased by less than 0.5 μg m^{-3}, improving all error statistics reported in Fig. 4 f. An increase of 2.75 μg m^{-3} is observed for O_3 concentrations. It is also worth to mention that the variations in pollutant concentrations are small (over the land below 10% generally) in comparison to the variations of PBL height, therefore other factors such as emissions spatial distribution, meteorology (e.g. advection and vertical dispersion, especially in low-wind areas), gas phase chemistry, aerosol physics and chemistry have to be investigated for improving model performances.

These results clearly show the importance of having good estimates of PBL heights but they also demonstrates that more investigations are necessary in order to identify the best parameterization of PBL heights but also vertical diffusivities and vertical advection schemes which improves the simulated concentrations over the whole Europe.

7.2 Influence of meteorology on NO_2 concentrations with CAMx results

Pollutant concentrations are strongly influenced by the reconstruction of meteorological fields. In this section a comparison of model performances in reproducing wind speed and NO_2 concentrations is presented and discussed. Furthermore, Planetary Boundary Layer (PBL) height data, collected at SIRTA site (Paris) have been used too. Being mainly related to emission processes, NO_2 has been selected as a tracer of the influence of dispersion on pollutant concentrations. The analysis has been performed over the Paris area since the hourly variation of the PBL is available. Two other limited areas, namely: the whole Germany (DE), the Po Valley (POV) have been selected to complement the analysis.

NO_2 observed data set has been set up from AirBase database (Mol and de Leeuw, 2005), selecting just background stations, having more than 75% valid data over the whole 2009. Modelled concentrations have been derived from the CAMx simulation results, while modelled meteorological fields have been derived from IFS.

In the case of the Paris area, the meteorological model showed a very good performance in reproducing the observed wind speed, whose temporal evolution clearly influences the corresponding temporal variability of NO_2 concentrations (Fig. 18). Also the PBL height is quite well reproduced by the model, though the model tends to underestimate the night-time minima and, conversely, to overestimate some diurnal peaks.

Within the Paris area NO_2 observations are quite well reproduced by CAMx, showing a low bias of the median value lower than 2 ppb, corresponding to less than 20% of the observed median concentration (Fig. 18). The availability of both wind speed and PBL height observations, allow the influence of both processes to be clearly detected. For example 3-4, 10 and 25 of March, the underestimation showed by CAMx seems well related to a corresponding overestimation of the PBL rather than the wind speed (Fig. 19). Conversely during night hours of March 5, CAMx results are more influenced by the wind speed.

The analysis has been completed comparing the diurnal cycle of both NO_2 and meteorological variables, reported in Fig. 20 and Fig. 21. At German sites NO_2 concentrations are slightly overestimated during night-time and underestimated during daytime. This behaviour does not seem strictly related to wind speed, particularly during night-time, thus being probably more related to vertical turbulence. At Po valley sites, NO_2 values are systematically underestimated, while wind speed is correctly reproduced, even partially underestimated during daytime hours. NO_2 modelled concentrations show a clear low bias during night-time, probably related to an imprecise reconstruction of the strong stable conditions that characterize this area during the cold season. The model discrepancies are enhanced during the morning hours, when the model is not able to capture the magnitude of the observed peak. The discrepancy is probably caused by a too rapid growth of the PBL during the first daytime hours. Late in the afternoon the NO_2 bias tends to decrease, probably thanks to a very quick collapse of PBL height after sunset.

At Paris sites, NO_2 modelled concentrations show a behaviour similar to the Po valley area. The availability of both wind speed and PBL height observations, allows most of the previous comments to be confirmed. Particularly it is worth noting

that at SIRTA site, PBL height shows a too rapid increase during morning hours followed by a too strong decrease just after sunset. However, underestimation of NOx emissions cannot be ruled out as depicted in Vaughan *et al.* (2016) or Chen and Borken-Kleefeld (2016), these works highlight the potential underestimation of NOx traffic emissions.

8 Discussion

The results from a mathematical model depend on three main factors: the model formulation (in terms of its assumptions, sub-models, numerical methods and their implementation in computer code); the model input data including boundary conditions; the skill of the model user particularly with respect to use of default values for certain inputs and parameters. When comparing results from a modelling exercise the performance, assessed from comparison between modelling results and data, is influenced by all three of these factors. It is therefore difficult to make judgments on the performance of a model without understanding the importance of configuration and use. In this model intercomparison exercise we have tried to achieve a greater focus on the effects of model formulation by standardising as far as possible the model input data and by running models for specific time periods having different meteorology and season (emissions and meteorology) to test responses over a range of input data. The comparison of results with observations has also been done in a standard way.

CMAQ shows the largest RMSE between predicted and observed values for NO_2 over all campaigns, LOTOS-EUROS shows the lowest RMSE for SO_2 over all campaigns, CAMx always exhibits the highest RMSE for SO_2 over all periods. This means that in several cases either the model formulation or the input setup influence the model performances more than specific features of the meteorological season.

For all pollutants and campaigns, there is not a strong correlation between the performances of the *ensemble* (through the RMSE of the difference between predictions of the ensemble and observed values) with the variability of models (through the coefficient of variation between individual model predictions and the ensemble predictions). This means if models are close to each other (low coefficient of variation), the mean of models can be far or close to the observed values, there are no specific rule. However, for SO_2 and PM2.5 a correlation of -0.2 to -0.3 is observed for three campaigns meaning that a large variability tends to improve the performance of the ensemble for these compounds. For the other compounds, O_3, NO_2 and PM10, the correlation is close to zero. The coefficient of variation is the lowest for ozone (below 10%) particularly in the afternoon hours (see supplementary material S7) and for the summer period 2006, while for SO_2 this coefficient is the highest generally between 30 and 40%. For PM this coefficient is about 10 to 20%, over several countries, the coefficient of variation is higher in the afternoon highlighting the difference between chemical schemes for the aerosol chemistry more active during daytimes, conversely, the low coefficient of variability for O_3 confirms a coherence of ozone chemistry scheme between models.

The intercomparison proved that CTMs are able to reproduce ozone concentrations, showing an average RMSE of individual models corresponding to 30% of the mean observed concentration for daily values. Modelled daily cycles are generally more spread during night-time than daytime hours. This means that, though most models shared the same meteorology, including

PBL height, they proved to be very sensitive to vertical dispersion and deposition parameterization, the two key processes governing O_3 concentration during nighttime. During daytime modelled concentrations are more similar, they show a different ranking with respect to night hours. This means, as expected, that during daytime vertical mixing reconstruction is more similar among models and chemical schemes exhibit a different efficiency in ozone production. This behaviour is not detectable in 2007, that was a cold and windy period, hampering the development of photochemical processes.

NO_2 performances are less robust than for O_3. The RMSE represents about 70% of the observed mean concentration, but the value is even higher in case of CMAQ. Bias is negative for most models, except CMAQ, adopting a different meteorology and MINNI, characterized by lower PBL heights. CHIMERE biases are closer to 0 than other models sharing the same meteorology, such as CAMx.

As for ozone, most of the discrepancies among models and with respect to observations take place during night-time, when the atmosphere is more stable. As most models share the same wind fields, the modelled spread in nighttime concentrations can be related to vertical dispersion. Such spread for primary species and particularly for CO can be considered as a measure of the uncertainty related to vertical mixing and qualitatively corresponds to 80-100% of the observed mean concentration. The height of the first level is also very important for the mixing and deposition processes, it ranges from 20 m for CAMx and CHIMERE to 90 m for EMEP. To be more representative of surface concentrations a correction is implemented for models having a coarse first surface layer (LOTOS-EUROS and EMEP). Daytime modelled concentrations are more similar among models and generally underestimated, though the modelled PBL field at noon seemed lower than the observed one. As already mentioned such a systematic discrepancy could be related to a measurement artefact, but also to photochemistry that could give rise to an excess of nitric acid. More accurate observations of Nitric acid and Nitrate would be required.

SO_2 concentrations show the worst performance, with RMSE values corresponding to 130-160% of the observed mean concentrations. The highest RMSE are shown by CAMx, MINNI and RCG. CAMx. It is worth noting that the modelled diurnal cycles show a weak morning peak, more typical of surface sources not observed in observations. Conversely, measured data present a diurnal peak, usually related to enhance downward mixing of aloft sources, where most of SO_2 is emitted. Discrepancies among models and with respect to observations can also be due to chemistry. For example in 2009, Bessagnet *et al.* (2014) reported for CHIMERE an underestimation of SO_2 concentrations on an hourly basis, while sulphate was overestimated; conversely RCG, adopting a more simplified approach for sulphur chemistry than CHIMERE, overestimated SO_2, while underestimated sulphates.

PM10 model performances are less homogenous within the four years than other pollutants. The campaigns 2006 and 2007 that were characterized by a more dispersive atmosphere show a mean RMSE around 10 µg m^{-3}, representing 55-65% of the mean observed concentration. Differently, the RMSE rises up to 15 µg m^{-3} for 2008 and 2009 campaigns representing more than 80% of the observed mean. The bias is better reproduced by EMEP and MINNI, while CAMx and CMAQ show the strongest underestimation. The analysis of each PM compound for the 2009 period (Bessagnet *et al.*, 2014) revealed that MINNI and EMEP were characterized by rather different scores, suggesting that their overall performance is influenced in a different way by both chemistry and meteorology. Particularly MINNI performance seems more driven by a reduced

dispersion often giving rise to higher concentrations than other models, while EMEP seems more able to capture the evolution of the single PM compound as shown in Bessagnet et al. (2014). CAMx and CMAQ often show the strongest negative bias. As for CAMx this result is probably driven by the combined effect of meteorology (also NO_2 is underestimated by CAMx) as well as the absence of some key processes such as sea salt and dust resuspension and a PM coarse chemistry. Differently CMAQ model was characterized by very high NO_2 concentrations stressing a less dispersive capability than other models. As for CMAQ, the low PM10 values are probably related to deposition processes. Indeed, for 2009 episode (Bessagnet *et al.*, 2014) CMAQ proved to be more efficient than the other models for dry deposition of both NO_X and SO_X compounds.

The observed diurnal cycles of PM10 are very flat for all campaigns with a small peak in the evening. The PM10 observations show an increase of concentrations in the afternoon while all models predict a decrease, indicating that all models are too sensitive to dynamical process (meteorology) and not sufficiently to the chemical formation. The analysis of individual compounds of PM will bring more details, it will be investigated in a companion paper.

Model performance for PM2.5 is on average slightly better than PM10, both in terms of bias and correlation. PM2.5 concentration is less affected by natural processes, which are more relevant for coarse PM, therefore the obtained results suggest that modelling natural processes still present some relevant weaknesses (Bessagnet *et al.*, 2014). Modelled diurnal cycles show improved performance in terms of bias, but not with respect to the daily evolution. Firstly, this result confirms that there are processes mainly affecting the coarse fraction that are still missing in some state of art CTMs, highlighted by the different biases between PM10 and PM2.5. Secondly, the differences in the daily pattern, particularly evident in 2006 where photochemistry is at its maximum, confirm that dilution processes during daytime hours are too efficient with respect to chemical processes, thus preventing the increase of modelled concentrations during afternoon hours.

Even if the meteorology was prescribed in the exercise, some variables related to dispersion modelling such as the vertical diffusion and the PBL height are often diagnosed in the model pre-processing. This step involves important differences in the dispersion as was shown for a tracer species like CO. Although most models used the same PBL from IFS (CHIMERE, CAMx, LOTOS-EUROS, RCG), the variability of models PBL (including other PBL parameterisation as used in EMEP, CMAQ and MINNI) shows important differences of PBL calculations over the ocean and the Mediterranean sea. IFS wind speeds are overestimated with a bias reaching 1 m s^{-1}, which can have a dramatic effect at low wind speed conditions.

The comparison of the meteorological fields pointed out that the reconstruction of the meteorological variables is still affected by relevant uncertainties. Wind speed simulated by IFS and COSMO showed a systematic difference along the whole day, with IFS providing an average wind speed that in 2007 and 2009 was 12% higher than COSMO. PBL reconstruction showed an even higher variability with a spread among the models corresponding to 27-29% of the mean midday PBL value of each campaign.

A comparison of modelling performances in reproducing wind speed and NO_2 concentrations was performed too, also including some analysis of the influence of Planetary Boundary Layer (PBL) height estimation. The comparison of modelled concentrations against wind speed and PBL heights confirmed that meteorology strongly influences CTMs performance.

Particularly the temporal evolution of wind speed is most responsible of model skilfulness in reproducing the daily variability of pollutant concentrations (*e.g.* the development of peak episodes), while the reconstruction of the PBL diurnal cycle seems more influencing in driving the corresponding pollutant diurnal cycle and hence the presence of systematic positive and negative bias detectable on daily basis.

9 Conclusions

One of the main outcomes of such a multi-seasonal intercomparison is that in most cases model performances are more influenced by the model setup than the season. Another general outcome stemming from the whole exercise is that model performances are more different from a pollutant to another than for the same pollutant within the different season. This confirms once again that on average and for the limited dataset used in this exercise, the model formulation (parameterization of chemical / physical processes, calculation of meteorological diagnosed variables) and set-up (number of vertical levels, value of key parameters, etc...) are more influencing than raw meteorological conditions on model performance. One of the few exceptions is shown by O_3 in 2009 where model results were characterized by RMSE values very similar to the other years, whereas bias was negative instead of positive as in the three previous years. But, as already pointed out, such a result was mainly driven by a relevant underestimation in the ozone boundary concentrations from MACC.

Even if the meteorology was prescribed, some variables like the planetary boundary layer (PBL) height, the vertical diffusion coefficient are diagnosed in the model pre-processors and explain the spread of models results. For ozone, this study shows the importance of boundary conditions on model calculations and then on the regime of the gas and particle chemistry. The worst performances are observed for sulphur dioxide concentrations that are poorly captured by the models. The performances of models are rather good and similar for the nitrogen dioxide. On average, the models provide a rather good picture of the particulate matter concentrations over Europe even if the highest concentrations are underestimated. For the PM, the mean diurnal cycles show a general tendency to overestimate the effect of the PBL height rise while the afternoon chemistry (formation of secondary species) is certainly underestimated, PM observations show very flat diurnal profiles for all seasons. In general the daytime PBL height is underestimated by all models, the largest variability of predicted PBL is observed over the ocean and seas. The temporal evolution of wind speed is most responsible of model skilfulness in reproducing the daily variability of pollutant concentrations (*e.g.* the development of peak episodes), while the reconstruction of the PBL diurnal cycle seems more influencing in driving the corresponding pollutant diurnal cycle and hence the presence of systematic positive and negative biases detectable on daily basis.

The study stresses the importance of emission sources particularly in wintertime, wood burning emissions are likely the most underestimated source, through the missing species called semi-volatile organic compounds. Road traffic emissions could

also be underestimated, gasoline and diesel vehicles are both concerned, and more generally all activity sectors involving combustion processes can be concerned. In this study, the importance of meteorological data is highlighted, the difficulties for meteorological models to simulate meteorological variables like wind speed and PBL height during stable conditions can lead to dramatic consequences on air quality modelling. Developments in air quality modelling have not only to focus on processes but also on emissions and meteorological input data. To complement the analysis, companion papers will focus on depositions of sulphur/nitrogen compounds and on the behaviour of models for particulate matter species. This ensemble of analyses will help to prioritize the improvement of air quality models used in the frame of the CLRTAP.

Acknowledgements

The EBAS database has largely been funded by the CLRTAP-EMEP program, AMAP and by NILU internal resources. Specific developments have been possible due to projects like EUSAAR (EBAS web interface), EBAS-Online (upgrading of database platform) and HTAP (import and export routines to build a secondary repository for in support of www.htap.org. A large number of specific projects have supported development of data and metadata reporting schemes in dialog with data providers (CREATE, ACTRIS and others). For a complete list of programs and projects for which EBAS serves as a database, please consult the information box in the Framework filter of the web interface. These are all highly acknowledged for their support particularly Anne-G. Hjellbrekke. INERIS was financed by the French Ministry in charge of Ecology. RSE contribution to this work has been financed by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE S.p.A. and the Ministry of Economic Development - General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency in compliance with the Decree of March 8, 2006. PSI contribution in this work was funded by the Swiss Federal Office for the Environment (FOEN).

References

Aas, W., Tsyro, S., Bieber, E., Bergström, R., Ceburnis, D., Ellermann, T., Fagerli, H., Frölich, M., Gehrig, R., Makkonen, U., Nemitz, E., Otjes, R., Perez, N., Perrino, C., Prévôt, A. S. H., Putaud, J.-P., Simpson, D., Spindler, G., Vana, M., and Yttri, K. E.: Lessons learnt from the first EMEP intensive measurement periods, Atmos. Chem. Phys., 12, 8073-8094, 2012)

Aksoyoglu, S., Keller, J., Barmpadimos, I., Oderbolz, D., Lanz, V. A., Prévôt, A. S. H., and Baltensperger, U.: Aerosol modelling in Europe with a focus on Switzerland during summer and winter episodes, Atmos. Chem. Phys., 11, 7355-7373, 2011.

Ansari, A.S., Pandis, S.: Response of inorganic PM to precursor concentrations. Environ. Sci. Technol, 32, 2706-2714, 1998.

ARIANET: FARM (Flexible Air quality Regional Model) – Model formulation and user manual – Version 2.2. Arianet report R2004.04, Milano, 2004.

Banzhaf, S., M. Schaap, Kerschbaumer, A., Reimer, E., Stern, R., van der Swaluw, E., Builtjes, P.: Implementation and evaluation of pH-dependent cloud chemistry and wet deposition in the chemical transport model REM-Calgrid, Atmos. Environ., 49, 378–390, 2012.

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

Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M., Simmons, A. J., Suttie, M., and the GEMS-AER team: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation, J. Geophys. Res., 114, 2009.

Bessagnet B., Menut, L., Curci, G., Hodzic, A., Guillaume, B., Liousse, C., Moukhtar, S., Pun, B., Seigneur, C., Schulz, M.: Regional modelling of carbonaceous aerosols over Europe - Focus on Secondary Organic Aerosols, Journal of Atmospheric Chemistry, 61, 175-202, 2009.

Bessagnet B., Colette, A., Meleux, F., Rouïl, L, Ung, A., Favez, O., Thunis, P., Cuvelier, C., Tsyro, S., Stern, R., Manders, A., Kranenburg, R., Aulinger, A., Bieser, J., Mircea, M., Briganti, G., Cappelletti, A., Calori, G., Finardi, S., Silibello, C., Ciarelli, G., Aksoyoglu, S., Prévot, A., Pay, M.-T., Baldasano, J., García Vivanco, M., Garrido, J. L., Palomino, I., Martín, F., Pirovano, G., Roberts, P., Gonzalez, L., White, L., Menut, L., Dupont, J.-C., Carnevale, C., Pederzoli, A.: The EURODELTA III exercise – Model evaluation with observations issued from the 2009 EMEP intensive period and standard measurements in Feb/Mar 2009, Technical EMEP report 1/2014, 2014.

Bettems, J.-M.: The COSMO Priority Project 'COLOBOC': Final Report, Technical Report No. 27. http://www.cosmo-model.org/content/model/documentation/techReports/docs/techReport27.pdf, 2015.

Bieser, J., Aulinger, A., Matthias, V., Quante, M., and Denier van der Gon, H.: Vertical emission profiles for Europe based on plume rise calculations., Environ. Pollut., 159, 2935–2946, doi:10.1016/j.envpol.2011.04.030, 2011.

Binkowski, F. and Shankar, U.: The Regional Particulate Matter Model .1. Model description and preliminary results, J. Geophys. Res., 100, 26191–26209, 1995.

Binkowski, F. S.: The aerosol portion of Models-3 CMAQ. In Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. Part II: Chapters 9-18. D.W. Byun, and J.K.S. Ching (Eds.). EPA-600/R-99/030, National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, 10-1-10-16, 1999.

Binkowski, F S., Roselle, S. J.: Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1: Model Description. J. Geophys. Res., Vol 108, No D6, 4183 doi:10.1029/2001JD001409, 2003.

Bott, A.: A Positive Definite Advection Scheme Obtained by Nonlinear Renormalization of the Advective Fluxes. Mon. Wea. Rev., 117, 1006-1015, 1989.

Businger, J.A.: Comments on free convection in the turbulent Ekman layer of the atmosphere, J. Atmos. Sci. 28, 298-299, 1971.

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

Carlton, A. G., Bhave, P. V., Napelenok, S. L., Edney, E. O., Sarwar, G., Pinder, R. W., Pouliot, G. A., and Houyoux, M.: Model representation of secondary organic aerosol in CMAQv4.7, Environ. Sci. Technol., 44, 8553-8560, 2010.

Carson, D. J.: The development of a dry inversion-capped convectively unstable boundary layer. Q. J. R. Met. Soc., 99, 450-467, 1973.

Carter W. P. L.: Documentation of the SAPRC-99 Chemical Mechanism for VOC Reactivity Assessment," Draft report to the California Air Resources Board, Contracts 92-329 and 95-308, May 8, available at http://www.cert.ucr.edu/~carter/absts.htm#saprc99), 2000a.

Carter W. P. L.: Implementation of the SAPRC-99 Chemical Mechanism into the Models-3 Framework," Report to the United States Environmental Protection Agency, January 29. Available at http://www.cert.ucr.edu/~carter/absts.htm#s99mod3), 2000b.

Chang, J. S., R.A. Brost, I.S.A. Isaksen, S. Madronich, P. Middleton, W.R. Stockwell, and C.J, Walcek: A Three-dimensional Eulerian Acid Deposition Model: Physical Concepts and Formulation. J. Geophys. Res., 92, 14,681-14,700, 1987.

Chen Y. and Borken-Kleefeld J. :NOx Emissions from Diesel Passenger Cars Worsen with Age. Environmental Science & Technology, 50 (1). pp. 3327-5851, 2016.

Ciarelli, G., Aksoyoglu, S., Crippa, M., Jimenez, J. L., Nemitz, E., Sellegri, K., Äijälä, M., Carbone, S., Mohr, C., O'Dowd, C., Poulain, L., Baltensperger, U., and Prévôt, A. S. H.: Evaluation of European air quality modelled by CAMx including the volatility basis set scheme, Atmos. Chem. Phys., 16, 10313-10332, 10.5194/acp-16-10313-2016, 2016.

Claiborn, C., Lamb, B., Miller, A., Beseda, J., Clode, B., Vaughan, J., Kang, L., and Nevine, C. : Regional measurements and modelling of windblown agricultural dust: The Columbia Plateau PM10 Program, J. Geophys. Res., 103(D16), 19 753-19 767, 1998.

Clarke, A. D., Owens, S. R., and Zhou, J. C.: An ultrafine sea-salt flux from breaking waves Implications for cloud condensation nuclei in the remote marine atmosphere, J. Geophys. Res., 111, D06202, doi:10.1029/2005JD006565, 2006.

Colella, P., Woodward P. R.:The piecewise parabolic method (PPM) for gas dynamical simulation. J. Comput. Phys., 54, 174-201, 1984.

Crippa, M., Canonaco, F., Lanz, V. A., Äijälä, M., Allan, J. D., Carbone, S., Capes, G., Ceburnis, D., Dall'Osto, M., Day, D. A., DeCarlo, P. F., Ehn, M., Eriksson, A., Freney, E., Hildebrandt Ruiz, L., Hillamo, R., Jimenez, J. L., Junninen, H., Kiendler-Scharr, A., Kortelainen, A. M., Kulmala, M., Laaksonen, A., Mensah, A. A., Mohr, C., Nemitz, E., O'Dowd, C., Ovadnevaite, J., Pandis, S. N., Petäjä, T., Poulain, L., Saarikoski, S., Sellegri, K., Swietlicki, E., Tiitta, P., Worsnop, D. R., Baltensperger, U., and Prévôt, A. S. H.: Organic aerosol components derived from 25 AMS data sets across Europe using a consistent ME-2 based source apportionment approach, Atmos. Chem. Phys., 14, 6159-6176, 2014.

Curci, G., Ferrero, L., Tuccella, P., Barnaba, F., Angelini, F., Bolzacchini, E., Carbone, C., Denier van der Gon, H. A. C., Facchini, M. C., Gobbi, G. P., Kuenen, J. P. P., Landi, T. C., Perrino, C., Perrone, M. G., Sangiorgi, G., and Stocchi, P.: How

much is particulate matter near the ground influenced by upper-level processes within and above the PBL? A summertime case study in Milan (Italy) evidences the distinctive role of nitrate, Atmos. Chem. Phys., 15, 2629-2649, 2015.

Denier van der Gon, H., Jozwicka, M., Hendriks, E., Gondwe, M., and Schaap, M.: Mineral dust as a component of particulate matter, Tno, bop - wp2 - report, report 500099003, TNO Delft, The Netherlands, www.pbl.nl, iSSN:1875–2322 (print) ISSN: 1875- 2314 (on line), 2010.

Denier van der Gon, H. A. C., Bergström, R., Fountoukis, C., Johansson, C., Pandis, S. N., Simpson, D., and Visschedijk, A. J. H.: Particulate emissions from residential wood combustion in Europe – revised estimates and an evaluation, Atmos. Chem. Phys., 15, 6503-6519, doi:10.5194/acp-15-6503-2015, 2015.

Doms, G., J., F., Heise, E., Herzog, H.-J., Mrionow, D., Raschendorfer, M., Reinhart, T., Ritter, B., Schrodin, R., Schulz, J.-P. and Vogel, G.: A Description of the Nonhydrostatic Regional COSMO Model. Part II: Physical Parameterization, Tech. Rep., Deutscher Wetterdienst, http://www.cosmo-model.org/content/ model/documentation/core/cosmoPhysParamtr.pdf, 2011.

EC: 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, 2008.

EC: Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control, 1996.

ENVIRON: User's guide to the Comprehensive Air Quality model with extensions, (CAMx) version 5.40 (September, 2011), http://www.camx.com, 2011

Emberson, L.D., Ashmore, M.R., Simpson, D., Tuovinen, J.-P., Cambridge, H.M.: Towards a model of ozone deposition and stomatal uptake over Europe. EMEP/MSC-W 6/2000, Norwegian Meteorological Institute, Oslo, Norway, 57 pp, 2000a.

Emberson, L.D., Ashmore, M.R., Simpson, D., Tuovinen, J.-P., Cambridge, H.M.: Modelling stomatal ozone flux across Europe. Water, Air and Soil Pollution 109, 403-413, 2000b.

EMEP: Transboundary acidification, eutrophication and ground level ozone in Europe. Part I: Unified EMEP model description. EMEP status Report 1/2003, 2003.

Gong, S., Barrie, L., and Blanchet, J.: Modelling sea-salt aerosols in the atmosphere .1. Model development, J. Geophys. Res., 102, 3805–3818, doi:10.1029/96JD02953, 1997.

Gong, S.: A parameterization of sea-salt aerosol source function for sub- and super-micron particles, Global Biogeochemical Cycles, Volume 17, Issue 4, DOI: 10.1029/2003GB002079, 2003.

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

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modellingbiogenic emissions, Geosci. Model Dev., 5, 1471-1492, doi:10.5194/gmd-5-1471-2012, 2012.

Guinot, B., Cachier, H., and Oikonomou, K.: Geochemical perspectives from a new aerosol chemical mass closure, Atmos. Chem. Phys., 7, 1657-1670, doi:10.5194/acp-7-1657-2007, 2007.

Hanna, S. R., Schulman, L. L., Paine, R. J., Pleim, J. E., and Baer, M.: Development and evaluation of the offshore and coastal dispersion model, J. Air Pollut. Control Assoc., 35, 1039–1047, 1985.

Hayami, H., Sakurai, T., Han, Z., Ueda, H., Carmichael, G.R., Streets, D., Holloway, T., Wang, Z., Thongboonchoo, N., Engardt, M., Bennet, C., Fung, C., Chang, A., Park, S. U., Kajino, M., Sartelet, K., Matsuda, K., Amann, M.: MICS-Asia II: model intercomparison and evaluation of particulate sulfate, nitrate and ammonium. Atmos. Environ., 42, 3510e3527, 2008.

Holtslag, A. A. M. and van Ulden, A. P.: A simple scheme for daytime estimates of the surface fluxes from routine weather data. J. Clim. and Appl. Meteor., 22, 517-529, 1983.

Hosker, R. P.: A comparison of estimation procedures for over-water plume dispersion. ATDL contribution file ; no.99. Oak Ridge, Tenn.: Air Resources Atmospheric Turbulence and Diffusion Laboratory, 1974.

Ingleby, B., Isaksen, L. And Dahoui, M.: Comparison of Met Office and ECMWF Background Fields with Conventional Observations, Sixth WMO Symposium on Data Assimilation, 7-11 October 2013, 2013.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 4073-4109, doi:10.5194/acp-13-4073-2013, 2013.

Jathar, S. H., Gordon, T. D., Hennigan, C. J., Pye, H. O. T., Pouliot, G.A., Adams, P. J., Donahue, N. M., and Robinson, A. L.: Unspeciated organic emissions from combustion sources and their influence on the secondary organic aerosol budget in the United States, Proceedings of the National Academy of Sciences, 111 (29), 10473-10478, 2014.

Jeričevič, A., Kraljevič, L., Grisogono, B., Fagerli, H., and Večenaj, Ž.: Parameterisation of vertical diffusion and the atmospheric boundary layer height determination in the EMEP model, Atmos. Chem. Phys., 10, 341-364; doi:10.5194/acp-10-341-2010, 2010.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527-554, 2012.

Kiesewetter, G., Borken-Kleefeld, J., Schöpp, W., Heyes, C., Thunis, P., Bessagnet, B., Terrenoire, E., Fagerli, H., Nyiri, A., and Amann, M.: Modelling street level PM10 concentrations across Europe: source apportionment and possible futures, Atmos. Chem. Phys., 15, 1539-1553, 2015.

Klimont, Z., Kupiainen, K., Heyes, C., Cofala, J., Rafaj, P., Höglund-Isaksson, L., Borken, J., Schöpp, W., Winiwarter, W., Purohit, P., Bertok, I., and Sander, R.: ECLIPSE V4a: Global emission data set developed with the GAINS model for the period 2005 to 2050: key features and principal data sources, International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, 2361 Laxenburg, Austria, 8 pp., available at: http:

//eccad.sedoo.fr/eccad_extract_interface/JSF/page_login.jsf, 2013.Knote C., Brunner D., Vogel H., Allan, Asmi J.A., Aijala M., Carbone S., van der Gon H. D., Kiendler-Scharr A., Mohr C., Poulain L., Prevot A. S. H., Swietlicki E., and Vogel B.: Towards an online-coupled chemistry-climate model: evaluation of trace gases and aerosols in COSMO-ART, Geosci. Model Dev., 4, 1077–1102, 2011.

Koeble, R. and Seufert, G.: Novel Maps for Forest Tree Species in Europe, in: A Changing Atmosphere, 8th European Symposium on the Physico-Chemical Behaviour of Atmospheric Pollutants, Torino, Italy, 17–20 September, 2001, http://ies.jrc.ec.europa.eu/Units/cc/events/torino2001/torinocd/ Documents/Terrestrial/TP35.htm, 2001.

Köhler, I., Sausen, R., and Klenner, G.: NOx production from lightning, The impact of NOx emissions from aircraft upon the atmosphere at flight altitudes 8–15 km (AERONOX), edited by: Schumann, U., final report to the Commission of the European Communities, Deutch Luft und Raumfart, Oberpfaffenhofen, Germany, 1995.

Koo, B., Knipping, E., and Yarwood, G.: 1.5-Dimensional Volatility Basis Set Approach for Modeling Organic Aerosol in CAMx and CMAQ, Atmospheric Environment, 95, 158-164, 2014.

Kuenen, J., H. Denier van der Gon, A. Visschedijk, H. van der Burgh, R. van Gijlswijk: MACC European emission inventory 2003-2007, TNO report, TNO-060-UT-2011-00588, 2011.

Lange R.: Transferrability of a three-dimensional air quality model between two different sites in complex terrain. J. Appl. Meteorol., 78, 665-679, 1989.

Loosemore, G. A. and Hunt, J. R.: Dust resuspension without saltation, J. Geophys. Res., 105(D16), 20663-20671, 2000.

Lindfors V. and Laurila T.: Biogenic volatile organic compound (VOC) emissions from forests in Finland, Bor. Env. Res., 5: 95–113, 2000.

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

Mailler, S., Khvorostyanov, D., and Menut, L.: Impact of the vertical emission profiles on background gas-phase pollution simulated from the EMEP emissions over Europe, Atmos. Chem. Phys., 13, 5987-5998, doi:10.5194/acp-13-5987-2013, 2013.

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

Matthias V., Aulinger A., Quante M.: Adapting CMAQ to investigate air pollution in North Sea coastal regions . Environmental Modelling & Software, 23, 356-368, 2008.

Maul, P.R.: Atmospheric transport of sulfur compound pollutants. Central Electricity Generating Bureau MID, SSD/80/0026/R, Nottingham. England, 1980.

Menut L, B. Bessagnet, D. Khvorostyanov, M. Beekmann, N. Blond, A. Colette, I. Coll, G. Curci, G. Foret, A. Hodzic, S. Mailler, F. Meleux, J. L. Monge, I. Pison, G. Siour, S. Turquety, M. Valari, R. Vautard and M. G.Vivanco: CHIMERE 2013: a model for regional atmospheric composition modelling, Geoscientific Model Development, 6, 981-1028, 2013.

Minguillón, M. C., Perron, N., Querol, X., Szidat, S., Fahrni, S. M., Alastuey, A., Jimenez, J. L., Mohr, C., Ortega, A. M., Day, D. A., Lanz, V. A., Wacker, L., Reche, C., Cusack, M., Amato, F., Kiss, G., Hoffer, A., Decesari, S., Moretti, F., Hillamo, R., Teinilä, K., Seco, R., Peñuelas, J., Metzger, A., Schallhart, S., Müller, M., Hansel, A., Burkhart, J. F., Baltensperger, U., and Prévôt, A. S. H.: Fossil versus contemporary sources of fine elemental and organic carbonaceous particulate matter during the DAURE campaign in Northeast Spain, Atmos. Chem. Phys., 11, 12067-12084, doi:10.5194/acp-11-12067-2011, 2011.

Mol, W. J. A., Leeuw F. A. A. M. de: AirBase: A Valuable Tool in Air Quality Assessments in: The Proceedings of the 5th International Conference on Urban Air Quality, Valencia Spain 29-31 March 2005, Editors R.S. Sokhi, M.M. Millán and N. Moussiopoulos, 2005.

Monahan, E. C.: In The Role of Air-Sea Exchange in Geochemical Cycling, chapter The ocean as a source of atmospheric particles, pages 129–163. Kluwer Academic Publishers, Dordrecht, Holland, 1986.

Nenes, A, C. Pilinis, and S.N. Pandis: ISORROPIA: A New Thermodynamic Model for Multiphase Multicomponent Inorganic Aerosols. Aquatic Geochemistry, 4, 123-152, 1998.

Nenes, A., Pilinis, C., and Pandis, S. N.: Continued development and testing of a new thermodynamic aerosol module for urban and regional air quality models, Atmos. Environ., 33, 1553–1560, 1999.

Nolle, N., Ellul, R., Heinrich, G., Güsten, H.: A long-term study of background ozone concentrations in the central Mediterranean—diurnal and seasonal variations on the island of Gozo, Atmospheric Environment, Volume 36, Issue 8, March 2002, Pages 1391-1402, 2002.

O'Brien J. J.: A note on the vertical structure of the eddy exchange coefficient in the planetary boundary layer. J Atmos Sci, 27:1213–1215, 1970.

Pal, S., Haeffelin, M., Batchvarova E.: Exploring a geophysical process-based attribution technique for the determination of the atmospheric boundary layer depth using aerosol lidar and near-surface meteorological measurements, in press, Journal of Geophysical Research (Atmospheres), 118, 1–19, doi:10.1002/jgrd.50710, 2013.

Pay, M.T., Jiménez-Guerrero. P., Baldasano, J.M.: Assessing sensitivity regimes of secondary inorganic aerosol formation in Europe with the CALIOPE-EU modellingsystem. Atmos. Environ., 51, 146-164, 2012.

Perniogotti D., Thunis, P., Cuvelier, C., Georgieva, E., Gsella, A., De Meij, A, Pirovano, G., Balzarini, A., Riva, G. M., Carnevale, C., Pisoni, E., Volta, M., Bessagnet, B., Kerschbaumer, A., Viaene, P., De Ridder, K., Nyiri, A., Wind, P.: POMI: a model inter-comparison exercise over the Po Valley, Air Qual Atmos Health, DOI 10.1007/s11869-013-0211-1, 2013.

Pleim, J. E.: A new combined local and non-local PBL model for meteorology and air quality modeling. CMAS conference paper. North Carolina. http://www.cmascenter.org/conference/2006/abstracts/pleim_session1.pdf, 2006.

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

Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. Journal of applied Meteorology and Climatology 46, 1396-1408, 2007b.

Putaud, J.-P., Van Dingenen, R., Dell'Acqua, A., Raes, F., Matta, E., Decesari, S., Facchini, M. C., and Fuzzi, S.: Size-segregated aerosol mass closure and chemical composition in Monte Cimone (I) during MINATROC, Atmos. Chem. Phys., 4, 889-902, 2004.

Prank, M., Sofiev, M., Tsyro, S., Hendriks, C., Semeena, V., Vazhappilly Francis, X., Butler, T., Denier van der Gon, H., Friedrich, R., Hendricks, J., Kong, X., Lawrence, M., Righi, M., Samaras, Z., Sausen, R., Kukkonen, J., and Sokhi, R.: Evaluation of the performance of four chemical transport models in predicting the aerosol chemical composition in Europe in 2005, Atmos. Chem. Phys., 16, 6041-6070, doi:10.5194/acp-16-6041-2016, 2016.

Reimer, E. and Scherer, B.: An operational meteorological diagnostic system for regional air pollution analysis and long term modelling, In: Air Pollution Modelling and its Application IX, edited by: H. v. Dop, and G. Kallos (Eds.), NATO Challenges of Modern Society, Kluwer Academic/Plenum Publisher, New York, 1992.

Sauter, F., van der Swaluw, E., Manders-Groot, A., Wichink Kruit, R., Segers, A., Eskes, H.: LOTOS-EUROS v 1.8 Reference Guide, TNO-060-UT-2012-01451, TNO report, The Netherlands. http://www.lotos-euros.nl/doc/LOTOS-EUROS-v18-reference-guide.pdf, 2014.

Schaap, M., Manders, A.M.M., Hendriks, E.C.J., Cnosse, J.M, Segers, A.J.S., Denier van der Gon, H.A.C., Jozwicka, M., Sauter, F., Velders, G., Matthijsen, J., Builtjes, P.J. H.: Regional modelling of particulate matter for the Netherlands, BOP report 500099008, PBL, the Netherlands, 2009.

Schaap, M., C. Cuvelier, C. Hendriks, B. Bessagnet, J.M. Baldasano, A. Colette, P. Thunis, D. Karam, H. Fagerli, A. Graff, R. Kranenburg, A. Nyiri, M.T. Pay, L. Rouïl, M. Schulz, D. Simpson, R. Stern, E. Terrenoire, P. Wind: Performance of European chemistry transport models as function of horizontal resolution, Atmospheric Environment, 112, 90-105, 2015.

Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A.: Modelling the formation of secondary organic within a comprehensive air quality model system. J. Geophys. Res., 106, 28275-28293, 2001.

Seinfeld, J. H. and S. N. Pandis: Atmospheric Chemistry and Physics, From Air Pollution to Climate Change. John Wiley and Sons, Inc., NY, 1998.

Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreiro, A., Guenther, A., Hewitt, C. N., Janson, R., Khalil, M. A. K., Owen, S., Pierce, T. E., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrason, L., and Öquist, M. G.: Inventorying emissions from Nature in Europe, J. Geophys. Res., 104, 8113– 8152, 1999.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., and Wind, P.: The EMEP MSC-W chemical transport model – technical description. Atmos. Chem. Phys., 12, 7825–7865, 2012.

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

Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Moran, M. D., Appel, K. W., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Miranda,

A. I., Nopmongcol, U., Prank, M., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., Galmarini, S.: Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII, Atmos. Environ., Volume 53, Pages 75-92, ISSN 1352-2310, 2012.

Sorensen, J. H.: Sensitivity of the DERMA long-range gaussian dispersion model to meteorological input and diffusion parameters. Atmos. Environ., 32 (24), 4195-4206, 1998.

Steinbacher, M., C. Zellweger, B. Schwarzenbach, S. Bugmann, B. Buchmann, C. Ordonez, A. S. H. Prevot,and C. Hueglin, Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, J. Geophys. Res., 112, D11307, doi:10.1029/2006JD007971, 2007.

Stern, R., Yamartino, R., Graff, A.: Analyzing the response of a chemical transport model to emissions reductions utilizing various grid resolutions. In: Twenty-eighth ITM on Air Pollution Modelling and its Application, Leipzig, Germany, 15–19 May 2006, 2006.

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

Strader, R., F. Lurmann, Pandis, S. N.: Evaluation of secondary organic aerosol formation in winter. Atmos. Environ., 33, 4849-4863, 1999.

Terrenoire, E., Bessagnet, B., Rouïl, L., Tognet, F., Pirovano, G., Létinois, L., Beauchamp, M., Colette, A., Thunis, P., Amann, M., and Menut, L.: High-resolution air quality simulation over Europe with the chemistry transport model CHIMERE, Geosci. Model Dev., 8, 21-42, 2015.

Thunis P., Cuvelier, C., Roberts, P., White, L., Post, L., Tarrason, L., Tsyro, S., Stern, R., Keschbaumer, A., Rouïl, L., Bessagnet, B., Bergstrom, R., Schaap, M., Boersen, G., Builtjes, P.: EuroDelta-II, Evaluation of a Sectorial Approach to Integrated Assessment Modelling including the Mediterranean Sea. JRC Scientific and Technical Reports – EUR 23444 EN 2008, 2008.

Thunis, P., A. Pederzoli, D. Pernigotti: Performance criteria to evaluate air quality modelling applications., Atmos. Environ., 59, 476-482, 2012.

Tørseth, K., Aas, W., Breivik, K., Fjæraa, A. M., Fiebig, M., Hjellbrekke, A. G., Lund Myhre, C., Solberg, S., and Yttri, K. E.: Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972–2009, Atmos. Chem. Phys., 12, 5447-5481, doi:10.5194/acp-12-5447-2012, 2012.

Troen, I. and Mahrt, L.: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. Bound.-Layer Meteorol., 37, 129-148, 1986.

Tsyro, S., Aas, W., Soares, J., Sofiev, M., Berge, H., Spindler, G.: Modelling of sea salt concentrations over Europe: key uncertainties and comparison with observations, Atmos. Chem. Phys., 11, 10367–10388, doi:10.5194/acp-11-10367-2011, 2011.

Tuovinen, J.-P., Ashmore, M., Emberson, L., Simpson, D.: Testing and improving the EMEP ozone deposition module, Atmos. Environ., 38, 2373–2385, 2004.

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

van Leer B.: Multidimensional explicit difference schemes for hyperbolic conservation laws, in Computing Methods in Applied Sciences and Engineering VI, edited by R. Growinski and J. L. Lions Elsevier, Amsterdam, 1984.

van Loon M., R. Vautard, M. Schaap, R. Bergström, B. Bessagnet, J. Brandt, P.J.H. Builtjes, J.H. Christensen, C. Cuvelier, A. Graff, J.E. Jonson, M. Krol, J. Langner, P. Roberts, L. Rouïl, R. Stern, L. Tarrasón, P. Thunis, E. Vignati, L. White, P. Wind: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, Atmos. Environ., Volume 41, Issue 10, Pages 2083-2097, 2007.

Van Zanten, M.C., Sauter, F.J., Wichink Kruit, R.J., Van Jaarsveld, J.A., Van Pul, W.A.J.: Description of the DEPAC module: Dry deposition modelling with DEPAC_GCN2010. RIVM report 680180001/2010, Bilthoven, the Netherlands, 2010.

Vaughan, A. R., Lee, J. D., Misztal, P. K., Metzger, S., Shaw, M. D., Lewis, A. C., Purvis, R. M., Carslaw, D. C., Goldstein, A. H., Hewitt, C., N., Davison, B. D., Beevers, S. D., Karl, T. G. Spatially resolved flux measurements of NOx from London suggest significantly higher emissions than predicted by inventories. Faraday Discussions, DOI: 10.1039/c5fd00170f, 2016.

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

Vautard, R., Builtjes, P. H. J., Thunis, P., Cuvelier, C., Bedogni, M., Bessagnet, B., Honoré, C., Moussiopoulos, N., Pirovano, G., Schaap, M., Stern, R., Tarrason, L., Wind, P.: Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry transport models over four European cities within the CityDelta project, Atmos. Environ., Volume 41, Issue 1, Pages 173-188, ISSN 1352-2310, http://dx.doi.org/10.1016/j.atmosenv.2006.07.039, 2007.

Vautard R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Foltescu, V., Graff, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouïl, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., Wind, P.: Skill and uncertainty of a regional air quality model ensemble, Atmos. Environ., Volume 43, Issue 31, Pages 4822-4832, ISSN 1352-2310, 10.1016/j.atmosenv.2008.09.083, 2009.

Venkatram, A. and Pleim, J.: The electrical analogy does not apply to modelling dry deposition of particles, Atmos. Environ., 33, 3075-3076, 1999.

Venkatram, A.: Estimation of Turbulence Velocity Scales in Stable and Unstable Boundary Layer for Dispersion Applications. Eleventh NATO-CCMS International Technical Meeting on Air Pollution Modelling and its Application, 1980.

Vestreng, V., G. Myhre, H. Fagerli, S. Reis, and L. Tarrason: Twenty-five years of continuous sulphur dioxide emission reduction in Europe, Atmos. Chem. Phys., 7, 3663–3681, 2007.

Vukovich, J. and Pierce, T.: The Implementation of BEIS3 within the SMOKE Modeling Framework", In Proceedings of the 11th International Emissions Inventory Conference, Atlanta, Georgia, April 15-18, 2002, (Available online: www.epa.gov/ttn/chief/conference/ei11/modeling/vukovich.pdf), 2002.

Walcek, C. J.: Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection. Journal of Geophysical Research D: Atmosphere 105 (D7), 9335-9348, 2000.

Walcek, C. J. and Taylor, G. R.: A theoretical method for computing vertical distribution of acidity and sulphate production within cumulus clouds. J. Atmos. Sci., 43, 339-355, 1986.

Wesely, M. L.: Parameterization of Surface Resistances to Gaseous Dry Deposition in Regional-Scale Numerical Models. Atmos. Environ., 23, 1293-1304, 1989.

Wyngaard, J. C.: Atmospheric Turbulence. Ann. Rev. Fluid Mech., 24, 205-33, 1992.

Yamartino, R. J., Scire, J., Carmichael, G. R., Chang, Y. S.: The CALGRID mesoscale photochemical grid model---I. Model formulation, Atmos. Environ., 26A, 1493-1512, 1992.

Yamartino, R. J.: Nonnegative, conserved scalar transport using grid-cell-centered, spectrally constrained Blackman cubics for applications on a variable-thickness mesh. Mon. Wea. Rev., 121, 753-763, 1993.

Yamartino, R.J.: Refined 3d Transport and Horizontal Diffusion for the REM/CALGRID Air Quality Model. Freie Universiträt Berlin, Institut für Meteorologie, 2003.

Yarwood. G., Rao, S., Yocke, M., Whitten, G. Z.: Updates to the Carbon Bond chemical mechanism: CB05. Final Report prepared for US EPA. Available at http://www. camx.com/publ/pdfs/CB05_Final_Report_120805.pdf, 2005.

Zender, C. S., Bian, H., Newman, D.: Mineral Dust Entrainment And Deposition (DEAD) model: Description and 1990s dust climatology, J. Geophys. Res., 108(D14), 4416, 2003.

Zhang, L., Gong, S., Padro, J., Barrie, L.: A size-segregated particle dry deposition scheme for an atmospheric aerosol module. Atmos. Environ., 35, 549-560, 2001.

Zhang, L., Brook, J. R., Vet, R.: A revised parameterization for gaseous dry deposition in air-quality models. Atmos. Chem. Phys., 3, 2067–2082, 2003.

Zhang, K. M., Knipping, E. M., Wexler, A. S., Bhave, P. V., Tonnesen, G. S.: Size distribution of sea-salt emissions as a function of relative humidity, Atmos. Environ., 39(18), 3373-3379, 2005.

Zilitinkevich, S., Baklanov, A.: Calculation of the height of the stable boundary layer in practical applications. Boundary Layer Meteorology, 105(3), 389-409, 2002.

Zyryanov, D., Foret, G., Eremenko, M., Beekmann, M., Cammas, J.-P., D'Isidoro, M., Elbern, H., Flemming, J., Friese, E., Kioutsioutkis, I., Maurizi, A., Melas, D., Meleux, F., Menut, L., Moinat, P., Peuch, V.-H., Poupkou, A., Razinger, M., Schultz, M., Stein, O., Suttie, A. M., Valdebenito, A., Zerefos, C., Dufour, G., Bergametti, G., and Flaud, J.-M.: 3-D

evaluation of tropospheric ozone simulations by an ensemble of regional Chemistry Transport Model. Atmos. Chem. Phys., 12, 3219-3240, 2012.

List of figures

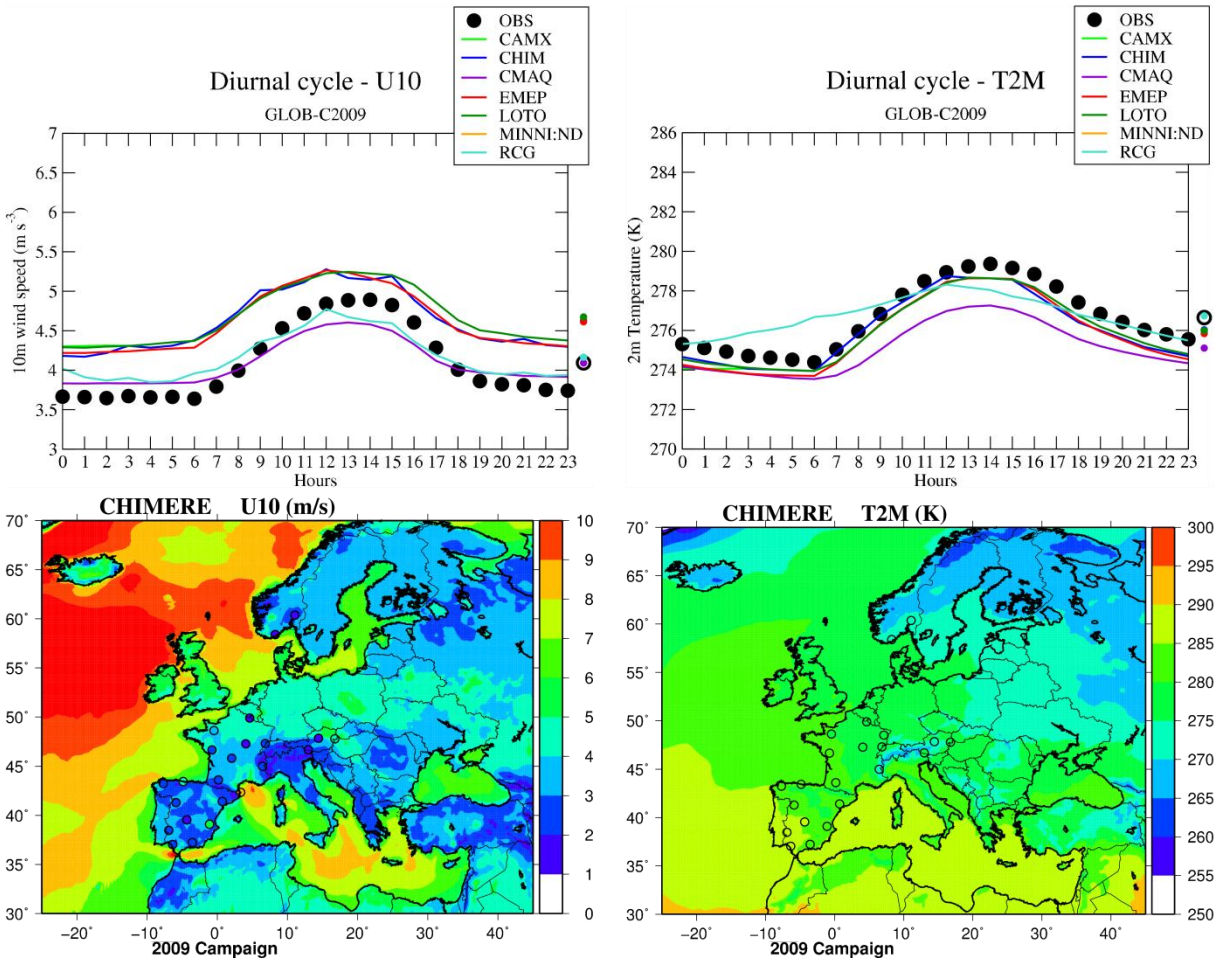

Fig. 1: Comparisons of observed *versus* predicted meteorological variables (U10, T2M) for the 2009 campaign. *Top left panel*: mean diurnal cycle of the 10 m wind speed, *top right panel*: mean diurnal cycle of the 2 meter temperature, *bottom left panel*: mean 10 meters wind speed for CHIMERE, *bottom right panel*: mean 2 meters temperature for CHIMERE (Some observations at EMEP stations are provided with coloured circles over the maps). Red color is assigned for values exceeding the colour scale.

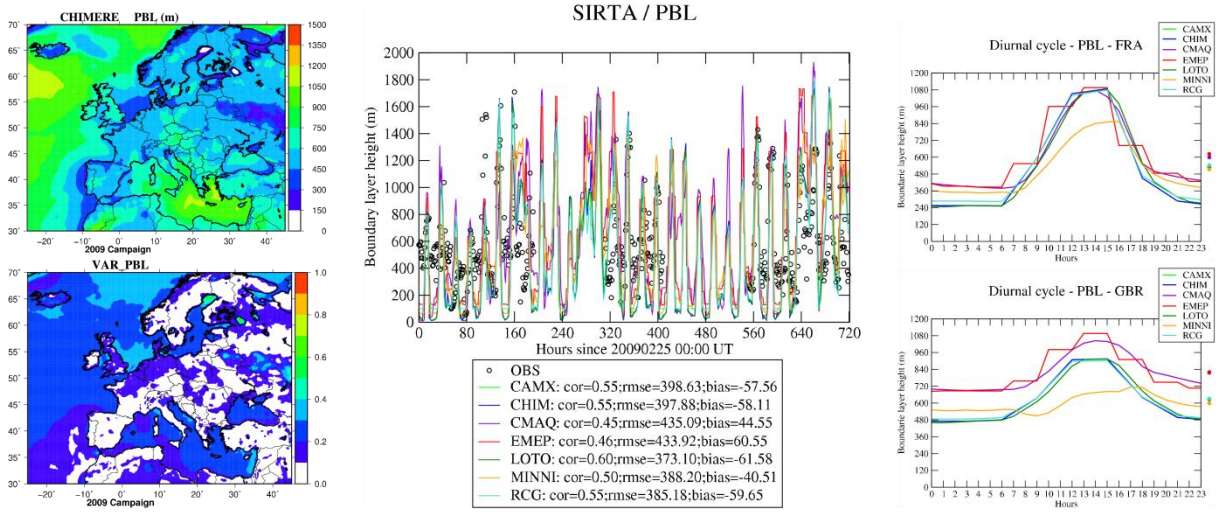

Fig. 2: Spatial representations and time variations of the PBL height for the 2009 campaign. *Top left panel:* Mean height of the CHIMERE PBL height issued from ECMWF data. *Bottom left panel:* Mean coefficient of variation for the PBL height. *Central panel:* hourly variation of the PBL height at the SIRTA station. *Top right panel*: Average diurnal cycle of the PBL height predicted by the models in France. *Bottom right panel*: Average diurnal cycle of the PBL height predicted by the models in Great Britain.

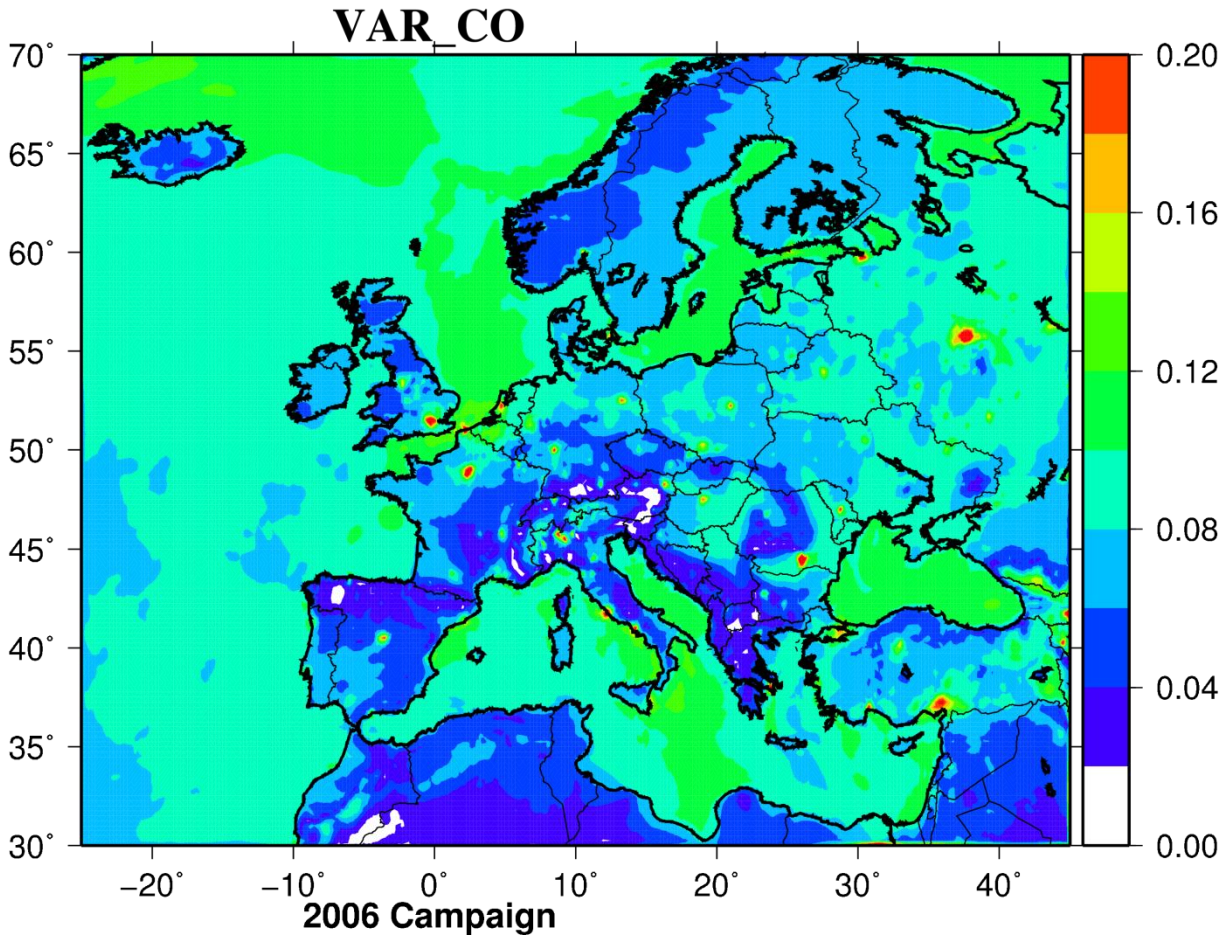

Fig. 3: Mean coefficient of variation of the CO concentrations predicted by the models for the 2006 campaign (no unit). Red color is assigned for values exceeding the color scale.

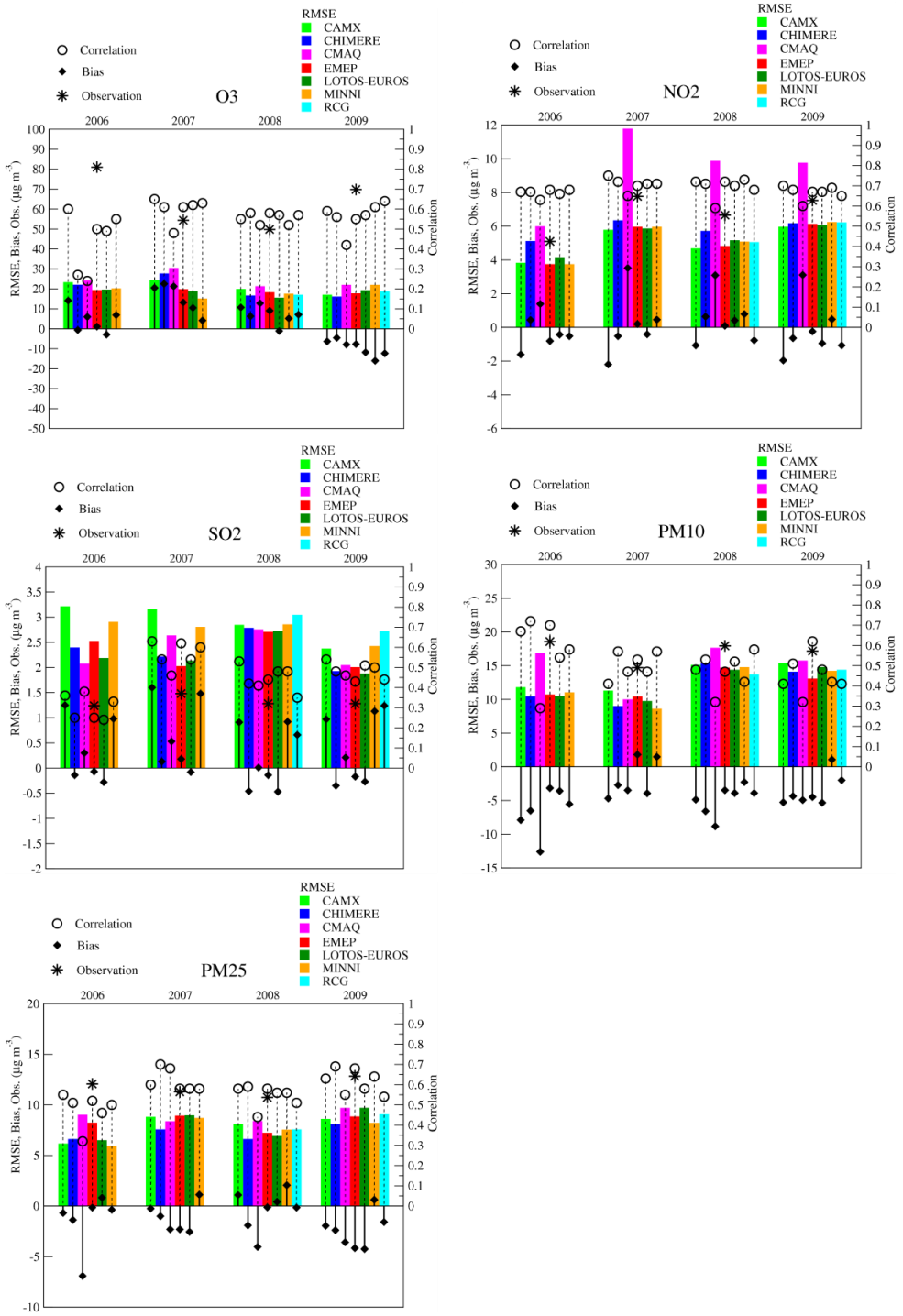

Fig. 4: Overall performance of models for Ozone, Nitrogen dioxide, Sulphur dioxide, PM10 and PM2.5 daily mean concentrations for all campaigns.

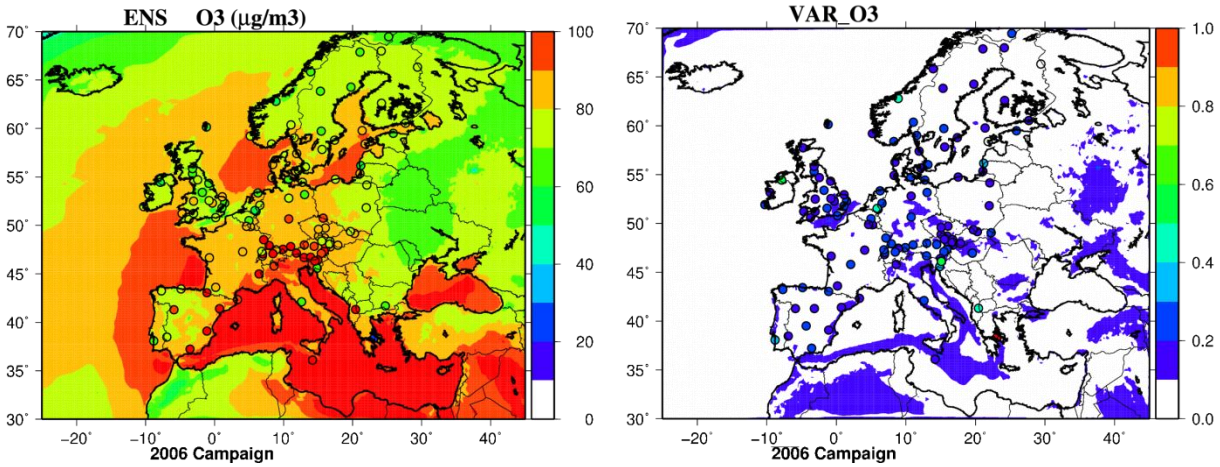

Fig. 5: ***Left column*****: Mean ozone concentrations (µg m^{-3}) of the "ensemble" (ENS) for the 2006 campaign with corresponding observations (coloured dots).** ***Right column*****: coefficient of variation of models (no unit) constituting the ensemble with corresponding normalized root mean square errors of the "ensemble" (coloured dots). Red color is assigned for values exceeding the color scale.**

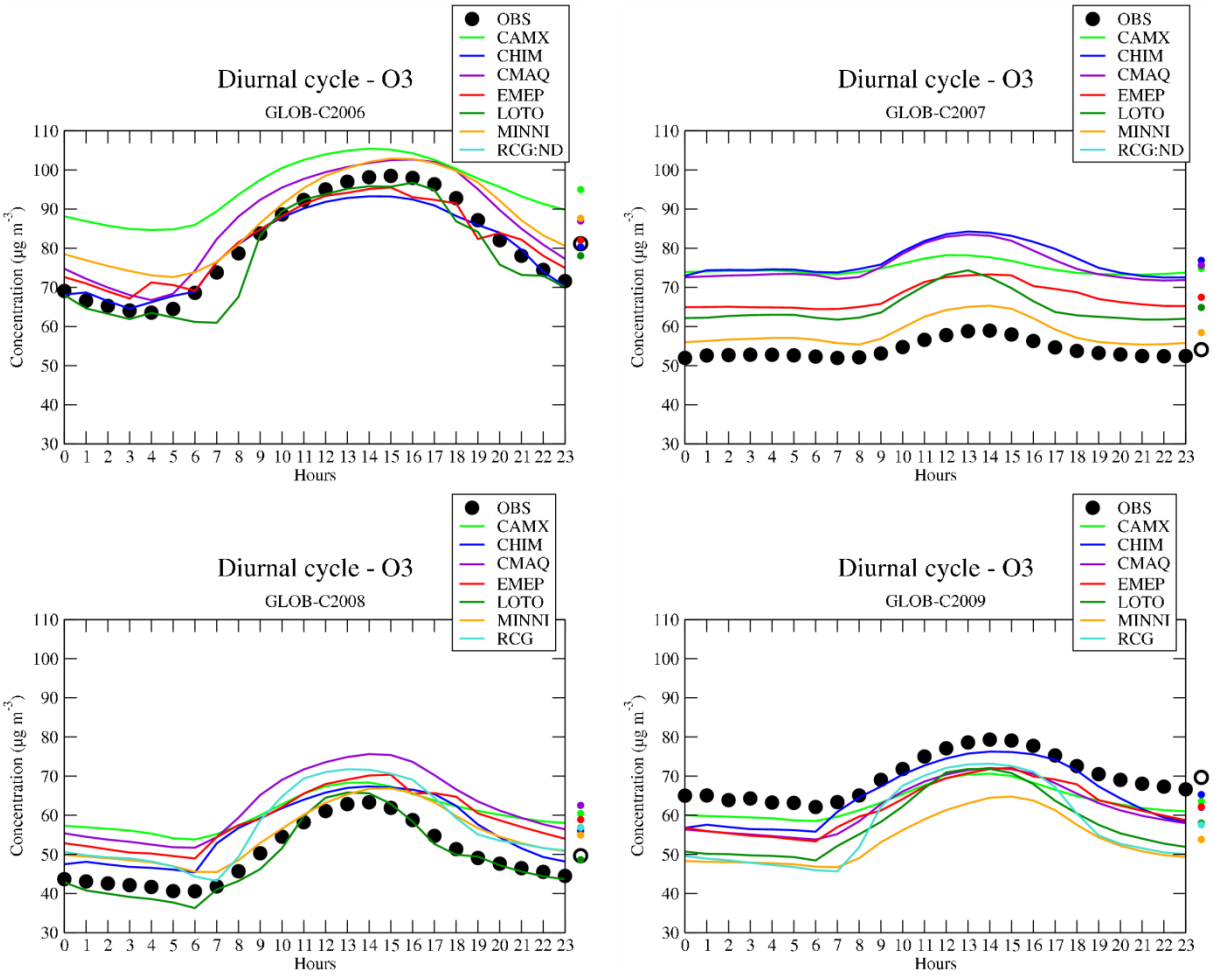

Fig. 6: Mean ozone diurnal cycles for all campaigns simulated by the models compared with observations. Averaged concentrations are provided on the right side of the charts.

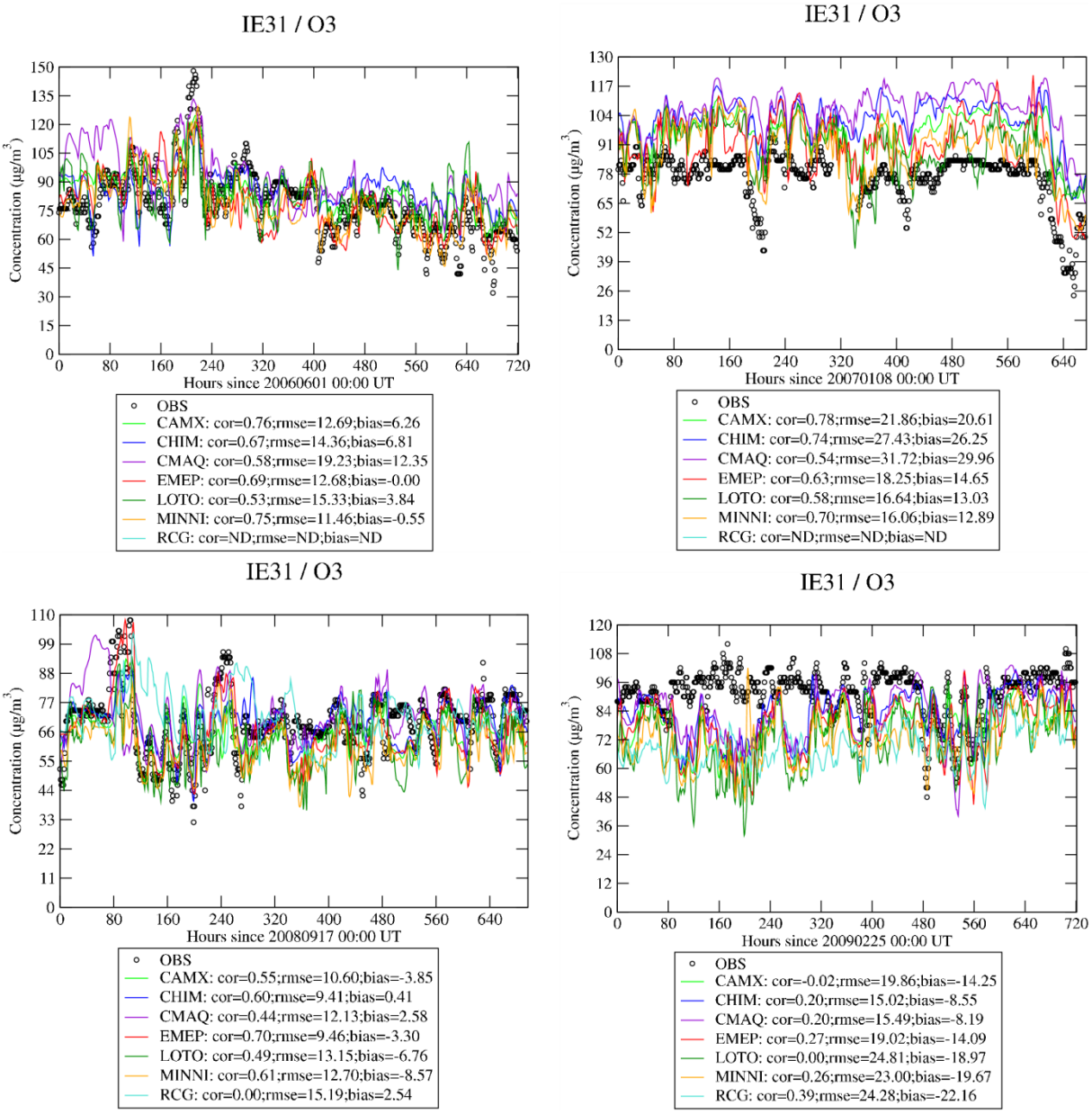

Fig. 7: Timeseries of hourly concentrations at Mace Head for all models and campaigns

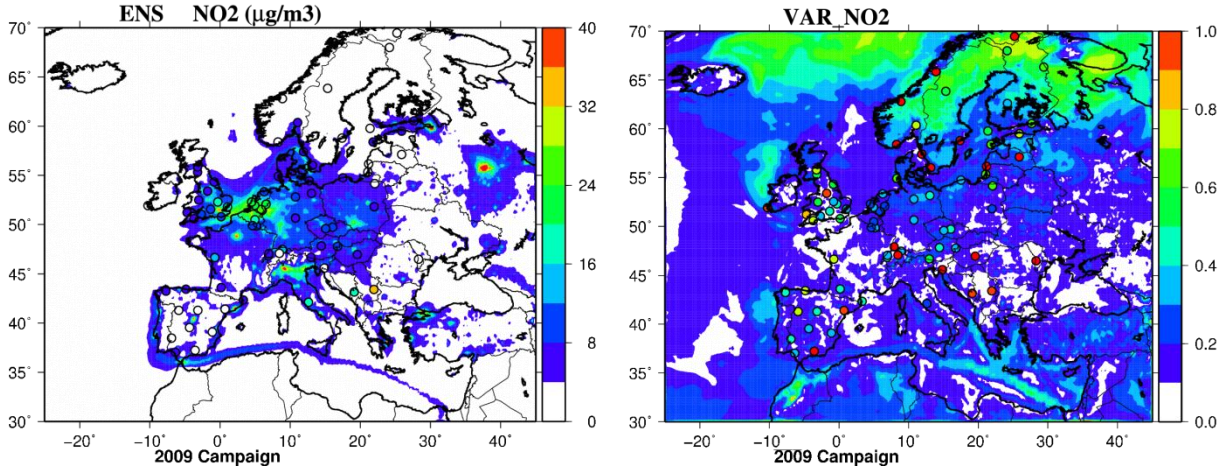

Fig. 8: ***Left column*****: Mean nitrogen dioxide concentrations (µg m^{-3}) of the "ensemble" (ENS) for the 2009 campaign with corresponding observations (coloured dots).** ***Right column*****: coefficient of variation of models (no unit) constituting the ensemble with corresponding normalized root mean square errors of the "ensemble" (coloured dots). Red color is assigned for values exceeding the color scale.**

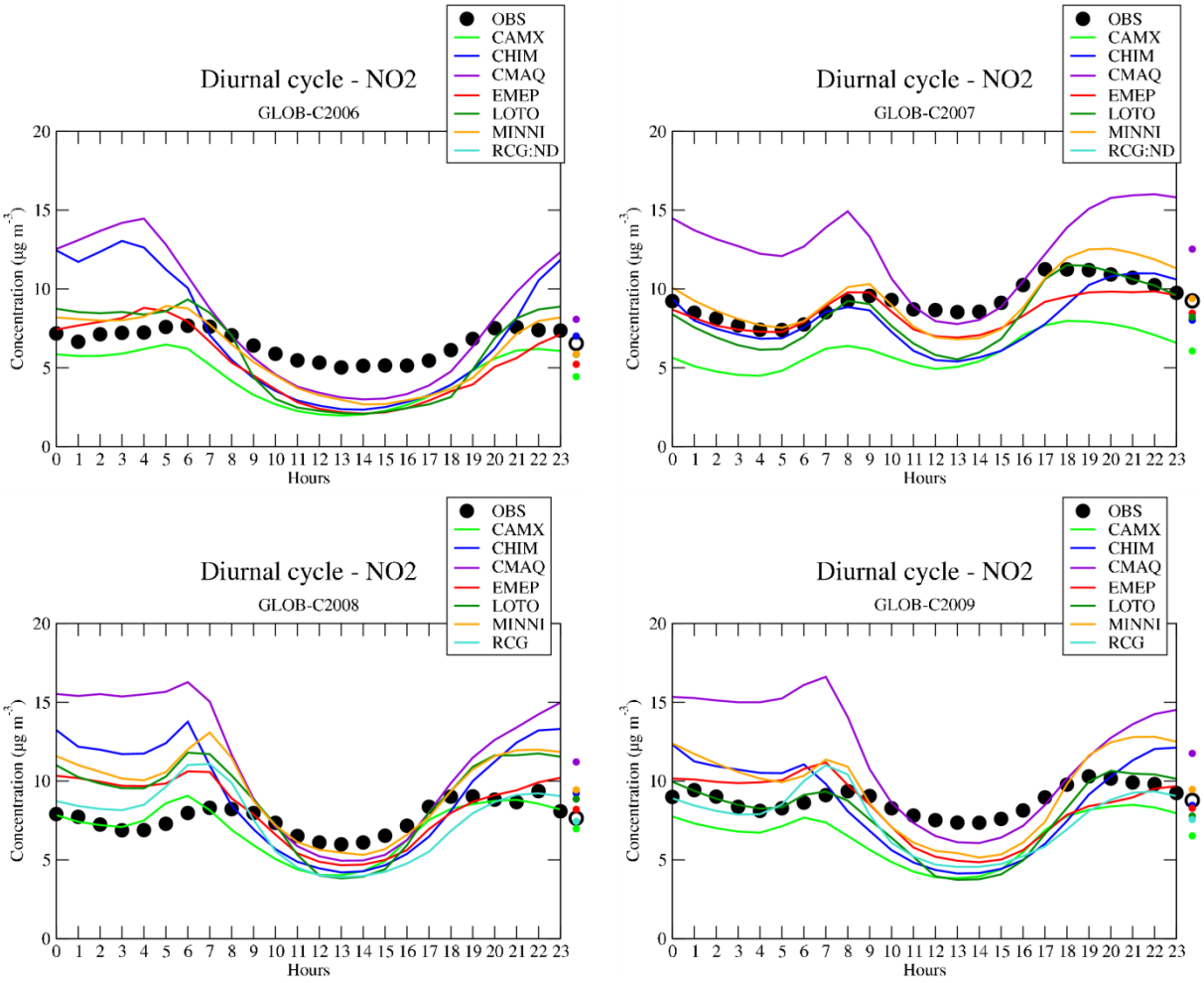

Fig. 9: Mean diurnal cycles of nitrogen dioxide for all campaigns simulated by the models compared with observations. Averaged concentrations are provided on the right side of the charts.

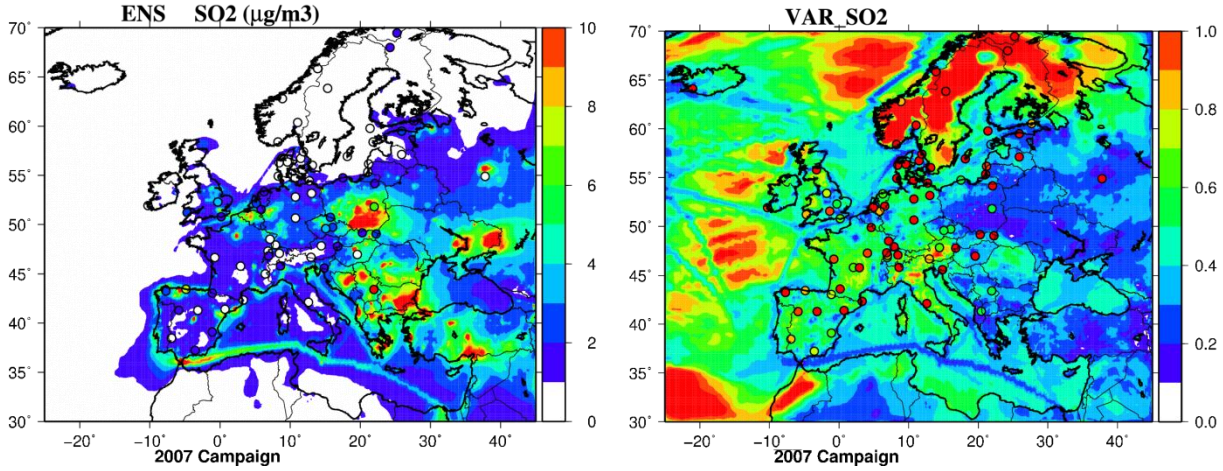

Fig. 10: ***Left column*****: Mean SO_2 concentrations (μg m^{-3}) of the "ensemble" (ENS) for the 2007 campaign with corresponding observations (coloured dots).** ***Right column*****: coefficient of variation of models (no unit) constituting the ensemble with corresponding normalized root mean square errors of the "ensemble" (coloured dots). Red color is assigned for values exceeding the color scale.**

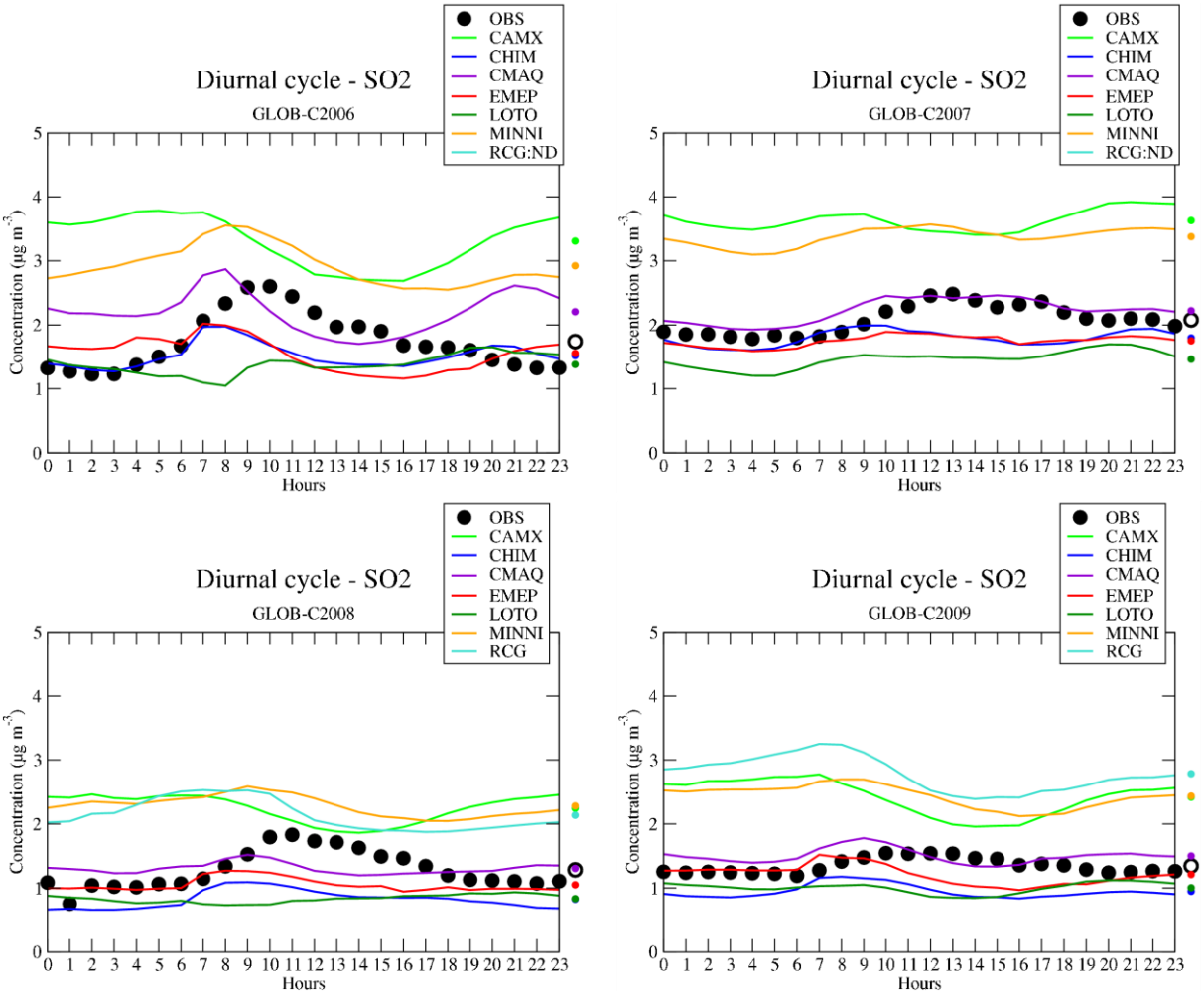

Fig. 11: Mean SO2 diurnal cycles for all campaigns simulated by the models compared with observations. Averaged concentrations are provided on the right side of the charts.

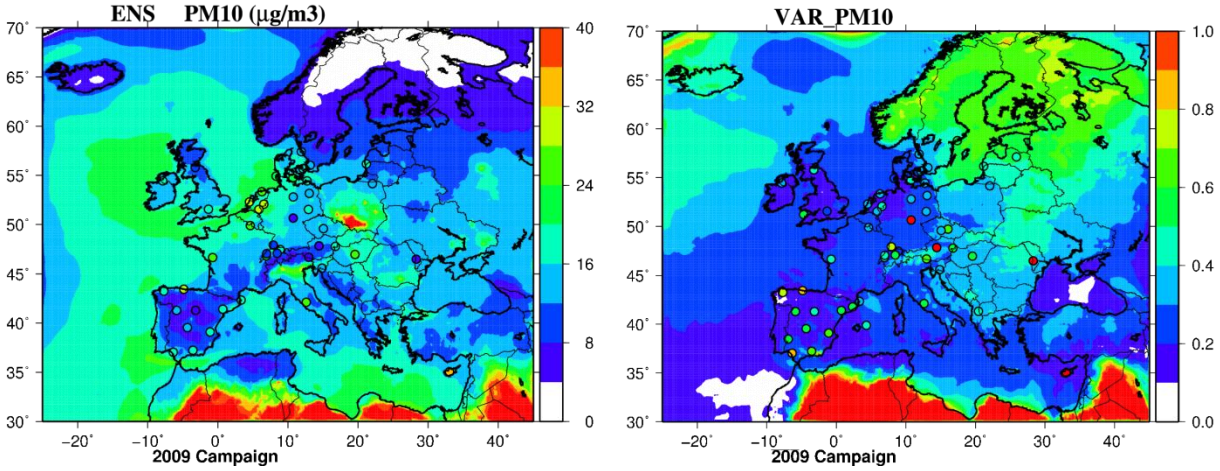

Fig. 12: ***Left column*****: Mean PM10 concentrations (µg m^{-3}) of the "ensemble" (ENS) for the 2009 campaign with corresponding observations (coloured dots).** ***Right column*****: coefficient of variation of models (no unit) constituting the ensemble with corresponding normalized root mean square errors of the "ensemble" (coloured dots). Red color is assigned for values exceeding the color scale.**

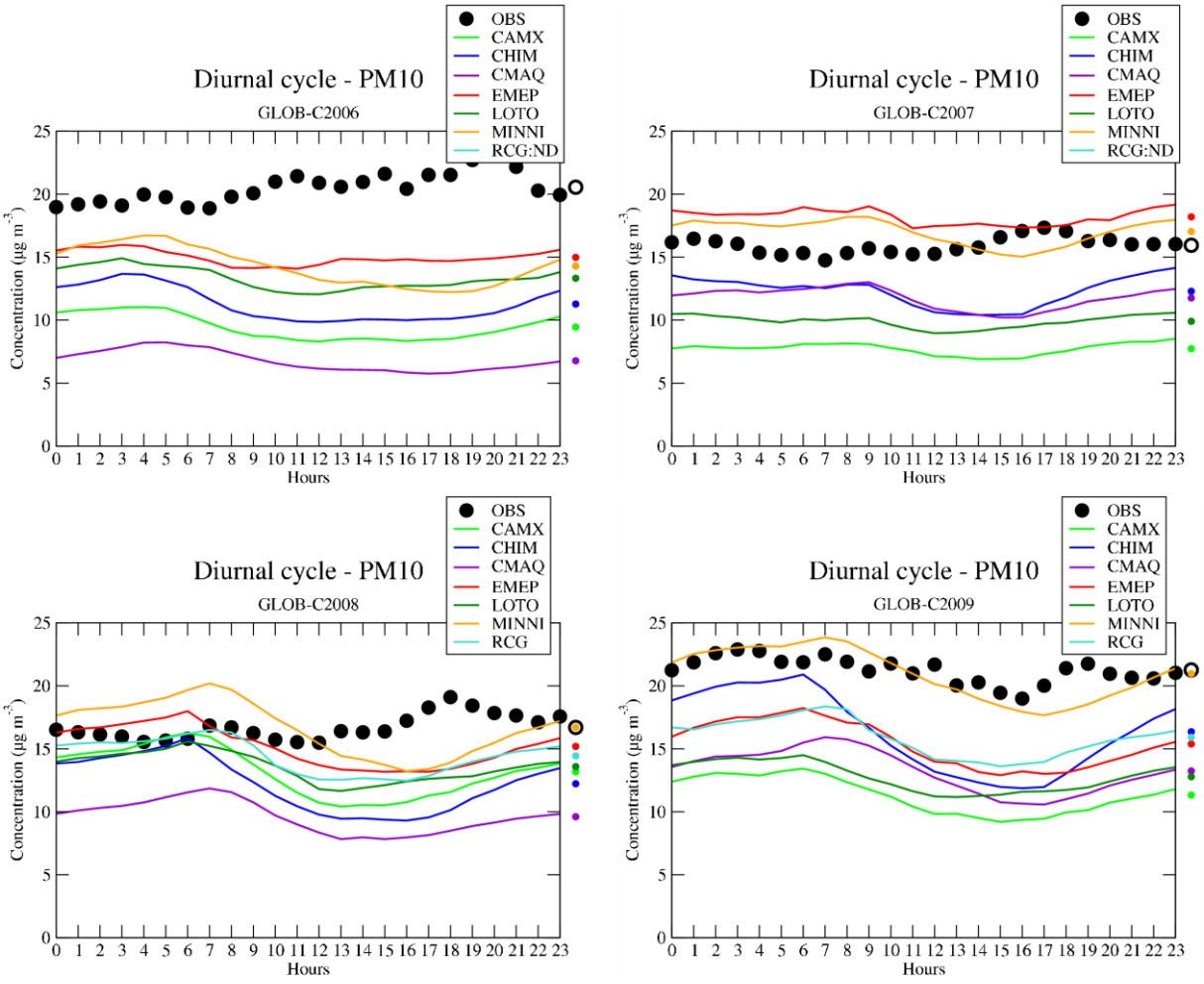

Fig. 13: Mean diurnal cycles of PM10 for all campaigns simulated by the models compared with observations. Averaged concentrations are provided on the right side of the charts.

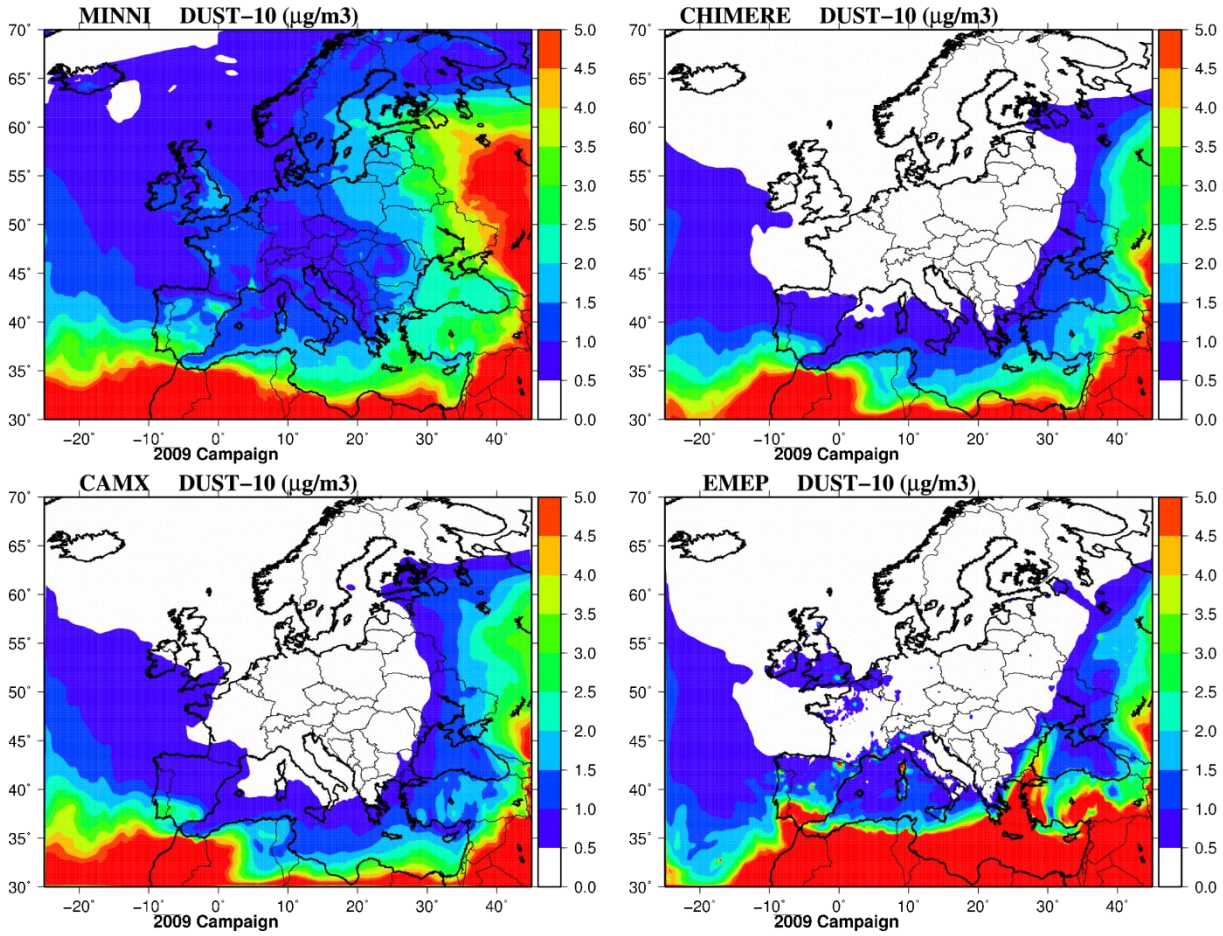

Fig. 14: Mean dust concentrations (μg m^{-3}) in the PM10 fraction for the 2009 campaign computed by the MINNI, CHIMERE, CAMx and EMEP models.

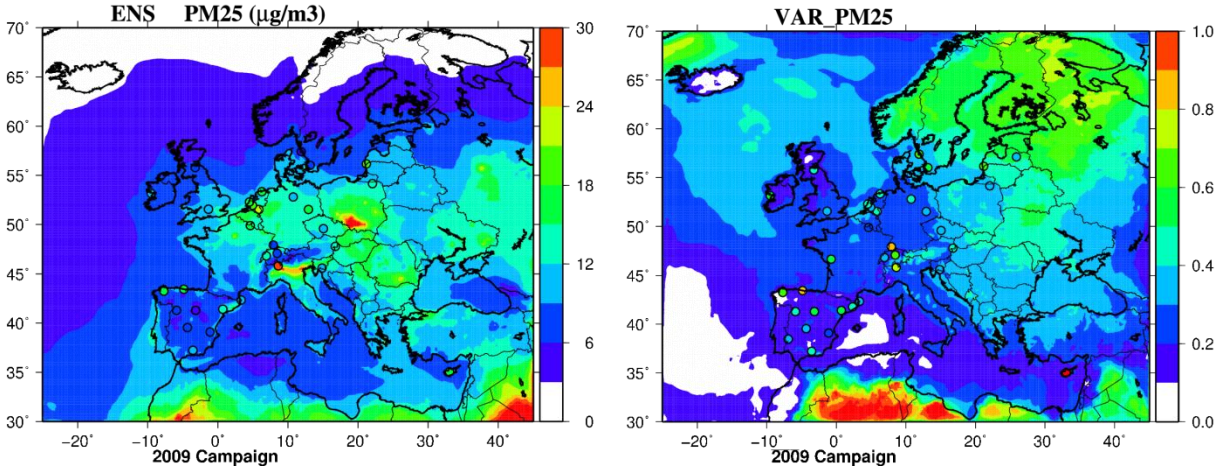

Fig. 15: ***Left column*****: Mean PM2.5 concentrations (µg m^{-3}) of the "ensemble" (ENS) for the 2009 campaign with corresponding observations (coloured dots).** ***Right column*****: coefficient of variation of models (no unit) constituting the ensemble with corresponding normalized root mean square errors of the "ensemble" (coloured dots). Red color is assigned for values exceeding the color scale.**

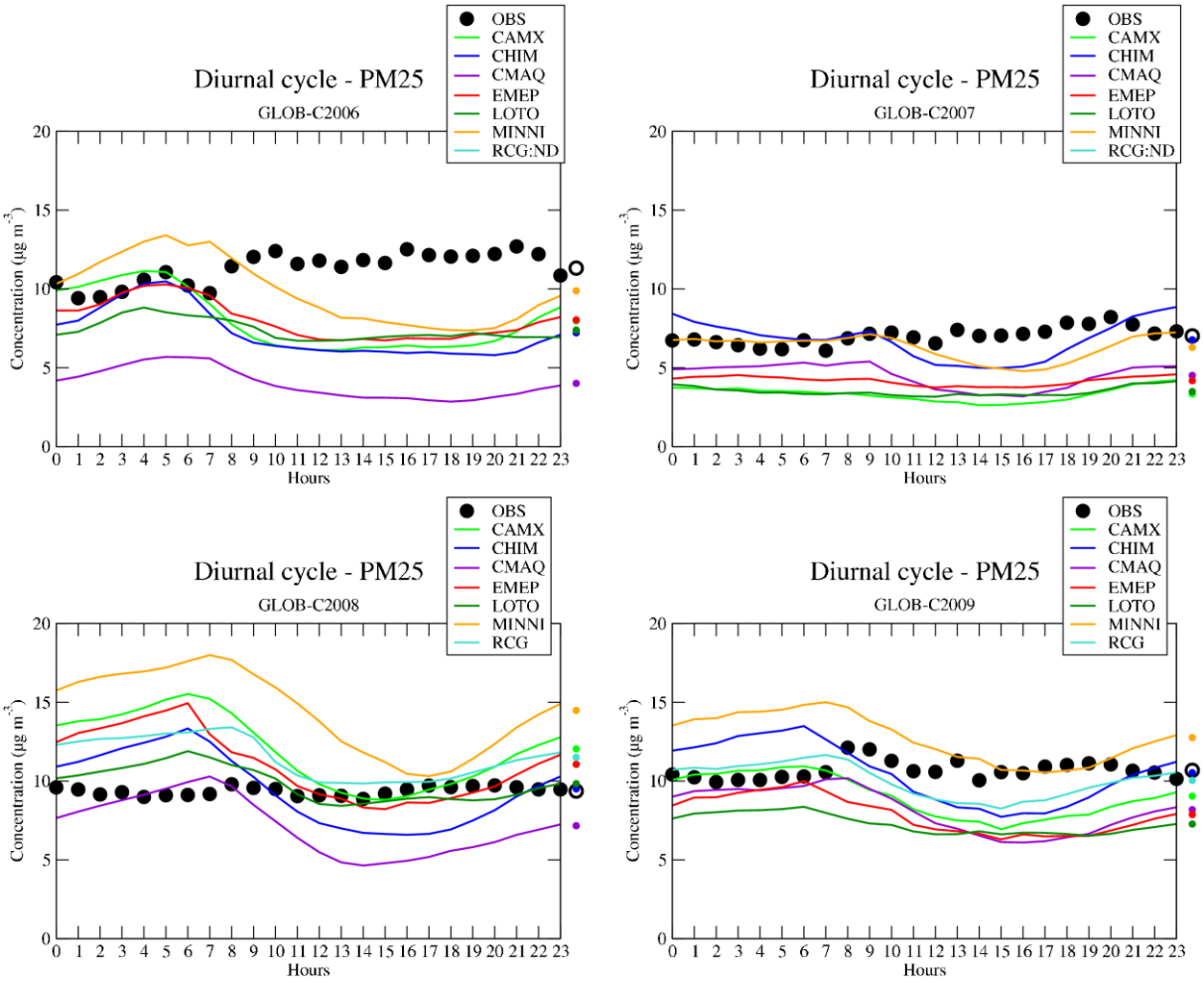

Fig. 16: Mean diurnal cycles of PM2.5 for all campaigns simulated by the models compared with observations. Averaged concentrations are provided on the right side of the charts.

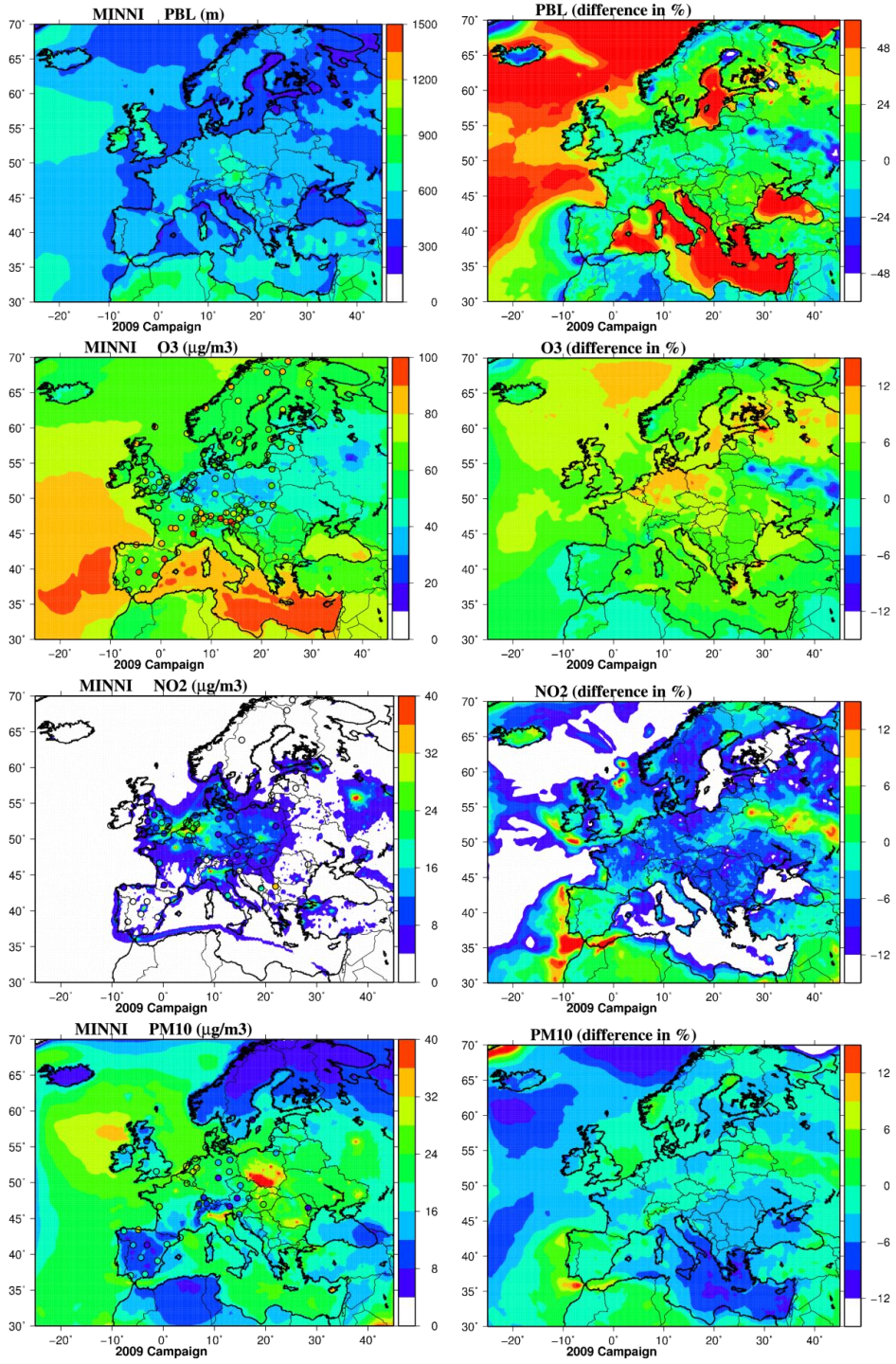

Fig. 17: Left graphs show the average PBL heights and the average concentrations for O_3, NO_2 and PM10 using original MINNI's parameterizations. Right graphs show the percentage difference between the average concentrations calculated with PBL heights given by IFS (PBL_{IFS}) and by MINNI's parameterizations (PBL_{MINNI}). Red color is assigned for values exceeding the color scale.

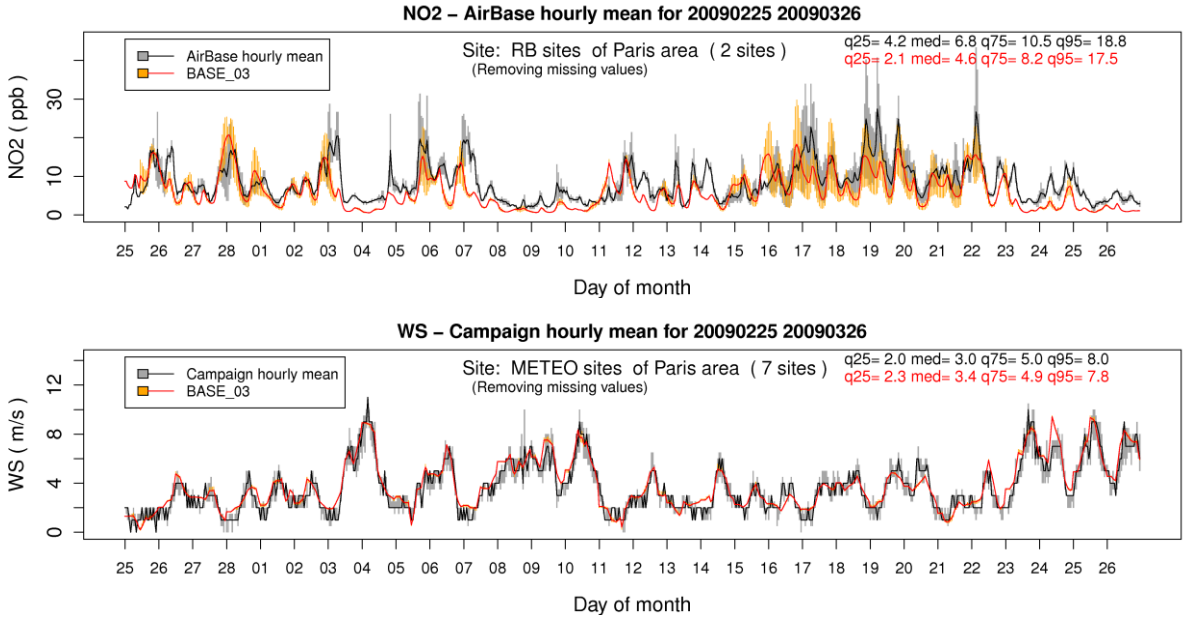

Fig. 18: Time series of hourly Box plots showing the distribution of the observed and computed NO_2 concentration (top) and wind speed (bottom) for CAMx (meteorology from IFS). Observations are in black/grey; modelled values in red/orange. Bars show the 25^{th} -75^{th} quantile interval, while the median is displayed by the continuous line. The 25^{th}, 50^{th}, 75^{th}, and 95^{th} quantile of the whole campaign are reported too. Comparison of computed and observed boxplot time series evaluated at Airbase and meteorological sites, available over the Paris area.

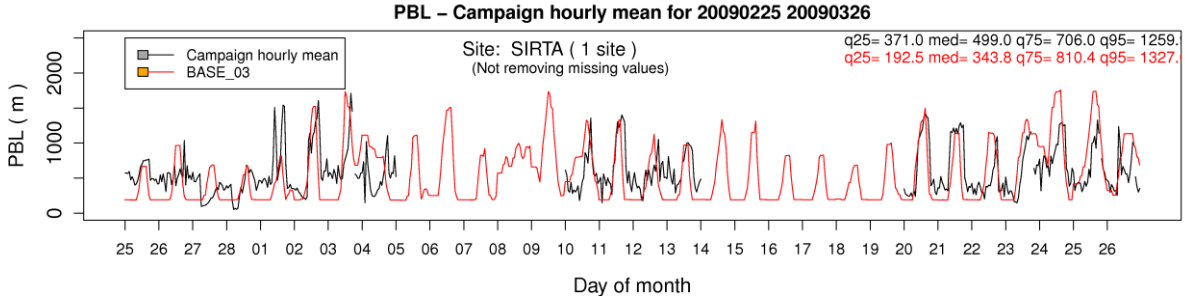

Fig. 19: Time series of hourly Box plots showing the distribution of the observed and computed PBL height. Observations are in black/grey; modelled values in red/orange. Bars show the 25^{th} -75^{th} quantile interval, while the median is displayed by the continuous line. The 25^{th}, 50^{th}, 75^{th}, and 95^{th} quantile of the whole campaign are reported too. Comparison of computed and observed boxplot time series evaluated at SIRTA site.

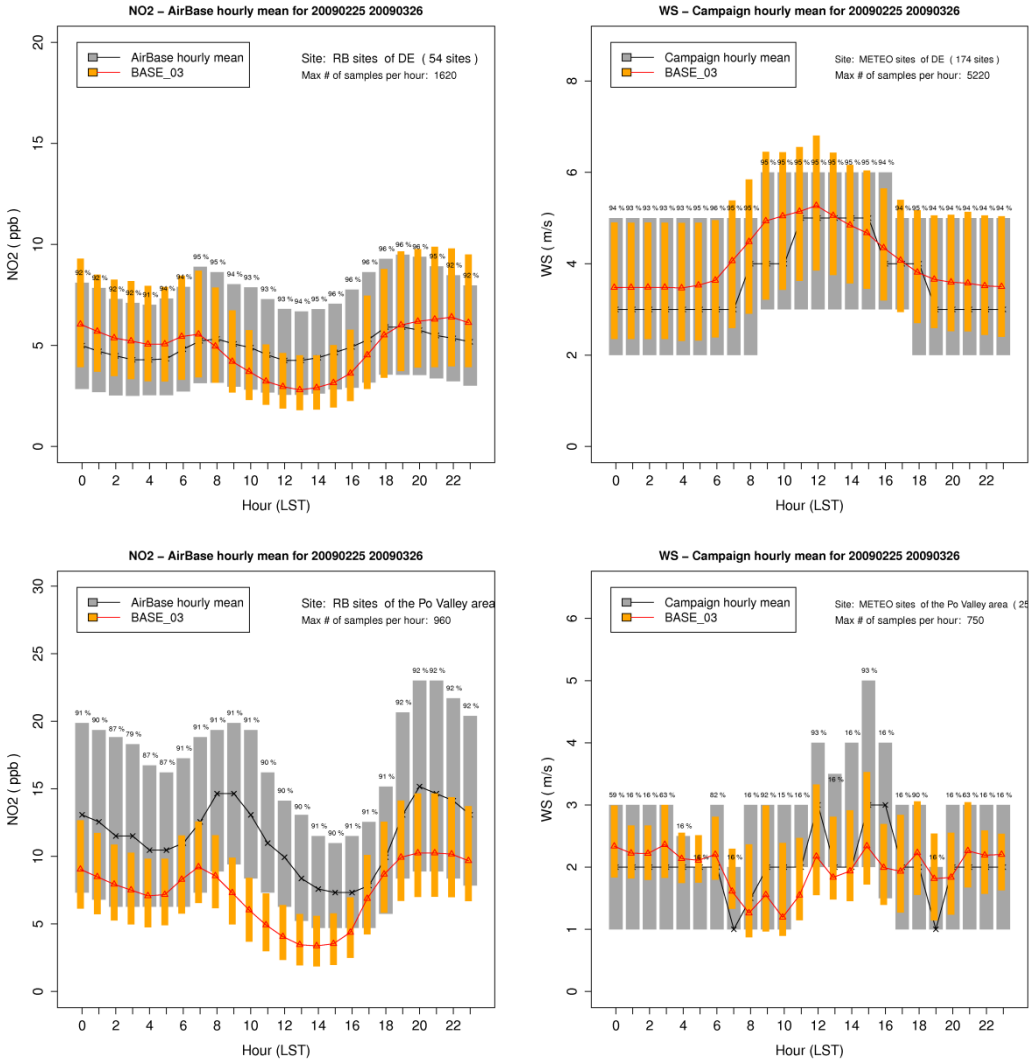

Fig. 20: Time series of hourly Box plots showing the distribution of the diurnal cycle observed and computed NO_2 concentration (left) and wind speed (right) over Germany (top panels) and Po valley (bottom panels). Observations are in black/grey; modelled values in red/orange. Bars show the 25th -75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole campaign are reported too. Comparison of computed and observed boxplot time series evaluated at AirBase and meteorological sites, available over Germany and Po valley. Hour is in UTC time.

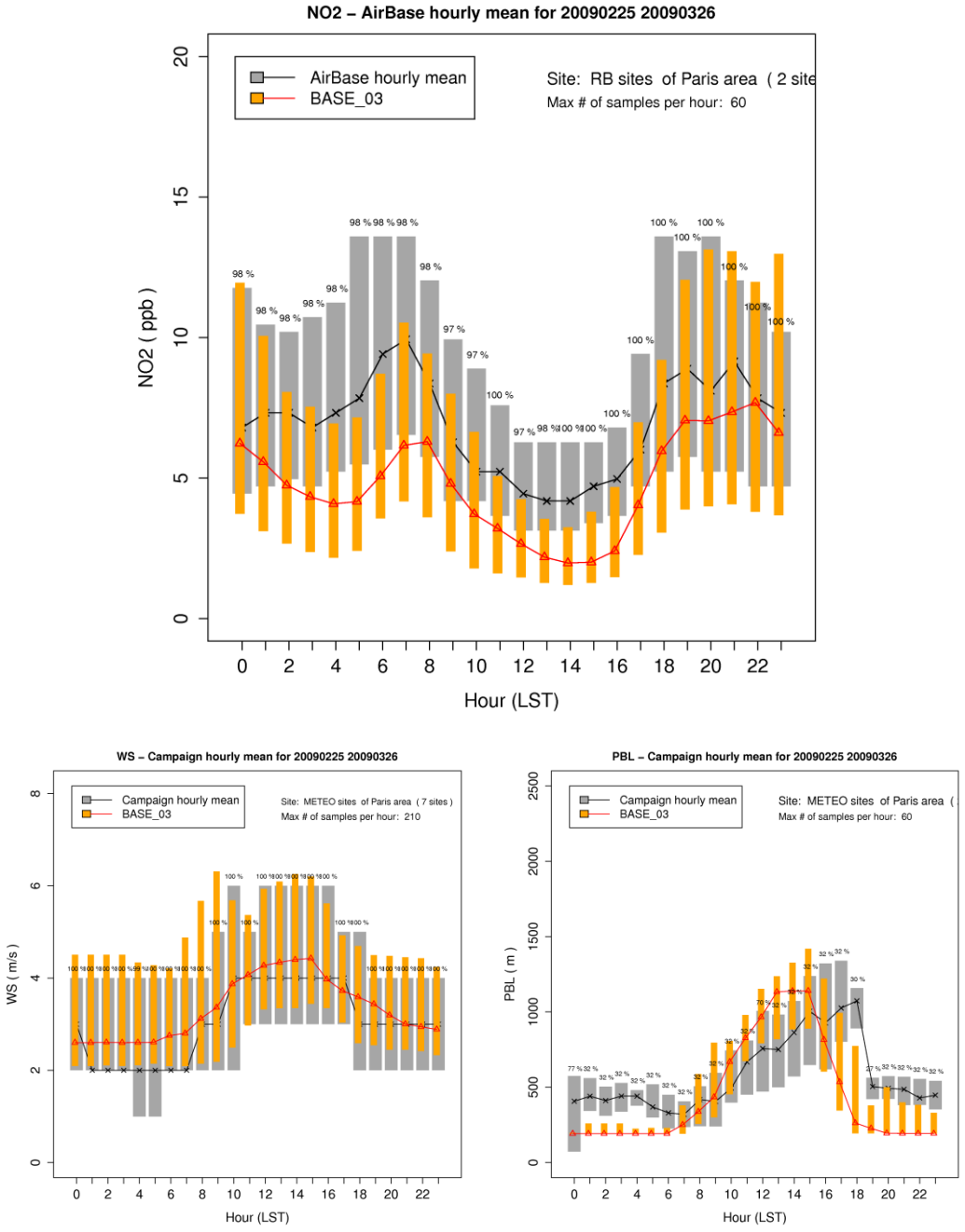

Fig. 21: Time series of hourly Box plots showing the distribution of the diurnal cycle observed and computed NO_2 concentration (top), wind speed (bottom left) and PBL height (bottom right). Observations are in black/grey; modelled values in red/orange. Bars show the 25th -75th quantile interval, while the median is displayed by the continuous line. The 25th, 50th, 75th, and 95th quantile of the whole campaign are reported too. Comparison of computed and observed boxplot time series evaluated at AirBase and meteorological sites, available over the Paris area. Hour is in UTC time.

List of tables

Table 1: Models involved in the study

Teams	Models with references	Model acronym in this study	Simulated periods
PSI/RSE	**CAMx** (ENVIRON, 2011)	CAMX	2006, 2007, 2008, 2009
INERIS	**CHIMERE** (Menut *et al*., 2013)	CHIM	2006, 2007, 2008, 2009
HZG	**CMAQ** (Byun *et al*., 2006; Matthias et al., 2008)	CMAQ	2006, 2007, 2008, 2009
MSC-W - Met.NO	**EMEP** (Simpson *et al*., 2012)	EMEP	2006, 2007, 2008, 2009
TNO	**LOTOS-EUROS** (Sauter *et al.*, 2014)	LOTO	2006, 2007, 2008, 2009
ENEA/ARIANET	**MINNI** (ARIANET, 2004)	MINNI	2006, 2007, 2008, 2009
FUB	**RCG** (Stern *et al*., 2006)	RCG	2008, 2009

Table 2: Synthetic description of models (part 1)

	EMEP	CHIMERE	LOTOS-EUROS	RCG	CMAQ	MINNI	CAMx
version	rv4.1.3	Chimere2013	v1.8	v2.1	V4.7.1	FARM V3.1.12	V5.40
			VERTICAL MODEL STRUCTURE				
Vertical layers	20 sigma	9 sigma	4 (3 dynamic layers and a surface layer)	6 fixed terrain following layers	30 sigma	16 fixed terrain-following layers	33 sigma
Vertical extent (hpa or m)	100 hPa	500 hPa	3500 m	3000 m	100 hPa	10000 m	8000 m
First layer depth	90 m	20 m	25 m	25 m	42 m	40 m	20 m
Correction of first level concentration	Yes	No	Yes	No	No	No	No
			NATURAL EMISSIONS				
Biogenic VOC	Based upon maps of 115 species from Koeble and Seufert (2001), and hourly temperature and light. See Simpson *et al.* (2012)	MEGAN model v2.04	Based upon maps of 115 species from Koeble and Seufert (2001), and hourly temperature and light (Beltman *et al.*, 2013)	Based upon maps of 115 species from Koeble and Seufert (2001), and hourly temperature and light.using emissions factors of Simpson *et al.* (1999)	BEIS 3.14 emission inventory (Vukovich and Pierce, 2002)	MEGAN model v2.04	MEGAN model v2.1
Soil NO	After Simpson *et al.* (2012)	MEGAN model v2.04	Not used here	From Simpson *et al.* (1999)	BEIS 3.1.4	MEGAN v2.04	MEGAN model v2.1
Lightning emissions	Climatological fields, Köhler *et al.* (1995)	No	No	No	No	No	No
Sea salt	Monahan (1986) and Martensson (2003), see Tsyro *et al.* (2011).	Monahan *et al.* (1986)	Martensson *et al.* (2003) and Monahan *et al.* (1986)	Gong *et al.* (1997) and Monahan *et al.* (1986)	Zhang *et al.* (2005) and Clarke *et al.* (2006)	Zhang *et al.* (2005)	Not used
Windblown Dust	After Simpson *et al.* (2012)	No	Denier van der Gon *et al.* (2009).	Loosemore and Hunt (2000), Claiborn *et al.* (1998)	No	Vautard *et al.* (2005)	No
Road traffic suspension	Denier van der Gon *et al.* (2009).	No	No	No	No	No	No
			LANDUSE				
Landuse database	CCE/SEI for Europe, elsewhere GLC2000	GLOBCOVER (24 classes)	Corine Land Cover 2000 (13 classes)	Corine Land Cover 2000 (13 classes)	Corine Land Cover 2006 (44 classes)	Corine Land Cover 2006 (22 classes)	USGS data
Resolution	Flexible, CCE/SEI ~ 5 km	About 300 m	1/60 x 1/60 degrees	1/60 x 1/60 degrees	About 250 m	About 250 m	10 minutes

Table 3: Synthetic description of models (part 2)

	EMEP	CHIM	LOTO	RCG	CMAQ	MINNI	CAMX
				METEOROLOGY			
Driver	ECMWF IFS	ECMWF IFS + urban mixing	ECMWF IFS	ECMWF IFS + Observations	COSMO CLM	ECMWF IFS	ECMWF IFS
Resolution	0.25°x0.25°	0.25°x0.25°	0.25°x0.25°	0.25°x0.25°	24 km x 24 km (Lambert Conformal Conic Projection)	0.25°x0.25°	0.25°x0.25°
				PROCESSES			
Advection scheme	Bott (1989a,b)	Van Leer (1984)	Walcek (2000)	Walcek (2000) modified by Yamartino (2003).	Blackman cubic polynomials (Yamartino, 1993	Blackman cubic polynomials (Yamartino, 1993)	Bott (1989a,b)
Vertical diffusion	Kz approach following O'Brien (1970) and on Jeričevič *et al.* (2010) for stable and neutral conditions	Kz approach following (Troen and Mart, 1986) IFS PBL	Kz approach IFS PBL	Kz-approach and IFS PBL	ACM2 PBL scheme (Pleim, 2007a)	Kz following Lange (1989). PBL from Maul (1980) version of Carson (1973) algorithm for day times.	Kz approach following O'Brien (1970) IFS PBL
Dry deposition scheme	resistance approach for gases, Venkatram and Pleim (1999) for aerosols, Simpson *et al.* (2012)	resistance approach Emberson (2000a,b)	Resistance approach,DEPAC3.11for gases, Van Zanten *et al.* (2010) and Zhang et al (2001) for aerosols	resistance approach, DEPAC-module	Resistance approach, Venkatram and Pleim (1999)	Resistance model (Walcek and Taylor, 1986; Wesely, 1989)	Resistance model for gases (Zhang *et al.*,2003) and aerosols (Zhang *et al*., 2001)
Compensation points	No, but zero NH_3 deposition over growing crops	No	Only for NH_3 (for stomatal, external leaf surface and soil = 0)	No	No	No	No
Stomatal resistance	DO3SE-EMEP: Emberson *et al.* (2000a,b), Tuovinen *et al.* (2004), Simpson *et al.* (2012)	Emberson (2000a,b)	Emberson (2000a,b)	Wesely (1989)	Wesely (1989)	Wesely (1989)	Wesely (1989)
Wet deposition of gases	In-cloud and sub-cloud scavenging coefficients	In-cloud and sub-cloud scavenging coefficients	sub-cloud scavenging coefficient	pH dependent scavenging coefficients	In-cloud and sub-cloud scavenging which depends on Henry's law constants, dissociation constants and cloud water pH. Chang *et al.* (1987)	In-cloud and sub-cloud scavenging coefficients (EMEP, 2003)	In-cloud and sub-cloud scavenging model for gases and aerosols (Seinfeld and Pandis, 1998)
Wet deposition of particles	In-cloud and sub-cloud scavenging	In-cloud and sub-cloud scavenging	Sub-cloud scavenging coefficient	Sub-cloud scavenging coefficients	In-cloud and sub-cloud scavenging	In-cloud and sub-cloud scavenging coefficients	In-cloud and sub-cloud scavenging model for gases and aerosols (Seinfeld and Pandis, 1998)
Gas phase chemistry	EmChem09 (Simpson *et al*.	MELCHIOR	TNO CBM-IV	CBM-IV	CB-05 with chlorine chemistry	SAPRC99 (Carter, 2000a,b)	CB-05 (Yarwood *et al.*, 2005)

	2012)				extensions (Yarwood *et al.*, 2005)		
Cloud chemistry	Aqueous SO_2 chemistry	Aqueous SO_2 chemistry and ph computation	No	Simplified aqueous SO_2 chemistry	Aqueous SO_2 chemistry (Walcek and Taylor, 1986)	Aqueous SO_2 chemistry (Seinfeld and Pandis, 1998)	Aqueous SO_2 chemistry RADM-AQ (Chang *et al.*, 1987)
Coarse nitrate	Yes	No reactions with Ca or Na but coarse might exist with transfer from smaller particles	Yes	Yes	No	No	No
Secondary Inorganic equilibrium	MARS (Binkowski and Shankar,1995)	ISORROPIA (Nenes *et al.*, 1999)	ISORROPIA v.2	ISORROPIA	ISORROPIAv1.7	ISORROPIA v1.7 (Nenes *et al.*, 1998)	ISORROPIA (Nenes *et al.*, 1998)
SOA formation	VBS-NPAS – Simpson *et al.* (2012)	After Bessagnet *et al.* (2009)	Based on Bergström et al (2012)	SORGAM module (Schell *et al.*, 2001)	SORGAM module (Schell *et al.*, 2001)	SORGAM module (Schell *et al.*, 2001)	CAMx-VBS (beta version) (Koo *et al.*, 2014)
VBS	Yes, Bergström *et al* (2012), Simpson *et al.* (2012)	No	Yes, based on Bergström et al (2012)	No	No	No	Yes based on Koo *et al.* (2014)
Aerosol model	Bulk- approach (2 modes)	8 bins (40 nm to 10 µm)	Bulk- approach (2 modes)	Bulk approach (2 modes)	AERO5 (Carlton *et al.*, 2010), Log-normal approach (3 modes)	AERO3 (Binkowski, 1999); 3 modes: Aitken, accumulation, coarse	Bulk- approach (2 modes)
Aerosol physics	No dynamics	Coagulation/condensation/nucleation	No dynamics	No dynamics	Coagulation/condensation/nucleation	Coagulation/condensation/nucleation	No dynamics

Table 4: Error statistics used to evaluate model performance (M and O refer respectively with Model and Observations data, and N is the number of observations)

Mean Bias	$(\bar{M} - \bar{O})$ with $\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i$ and $\bar{O} = \frac{1}{N}\sum_{i=1}^{N} O_i$
Normalised Mean Bias	$NMB = (\bar{M} - \bar{O})/\bar{O}$
Mean Bias	$MB = (\bar{M} - \bar{O})$
Mean Gross Error	$MGE = \frac{1}{N}\sum_{i=1}^{N} \lvert M_i - O_i \rvert$
Standard Deviation	$SD_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$ with X=O or M
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}$
Normalized Root Mean Square Error	$NMSE = \frac{1}{\bar{M}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}$
Correlation Coefficient	$R = \left(\sum_{i=1}^{N}(M_i - \bar{M})(O_i - \bar{O})\right) \Big/ \left(\sqrt{\sum_{i=1}^{N}(M_i - \bar{M})^2 \times \sum_{i=1}^{N}(O_i - \bar{O})^2}\right)$

Table 5: PM10 and PM2.5 spatial correlations for all campaigns

	2006		**2007**		**2008**		**2009**	
	PM10	***PM2.5***	**PM10**	***PM2.5***	**PM10**	***PM2.5***	**PM10**	***PM2.5***
CAMx	0.58	*0.32*	0.24	*0.60*	0.32	*0.47*	0.07	*0.46*
CHIMERE	0.65	*0.32*	0.58	*0.78*	0.39	*0.42*	0.55	*0.66*
CMAQ	0.50	*0.19*	0.50	*0.80*	0.11	*0.42*	0.11	*0.37*
EMEP	0.75	*0.24*	0.56	*0.62*	0.34	*0.48*	0.68	*0.61*
LOTOS-EUROS	0.34	*0.05*	0.50	*0.61*	0.27	*0.37*	0.50	*0.37*
MINNI	0.61	*0.43*	0.55	*0.58*	0.20	*0.45*	0.32	*0.51*
RCG	ND	*ND*	ND	*ND*	0.62	*0.32*	0.44	*0.36*