Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

Evaluation and Error Apportionment of an Ensemble of Atmospheric Chemistry Transport Modelling Systems: Multi-variable Temporal and Spatial Breakdown

Efisio Solazzo[1*], Roberto Bianconi[2], Christian Hogrefe[3], Gabriele Curci[4,5], Ummugulsum Alyuz[6], Alessandra Balzarini[7], Rocío Baró[8], Roberto Bellasio[2], Johannes Bieser[9], Jørgen Brandt[10], Jesper H. Christensen[10], Augistin Colette[11], Xavier Francis[12], Andrea Fraser[13], Marta Garcia Vivanco[11,14], Pedro Jiménez-Guerrero[8], Ulas Im[10], Astrid Manders[15], Uarporn Nopmongcol[16], Nutthida Kitwiroon [17], Guido Pirovano[7], Luca Pozzoli[6,1], Marje Prank[18], Ranjeet S. Sokhi[12], Paolo Tuccella[5], Alper Unal[6], Greg Yarwood[16], Stefano Galmarini[1]

1 European Commission, Joint Research Centre (JRC), Directorate for Energy, Transport and Climate, Air and Climate Unit, Ispra (VA), Italy
2 Enviroware srl, Concorezzo, MB, Italy
3 Atmospheric Model Application and Analysis Branch - Computational Exposure Division - NERL, ORD, U.S. EPA
4 CETEMPS, University of L'Aquila, Italy
5 Dept. Physical and Chemical Sciences, University of L'Aquila, Italy
6 Eurasia Institute of Earth Sciences, Istanbul Technical University, Turkey
7 Ricerca sul Sistema Energetico (RSE SpA), Milano, Italy
8 University of Murcia, Department of Physics, Physics of the Earth, Campus de Espinardo, Ed. CIOyN, 30100 Murcia, Spain
9 Institute of Coastal Research, Chemistry Transport Modelling Group, Helmholtz-Zentrum Geesthacht, Germany
10 Aarhus University, Department of Environmental Science, Frederiksborgvej 399, 4000 Roskilde, Denmark
11 INERIS, Institut National de l'Environnement Industriel et des Risques, Parc Alata, 60550 Verneuil-en-Halatte, France
12 Centre for Atmospheric and Instrumentation Research (CAIR), University of Hertfordshire, Hatfiled, UK
13 Ricardo Energy & Environment, Gemini Building, Fermi Avenue, Harwell, Oxon, OX11 0QR, UK
14 CIEMAT. Avda. Complutense, 40. 28040. Madrid, Spain
15 Netherlands Organization for Applied Scientific Research (TNO), Utrecht, The Netherlands
16 Ramboll Environ, 773 San Marin Drive, Suite 2115, Novato, CA 94998, USA
17 Environmental Research Group, Kings' College London, London, United Kingdom
18 Finnish Meteorological Institute, Atmospheric Composition Research Unit, Helsinki, Finland

*Author for correspondence: E.Solazzo, efisio.solazzo@jrc.ec.europa.eu, Phone: +390332789944

Abstract. Through the comparison of several regional-scale chemistry transport modelling systems that simulate meteorology and air quality over the European and American continents, this study aims at *i)* apportioning the error to the responsible processes using time-scale analysis, *ii)* helping to detect causes of models error, and *iii)* identifying the processes and scales most urgently requiring dedicated investigations.

The analysis is conducted within the framework of the third phase of the Air Quality Model Evaluation International Initiative (AQMEII) and tackles model performance gauging through measurement-to-model comparison, error decomposition and time series analysis of the models biases for several fields (ozone, CO, SO_2, NO, NO_2, PM_{10}, $PM_{2.5}$, wind speed, and temperature). The operational metrics (magnitude of the error, sign of the bias, associativity) provide an overall sense of model strengths and deficiencies, while apportioning the error to its constituent parts (bias, variance and covariance) can help to assess the nature and quality of the error. Each of the error components is analysed independently and apportioned to specific processes based on the corresponding timescale (long scale, synoptic, diurnal, and intra-day) using the *error apportionment* technique devised in the former phases of AQMEII.

The application of the error apportionment method to the AQMEII Phase 3 simulations provides several key insights. In addition to reaffirming the strong impact of model inputs (emissions and boundary conditions) and poor representation of the stable boundary layer on model bias, results also highlighted the high inter-dependencies among meteorological and chemical variables, as well as among their errors. This indicates that

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

the evaluation of air quality model performance for individual pollutants needs to be supported by complementary analysis of meteorological fields and chemical precursors to provide results that are more insightful from a model development perspective. The error embedded in the emissions is dominant for primary species (CO, PM, NO) and largely outweighs the error from any other source. The uncertainty in meteorological fields is most relevant to ozone. Some further aspects emerged whose interpretation requires additional consideration, such as, among others, the uniformity of the synoptic error being region and model-independent, observed for several pollutants; the source of unexplained variance for the diurnal component; and the type of error caused by deposition and at which scale.

1. Introduction

The Air Quality Model Evaluation International Initiative (AQMEII, Rao et al., 2010) has been active since 2008 with the aim of promoting the research on regional air quality model evaluation across the modelling communities of Europe and North America. It is coordinated by the European Joint Research Centre (JRC) and the U.S. Environmental Protection Agency (EPA) and it has now reached its third phase, referred to as AQMEII3 hereafter. The experience gathered in the first two phases consisted of important advancement in the model evaluation research as well as establishing a large community of participating regional modeling groups, and have made AQMEII a natural candidate to collaborate with the Hemispheric Transport of Air Pollution (HTAP) initiative. HTAP, a taskforce of the Long Range Transport of Air Pollution program (LTRAP) acting within the UNECE program, relies on a community of global scale chemical transport models to investigate the fate of air pollutants emitted in the Northern hemisphere and determine the contribution of remote sources as well as their impacts to the background concentration in different parts of the globe. HTAP is in its second phase and the activities undertaken during this second phase include coordinating simulations by both global and regional scale models. The regions of interest in the Northern hemisphere are North America, Europe and South East Asia. The regional-scale modelling component of this activity for Europe and North America is being coordinated by AQMEII while the Asian component is being coordinated by MICs-ASIA (Model Intercomparison Study-Asia). Global models participating in HTAP are used by the AQMEII regional models as boundary conditions and special attention has been given to the emission inventory to ensure that it is consistent between the global and regional-scale simulations as described in Janssens-Maenhout et al. (2015). The activity described here relates to the evaluation of the base case scenario set up within the context of HTAP and AQMEII (a description of the HTAP program can be found at www.htap.org).

Following the simulation strategy developed over the first two phases of the AQMEII activity, two continental-scale domains have been used in the exercise - one over Europe (EU) and one over North America (NA) (Figure 1). The modelling groups participating in AQMEII3 performed air quality (AQ) simulations over one or both of these domains. Each group has been provided the same inputs for anthropogenic emissions and boundary conditions and has been left the choice of the optimal configuration of the modelling systems, including meteorology, grid spacing, and natural emissions. To facilitate the cross-comparison among models, the modelled outputs have been successively interpolated to a common regular grid of 0.25° spacing over both continents. The comparison with observational data is performed by interpolating (or by simply taking the value from the grid cell where the monitoring sites are situated) the model values to prescribed observation stations (receptors) for surface measurements and at specified vertical heights for comparisons against measured profiles. As in the previous two phases of AQMEII, the ENSEMBLE system (Galmarini et al., 2012) hosted by the JRC has been used to accommodate all of the data and to pair modelled to observational values in time and space to provide direct comparison and statistical analysis.

The model evaluation approach proposed and applied in this study combines aspects of operational and diagnostic evaluation as defined by Dennis et al. (2010). It makes use of the classical statistical indicators typically employed for operational evaluation based on the direct comparison with observations, but also provides more indications on the processes contributing to model errors, which is the focus of diagnostic

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

model evaluation (Solazzo and Galmarini, 2016). The data used in the analysis are not process specific but are ordinary time series of modelled and monitoring data which are decomposed into four spectral components: ID (intra-day), DU (diurnal), SY (synoptic), and LT (long-term), each determined by different physical and chemical processes (Rao et al., 1997). The error apportionment applied to each spectral component can provide indications on the possible sources of error. The scope, as also highlighted by Gupta et al. (2009), is to move beyond the usual aggregate metrics that only offer a statistical interpretation, towards the use of measures selected for the quality of the information they can provide to model developers and users.

The evaluation of the AQMEII3 suite of model runs is carried out for surface temperature (Temp) and wind speed (WS), and for the species CO, NO, NO_2, ozone, SO_2, PM_{10} (EU) and $PM_{2.5}$ (NA). Additional analyses making use of emission reduction scenarios (CO and NO) and vertical profiles (Temp, WS, ozone) are also presented.

The main scope of the analysis is to present a detailed overview of the skill of AQ models when compared against measurements, for several regulatory pollutants and their precursors. For each species, the error is

1. quantified seasonally for three sub-regions of each continent;
2. qualified in terms of bias, variance, or covariance type of error, and
3. apportioned to the atmospheric time-scale, i.e. ID, DU, SY, or LT.

Given the large amount of models and species for two continents and the screening scopes of this work, maps of model metrics at individual receptors are omitted. Instead, spatial averaging over pre-selected homogenous sets of measurement points is presented. Investigation of signal associativity through clustering analysis has been performed for ozone and PM (PM_{10} for EU and $PM_{2.5}$ for NA) over both continents following the procedure outlined by Solazzo and Galmarini (2015), allowing the detection of three sub-regions (hereafter referred to as EU1, EU2, EU3 and NA1, NA2 NA3) (Figure 1) where the LT and SY components have shown robust clustering features. For consistency and to facilitate the interpretation of the results, the same sub-regions have been adopted for all species.

The error break-down, the time series decomposition, and the models and observational data used are presented in Section 2. In Section 3, the results of the error apportionment analysis are presented and discussed. A novel analysis based on the autocorrelation function (acf) of the LT component is presented in Section 4 for ozone. Conclusions are drawn in Section 5.

2. Methodology

The first step of the analysis is the spectral decomposition of the time series of modelled and observed species, as outlined in the methodology proposed in Solazzo and Galmarini (2016). Because each spectral component represents a range of processes in a specific spectral range, the deviation of the modelled from the observed spectral component is informative about the process(es) causing the error. The second step is to separate the mean square error (MSE) of each spectral component into its constituent parts: the bias, variance and covariance. These time-scale specific errors, expressed in terms of bias, variance, and covariance then allow a more precise diagnosis of their cause.

2.1 Error break down

The MSE is the squared difference of the modelled and observed values:

$$MSE = E(mod - obs)^2 = \frac{\sum_{i=1}^{nt}(mod_i - obs_i)^2}{n_t} \quad \text{EQ 1}$$

where E(·) denotes expectation and n_t is the length of the time series. The bias is:

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

$$bias = E(mod - obs) \quad \text{EQ 2}$$

i.e., $bias = \overline{mod} - \overline{obs}$ (the overbar indicates temporal averaging). The following relationship holds:

$$MSE = var(mod - obs) + bias^2 \quad \text{EQ 3}$$

(var(·) is the variance operator). By applying known the known property of the variance for correlated fields:

$$var(mod - obs) = var(mod) + var(obs) - 2cov(mod, obs) \quad \text{EQ 4}$$

the MSE can be expressed as:

$$MSE = bias^2 + var(mod) + var(obs) - 2cov(mod, obs), \quad \text{EQ 5}$$

where the covariance term (last term on the right hand side of Eq 5) accounts for the degree of correlation between the modelled and observed time series. Following Solazzo and Galmarini (2016), the MSE Eq 5 is rewritten as:

$$MSE = \left(\overline{mod} - \overline{obs}\right)^2 + (\sigma_{mod} - r\sigma_{obs})^2 + mMSE \quad \text{EQ 6}$$

where

$$mMSE = \sigma_{obs}^2(1 - r^2) \quad \text{EQ 7}$$

is the minimum error achievable by an accurate (unbiased, $\overline{mod} = \overline{obs}$) and precise ($\sigma_{mod} = \sigma_{obs}$) modelling system (*r* is the linear correlation coefficient). *mMSE* is the unexplained portion of the error and reflects the amount of observed variance not explained by the models (Solazzo and Galmarini, 2016). The *mMSE* type of error is caused by the variability of the observation not reproduced by the models, which includes incommensurability, noise, and timing of the signal summarised by the coefficient of determination (Solazzo and Galmarini, 2016), as well as by the error induced by the meteorological drivers (for primary and secondary species) and by the short and long range transport of precursors (for secondary species such as ozone)).

The decomposition in Eq 6 includes all the operational metrics commonly adopted to evaluate the AQ models (bias, variance, correlation coefficient, and their sum, the MSE), and is thus suitable to be used as compact estimator of model performance.

2.2. Spectral decomposition and error attribution

Spectral filtering has been applied to the measured and modelled hourly-averaged time series at the monitoring sites using the Kolmogorov-Zurbenko (kz) low-pass filter (Zurbenko, 1986). This allows to separate different phenomena having distinct signals, such as long-term and short-term fluctuations in the observed and modelled time series (Rao et al., 1997). Applications of the *kz* filter to ozone have been described in a number of previous studies (Rao et al., 1997; Wise and Comrie, 2005; Hogrefe et al., 2000; 2014; Galmarini et al., 2013; Kang et al., 2013; Solazzo and Galmarini, 2015 and 2016).

The kz filter depends on the length of the moving average window *m* and the number of iterations *k* ($kz_{m,k}$) (*k* also indicates the level of noise suppression). Since the kz is a low-pass filter, the filtered time series consists of the low-frequency component, while the difference between two filtered time series (with different *k* and *m*) provides a band-pass filter. This latter property has been used in this study, as well as in a number of previous studies, to decompose the modelled and observed time series as:

$$FT(S) = LT(S) + SY(S) + DU(S) + ID(S) \quad \text{EQ 8}$$

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

where *S* is the time series of the species being analysed and FT is the full (un-decomposed) time series.

The base line component LT is the long term component (periods longer than 21 days) and accounts for the temporal fluctuations determined by low frequencies, such as boundary conditions and seasonal variation in emissions and photo-chemistry. SY is the synoptic component containing fluctuations related to weather-processes and precursor emissions occurring on scales between 2.5 and 21 days. The DU (diurnal) component accounts for fluctuations due to diurnal periodicity occurring on temporal scales between 0.5 and 2.5 days, and ID is the intra-day component, accounting for fast-acting, local-level processes (time scale less than 12 hours) (the spectral components have the same units as the un-decomposed time series).

The decomposition Eq 8 is such that the un-decomposed time series is perfectly returned by the summation (or by the exponential product, see Appendix 1 for details) of the components. The band-pass nature of the SY, DU, and ID components is such that they only describe the processes in the time window the filter allows the signal to 'pass'. For instance, the DU component is insensitive to processes outside the range between 0.5 and 2.5 days.

Because the kz filter was originally developed to deal with ozone, the parameters *k* and *m* (Appendix 1) are specifically tailored for ozone, taking into consideration its chemistry and life-time. In this study we have applied the kz filter to other species and kept the same values for *k* and *m* for consistency and to facilitate the comparison of the results. Although some species (e.g. PM, CO, SO_2) may be less sensitive to day/night cycles than ozone, the distinction between DU and ID are still revealing of emission patterns like vehicular traffic and industrial activities as well as diurnal variations in vertical mixing. Moreover, the SY and LT are associated with transport and other weather processes common to all species.

Two aspects of the signal filtering having a profound impact on model evaluation are:

1. The non-orthogonality of the spectral components is one of the major drawbacks of the signal decomposition. A clear-cut separation of the components of Eq 8 is not achievable, since the separation is a non-linear function of the parameters *m* and *k* (Rao et al., 1997; Kang et al., 2013) and the leakage among components mixes together in each component different physical processes. Galmarini et al. (2013) found that the explained variance by the spectral components accounts for 75 to 80% of the total variance while the remaining portion of the variance is due to the interactions between the estimated components. The effect of these interactions on the error apportionment pursued in this study is outlined and quantified in section 3. Other spectral techniques could be used but either they not guarantee the absence of signal leakage (e.g. anomaly perturbation method) or require special treatment of missing data (e.g. wavelet transform method) (Rao et al., 1997; Eskridge et al., 1997).

2. The bias is calculated as the distance between the time average modelled and observed time series. In such a 'time average' sense, the base line LT is the only biased component, containing the entire bias of the original time series. The other components are zero-mean fluctuations about LT and are unbiased. Although inaccuracy at each time step can also derive from the SY, DU and ID components (Johnson, 2008), in this study the signal is taken as time-averaged over a finite period, and therefore the entire bias is apportioned to the base-line (LT) component.

2.3 Models and observational Data

Table 1 summarises the modelling systems participating in AQMEII3. Twelve modelling groups produced outputs over EU and four over NA (although not all fields were made available by all groups). Sensitivity simulations performed by two groups, in which alternate emission inventories were used, raises the number of EU contributions to fourteen.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

The 'standard' emission inventories are those developed for the second phase of AQMEII for EU and NA and extensively described in Pouliot et al. (2015). For EU, the 2009 inventory of anthropogenic emissions was used, although biogenic emissions (meteorology-dependent) were specifically calculated for the year of 2010 by several groups. In regions not covered by the standard inventory, such as North Africa, five modelling systems (Table 1) have complemented the standard inventory with the HTAPv2.2 (Janssens-Maenhout et al., 2015) datasets. The two inventories are the same over EU and in the MACC inventory the non-European emissions are not included. Other small differences might exist among the two inventories (like in the shipping emissions), but we consider them to be of small impact for the spatial averaged analysis carried out in this study. Emissions from lightning and volcanic sources are not contained in the EU and NA emissions inventories, since not all participating models include robust methods for estimating these emissions.

Two EU modelling systems (CHIMERE, SILAM) made results available with both inventories. For both continents the regional scale emission inventories where embedded in the global scale inventory (Janssens-Maenhout et al., 2015) to guarantee coherence and harmonization of the information used by the two communities. The ability of some modelling groups to perform sensitivity simulations with both the TNO MAC and the HTAP v2.2 information allowed also to determine the impact of North African emissions on the European domain. For Chimere, the MACC inventory over France and the UK was spatially redistributed considering national inventories (having higher spatial resolution), while for the other countries it was redistributed by considering point source locations, land-use and population. For processing the HTAP inventory, population was not used as a parameter for spatially distributing the emissions.

For the NA domain, the 2008 National Emission Inventory was used as the basis for the 2010 emissions, providing the inputs and datasets for processing with the SMOKE emissions processing system (Mason et al., 2012). Year specific updates for the year of 2010 were made for several sectors, including mobile sources, power plants, wildfires, and biogenic emissions. Additional details can be found in Im at al. (2015a,b) and Pouliot et al. (2015).

Chemical boundary conditions were provided by the Composition – Integrated Forecast System (C-IFS) model (Flemming et al., 2015), including ozone, NO_x, CO, CH_4, SO_2, NMVOCs, dust, organic matter, black carbon and sulphate. Sea salt at the boundaries, although provided, was not used due to unrealistically high values.

[Table 1 here]

2.3.1 Model Features

This section presents the main features of the modelling systems participating to AQMEII3. Complementary information is provided in Table 1.

The FMI (Finnish Meteorological Institute) has taken part with the ECMWF-SILAM system (ECMWF-SILAM_M and ECMWF-SILAM_H of Table 1, indicating the instances of the SILAM model using the MACC and the HTAP emission inventory, respectively) (ECMWF: European Centre for Medium-Range Weather Forecasts). SILAM v5.4 (Sofiev et al., 2015) has been used, with meteorological input extracted from the ECMWF operational archives. The thickness of the first layer is 30m. The simulation included sea-salt emissions as in Sofiev et al. (2011) (but not from the boundaries), biogenic VOC (volatile organic compounds) emissions as in Poupkou et al. (2010) and wild-land fire emissions as in Soares et al. (2015). The wind-blown dust is only included from the lateral boundary conditions. Anthropogenic NO_x emissions have been treated as 10% NO_2 and 90% NO. The volatility distribution of anthropogenic OC was taken from Shrivastava et al. (2011). The gas phase chemistry was simulated with CBM-IV, with reaction rates updated according to the recommendations of IUPAC (http://iupac.pole-ether.fr) and JPL (http://jpldataeval.jpl.nasa.gov). The secondary inorganic aerosol formation was computed with updated DMAT scheme (Sofiev, 2000) and secondary organic aerosol formation

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

with the Volatility Basis Set (VBS, Ahmadov et al., 2012). A known deficiency of the SILAM version used in this study is the overestimation of ozone dry deposition.

The LOTOS-EUROS modelling system (Schaap et al. 2008, Sauter et al. 2012) has been applied by TNO (the Netherlands Organization for Applied Scientific Research), using version v1.10.1. The meteorological inputs have been extracted from the ECMWF operational archives. For biogenic emissions the approach as described in Beltman et al. (2013) has been used. Gas-phase chemistry is based on CBM-IV (modified reaction rates, see Sauter et al., 2012), secondary inorganic aerosol (SIA) formation on Isorropia II (Fountoukis and Nenes, 2009) and for semivolatile species the VBS approach was used (Donahue et al. 2006, Bergström et al. 2012), with 100% of the emitted OC mass in the 4 lowest volatility classes that are predominantly solid and an additional 150% in the five higher volatility bins. Modelled terpene emissions were reduced by 50% to limit their contribution to SOA (secondary organic aerosol) formation which was found to be too high otherwise. This is justified since contributions of terpene to SOA formation is known to be very uncertain and at the same time the model is very sensitive to terpene emissions (Bergström et al., 2012). 3% of the total anthropogenic NO_x emissions were attributed to NO_2 while 97% were attributed to NO. No NO_x emissions from soil were taken into account. The model includes pH dependent conversion rates for SO_2 (Banzhaf et al., 2012), while only below-cloud scavenging is used for wet deposition. Mineral dust emissions were calculated on-line, including emissions from road resuspension and agricultural activities, according to Schaap et al. (2009). For sea spray the parameterizations by Monahan et al. (1986) and Martensson et al. (2003) were used. A specific feature of LOTOS-EUROS is that it only covers the lower 3.5 km of the atmosphere, with a static 25 m surface layer, a dynamic mixing layer and two dynamic reservoir layers. This makes the model relatively fast in terms of computation time but has implications for the vertical mixing of species for instances where the mixing layer rapidly changes in height.

WRF-WRF/Chem1 is applied by the University of L'Aquila (Italy). The version 3.6 of the Weather Research and Forecasting model with Chemistry model (WRF/Chem) (Grell et al., 2005) has been used for AQMEII3. This version of the model has been modified to include the new chemistry option implemented by Tuccella et al. (2015) that includes in the simulation of direct and indirect aerosol effects a better representation of the secondary organic aerosol mass, calculated as in Ahmadov et al. (2012). Here only direct effects have been included in the simulation, for computational expediency. The model uses RACM-ESRL gas phase chemical mechanism (Kim et al., 2009), an updated version of the Regional Atmospheric Chemistry Mechanism (RACM) (Stockwell et al., 1997). The inorganic aerosols are treated with the Modal Aerosol Dynamics Model for Europe (MADE) (Ackermann et al., 1998). The parameterization for SOA production is based on the VBS approach. The aerosol direct and semi direct effects are taken in account following Fast et al. (2006). Cloud chemistry in the convective updraft is modelled using the scheme of Walcek and Taylor (1986), while the aqueous phase oxidation of SO_2 by H_2O_2 in the grid-resolved clouds is parameterized with the scheme used in GOCART (Goddard Chemistry Aerosol Radiation and Transport). Wet deposition from convective and resolved precipitation is included following Grell and Freitas (2014). The photolysis frequencies are calculated with the Fast-J scheme (Fast et al., 2006), the dry deposition velocities are simulated with the parameterization developed by Wesely (1989). Dry deposition and photolysis schemes were modified to take in account the effects of the soil snow coverage following Ahmadov et al. (2015). The anthropogenic emissions are taken from TNO-MACC inventory for 2009 (Kuenen et al., 2014) and have been adapted to the chemical mechanism used following the method of Tuccella et al. (2012). The biogenic emissions have been calculated online by using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2006). Anthropogenic NO_x sources were assumed 95% of NO and 5% of NO_2. The main physical parameterization used include the Rapid Radiative Transfer Method for Global (RRTMG) for solar and infrared radiation (Iacono et al. 2008), Morrison microphysics (Morrison et al., 2010), the Mellor-Yamada Nakanishi-Niino (MYNN) planetary boundary layer (PBL) scheme (Nakanishi-Niino, 2006), the NOAH land-surface model (Chen and Dudhia, 2001) and the Grell-Freitas scheme for cumulus clouds (Grell and Freitas, 2014). The meteorological analysis used to initialize WRF are provided by the ECMWF with a horizontal resolution of 0.5° every 6 hours. Chemical

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

boundary conditions are taken from C-IFS. A series of 72-hour simulations has been performed on each day starting at 00 UTC. Each run is preceded by a pre-forecast of 12 hours (from 12 to 00 UTC) only with meteorology, in which the model is nudged toward analysis above the PBL in order to prevent a drift from synoptic circulation patterns. The last hour of this spin-up is then used as meteorological initial condition for WRF/Chem. The chemical state is restarted from the previous 72-hours run.

WRF-WRF/Chem2 applied by the University of Murcia (Spain) relies on the WRF-Chem model (Grell et al., 2005). The following physics options have been applied for the simulations: RRTMG long-wave and short-wave radiation scheme; Lin microphysics (Lin et al., 1993), the Yonsei University (YSU) PBL scheme (Hong et al., 2006), the NOAH land-surface model and the updated version of the Grell-Devenyi scheme (Grell and Devenyi, 2002) with radiative feedback. Chemical options include: RADM2 chemical mechanism (Stockwell et al., 1990); MADE/SORGAM aerosol module (Schell et al., 2001) including some aqueous reactions; Fast-J photolysis scheme. The modelling domain covers Europe and a portion of Northern Africa.

Simulations of WRF-CAMx over EU have been performed by RSE (Italy) using CAMx version 6.10 (Environ, 2014) with Carbon Bond 2005 (CB05) gas phase chemistry (Yarwood et al., 2005) and the Coarse-Fine (CF) aerosol module. Input meteorological data were generated by WRF-Chem model version 3.4.1 (Skamarock et al., 2008a,b), driven by ECMWF analysis fields. Grid nudging of wind speed, temperature and water vapour mixing ratio has been employed within the PBL, with a nudging coefficient of 0.0003 sec-1. WRF-Chem has been adopted to predict GOCART dust emissions (Ginoux et al., 2001) along with the meteorology. The WRFCAMx pre-processor (version 4.2; ENVIRON, 2014) was used to create CAMx ready input files collapsing the 33 vertical layers used by WRF to 14 layers in CAMx but keeping identical the layers up to 230 m above ground level. Anthropogenic emissions were derived by the TNO-MACC data applying a NO_2/NO_x ratio of 5% for each emission category. Biogenic VOC emissions were computed by applying the MEGAN emission model v2.04. Sea salt emissions were computed using published algorithms (de Leeuw et al., 2000; Gong, 2003).

Aarhus University (Denmark) applied the WRF-DEHM modelling system over EU and NA. The DEHM model used anthropogenic emissions from the EDGAR-HTAP database and biogenic emissions are calculated using the MEGAN model. The gas-phase chemistry module includes 58 chemical species, 9 primary particles and 122 chemical reactions (Brandt et al., 2012). Secondary organic aerosols (SOA) are calculated following the two-product approach assuming that hydrocarbons undergo oxidation through O_3, OH and NO_3 and for only two semi-volatile gas products (Zare et al., 2014). However, the module is simple as it does not include aging processes and further reactions in the gas and particulate phase (Zare et al., 2014). Other modelling options include the Noah Land Surface Model (Chen and Dudhia, 2001), Eta similarity surface layer (Janjic, 2002), the Mellor-Yamada-Janjic (Eta operational) boundary layer scheme (Mellor and Yamada, 1982), the Kain-Fritsch (Kain, 2004) scheme for cumulus parameterisation, the WRF Single-Moment 5-class Microphysics scheme (Hong et al., 2004), and the CAM scheme for both long and short radiation (Collins et al., 2004).

WRF-CMAQ1 has been applied by the ITU (Istanbul Technical University) over EU. The WRFv3.5 model has been used with the following physical options: WSM3 microphysics scheme (Hong et al., 2004), RRTM (long-wave radiation scheme, Dudhia shortwave radiation scheme (Dudhia, 1989), NOAH land surface model, Yonsei University PBL scheme and Kain–Fritsch cumulus parameterization scheme (KF2, Kain, 2004). The NCEP (National Centers for Environmental Prediction) FNL Operational Model Global Tropospheric Analyses has been used for boundary conditions and nudging the meteorological simulation. The MCIP version 3.6 (Otte and Pleim, 2010) has been used to process WRF output for CMAQ. The MEGANv2.1 (Guenther et al., 2012) model has been used to calculate the biogenic VOC emissions from vegetation, using surface temperature and radiation from MCIP output. CMAQv4.7.1 (Foley et al., 2010) was configured with the CB05 chemical mechanism and the AERO5 module (Foley et al., 2010) for the simulation of gas-phase chemistry and aerosol and aqueous chemistry, respectively. 95% of NO_x anthropogenic emissions were considered as NO.

The WRF-CMAQ2 system has been applied by Ricardo Energy & Environment (Ricardo-E&E) over EU. It has been configured using WRFv3.5.1 and CMAQ v5.0.2. The WRF model adopted the KF2 cumulus cloud parameterization and Morrison microphysics scheme (Morrison et al., 2009), the ACM2 (Asymmetric Convective Model version 2, Pleim, 2007) for the PBL, the Pleim-Xiu land-surface model (Xiu and Pleim, 2001), and the RRTMG radiative module. The NCEP FNL Operational Model Global Tropospheric Analyses has been used to generate boundary conditions for the European meteorological simulation. Nudging of temperature, wind speed, and water vapour mixing ratio has been applied above the PBL (Gilliam et al., 2012). The CMAQ model adopted the CB05-TUCL chemical mechanism (Whitten et al., 2010; Sarwar et al., 2011a), the AERO6 three mode aerosol module (Appel et al., 2013). The MCIP version 4.2 has been used to process WRF output for CMAQ. The MEGANv2.0.4 model has been used to calculate the biogenic VOC emissions from vegetation, using surface temperature and radiation from MCIP output. For road transport, 86% of NOx anthropogenic emissions were considered as NO and 95% of NOx anthropogenic emissions were considered as NO for all other emissions.

The WRF-CMAQ3 modelling system has been applied by the University of Hertfordshire and utilized the uncoupled version of the WRF-v3.4.1 model and CMAQ v5.0.2. The WRF simulations were performed using 18km x 18km horizontal grid resolution with 36 vertical sigma layers. The simulations used Unified Noah Land Surface Model as the land surface scheme, Pleim-Xiu Scheme for the surface layer, RRTMG as the long-wave and shortwave radiation scheme, Morrison 2-moment scheme for microphysics parameterization, KF2 scheme for cumulus parameterization, and ACM2 scheme for PBL parameterization. Meteorological initial and lateral boundary conditions were derived from the ECMWF analysis. In order to constrain the meteorological model towards the analyses a grid nudging technique was employed every 6 hours of WRF simulation. The results from WRF simulations were pre-processed for CMAQ using Meteorology-Chemistry Interface Process (MCIP) version 3.6 (Otte et al., 2005). In CMAQ model, the gas phase chemical mechanism was based on carbon bond chemical mechanism version 5 (Foley et al., 2010) with updated toluene and chlorine chemistry (CB05-TUCL) and the aerosol chemical reaction were treated with AERO6 module. The CMAQ model consisted of 35 vertical layers and extending up to ~16 km height with the thickness of lowest layer is approximately 20 m. The EDGAR HTAP V2 emissions (0.1^{o} x 0.1^{o}) as well as TNO emissions data (~7 km x 7 km) were used as anthropogenic area and point sources emission data respectively in CMAQ. The biogenic emissions were derived from MEGAN.

The WRF-CMAQ4 simulation has been performed by the Kings College (UK) using CMAQ v5.0.2 (Byun and Schere, 2006) with CB05 chemical mechanism that includes aqueous and aerosol chemistry. The CMAQ model is driven by meteorological fields from the WRF v3.4.1. The lateral boundary conditions for WRF are taken from the Global Forecast System (GFS) model with 6-hr interval and 1° grid resolution. The WRF physic schemes include RRTM radiation module KF2 cumulus parameterization, WSM6 microphysics (Hong and Lim, 2006), Pleim-Xiu surface layer scheme (Pleim and Xin, 2003), RUC land surface model (Benjamin et al., 2004), and ACM2 PBL scheme. The anthropogenic emissions for most part of the model domain are from MACC and the missing information have been filled with the emissions provided by EDGAR/HTAP. The biogenic emissions were estimated using the Biogenic Emission Inventory System version 3 (BEIS3) model in SMOKE v2.6 (https://www.cmascenter.org/smoke). The dust (Tong, et al, 2011) and sea-salt (Gantt et al., 2015) emissions are generated using CMAQ inline modules. The ratio for NO_2/NO_x emissions is ~10% (Bieser et al., 2011a).

The INERIS and CIEMAT institutes jointly applied the ECMWF-Chimere system. CHIMERE (version CHIMERE 2013) has been run for a 0.25x0.25 horizontal resolution and 9 vertical levels, extending up to 500 hPa with a first (lower)-layer depth of 20 m, using the meteorology provided by ECMWF IFS (Integrated Forecast System). Biogenic VOC emissions from vegetation and soil NO emissions have been calculated with the MEGAN model (version 2.04; Guenther et al., 2006, 2012). Sea salt emissions inside the domain have been calculated according to Monahan (1986). No sea salt condition was considered at the boundaries. The wind-blown dust is only included from the lateral boundary conditions. CHIMERE uses the MELCHIOR2 chemical mechanism (Lattuati, 1997) and ammonium nitrate equilibrium was calculated with ISORROPIA (Nenes et al., 1999). Dry

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

deposition is based on the resistance approach (Emberson 2000a,b) and both in-cloud and sub-cloud scavenging have been considered for wet deposition.

HZG has used the COSMO-CLM meteorological model to drive the CMAQ model. For AQMEII3 the CMAQ version 5.0.1 was used, with the CB05-TUCL scheme and the multi-pollutant aerosol module AERO6. CMAQ is run on a 24x24km² horizontal grid, using 30 vertical layers up to 50hPa (lowest layer of approximately 40m). CMAQ was run using the optional in-line calculation of dry deposition velocities. Wet deposition processes include in-cloud and sub-cloud scavenging processes. All atmospheric parameters were taken from regional atmospheric simulations with the COSMO-CLM (CCLM) mesoscale meteorological model (version 4.8) for the year 2010 (Geyer, 2014) using NCEP forcing data employing a spectral nudging method for large-scale effects (Kalnay et al., 1996). CCLM is the climate version of the regional scale meteorological community model COSMO (Rockel et al., 2008; Steppeler et al., 2003; Schaettler et al. 2008). CCLM uses the TERRA-ML land surface model (Schrodin and Heise, 2001), a TKE closure scheme for the PBL (Doms et al., 2011), cloud microphysics after Seifert and Beheng (2001), the Tiedtke scheme (Tiedtke, 1989) for cumulus clouds and a long wave radiation scheme following Ritter and Geleyn (1992). The meteorological fields were afterwards processed to match the 24x24km² CMAQ grid using the LM-MCIP pre-processor. The emission input for CCLM-CMAQ is based on the EDGAR HTAPv2 database, interpolated to the CMAQ model grid and aggregated following the SNAP emission sector nomenclature. Sector specific hourly temporal profiles and speciation factors of PM and VOC species were applied by the SMOKE for Europe emissions model (Bieser et al., 2011a). The temporal profiles used were fixed monthly, weekly, and diurnal profiles. NOx emissions were split using a NO/NO_2 ratio of 0.9/0.1 for mobile sources and a fixed ratio of 0.9/0.1 for all other source sectors. Biogenic emissions and NO emissions from soil were calculated using BEISv3.14. Sea-salt emissions are calculated in-line by CMAQ including sulphate emissions based on an average sulphate content of 7.7%. Finally, fixed vertical profiles were applied for each source sector (Bieser et al., 2011b).

The WRF-CMAQ system applied over NA by the US EPA (Environmental Protection Agency) has been configured using WRFv3.4 and CMAQv5.0.2 (Appel et al., 2013; see also Foley et al., 2010 and Byun and Schere, 2006). The options used in these WRF and CMAQ simulations are identical to those described in Hogrefe et al. (2015) except that the current simulations were performed in offline rather than two-way coupled mode. Temperature, wind speed, and water vapor mixing ratio were nudged above the PBL following the approach described in Gilliam et al. (2012). Soil temperature and moisture were nudged following Pleim and Xiu (2003) and Pleim and Gilliam (2009). The NO_2/NO_x split applied during SMOKE emissions processing varies for different categories. For many categories is the assumed split 90% NO / 10% NO_2, but for mobile sources the split varies for different types of vehicles and different emission processes.

Ramboll Environ used CAMx (version 6.2, Ramboll Environ, 2015) for simulations over NA, with CB05 chemical mechanism for gas-phase. The modeling domain covers the CONUS US with 459 by 299 grid cells of 12 by 12 km size and 26 vertical layers. Height of first layer is 20 m. Biogenic emissions were obtained from the MEGAN model version 2.1 (Guenther et al., 2006). Meteorological fields were produced by the US EPA (Environmental Protection Agency) using WRF model and reformatted using the WRFCAMx pre-processor to be readily used by the CAMx model.

2.3.2 Observational Data Used

The observational data used in this study is the same as the dataset used in second phase of AQMEII (Im et al., 2015a,b) and was derived from the surface air quality monitoring networks operating in EU and NA. In EU, surface data were provided by the European Monitoring and Evaluation Programme (EMEP; http://www.emep.int/) and the European Air Quality Database (AirBase; http://acm.eionet.europa.eu/databases/airbase/). In NA observational data were obtained from the NAtChem (Canadian National Atmospheric Chemistry) Database and from the Analysis Facility operated by Environment Canada (http://www.ec.gc.ca/natchem/). For the purposes of comparing the models against observations, only

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

stations with data completeness greater than 75% for the whole year and elevation above ground below 1000 m have been included in the analysis. Stations with continuous missing records for periods longer than 15 days have been removed from the dataset.

In addition, we also make use of vertical profiles of ozone, temperature and wind speed measured by ozonesondes. Ozonesonde data have been extracted from the World Meteorological Organization (WMO) World Ozone, and Ultraviolet Radiation Data Centre (Toronto, Canada) and made available to the AQMEII community. These measurements report vertical profiles of ozone at several vertical levels. Further details on these data are given in Solazzo et al. (2013).

Time-averaged statistics have been calculated after the spatial aggregation of the modelled and observed time series and prior to the spectral decomposition (the original time series have been spatially averaged first and then this spatial average time series has been spectrally decomposed). As a consequence of the spatial averaging, the relative importance of the ID component is likely reduced, since the ID fluctuations are highly variable in space (Hogrefe et al., 2014). Further, no land-use type filtering has been applied to the stations used for evaluation. While this choice has limited impact on the SY and LT components (Solazzo and Galmarini, 2015; Galmarini et al., 2013), the DU components of some species (such as PM, NO_x) might be strongly influenced by the vicinity of urban stations to emissions sources.

Details of the modelled regions and number of receptors are reported in Table 2.

[Table 2 here]

3. Results

The analyses presented in this section focus on evaluating the performance of the models. The accuracy of the spectral components is first analysed in terms of the root MSE and quantified on a seasonal basis. The season most affected by error is then further investigated by applying the error apportionment (Eq 6) to the spectral components. Results are presented for one sub-region only (EU2 and NA1 or NA2) in the main portion of the manuscript while results for the other sub-regions are included in the supplementary material.

The combination of the spectral decomposition and error apportionment approaches has the effect of neglecting the error associated with the cross components (twelve spectral interaction terms, see Solazzo and Galmarini (2016) for details) since the apportionment only deals with the error of the 'diagonal' components LT, DU, SY, ID. The reason is that while the contribution of the cross components to the overall error can be quantified, the associated time series needed to carry out the apportionment analysis cannot. The neglected part of the error is quantified in Table S1. In some instances, such portion can be as high as 20% of the total error for ozone.

Tables summarising the operational statistics (MB: Mean Bias; *r*: Pearson Correlation coefficient; RMSE: Root Mean Square Error) are reported in the Supplementary material and have been calculated using the 'openair' package (Carslaw and Ropkins, 2012).

3.1 Meteorological drivers: Temperature and Wind Speed

3.1.1 Near-surface Model Evaluation

The RMSE for surface temperature and wind speed is reported in Figure 2 (EU) and Figure 3 (NA). For EU (Figure 2a), the RMSE of the full (i.e. not spectrally decomposed and denoted as "FT" in the plots) time series of temperature for the entire year is, on a seasonal average, on the order of ~0.5-2K (but often exceeding 3K in EU3), with higher values typically occurring in spring and winter. The CHIMERE and SILAM models (both directly driven by the global meteorological fields provided by ECMWF) report the smallest error in EU1 and EU2, while the WRF/Chem2 model has the largest error in all sub-regions (up to ~5K for EU3 in summer) which

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

is largely caused by the unusually large error in the SY component when compared to other models. The RMSE of the LT component resembles the behaviour of the full time series, with the highest error in spring and winter (on average). The RMSE of the SY component is below ~2K (slightly higher in EU3) except for WRF/Chem2, whereas the DU component shows a more marked regional dependence, with the EU3 sub-region reporting, on average, approximately 50% higher seasonal error than the other two sub-regions, more pronounced in summer.

The bias is predominantly negative (model underestimation) for all EU models and sub-regions, except for WRF-CMAQ4 in EU3, where the model overestimates the measured temperature in summer and winter. Finally, the correlation coefficient is higher than 0.90 for the majority of models and spectral components (Table S2).

For NA (Figure 3a) the temperature RMSE of the WRF-DEHM and CCLM-CMAQ models (peaking in winter and autumn) is ~ 1-1.5K larger than the WRF-CMAQ model. The error of the SY component is of ~0.5K, while that of the DU component is significantly higher (between 0. 5K and 2K). The WRF-CMAQ model has a small bias (LT error small) so that the overall error is dominated by the error in the DU component. The bias is negative for the WRF-DEHM model in all sub-regions and has the same sign for CCLM-CMAQ and WRF-CMAQ, i.e. negative in spring and positive in the other seasons (although for NA2 and NA3 WRF-CMAQ reports a slightly negative bias also in winter) (Table S2).

The RMSE of the surface WS for EU shows large model-to-model variability, more markedly for the LT and SY components (all sub-regions, Figure 2b), whereas the error of the DU component is more evenly distributed across models (and significantly higher in EU3, where low-wind speed conditions are predominant). Although the meteorological fields are assimilated within the models (either from NCEP or from ECMWF, see Table 2), there are profound differences in the way these fields are ingested and interpolated to the model grid, as well as differences in the parameterisation of the boundary and surface layer which impact the modelled wind speed and temperature. For example, the two instances of WRF/Chem applied the assimilation of the meteorological fields (wind speed, temperature, and relative humidity) of global meteorological fields only above the PBL, whereas other models (e.g. WRF-CAMx) assimilated the global data also within the PBL. For the models directly driven by the global fields, (e.g SILAM, Chimere) the seasonal error for WS (~0.5-1 ms^{-1}) and temperature (0.4-1.2K) (Figure 2a,b) can be considered as the uppermost limit the accuracy of the models can achieve. Thus, the assimilation and interpolation methods errors (which are specific to the configuration of the meteorological model) can add up more than 1.5K and 2ms^{-1} to the total error.

The full WS time series of the WRF-DEHM, WRF/Chem1 and WRF/Chem2 models report the largest error (in excess of 1.5m/s), and the WRF-CAMx model even up to 2.4 m/s in winter (all sub-regions, Figure 2b). On average, the remaining models have an error of 0.5-0.7m/s. Most of the error is apportioned to the LT component, with the SY and DU below 0.3 m/s (except for WRF-CAMx and the other models mentioned above).

The WS bias is positive for all models (model over-prediction), for all seasons and sub-regions (only exception is the CCLM-CMAQ model, biased low during spring and summer in EU3 and WRF-CMAQ2 during summer in EU1). The correlation coefficient is above 0.9 for the majority of models and components (except for the models affected by large errors such as the WRF-CAMx model). In general, *r* is slightly lower in EU3, and is at maximum for the SY component (Table S3).

For NA (Figure 3b), the WRF-DEHM model reports an error of ~1-1.2 m/s during all seasons and sub-regions, while the error of the WRF-CMAQ model ranges between 0.45 and 0.75 m/s for all seasons and sub-regions. The error of the SY and DU components is small (below 0.3m/s for each season) for both models. Both models are biased high (all instances) and the correlation coefficient is in the order of ~0.9 or above (Table S3).

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

3.1.2 Vertical profiles

Vertical profiles of mean bias for Temp and WS are reported in Figure 4 to Figure 7. The modelled profiles have been evaluated using ozonesondes measurements. The frequency and local time of the launches are summarised in Table 3. The launches in EU predominately occurred during daylight hours, whereas for NA measurements are available also for night-time and late afternoon. The sign and magnitude of the bias are informative about error in the PBL processes, which will help the discussion on the error of the modelled pollutants (section 3.3).

The bias for temperature in EU ranges between -3K (CCLM-CMAQ at station 308, Figure 5) and +2K (WRF-CMAQ4 at station 308 and SILAM at station 156) at the surface. In most cases the temperature bias profiles fluctuate around zero (station 053, located between EU1 and EU2; station 043; station 242 in EU2, and partially station 316 in EU2), whereas for some stations the bias keeps the same sign throughout the troposphere, negative for station 156 (launches at 10-12 LT) and positive for station 099 (early morning launches). The difference in altitudes (491 m asl the former and 1000 m asl the latter) and the complex terrain of the alpine region might also be factors for the large model differences at these two (relatively close) stations.

Vertical profiles of Temp in NA (Figure 6) shows strong surface bias (negative) at station 021 and 457 (both close to the western border of the domain), for both models. At station 021 (data collected under daylight conditions) the bias becomes positive and small in magnitude above the PBL, whereas at station 457 (data collected under night-time conditions) the bias keeps the same sign throughout the troposphere. At the other stations, the bias within the PBL is overall small and either positive (107, 456) or slightly negative (stations 458, 338).

Bias profiles for WS at eight ozonesondes stations in EU (Figure 4) show a tendency of overestimation in the PBL and of underestimation above ~1000m, although there are some exceptions for different models and/or launching stations. The WRF/Chem1 has the largest positive bias at all sites, with the bias staying positive well above the PBL at all stations in contrast with all other models (WRF/Chem1 model adopted the assimilation of meteorological fields only above the PBL, and only during the first 12 hours of meteorological spin-up). WS overestimation by WRF/Chem is a known concern (e.g. Tuccella et al., 2012b; Jimenez and Dudhia, 2012; Mass and Ovens, 2011) and it is likely to have a major impact on the dispersion of pollutants. As for EU, the WS bias profiles in NA are biased high near the surface (except for the station 338 and, partially, station 021) (Figure 6). Above the PBL the tendency is to underestimate the WS (up to ~1.5m/s), although less dramatically than in EU. As both NA models are driven by WRF for meteorology, the WS profiles are alike and the magnitude of the bias very similar.

3.2 Dry Deposition

The simulated annual accumulated dry deposition per unit area over the continental areas for NO_2, ozone, and $PM_{2.5}$ is reported in Figure 8 for EU and NA. The graphs report the modelled values only (no observations are readily available). The model-to-model variability in dry deposition is mainly attributable to land cover and model grid size, as the majority of the models employ variations of the resistance scheme (Table 1). As recently noted by Valmartin et al. (2014), developments of the dry deposition schemes can have a profound impact on the overall model bias and on the accuracy of the modelled cycle of the pollutants.

The deposition of NO_2 is very similar among all the models for both continents, with the only exception of the WRF-DEHM model in EU and NA, whose median and 75th percentile values are below 0.5 and 1.5 kg/km2, respectively. For ozone, the medians of the distribution are in the range ~80-200 kg/km^2 for EU (nine models), whereas the 75th percentile shows larger variability, ranging between ~150kg/km^2 for WRF-CMAQ1 and ~500kg/km^2 for the WRF-DEHM and WRF/Chem1 models. The median difference for ozone is more marked in

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

NA (Figure 8b), between ~200 and ~300 kg/km^2 (two models), with a notably relative impact of 50% and 33% of the median values for WRF-CMAQ and WRF-DEHM respectively.

Finally, deposited $PM_{2.5}$ in EU is modelled with varying magnitudes, from below 5 kg/km^2 (SILAM, WRF-CMAQ1, WRF-CMAQ2) up to 35 kg/km^2 (WRF/Chem1) (median values). The median values for the two NA models are very similar ~25kg/km^2, but there is a large discrepancy between the 75th percentile values, with that of WRF-CMAQ (~170kg/km^2) more than four times higher than values predicted by WRF-DEHM.

3.3 Chemical Species: Mean Square Error and Error Apportionment

3.3.1 CO

CO is a moderately long-lived primary pollutant principally produced by incomplete combustion of fossil fuels, wildfires and, on the global scale, by the oxidation of methane. CO also acts as precursor to ozone. Results of the AQMEII3 models for CO are reported in Figure 9 and Figure 10, and in Table S5.

In general, there are profound differences between the CO statistics for EU and NA, with the latter showing a more marked temporal and spatial dependency as well as model-to-model variability (the yearly mean observed values of CO in EU and NA are of 336 ppb and of 248 ppb, respectively). The EU error (Figure 9a) is, generally, uniform across models and sub-regions, approximately three times higher in winter than in summer. The magnitude of the SY and DU errors is comparable (~15-25 ppb on average in EU1 and EU2, sensibly higher in EU3). Also for NA (Figure 9b) the DU and SY errors are similar, but varying by model, sub-region, and season.

The homogeneity of error in EU suggests that it is originated by a common source. Previous investigations (Innes et al., 2013; Giordano et al., 2015) indicate that the boundary conditions have a limited contribution to the bias of CO within the interior of the domain, where the emissions are far more important. In particular, the MACC inventory used by the EU regional models likely underestimates the CO emissions (especially in winter) (Giordano et al., 2015). We conclude that most probably the cause of model bias for CO is attributable to the emissions and, to a lesser extent, the generally overestimated surface wind speed (section 3.1.1). Sensitivity of the model error to emission changes for CO is discussed in the next section.

The correlation coefficient for EU generally peaks in spring (LT component) while it is at a minimum for the LT component in winter and overall poor for the DU and SY components. In contrast, for NA the minimum correlation coefficient is observed in spring/summer (LT component), with the correlation for DU component having a mixed behaviour depending on the sub-region, but it is typically low in summer (Table S5 of the supplementary material).

The winter LT error for EU is of ~140-220ppb in EU1 and EU2, and up to 600ppb in EU3, typically higher than in NA (~100 ppb, peaking in autumn and mostly due to model underestimation), while the opposite holds for the DU and ID error which are significantly lower in EU (Figure 10) than in NA (except for EU3). Since CO is a primary pollutant, its error is affected by the diurnal dynamics of the PBL height, which is most problematic in winter, when modelled PBL has the tendency to become too stable too early, anticipating the evening transition (Pleim et al., 2016). In fact the biases of CO and that of PM_{10} (another primary pollutant) in winter are highly correlated for almost all models (not shown), indicating a common causes of the error.

The error due to variance in EU (under-estimated by the models) and *mMSE* are significant in the DU and SY components in winter (Figure 10a). In particular, the variance error of winter DU is small compared to the *mMSE*, which accounts for almost the entire DU error, up to over 30 ppb. For SY, the model SILAM_H shows an *mMSE* error of over 75 ppb, the variance part being approximately null. On average, the DU and SY errors are approximately similar for all EU models (~45ppb for DU and ~65ppb for SY), indicating some common error source such as missing sources and process and strong emission underestimation at these time-scales. A further reason could stem from the lack of temperature dependent emissions (the current emission inventory

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

processing approach employs constant temporal emission profiles, and therefore cold/warm episodes are not incorporated in the modelled emissions while these episodes do affect real-world emissions). The lack of temperature-dependant emission is likely to have a strong effect for CO, as about 50% CO emissions comes from
residential heating (at least in mid/north European countries). A test to this hypothesis is currently under investigation by running the CCLM-CMAQ model with a set of emissions using temperature data for the temporal disaggregation for residential heating emissions.

While the SY error is comparable for the two continents, the DU and ID errors are remarkably higher in NA (all sub-regions, also due to an excess of variance) and for several instances comparable or even higher than the LT error. With the exception of the WRF-DEHM model (variance error negligible), the DU and ID error for the NA models are due to both *mMSE* and variance.

3.3.1.A SENSITIVITY SIMULATIONS WITH REDUCED EMISSIONS AND BOUNDARY CONDITIONS

Additional sensitivity runs have been carried out by the majority of modelling groups, in which the amount of anthropogenic emissions are reduced by 20% in both the boundary conditions and the modelling domain. It is instructive to assess the error variation between the sensitivity runs (denoted as 's20%') and the base case for primary species such as CO:

$$\%RMSE = 100 * \frac{RMSE_{CO}^{s20\%} - RMSE_{CO}^{base}}{RMSE_{CO}^{base}}$$

Figure 11 reports the error variation for central Europe (sub-region EU2), where the effect of local CO outweighs the influence of the CO entering from the boundaries (similar plots for the other two EU sub-regions are reported in the Supplement). A decrease of 20% CO produces a RMSE variation of ~10% (averaged over models and components). A naïve projection indicates that a reduction of 100% (thus removing CO from emissions and boundary conditions altogether) would produce a variation of the error of ~50%. The sign of the error variation indicates that there are circumstances where a reduction of the base case emissions is actually beneficial as the error is reduced (even substantially in the instances where the emissions were overestimated in the base case).

The DU component for CO is the most sensitive to emissions changes with an average of ~24% error variation in summer. The SILAM model is the most sensitive to changes in the amount of pollutants entering the domain. Striking error differences with respect to the base case are detected for summer CO (DU error improved by 50%), possibly pointing to false peaks in the base case that contribute heavily to the RMSE (as suggested by the low correlation coefficient, Table S5). The reduction of the emission by 20% lowers the peaks and could be the explanation for the improvement observed for the 's20%' scenario for SILAM.

3.3.2 NO

NO is emitted by both natural and anthropogenic sources and its chemistry patterns are closely connected to those of NO_2 and ozone. Due to the ozone-NO titration reaction (timescale < 1 hour at all temperatures), the uncertainty in emissions, transport, and vertical mixing dominates the uncertainty in chemistry. As no observational data was available for NA, the discussion is limited to EU. The European Environment Agency (EEA) reports an estimated uncertainty for NO_x emission of ~20% (EEA, 2011); Vestreng et al. (2009) found ±8-25% uncertainties in EU NO_x emissions, in line with other similar bottom-up uncertainty studies (see Pouliot et al., 2015). A further source of uncertainty and model to model difference is the vertical emission profiles adopted and how this is interpolated to the vertical grids used by the models. Within the SILAM model, for example, the vehicular traffic emissions are released largely at the bottom of the first layer and this sub-grid information about the vertical location of the plume used in the vertical transport scheme further supresses the mixing to the upper layers, thus keeping the surface concentrations higher.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

The analysis of the RMSE for NO in Figure 12a shows how the largest modelling error for NO occurs in winter and autumn, similar in magnitude for EU1 and EU2 (~7ppb), while is more than double in EU3 (up to 30 ppb). The DU and SY errors are comparable in magnitude (although the DU error is slightly higher), and are approximately evenly distributed among the models. Also for NO the error of the SY component is model-independent, as noted for CO and as will be discussed for ozone and PM_{10}. Because it is mainly composed by *mMSE* error (Figure 12b) it can be hypothesized that the unexplained meteorological variance is responsible for the majority of the SY error.

The winter bias and variance errors are predominantly negative, indicating model underestimation and reduced variability. The opposite holds for the two instances of SILAM, for which the bias and variance are positive (all sub-regions). This can be associated with the underestimated ozone concentrations in this model also the applied vertical emission profiles mentioned earlier for this model could have an influence. The correlation coefficient varies greatly by model, by components and by season and typically degrades for the summer seasons (LT component, most models). The SY component also exhibits low values of *r*, especially in summer for EU1 and autumn (Table S6). The large variability of the correlation coefficient indicates that the models are not able to capture the fluctuations of this important precursor at all scales.

From the error decomposition plots (Figure 12b) it emerges that

- the LT components shows a *mMSE* error approximately uniform for all modelling systems (between ~3 and 4 ppb);
- in the majority of the cases the *mMSE* error dominates the ID, DU and SY components;
- the SY component has an error comparable to that of DU for the *mMSE* part, but overall higher due to a predominant lack of variance (as high as 50% of the total SY error for some models).

Due to its fast chemistry and short travelling distance, the error of representativity for NO (mismatch of the area of representativeness between models with grid spacing of ~15 km up to 50 km and point measurements) is likely more significant than for other pollutants with longer life-time. NO is almost a primary pollutant with negligible deposition (Wesely and Hicks, 2000) and small influence of the boundary conditions (Giordano et al., 2015), therefore observational sites are affected by local scale effects in the range of a few kilometres, below the grid spacing of the majority of the models. This has the effect of higher observed mean values compared to the models (enhancing the bias error) and stronger variability in the observations than the models (variance error).

The correlation between the bias of NO with the bias of the other species reveals strong links at several temporal scales (less for the DU time scale though) and also in terms of processes, although it varies greatly by model. For instance, *corr*($bias_{NO}$, $bias_{O3}$) is overall strong (and negative) for the majority of the models, but for different time scales, i.e. stronger for the SY components for some models (e.g. LOTOS-EUROS), or for the LT (SILAM), or for the DU (CHIMERE). Additional analysis are envisioned to determine the causes of such a behaviour.

3.3.2.A Sensitivity simulations with reduced emissions and boundary conditions

The analysis discussed in Section 3.3.1.A is repeated here for NO and results are presented in Figure 13. A decrease by 20% of the amount of NO in the domain produces a variation of RMSE of ~8% (averaged over models and spectral components). A naïve projection indicates that a reduction of 100% (thus removing the production of NO from emissions and boundary conditions) would produce a variation of the error of ~35%. Such an amount is less than that found for CO (~50%, section 3.3.1.A), which is consistent with the photochemical processes involving NO but not CO.

The LT component is the most sensitive to changes for NO, with an average of ~17% error variation ((and up to 20% in autumn, both positive and negative). Again, the SILAM model is the most sensitive to changes in the

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

amount of pollutants entering the domain. Remarkable differences between the 's20%' scenario and the base case are detected for summer and autumn (LT error variation of 100%) (Figure 13). The improvement of the error of SILAM (and of the other models) for the 's20%' scenario is due to the overestimation of NO mean concentration in the base case (positive bias, Table S6).

3.3.3 NO_2

Primary NO_2 is emitted by a variety of combustion sources and plays a major role in atmospheric reactions that produce ground-level ozone. NO_2 is also a precursor to nitrates, which contribute to PM formation. As for NO, only a small portion of the total error is expected to stem from the boundary conditions. The AQMEII3 modelling systems attribute a fraction of NO_2 emission ranging between 3% and 10% of the total NO_x emissions (some models treat the NO_2 emission from the transport sector differently, see Table 1). The results of the error analysis discussed hereafter do not reveal, though, grouping of model behaviour consistent with the choice of the NO_2 to NO_x emissions ratio.

The RMSE distribution (Figure 14a,b) shows a marked model-to-model variability in the LT and DU components, while it is more uniform for the SY component, also in the seasonal stratification. Moreover, the error distribution shows to be weakly dependant on the specific sub-region (for both continents, especially for the DU component), suggesting that regional features (e.g. differences in climate between the regions) have little impact on NO_2 performance, which is most affected by chemistry and error in the meteorology. Local-scale features (e.g. representation of urban / rural emission differences) may still be important, but they may have similar errors in all regions.

The largest error occurs in winter (both continents), and is shared approximately equally between the SY and DU components (for some models the SY and LT errors are comparable due to the little bias).

The bias is the main contributor to the NO_2 error and stems from a model under-prediction of the mean observed concentration (but, with the exception of the winter season, is positive for WRF-CMAQ in NA) (Table S7). However, the tendency of NO_2 measurements to be likely overestimated by the majority of commercially available instruments for detecting NO_x (Steinbacher et al., 2007) needs to be taken in to account. The magnitude of the bias higher in EU (from ~1.3pbb of WRF-CMAQ1 in EU1 to ~-12.5ppb of CCLM-CMAQ in EU3) than in NA (the maximum being ~5.5ppb in NA3 by the WRF-DEHM model), with the mean observed values being of 11.5ppb and 10.5ppb for EU and NA, respectively.

The correlation coefficient is typically higher in spring/autumn and poorer in summer/winter (in summer there are several instances of negative correlation) (Table S7). The LT component for EU, and the LT and SY components for NA, are those with higher correlation coefficients, while SY and DU are the poorest in EU and DU the poorest in NA (but still higher than ~0.4).

The median value of the modelled accumulated deposition per unit area (Figure 8) for NO_2 ranges from ~0.4 to ~1.9 kg/km^2 for EU (nine models) and from ~0.3 to 2.3 kg/km^2 for NA (two models). With the exception of the WRF-DEHM model (similar values for EU and NA of 0.3-0.4 kg/km^2), the modelled values for NO_2 deposition are uniform across the EU models, while the deviation between the two NA models for deposition is not negligible, also in light of the different native grid sizes of 50km and 12km (WRF-DEHM and WRF-CMAQ, respectively). Therefore, for the majority of the EU models model-to-model differences in the error are unlikely due to significant difference in the deposition, while it remains a possibility for NA.

The magnitude of the error for NO_2 resembles that of NO and ozone, although the apportionment reveals significant differences. In fact, while NO includes variance error and a uniform share of *mMSE*, the LT error of NO_2 for winter is almost completely determined by the bias, for both continents (Figure 15 and Figure 16). The other NO_2 spectral components (ID, DU, SY) reveal more profound difference with respect to NO, both in terms of bias and of error apportionment. The ID error for NO_2 is even smaller than that of NO (less than 1 ppb) and

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

745 can be regarded as noise. Also the DU (~1.5 ppb) and SY (~1 ppb) errors are considerably smaller than for NO
746 (both continents), although the DU error presents some excess of variance for WRF-CMAQ3 and the two
747 instances of the CHIMERE model (Figure 15).

748 The model-to-model variability of RMSE for the LT component Figure 15) is very similar to that of NO (Figure
749 12), while the DU variability resembles that of ozone (Figure 18), although for NO_2 the DU error is lower in
750 magnitude and more uniform across seasons.

751 Moreover, NO_x observations are strongly affected by local emissions and thus the error may stem from the
752 incommensurability of comparing grid-averaged values against point measurements highly affected by local-
753 scale emissions. However, the error apportionment analysis carried out separately for 'rural' and 'urban'
754 background stations (the area type classification is taken for the stations metadata) does not reveal any
755 relevant differences (Figure 15 for EU2 and Figure 16 for NA1), if not a slight increase of the variance error
756 over both continents.

757 3.3.4 OZONE
758 Due to the adverse effects on human health and to the impact on climate, tropospheric ozone is regulated in
759 EU and NA and substantial efforts are made to improve the models' predictive skill for this pollutant.
760 Tropospheric ozone can be either transported from regions outside the modelled domain, be the result of
761 stratosphere/troposphere exchange, or be produced locally by photochemistry through oxidation of VOCs
762 (volatile organic compounds) and CO in the presence of NO_x and sunlight. Due to its photochemical nature,
763 ozone production is directly influenced by temperature through speeding up the rates of the chemical
764 reactions and increasing the emissions of VOCs (e.g. isoprene) from vegetation (Jacob and Winner, 2009).
765 Along with dry deposition, chemistry can act as local sink to ozone depending on the photochemical regime.

766 Results of the AQMEII3 modes for ozone are reported in Figure 17and Figure 18, and in Table S4. Overall, the
767 correlation between modelled and observed ozone time series is higher for the winter and fall seasons than
768 the spring and summer seasons in EU, while the opposite holds true in NA where the maximum correlation is
769 observed in summer (all sub-regions) (Table S4). In EU, the RMSE is generally lower in winter than in the warm
770 seasons (summer and spring) (RMSE in summer ranges between 4.3 ppb of WRF/Chem1 in EU1 and 21 ppb of
771 WRF-CAMQ1 in EU3), with the exception of the CCLM-CMAQ model for which the RMSE peaks in autumn (all
772 sub-regions).

773 Due to the strong and well defined diurnal cycle characterized by ozone formation and loss, the correlation
774 coefficient is generally higher for the DU component, while it tends to be lowest for the SY component (Table
775 S4 and Figure 18). The SY component often exhibits the lowest correlation among all components, especially in
776 summer (EU) and spring (NA), possibly due to the combined effect of transport of precursors, deposition and
777 chemistry (formation/depletion of ozone from precursor emission in the regions where the ozone is
778 transported) (Bowdalo et al., 2016). However, the SY error is generally small (~2-3 ppb, although higher for
779 EU3, where the SY error is double that of the other sub-regions) and is mostly due to *mMSE*, it is thus
780 characterised by poor coefficients of determination and underestimated variability (Eq 7). Therefore, the SY
781 component suffers from low precision (for some models $r < 0.3$) meaning that the variability of the synoptic
782 mechanisms needs further attention, especially in the meteorological conditions leading to high ozone level
783 episodes, especially in relation to temperature, cloudiness, and radiation. The WRF/Chem2 model (having the
784 highest error for temperature, Figure 2b) reports the largest SY error for ozone (especially the variance part).
785 For this model, the correlation between the ozone and the Temp error for the SY component $corr(bias_{O3},$
786 $bias_{Temp})_{SY}$ is 0.44 for the summer months in EU2 (not shown), among the highest, which helps explain part of
787 the SY error for ozone. In order to characterise better the *mMSE* part of the error for the periodic components,
788 such as DU and SY, analysis of the phase and amplitude are foreseen.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

The error of the DU component is largely due to the *mMSE* term (Figure 18a) which is comparable for all models in the range of 2-5 ppb, with some significant excess of variance for WRF-CMAQ2 and WRF-CMAQ3 in EU2 (~5 ppb). One possible reason is the dynamics of the nocturnal PBL as well as the timing of the ozone cycle, with an either too fast or too slow modelled ozone peak (e.g. Pirovano et al., 2012). Limitations of the models to reproduce the amplitude and phase of the daily ozone cycle were already highlighted in the first and second phase of AQMEII, mostly related to the representation of night-time and stable conditions. Further, the variance error for WRF-CMAQ2 and WRF-CMAQ3 can be induced by the bias of temperature and/or concentration of ozone precursors. For WRF-CMAQ2 (WRF-CMAQ3), *corr*($bias_{O3}$, $bias_{Temp}$)$_{DU}$ is 0.88 (0.94) and *corr*($bias_{O3}$, $bias_{NO2}$)$_{DU}$ is 0.86 (0.83) (summer months, EU2) (not shown), which indicates that the bias of the Temp and NO_2 fields are strongly associated with the error of ozone at the DU scale. According to Bowdalo et al. (2016) the bias of the ozone amplitude cycle linearly evolves with NO_x emissions, suggesting that correction of the error for ozone needs to start from NO_x emissions. Otero et al. (2106) have shown that meteorological drivers account for most of the explained variance of ozone, especially over central and northwest Europe. One of the main drivers of ozone is the daily maximum temperature, in relation to the effect of temperature on emissions of VOCs. Therefore, while part of the bias error is likely due to NO_x emissions, the *mMSE* and variance error are also induced by error in meteorology. Other sources of biases are transcontinental transport in winter (Hogrefe et al., 2011) and missing processes during spring and summer, such as the large scale effect of forested areas.

The error in the LT component is dominated by the bias error (Figure 18) (almost completely for NA) although with significant exceptions in EU (for CCLM-CMAQ the *mMSE* error of the LT component is larger than the bias portion). The May-September ozone LT bias for EU2 peaks at 12-13 ppb (WRF-CMAQ1), while it is ~6 ppb in NA3 (but in excess of 20ppb in NA2 by the WRF-DEHM model) (the yearly average measured ozone mixing ratio is 26.5 and 29ppb for EU and NA, respectively). The bias of the precursors and of the meteorological fields is typically highly correlated with the bias of ozone. For instance, in EU2 for the WRF-CMAQ1 model *corr*($bias_{O3}$, $bias_{Temp}$)$_{LT}$ is 0.65 and *corr*($bias_{O3}$, $bias_{WS}$)$_{LT}$ is 0.81 (summer months).

Although the concentration peaks are associated with the ID and DU components, the contribution to the total error of the ID component is small (< 2ppb) due to the flattening of the spikes operated by the spatial averaging carried out prior of the spectral decomposition. The noise of the ID component is reflected by the correlation coefficient being lower than the correlation of the DU component. However, the acf (auto-correlation function) associated with the signal of the ID component is well structured and periodic (not shown), indicating that the ID component for ozone is not entirely a white noise-type of signal, but incorporates useful information, although there is the possibility that the ID periodicity is due to a periodic leakage from the DU component, due to the imperfect separation of the ID and DU components. This latter aspect will require additional investigation.

3.3.4.A Ozone Vertical Profiles

A further analysis aimed at detecting errors introduced by the vertical mixing is carried out by using modelled and observed ozone profiles from ozonesondes. A summary of the records provided by the ozonesondes for ozone are reported in Table 3. Plots of the simulated and observed ozone levels at fixed heights (through the ENSEMBLE system models and measurements are paired at the heights of 0, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000 m) are reported in Figure 19 and Figure 20. The ozonesonde data are mainly available during daylight, although two stations with night-time data are available for NA (Table 3).

Overall, the general tendency of the models in both continents is to underestimate the ozone levels above the PBL, suggesting that not enough ozone enters the continental domains through the inflow boundaries. The most significant underestimation (~10 ppb) is observed at the two stations closer to the west boundary for EU (stations 318 and 043). The boundary layer deficit of ozone is a long standing issue, as similar conclusions were derived for the first (Solazzo et al., 2013) and second (Im et al., 2015; Giordano et al., 2015) phase of AQMEII,

as well as in other studies (Katragkou et al., 2015), emphasizing the strong dependence of regional models on the lateral boundary, whose effects propagate far into the interior of the domain.

Towards the interior of the EU domain (stations 134, 157, 242) the profiles are in closer agreement with the observations, with the WRF-CMAQ1 model performing the best throughout the troposphere, possibly due to the overestimation of the entrainment of upper tropospheric ozone, as revealed by the strong gradient of WRF-CMAQ1 at 6000m (Figure 19).

For NA (Figure 20), the general tendency is of good agreement within the PBL and underestimation aloft for the WRF-CMAQ model and of overestimation (stations 107, 456, and 458 – afternoon/night launches) at the surface and mild underestimation above the PBL for the WRF-DEHM model.

3.3.4.B RELATIONSHIP BETWEEN THE BIAS OF OZONE, NO_x AND TEMPERATURE

The relationship between the bias of NO and the bias of ozone is reported in Figure 21 for the EU2 region (similar plots including the bias of NO_2 for EU and NA are provided in the supplementary material). A linear relationship between the biases of the two species is detectable, more evident in winter. Large, positive ozone bias is driven by underestimation of NO (a primary species) whereas the largest negative ozone bias correspond to the largest overestimation of NO. The role of the temperature bias is less clear, but the NO_2 and ozone relationship (Figure S7) suggests that large NO_2 bias is associated with temperature under-prediction. The partition of NO_x emission into primary NO and NO_2 seems to suggest that the models adopting a 95%-5% ratio suffer lower ozone bias (at least in winter), although in general the clustering of models based on the NO/NO_2 share of total NO_x emission is far from robust. A simple linear regression between NO bias and ozone bias (based on the yearly time series) among the EU models suggests that the NO_x and temperature biases can explain, on average, ~35% and ~16% of the variability of the ozone bias, respectively.

3.3.5 SO_2

SO_2 is another primary regulated pollutant which, in EU and NA, is mainly emitted from coal power plants and also from the residential heating and waste disposal sector. SO_2 acts as a precursor to sulphates, which are one of the main components of PM in the atmosphere. Any error in SO_2 is likely inherited by these secondary species. The EEA reports an estimated uncertainty for SO_2 emission of ~10% (EEA, 2011), therefore SO_2 emissions are expected to be more accurate than NO_x emissions. This is reflected in the low bias in both continents (~1-2ppb in winter, mostly due to model underestimation) (Figure 22 and Figure 23). The averaged observed concentration of SO_2 is of 1.92ppb and 2.7ppb in EU and NA, respectively.

The seasonal modelled error for SO_2 ranges, on average, between 0.65 and 1.3ppb in EU and between in excess of ~1 and 5ppb in NA (the maximum error in NA2), peaking in autumn.

In EU and NA1, the error of ID, DU and SY components is comparable for all seasons and, on average, below 0.6ppb. There are some exceptions, most notably the WRF-CMAQ3 model, which is the only one significantly biased high (Figure 23a) and shows an excess of variance significantly larger than the other models.

The large variability of the model-to-model error (especially in EU) and correlation coefficient in both continents is an indication that the mechanisms governing the initial mixing and subsequent transport and chemical transformation suffer from different sources of error, mostly covariance, at all scales. In no instance is the correlation coefficient consistently above 0.5 for all seasons and spectral components while there are several instances of negative correlation between the spectral components of observed and modelled SO_2 (e.g. CCLM-CMAQ model in EU and several others). The poor correlation coefficient of, especially, the ID and DU components for both continents, indicates that the peaks of the SO_2 concentration are not caught by the models, leading to low precision. Although the mean fluctuations are, generally, well reproduced (low variance error in both continents), it remains a significant portion of unexplained variance (*mMSE*) error, which can derive from meteorology and chemistry. Bieser et al. (2011b) showed that the height of the release and

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

vertical distribution of the SO_2 emission influence the SO_2/SO_4 ratio as the oxidation (aging) of SO_2 is more effective if the emissions are higher up. As power plants are the major source of SO_2 further analysis should investigate the impact of differences in the vertical emission distribution between models.

3.3.6 Particulate Matter

Particulate matter (PM), both in the fine and coarse fraction, is directly emitted by biomass and fossil fuel combustion in domestic and industrial activities, and also formed from precursors in the atmosphere.

From the AQMEII3 suite of model runs, the error for PM is evaluated for PM_{10} in EU and $PM_{2.5}$ in NA. The choice is dictated by the availability of hourly measurements in the two continents. The RMSE distribution is reported in Figure 24 (PM_{10} for EU) and Figure 25 ($PM_{2.5}$ for NA). The error distribution for EU reveals that, despite the large numbers of modelling options and parameters characterising the chemistry and physics of particles, the error distribution for DU and SY is homogeneous among the EU models. For these components the error is approximately uniform over seasons, although with some exceptions (significantly higher in EU3, although based on two receptors only). EU3 is a small area compared to EU1 and EU2, but is densely populated, intensively farmed, with a large amount of wood burning in winter, and agricultural area in summer. It is surrounded by mountains and stagnant flow conditions are predominant. It is, thus, a challenging area for current modelling systems, especially for primary species such as PM.

The LT component shows some significant model-to-model variations due to the WRF-CAMx and WRF-CMAQ1 models which have lower error in spring and summer compared to the other models, while the CCLM-CMAQ model has higher LT error in EU1.

The magnitude of the SY error in EU is, on average, of ~6 μg m^{-3} during winter, with a peak of 10.5 μg m^{-3} in EU2 (WRF-CAMx model). The magnitude of the DU error is lower (~2-2.5 μg m^{-3} in EU1 and EU2, and ~5-6 μg m^{-3} in EU3) with the largest share in autumn, spring, and winter and slightly lower in summer. The error of the LT component ranges between ~11-15 μg m^{-3} in EU1 and EU2 and up to 25 μg m^{-3} during winter in EU3.

The analysis of the correlation coefficient reveals that the model to model differences in the correlation coefficient with the observed component time series tend to be most pronounced for the DU and ID components, indicating that these two components are pivotal in determining the overall model skill in terms of capturing observed fluctuations in PM_{10} concentration. In more detail, the correlation is poor for the DU component (especially in EU2 and EU3, Table S9), possibly due to PBL dynamics and emission profiles (as discussed above for the RMSE at the DU scale). The LT component has correlation values highly varying among models and, for the same model, among seasons (e.g. the LT correlation of the WRF-CMAQ4 model in EU3 is ~0.9 during spring but only of 0.35 in summer).

In winter the LT and SY error is more severe likely due to the larger uncertainties in PM_{10} emissions of combustion processes (wood burning, residential heating) (Van der Gon et al., 2014), as well as due to the current limitations in modelling the vertical mixing during stable conditions, as mentioned for the gaseous species (especially CO, being another primary species). The majority of the EU models show an LT error in winter between 12 and 16 μg m^{-3}, eight models above 16 μg m^{-3} and only one (WRF-CAMx) below 10 μg m^{-3}. The SY winter error exceeds 5 μg m^{-3} for all models (all sub-regions) and three instances (WRF-CAMx, WRF/Chem1 and WRF/Chem2, this latter showing the highest accumulated deposition for $PM_{2.5}$, Figure 8b) report an error above 7.5 μg m^{-3}. All the remaining models have comparable *mMSE* and variance errors (Figure 26), and are biased low (model under-prediction), possibly due to missing PM source and overestimated surface wind speed. As for the WRF-CAMx model, the low bias on LT component and the relatively high error on covariance in SY fraction suggest that the model was able to capture the mean magnitude of PM concentration over the entire year, but failed in reconstructing the correct variability of the different episodes, whose timing is generally driven by the synoptic time scale.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

The $PM_{2.5}$ evaluation in NA is restricted to two models, WRF-DEHM and WRF-CMAQ, which show comparable error (Figure 25). The WRF-CMAQ (WRF-DEHM) model has an error ranging between ~3.5 (~2) and ~6 (~8.5) μg m^{-3}. The main contribution to the total error stems from the LT component (predominantly negative bias) and from the SY component (2-3 μg m^{-3}). The DU component contributes to about 1.5 μg m^{-3} (comparable *mMSE* and variance error).

Both NA models are biased low in summer (all sub-regions), which can be attributed to limitations in the SOA mechanism (Zare et al., 2014). Because of the higher contribution of primary $PM_{2.5}$ to total $PM_{2.5}$ during wintertime, differences in horizontal and vertical resolution (Table 1) likely contribute to the difference in wintertime LT bias. The correlation coefficient for the two models is in general higher in winter (full time series) and deteriorated for the DU component (all seasons and sub-regions).

As inferred for the species discussed above, the uniformity of model behaviour is indicative of errors stemming from external fields, likely emissions, where missing sources of PM can affect the error within certain time scales for all models. Further common causes of error are intrinsic to the model-observation comparison as modelled PMs is commonly dry while this is not always the conditions for the measurements. For instance, the filter-based gravimetric measurements as recommended by the European Committee for Standardization (CEN) are likely to retain part of the particle-bound water after the filter conditioning at a constant temperature of 20° C and relative humidity of 50%. Recent findings by Prank et al. (2016) report the aerosol water content from the gravimetric measurements to range between 5 and 20% for $PM_{2.5}$ and between 10 and 25% for PM_{10}. The particle-bound water was found to be associated with hygroscopic particles such as sulphate, nitrate, and organic compounds. This remaining water content can be up to approximately 10-35% depending on the chemical composition of aerosols being measured (Tsyro, 2005, Kajino, et al., 2006, Jones and Harrison, 2006). The water aerosols should therefore be accounted when compared with these measurements. Part of the problem lies in secondary organic aerosol. In winter, in particular for wood burning part of the emissions are condensable gases that rapidly change to the aerosol phase (Van der Gon et al 2014), but are missed since they are not part of the presently used PM emission inventory. In summer, biogenic emissions that contribute to SOA formation and their yields are quite uncertain. A good representation of SOA is still a problem for all models. In spring, the application of manure and fertilizer leads to peaks of NH_3 emissions and subsequent NH_4 aerosol formation, contributing to PM_{10} and $PM_{2.5}$. The timing of these emissions is parameterized based on long-time averages, whereas in practice they are strongly related to meteorology. This can explain part of the discrepancy on the diurnal to synoptic time scale (Hendriks et al 2015).

4. Memory of the signal and removal processes: the case of ozone

The evaluation of the removal processes (chemical transformation, transport, and deposition) is difficult to assess in isolation with respect to other sources of error because of the bias of the signal. In this section we propose a bias-independent spatial analysis aimed at the quantification of the 'memory' of the signal. The analysis seeks the time interval (or memory) after which the signal loses any memory of its past. The memory of the modelled and observed signals is then compared. The methodology consists of:

1. calculating the autocorrelation function (*acf*) of the modelled and observed LT component;

2. then, calculating the quantity $acf_{mod=0}$ and $acf_{obs=0}$, i.e. the lag (time interval) where the *acf* of the modelled and observed LT component falls to zero, and finally

3. determining the difference between the two, yielding the difference between the modelled and the observed memory of the signal:

$$\Delta_{memory} = acf_{mod=0} - acf_{obs=0} \quad \text{Eq 9}$$

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

966 The *acf* is simply a measure of the degree of associativity of a time series with its lagged version. The
967 associativity is typically measured through the correlation coefficient, and the lag extends from one time step
968 (one hour in the case of hourly time series) to, generally, a third of the length of the time series. Because the
969 correlation is bias-independent, we conclude that the *acf* is also bias-independent therefore information from
970 Δ_{memory} is useful for the interpretation of the variance and covariance errors discussed in section 3.1. The
971 memory of the signal is different from the persistence indicator (previous day concentration) as used e.g. by
972 Otero et al. (2016) for accounting for pollutant episodes. As we deal with the LT component of the signal, short
973 term and synoptic episodes are in fact filtered out in this analysis.

974 In the supplementary material Figure S9 and Figure S10, the *acf* for the network-wide spatial average and for
975 the full year is reported. The *acf* is calculated for the LT component of the observed (first panel) and modelled
976 ozone time series. The zero of the *acf* and the slope of the decay of *acf* of the observations (approximately a
977 straight line from 1 to 0 in 2000 hours) are replicated by the models with various degree of success (Figure
978 S10). Our intent is to apply this analysis to the seasonal ozone time series at each receptor, and derive useful
979 information about the modelled removal/production processes. The spatial analysis is proposed for ozone, for
980 the months of May to September (Figure 28 and Figure 29) and for the full year (supplementary material
981 Figure S9 and Figure S10).

982 The average life time of ozone in the troposphere is of approximately 20-30 days (Solomon et al., 2007). By
983 analysing the LT component (processes > ~21 days) we therefore screen out the daily removal/transformation
984 due to chemistry and can focus on seasonal transport, deposition of the free tropospheric ozone, long term
985 chemistry (seasonal changes in vegetation that affect biogenic VOCs emissions and ozone deposition, and also
986 the monthly variations applied to the anthropogenic emission) and influence of boundary conditions. The
987 structure of the *acf* also benefits from the removal of short time scale processes as it is less affected by noise
988 and the results are easier to interpret.

989 The spatially distributed Δ_{memory} shows some clear regional effects for the majority of the models. The
990 $\Delta_{memory} > 0$ along the Mediterranean coast of Spain and France, with some severe excess of ozone
991 production (or underestimation of sinks) in southern/central France for some models (SILAM, WRF-CAMx,
992 WRF-CMAQ1, WRF-CMAQ2 and especially the L.-Euros model, for which the *acf* at the French receptors did
993 not reach zero).

994 The region covering the Po valley, Austria and extending into the continental eastern EU is affected by
995 negative Δ_{memory} (sometimes a deficit of one month for some models). The negative memory indicates that
996 the observed signal is more persistent than the modelled one, and that long term weather transitions are
997 smoother in gradient and longer in duration, and thus that the seasonal modulation of the signal is
998 overestimated by the models, thus producing variance error. Coupling the two behaviours (excess of ozone in
999 south France and south Spain with the short memory from the interior of east EU extending to the Po valley),
1000 might indicate an easterly synoptic transport of ozone (or of LT ozone precursor, such as the impact of CH_4 and
1001 CO on OH and photochemistry) masses whose duration is underestimated by the models. The relationship
1002 between the sign of Δ_{memory} and the land use type (vegetation vs urban) is subject of on-going investigations
1003 in the attempt to determine the role of VOCs emissions and deposition over different land types.

1004 The central part of Germany is affected by positive (on average in the range of 7 to 10 days) Δ_{memory}, mostly
1005 visible for the HTAP-emission based SILAM and Chimere results in contrast with the MACC-emission based
1006 ones of the same models. When the HTAP inventory is used the largest differences are observed in the central
1007 EU regions, indicating that also the LT chemistry plays a role.

1008 The deposition aspect of removal can be equally important as transport and chemistry. The memory of the
1009 signal directly depends on the amount of ozone available and a large, negative Δ_{memory} might indicate that
1010 the deposition is too high.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

For NA (Figure 29), the feature common to all models is the excess of removal in the Southern Atlantic coast and across the Eastern Canadian border. In contrast, the central-east part of the US shows large positive Δ_{memory} values (up to ~1.3 month for the WRF-DEHM model), with the exception of the WRF-CMAQ model, which is overall in line with the observed memory of the signal in this part of the domain. This result agrees with the seasonal phase analysis for ozone in global models by Bowdalo et al. (2016), where a delay of up to 4 months was detected for east USA.

The west coast has a mixed behaviour, but prevalently Δ_{memory} is negative. The hypothesis that too little ozone enters the domain trough the boundary conditions is contradicted by the Δ_{memory} ~0 for the full year in the west coast (see Figure S10). A potential excess of transport in this region also seems to be contradicted by the large number of stations for which Δ_{memory} is positive. A possible conclusion is that localised biogenic emission sources, radiation budget, and deposition are the main factors responsible for the negative sign of Δ_{memory} in this region.

5. Conclusions

The work presented in this paper summarises the results of the ongoing third phase of the AQMEII activity focusing on AQ model evaluation, applied to the continental scale domains of Europe and North America. The evaluation of the AQMEII3 suite of model runs is carried out for surface temperature and wind speed, and for the species CO, NO, NO_2, ozone, SO_2, PM_{10} (EU) and $PM_{2.5}$ (NA). Additional analyses making use of emission reduction scenarios (CO and NO) and vertical profiles have also been performed.

This work is primarily meant to provide a wide overview of the performance of current regional AQ modelling systems and to set the basis for additional diagnostic analysis that is currently in progress.

The model evaluation is carried out by quantifying the components of the error (bias, variance, *mMSE*) at four time-scales (ID, DU, SY, LT) each describing physical processes in a specific time range. The bias and variance measure the departure from the first and second moment of the observed distribution (mean and standard deviation), while the *mMSE* accounts for the unexplained observed variability and relates to the ability of the models to reproduce timing and shape as measured by the correlation coefficient. The apportionment of the error to the relevant time-scales and the analysis of the quality of the error have revealed that the LT bias is, by far, the first cause of error, followed by the variance error (fluctuations about the mean value) of the DU component and the unexplained variance of the DU and SY components, depending on the species and season. In more detail:

- The mean concentration of the primary species (NO, CO, PM_{10}, SO_2) is underestimated by the vast majority of the models in both continents, more markedly during the winter and autumn seasons. The largest share of error for these species is the bias of the LT components, most probably due to emissions and the effects of comparing point measurements to volume averaged concentrations.
- The meteorological fields of temperature and wind speed are consistently biased low and high, respectively. Based on the results of the European models directly driven by the global fields for meteorology (e.g SILAM, Chimere) the error for wind speed is of ~0.5-1 ms^{-1} and of ~0.4-1.2K for temperature. These errors can be considered as the uppermost limit the accuracy of the models can currently achieve. The use of nudging and interpolation methods (specific to the configuration of the meteorological model) can add more than 1.5K and 2ms^{-1} to the total error. The analysis of the available vertical profiles suggests that the models overestimate the wind speed within the PBL and vice versa above the PBL, possibly inducing a net outward flux of pollutants at the PBL interface.
- Modelled CO is affected by high errors, uniformly across models and components, more pronounced in winter and predominantly driven by the negative bias of the LT component, followed by variance error of the SY component. Modelled NO and NO_2 also report negative bias but, in contrast to CO, there is

significant model-to-model difference in error variability, possibly due to the chemistry of NO_x. The SY and DU errors of NO are comparable in magnitude (3-5 ppb) and mostly due to *mMSE* error. Preliminary sensitivity investigations for CO and NO seem to suggest that at most ~50% and ~35% of the total error, respectively, could be due to emissions. Finally, based on spatially averaged analysis, the error for NO/NO_2 is the same for urban and rural stations (i.e. the error is insensitive to the area-type of the stations).

- The error analysis for ozone shows large model-to-model variability for all errors and spectral components, with the exception of the SY component for which the error is similar among models and possibly driven by the error in temperature and in the boundary conditions, as modelled vertical ozone profiles near the domain's boundaries are typically underestimated in both continents by all models. The bias is prevalently positive, while the variance error is generally small. While the bias error for ozone is likely driven by error in NO_x emissions, the error in meteorology may factor in determining the *mMSE* and variance error. In fact, there are several models for which the bias of temperature and the bias of NO_2 are strongly associated with the DU error of ozone. A simple linear regression between NO_x bias and ozone bias (based on the yearly time series) among the EU models suggests that the NO_x and temperature biases can explain, on average, ~35% and ~16% of the variability of the ozone bias, respectively. Ongoing analyses are focusing on explaining the origin of the *mMSE* error by investigating the phase shift between the modelled and observed DU and SY components as well as on looking at maximum daily values rather than to the full time series.
- PM analysis (PM_{10} for Europe and $PM_{2.5}$ for North America) reveals that, for Europe, the error distribution for DU and SY is homogeneous and season independent among the models, despite the large numbers of modelling options and parameters characterising the chemistry and physics of particles. A common source of model bias (model underestimation, especially in winter) for PM_{10} likely lies in the emissions (missing sources) and in the overestimation of surface wind speed, whereas variance error may stem from PBL dynamics under stable conditions and missing processes in the model (SOA formation is a known issue for all models). The analysis of $PM_{2.5}$ (based on two models only) shows an excess of variance and low correlation coefficient in the DU component, possibly due to the timing of the PM cycle. Further analyses dealing with the PM components are needed to draw further considerations.
- The analysis of the memory of the ozone signal has revealed a strong model deficit in continental Europe, where the seasonal modulation of ozone is overestimated by the majority of the models. The opposite holds true in the continental US.

Although remarkable progress has been made since the first phase of AQMEII, both in terms of model performance and also in terms of developing a more versatile and robust evaluation procedure, results of AQ model evaluation and inter-comparison remain generic as they fail to associate errors with processes, or at least to narrow down the list of processes responsible for model error. AQ models are meant to be applicable to a variety of geographic (and topographic) scenarios, under almost any type of weather, season, and emission conditions. For such a wide range of conditions the inherent non-linearity among processes are difficult to disentangle and specifically designed sensitivity runs seems the only viable alternative. A model evaluation strategy relying solely on the comparison of modelled vs. observed time series would never be able to quantify exactly the error induced e.g. by biogenic emissions, vertical emission profiles and their dependence on temperature, deposition, vertical mixing, chemistry, and the analysis approach presented in this work is no exception. In fact, the methodology devised to carry out the evaluation activity in this study has not succeeded in determining the 'actual' causes of model error, although providing much clearer indications of the processes responsible for the error with respect to conventional operational model evaluation.

The highly non-linear nature of current AQ models requires the study of the relationships among error fields, those of the meteorological drivers and those of the precursors. When the seasonal and spectral structures of these relationships is analysed together with the error of the input fields (emissions and boundary conditions), then it would be possible to diagnose and explain accurately the processes responsible for the error. Future evaluation activities should envision sensitivity simulations and process specific analyses.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

APPENDIX 1.

Following Hogrefe et al. (2000) and Galmarini et al. (2013) the time windows (*m*) and the smoothing parameter (*k*) have been selected as follow:

$$\begin{gathered} ID(t) = \mathbf{x}(t) - kz_{3,3}(\mathbf{x}(t)) \\ DU(t) = kz_{3,3}(\mathbf{x}(t)) - kz_{13,5}(\mathbf{x}(t)) \\ SY(t) = kz_{13,5}(\mathbf{x}(t)) - kz_{103,5}(\mathbf{x}(t)) \\ LT(t) = kz_{103,5}(\mathbf{x}(t)) \\ \mathbf{x}(t)=ID(t)+DU(t)+SY(t)+LT(t) \end{gathered} \qquad \text{Eq. S1}$$

where **x**(t) is the time series vector. The additive property of the components whose summation returns the original time series might be questioned. In the original work by Rao et al. (1997) it is highlighted the importance of log-transform the components to stabilize the variance. In the case of log-transformation the original time series is obtained by the product of exponential functions whose exponents are the spectral components. For the purposes of the error apportionment analysis presented here, the results of using additive time series component of log-transformed did not produce substantial differences.

A clear-cut separation of the components of Eq. S1 is not achievable, since the separation is a non-linear function of the parameters m and k (Rao et al., 1997). It follows that the components of Eq. S1 are not completely orthogonal and that there is some level of overlapping energy (Kang et al., 2013). Galmarini et al. (2013) found that the explained variance by the spectral components account for 75 to 80% of the total, the remaining portion being on account of the interactions between the components.

APPENDIX 2.

Statistical indicators:

Root Mean Square Error

$$RMSE = \left(\frac{\sum_{i=1}^{n}(M_i - O_i)^2}{n}\right)^{0.5}$$

Mean Bias (MB)

$$MB = \frac{1}{n}\sum_{i=1}^{n} M_i - O_i$$

Pearson correlation coefficient (r)

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{M_i - \overline{M}}{\sigma_M}\right)\left(\frac{O_i - \overline{O}}{\sigma_O}\right)$$

Where *M* and *O* are the *n*-element modelled and observed time series, respectively, σ is the standard deviation and the overbar indicates temporal averaging.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contribution of various groups to the third air Quality Model Evaluation international Initiative (AQMEII) activity. The following agencies have prepared the data sets used in this study: U.S. EPA (North American emissions processing and gridded meteorology); U.S. EPA, Environment Canada, Mexican Secretariat of the Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales-SEMARNAT) and National Institute of Ecology (Instituto Nacional de Ecología-INE) (North American national emissions inventories); TNO (European emissions processing); Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CEA/CNRS/UVSQ (gridded meteorology for Europe); ECMWF/MACC

(Chemical boundary conditions). Ambient North American concentration measurements were extracted from Environment Canada's National Atmospheric Chemistry Database (NAtChem) PM database and provided by several U.S. and Canadian agencies (AQS, CAPMoN, CASTNet, IMPROVE, NAPS, SEARCH and STN networks); North American precipitation-chemistry measurements were extracted from NAtChem's precipitation-chemistry data base and were provided by several U.S. and Canadian agencies (CAPMoN, NADP, NBPMN, NSPSN, and REPQ networks); the WMO World Ozone and Ultraviolet Data Centre (WOUDC) and its data-contributing agencies provided North American and European ozonesonde profiles; NASA's AErosol RObotic NETwork (AeroNet) and its data-contributing agencies provided North American and European AOD measurements; the MOZAIC Data Centre and its contributing airlines provided North American and European aircraft takeoff and landing vertical profiles; for European air quality data the folowing data centers were used: EMEP European Environment Agency/European Topic Center on Air and Climate Change/AirBase provided European air- and precipitation-chemistry data. The Finnish Meteorological Institute for providing biomass burning emission data for Europe. Data from meteorological station monitoring networks were provided by NOAA and Environment Canada (for the US and Canadian meteorological network data) and the National Center for Atmospheric Research (NCAR) data support section. Joint Research Center Ispra/Institute for Environment and Sustainability provided its ENSEMBLE system for model output harmonization and analyses and evaluation. Although this work has been reviewed and approved for publication by the U.S. Environmental Protection Agency, it does not necessarily reflect the views and policies of the agency.

References

Abdul-Razzak, H. and Ghan, S. J., 2000.: A parameterization of aerosol activation, 2, Multiple aerosol types, J. Geophys. Res., 105, 6837–6844, DOI: 10.1029/1999JD901161.

Ackermann, I. J., Hass, H., Memmsheimer, M., Ebel, A., Binkowski, F. S., Shankar, U., 1998. Modal aerosol dynamics model for Europe: development and first applications, Atmos. Environ., 32 (17), 2981-2999, DOI: 10.1016/S1352-2310(98)00006-5

Ahmadov, R., McKeen, S. A., Robinson, A., Bahreini, R., Middlebrook, A., de Gouw, J., Meagher, J., Hsie, E., Edgerton, E., Shaw, S., Trainer, M., 2012. A volatility basis set model for summertime secondary organic aerosols over the eastern United States in 2006, J. Geophys. Res., 117, D06301, doi:10.1029/2011JD016831.

Ahmadov, R., McKeen, S., Trainer, M., Banta, R., Brewer, A., Brown, S., Edwards, P. M., de Gouw, J. A., Frost, G. J., Gilman, J., Helmig, D., Johnson, B., Karion, A., Koss, A., Langford, A., Lerner, B., Olson, J., Oltmans, S., Peischl, J., Pétron, G., Pichugina, Y., Roberts, J. M., Ryerson, T., Schnell, R., Senff, C., Sweeney, C., Thompson, C., Veres, P. R., Warneke, C., Wild, R., Williams, E. J., Yuan, B., Zamora, R., 2015. Understanding high wintertime ozone pollution events in an oil- and natural gas-producing region of the western US, Atmos. Chem. Phys., 15, 411-429, doi:10.5194/acp-15-411-2015

Appel, K. W., Pouliot, G. A., Simon, H., Sarwar, G., Pye, H. O. T., Napelenok, S. L., Akhtar, F., and Roselle, S. J., 2013. Evaluation of dust and trace metal estimates from the Community Multiscale Air Quality (CMAQ) model version 5.0, Geosci. Model Dev., 6, 883-899, doi:10.5194/gmd-6-883-2013

Banzhaf, S., M. Schaap, A. Kerschbaumer, E. Reimer, R. Stern, E. van der Swaluw, Builtjes, P., 2012. Implementation and evaluation of pH-dependent cloud chemistry and wet deposition in the chemical transport model REM-Calgrid. Atmospheric Environment 49, 378–390. DOI: 10.1016/j.atmosenv.2011.10.069.

Benjamin, S. G., Grell, G. A. , Brown, J. M. Smirnova, T.G., 2004. Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. Mon. Wea. Rev., 132, 473-494.

Bergström, R., Denier van der Gon, H. A. C., Prévôt, A. S. H., Yttri, K. E., Simpson, D., 2012. Modelling of organic aerosols over Europe (2002–2007) using a volatility basis set (VBS) framework: application of different assumptions regarding the formation of secondary organic aerosol, Atmos. Chem. Phys., 12, 8499-8527, doi:10.5194/acp-12-8499-2012

Bergström, R., Denier van der Gon, H. A. C., Prévôt, A. S. H., Yttri, K. E., Simpson, D., 2012. Modelling of organic aerosols over Europe (2002–2007) using a volatility basis set (VBS) framework: application of different assumptions regarding the formation of secondary organic aerosol, Atmos. Chem. Phys., 12, 8499-8527, doi:10.5194/acp-12-8499-2012

Bieser, J., A. Aulinger, V. Matthias, M. Quante and P. Builtjes, 2011a. SMOKE for Europe adaptation, modification and evaluation of a comprehensive emission model for Europe, Geosci. Model Dev., 4, 4768.

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

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

Binkowski, F.S., S. Arunachalam, Z. Adelman, and J.P. Pinto, 2007: Examining Photolysis Rates with a Prototype Online Photolysis Module in CMAQ. J. Appl. Meteor. Climatol., 46, 1252–1256.

1187 Bowdalo, D.R., Evans, M.J., Sofen, E.D., 2016. Spectral analysis of atmospheric composition: application to surface ozone
1188 model-measurement comparison. Atmospheric Chemistry and Physics Discussion 172. doi:10.5194/acp-2016-172
1189 Brandt, J., J. D. Silver, L. M. Frohn, C. Geels, A. Gross, A. B. Hansen, K. M. Hansen, G. B. Hedegaard, C. A. Skjøth, H. Villadsen,
1190 A. Zare, and J. H. Christensen, 2012. An integrated model study for Europe and North America using the Danish
1191 Eulerian Hemispheric Model with focus on intercontinental transport. Atmospheric Environment 53, 156-176
1192 Byun, D.W., Ching, J.K.S.: Science Algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ)
1193 Modeling System. EPA/600/R-99/030, US EPA National Exposure Research Laboratory, Research Triangle Park, NC,
1194 1999.
1195 Byun, D.W., Schere, (2006). Review of the governing equations, computational algorithms, and other components of the
1196 Models-3 community Multiscale Air Quality (CMAQ) modeling system. Applied Mechanics Reviews. v59 i2, 51-77 Doms,
1197 G., (2011). A Description of the Nonhydrostatic Regional COSMO model. Part I: Dynamics and Numerics., Tech. rep.,
1198 Deutscher Wetterdienst, available at: http://www.cosmo-
1199 model.org/content/model/documentation/core/cosmoDyncsNumcs.pdf (last access: 8 April 2015)
1200 Carlton, A. G., P. Bhave, S. Napelenok, E. O. Edney, G. Sarwar, R. W. Pinder, G. Pouliot, and M. Houyoux. Model
1201 Representation of Secondary Organic Aerosol in CMAQ v4.7. Environmental Science & Technology, American Chemical
1202 Society, Washington, DC, 44(22):8553-8560, (2010)
1203 Carslaw, D.C., Ropkins, K., 2012. openair – an R package for air quality data analysis. Environmental Modelling and
1204 Software 27-28, 52-61.
1205 Chang, J.S., Brost, R.A., Isaksen, I.S.A., Madronich, S., Middleton, P., Stockwell, W.R., Walcek, C.J., (1987) A three
1206 dimensional Eulerian acid deposition model: physical concepts and formulation, Geophys. Res. 92, 14681-14700.
1207 Chapman, E. G., Gustafson Jr., W. I., Easter, R. C., Barnard, J. C., Ghan, S. J., Pekour, M. S., and Fast, J. D.: Coupling
1208 aerosolcloud-radiative processes in the WRF-Chem model: Investigating the radiative impact of elevated point sources,
1209 Atmos. Chem. Phys., 9, 945–964, doi:10.5194/acp-9-945-2009, 2009.
1210 Chen, F. and Dudhia, J., 2001 Coupling an advanced landsurface/hydrology model with the Penn State/ NCAR MM5
1211 modeling system. Part I: model description and implementation, Mon. Weather Rev., 129, 569–585.
1212 Chen, F., Dudhia, J., 2001. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling
1213 system. Part I: model implementation and sensitivity. Monthly Weather Review 129, 569-585.
1214 Christensen, J. H., J. Brandt, L. M. Frohn and H. Skov, 2004: "Modelling of mercury in the Arctic with the Danish Eulerian
1215 Hemispheric Model. Atmospheric Chemistry and Physics. Vol. 4, pp 2251-2257.
1216 Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., and Brown, P.
1217 D., Atmospheric radiative transfer modeling: a summary of the AER codes, J. Quant. Spectrosc. Ra., 91, 233–244, 2005
1218 Collins, W. D., P. J. Rasch, B. A. Boville, J. J. Hack, J. R. MaCaa, D. L. Williamson, J. T. Kiehl, B. P. Briegleb, C. Bitz, S. J. Lin, M.
1219 Zhang, and Y. Dai, 2004. Description of the NCAR community atmosphere model (cam 3.0), Technical report, National
1220 Center for Atmospheric Research. NCAR/TN–464+STR, NCAR TECHNICAL NOTE
1221 (http://www.cesm.ucar.edu/models/atm-cam/docs/description/description.pdf)
1222 de Leeuw, G., Neele, F.P., Hill, M., Smith, M.H., Vignati, E., 2000. Production of sea spray aerosol in the surf zone. J.
1223 Geophys. Res. 105 (D24), 29397e29409.
1224 Denier van der Gon, H. a. C., Bergström, R., Fountoukis, C., Johansson, C., Pandis, S. N., Simpson, D. and Visschedijk, a.:
1225 Particulate emissions from residential wood combustion in Europe – revised estimates and an evaluation, Atmos.
1226 Chem. Phys. Discuss., 14(23), 31719–31765, doi:10.5194/acpd-14-31719-2014, 2014
1227 Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S. T., Scheffe, R., Schere, K., Steyn, D., and
1228 Venkatram, A.: A framework for evaluating regional-scale numerical photochemical modeling systems, Environ. Fluid
1229 Mech., 10, 471–489, doi:10.1007/s10652-009-9163-2, 2010
1230 Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mrionow, D., Raschendorfer, M., Reinhart, T., Ritter, B., Schrodin, R., Schulz,
1231 J.-P., and Vogel, G., (2011). A Description of the Nonhydrostatic Regional COSMO Model. Part II: Physical
1232 Parameterization, Tech. rep., Deutscher Wetterdienst, available at:
1233 http://www.cosmomodel.org/content/model/documentation/core/cosmoPhysParamtr.pdf (last access: 8 April 2015)
1234 Donahue, N. M., A. L. Robinson, C. O. Stanier, and S. N. Pandis (2006). "Coupled partitioning, dilution, and chemical aging of
1235 semivolatile organics," Environ. Sci. Technol., 40.8, pp. 635–2643.
1236 Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two–
1237 dimensional model. J. Atmos. Sci., 46, 3077–3107.
1238 Dudhia-RRTM for short-long wave radiation (Duhhai, 1989 and Mlawer et al, 1997), Kain-Fritsch cumulus parameterization
1239 (Kain, 2004), WSM6 microphysics (Hong and Lim, 2006), Pleim-Xiu surface layer scheme (Pleim, 2006), RUC land surface
1240 model (Benjamin, et al, 2004), and ACM2 PBL (Pleim, 2007)."
1241 Easter, R. C., Ghan, S. J., Zhang, Y., Saylor, R. D., Chapman, E. G., Laulainen, N. S., Abdul-Razzak, H., Leung, L. R., Bian, X.,
1242 and Zaveri, R. A.: MIRAGE: Model Description and Evaluation of Aerosols and Trace Gases, J. Geophys. Res., 109,
1243 D20210, doi:10.1029/2004JD004571, 2004.
1244 Emberson, L.D., Ashmore, M.R., Simpson, D., Tuovinen, J.-P., Cambridge, H.M. (2000a) Towards a model of ozone
1245 deposition and stomatal uptake over Europe. EMEP/MSC-W 6/2000, Norwegian Meteorological Institute, Oslo,
1246 Norway, 57 pp.
1247 Emberson, L.D., Ashmore, M.R., Simpson, D., Tuovinen, J.-P., Cambridge, H.M. (2000b) Modelling stomatal ozone flux
1248 across Europe. Water, Air and Soil Pollution 109, 403-413.

1249 EMEP (2003) Transboundary acidification, eutrophication and ground level ozone in Europe. Part I: Unified EMEP model
1250 description. EMEP status Report 1/2003.
1251 Environ, 2014. CAMx (Comprehensive Air Quality Model with Extensions) User's Guide Version 6.1 ENVIRON International
1252 Corporation, Novato, CA.
1253 Eskridge, R.E., Ku, J.Y., Rao, S.T., Porter, P.S., Zurbenko, I.G., 1997. Separating different scales of motion in time series of
1254 meteorological data. Bull. Amer. Meteor. Soc. 78, 1473-1483.
1255 Fast, J.D., Gustafson Jr., W.I., Easter, R.C., Zaveri, R.A., Barnard, J.C., Chapman, E.G., Grell, G.A., Peckham, S.E., 2006.
1256 Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled
1257 meteorology-chemistry-aerosol model. J. Geophys. Res. 111, D21305. http://dx.doi.org/10.1029/2005JD006721.
1258 Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel,
1259 A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and
1260 Tsikerdekis, A., 2015 Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 8,
1261 975-1003, doi:10.5194/gmd-8-975-2015.
1262 Foley, K. M., Roselle, S. J., Appel, K. W., Bhave, P. V., Pleim, J. E., Otte, T. L., Mathur, R., Sarwar, G., Young, J. O., Gilliam, R.
1263 C., Nolte, C. G., Kelly, J. T., Gilliland, A. B., and Bash, J. O.: Incremental testing of the Community Multiscale Air Quality
1264 (CMAQ) modeling system version 4.7, Geosci. Model Dev., 3, 205–226, doi:10.5194/gmd-3-205-2010, 2010
1265 Fountoukis, C., Nenes, A., 2007. ISORROPIA II: a computationally efficient thermo-dynamic equilibrium model for K^+–Ca^{2+}–
1266 Mg^{2+}–NH_4^+–Na^+–SO_4^{2-}–NO_3^-–Cl^-–H_2O aerosols. Atmos. Chem.Phys. 7, 4639-4659.
1267 G. Pirovano, A. Balzarini, B. Bessagnet, C. Emery, G. Kallos, F. Meleux, C. Mitsakou, U. Nopmongcol, G.M. Riva, G. Yarwood,
1268 Investigating impacts of chemistry and transport model formulation on model performance at European scale,
1269 Atmospheric Environment, Volume 53, June 2012, Pages 93-109, ISSN 1352-2310,
1270 http://dx.doi.org/10.1016/j.atmosenv.2011.12.052.
1271 Galmarini, S., Bianconi, R., Appel, W., Solazzo, E., et al., 2012. ENSEMBLE and AMET: two systems and approaches to a
1272 harmonised, simplified and efficient assistance to air quality model developments and evaluation. Atmos. Environ. 53,
1273 51-59.
1274 Galmarini, S., Kioutsioukis, I., and Solazzo, E.: E pluribus unum: ensemble air quality predictions, Atmos. Chem. Phys., 13,
1275 7153–7182, doi:10.5194/acp-13-7153-2013, 2013
1276 Gantt, B, J. T. Kelly and J. O. Bash, 2015: Updating sea spray aerosol emissions in the Community Multiscale Air Quality
1277 (CMAQ) model version 5.0.2, Geosci. Model Dev. Discuss., 8, 39053939, doi:10.5194/gmdd-8-3905-2015.
1278 Geels, C., Christensen, J.H., Frohn, L.M., Brandt, J., 2002. Simulating spatiotemporal variations of atmospheric CO2 using a
1279 nested hemispheric model. Physics and Chemistry of the Earth, Parts A/B/C 27 (35), 1495e1505.
1280 Geyer, B. (2014). High-resolution atmospheric reconstruction for Europe 1948–2012: coastDat2, Earth Syst. Sci. Data, 6,
1281 147–164, doi:10.5194/essd-6-147-2014
1282 Gilliam, R.C., J. M. Godowitch, and S. T. Rao, Improving the horizontal transport in the lower troposphere with four
1283 dimensional data assimilation, Atmospheric Environment, Volume 53, June 2012, Pages 186-201, ISSN 1352-2310,
1284 http://dx.doi.org/10.1016/j.atmosenv.2011.10.064
1285 Gilliam, R.C., Pleim, J.E., 2010. Performance assessment of new land surface and planetary boundary layer physics in the
1286 WRF-ARW. J. Appl. Meteor. Climatol. 49, 760-774
1287 Ginoux, P., et al. (2001), Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res.,
1288 106(D17), 20,255–20,273.
1289 Giordano, L., Brunner, D., Flemming, J., Hogrefe, C., Im, U., Bianconi, R., and et al., 2015. Assessment of the MACC
1290 reanalysis and its influence as chemical boundary conditions for regional air quality modelling in AQMEII-2.
1291 Atmospheric Environment 115, 371-388.
1292 Gong, S.L., 2003. A parameterization of sea-salt aerosol source function for sub- and super-micron particles. Global
1293 Biogeochem. Cycles 17, 1097-1104.
1294 Grell, G. A., Peckham, S. E., McKeen, S., Schmitz, R., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled ''online''
1295 chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, doi:10.1016/j.atmosenv.2005.04.027, 2005.
1296 Grell, G.A., Freitas, S.R., 2014. A scale and aerosol aware stochastic convective parameterization for weather and air quality
1297 modelling. Atmos. Chem. Phys., 14, 5233–5250, doi:10.5194/acp-14-5233-2014
1298 Grell, G.A., Devenyi, D., 2002. A generalized approach to parameterizing convection combining ensemble and data
1299 assimilation techniques, Geophysical Research Letters, doi: 10.1029/2002GL015311.
1300 Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions
1301 of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling
1302 biogenic emissions, Geosci. Model Dev., 5, 1471-1492, doi:10.5194/gmd-5-1471-2012, 2012.
1303 Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene
1304 emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210,
1305 doi:10.5194/acp-6-3181-2006, 2006.
1306 Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance
1307 criteria: implications for improving hydrological modelling. Journal of Hydrology 377, 80-91.
1308 Hansen, K.M., Christensen, J.H., Brandt, J., Frohn, L.M., Geels, C., 2004. Modelling atmospheric transport of a-
1309 hexachlorocyclohexane in the Northern Hemisphere with a 3-D dynamic model: DEHM-POP. Atmospheric Chemistry
1310 and Physics 4, 1125e1137.

1311 Hendriks, C., Kranenburg, R., Kuenen, J.J.P. ,Van den Bril, B., Verguts, V., Schaap, M. Ammonia emission time profiles based
1312 on manure transport data improve ammonia modelling across north western Europe, Atmos. Environment., Volume
1313 131, Pages 83-96 (April 2016)
1314 Hogrefe, C., Rao, S. T., Zurbenko, I. G., Porter, P. S., 2000. Interpreting the information in ozone observations and model
1315 predictions relevant to regulatory policies in the Eastern United States, B. Am. Meteorol. Soc., 81, 2083-2106,
1316 doi:0.1175/1520-0477(2000)0812.3.CO;2.
1317 Hogrefe, C., Hao, W., Zalewsky, E. E., Ku, J.-Y., Lynn, B., Rosenzweig, C., Schultz, M. G., Rast, S., Newchurch, M. J., Wang, L.,
1318 Kinney, P. L., and Sistla, G. 2011. An analysis of long-term regional-scale ozone simulations over the Northeastern
1319 United States: Variability and trends. Atmospheric Chemistry and Physics 11, 567-582
1320 Hogrefe, C., Roselle, S., Mathur, R., Rao, S. T., Galmarini, S., 2014. Space-time analysis of the Air Quality Model Evaluation
1321 International Initiative (AQMEII) phase 1 air quality simulation, J.Air Waste Manage., 64, 388–405.
1322 Hong, S.–Y. and J.–O. J. Lim, 2006. The WRF single–moment 6–class microphysics scheme (WSM6). J. Korean Meteor. Soc.,
1323 42, 129–151.
1324 Hong, S.-Y., Dudhia, J., and Chen, S.-H., 2004. A revised approach to ice microphysical processes for the bulk
1325 parameterization of clouds and precipitation, Mon. Weather Rev., 132, 103–120.
1326 Hong, S.Y., Pan, H.L., 1996. Nonlocal boundary layer vertical diffusion in a medium- range forecast model. Mon. Weather
1327 Rev. 124 (10), 2322-2339.
1328 Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A., Collins, W.D., 2008. Radiative forcing by long-lived
1329 greenhouse gases: calculations with the AER radiative transfer models. J. Geophys. Res. 113,
1330 D13103. http://dx.doi.org/ 10.1029/2008JD009944.
1331 Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A.,Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C.,Curci, G.,
1332 Denier van der Gon, H., Flemming, J., Forkel, R.,Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak,L.,
1333 Jorba, O., Knote, C., Makar, P. A., Manders-Groot, A.,Neal, L., Pérez, J. L., Pirovano, G., Pouliot, G., San Jose, R.,
1334 Savage,N., Schroder,W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella,P., Wang, K., Werhahn, J., Wolke, R., Zabkar, R.,
1335 Zhang,Y., Zhang, J., Hogrefe, C., and Galmarini, S.: Evaluation of operational online coupled regional air quality models
1336 over Europe and North America in the context of AQMEII phase 2. Part II: particulate matter, Atmos. Environ., 115,
1337 421–441, 2015a.
1338 Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A.,Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C.,Curci, G.,
1339 Flemming, J., Forkel, R., Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote,C., Kuenen,
1340 J. J. P., Makar, P. A., Manders-Groot, A., Neal, L.,Pérez, J. L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N.,Schroder,
1341 W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P.,Werhahn, J.,Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang,J.,
1342 Hogrefe, C., and Galmarini, S.: Evaluation of operational online-coupled regional air quality models over Europe and
1343 NorthAmerica in the context of AQMEII phase 2. Part I: ozone, Atmos.Environ., 115, 404–420, 2015b.
1344 Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R.J., Errera, Q.,
1345 Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J.W., Kapsomenakis,
1346 J., Lefever, K., Leitao, J., Razinger, M., Richter, A., Schultz, M.G., Simmons, A.J., Suttie, M., Stein, O., Thepaut, J.-N.,
1347 Thouret, V., Vrekoussis, M., Zerefos, C., the MACC team, 2013. The MACC reanalysis: an 8 yr data set of atmospheric
1348 composition. Atmos. Chem. Phys. 13, 4073-4109
1349 Jacob, D. J. and Winner, D. A.: Effect of climate change on air quality, Atmospheric Environment, 43, 51 – 63, 2009
1350 Janjic, Z. I., 2002: Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso model, NCEP
1351 Office Note, No. 437, 61 pp
1352 Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa,
1353 J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.:
1354 HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air
1355 pollution, Atmos. Chem. Phys., 15, 11411-11432, doi:10.5194/acp-15-11411-2015, 2015.
1356 Jimenez, P. A. and Dudhia, J.: Improving the representation of resolved and unresolved topographic effects on surface wind
1357 in the WRF model, J. Appl. Meteor. Climatol., 51, 300–316, 2012
1358 Johnson, R.: Assessment of Bias with Emphasis on Method Comparison, Clin. Biochem., 29, S37–S42, 2008.
1359 Jones, A. M. and Harrison, R.M. 2006. Assessment of natural components of PM10 at UK urban and rural sites,
1360 *Atmospheric Environment*, 40, 7733–7741
1361 Kain, J. S., 2004. The Kain-Fritsch convective parameterization: An update, J. Appl. Meteorol., 43, 170–181
1362 Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. J. Appl. Meteor., 43, 170–181.
1363 Kajino, M., Winiwarter, W and Ueda, H. 2006. Modeling retained water content in measured aerosol mass, *Atmospheric*
1364 *Environment*, 40, 5202–5213.
1365 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J.,Zhu, Y.,
1366 Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A.,Reynolds, R.,
1367 Jenne, R., and Joseph, D., (1996). The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471
1368 Kang, D., Hogrefe, C., Foley, K. L., Napelenok, S. L., Mathur, R., and Rao, S. T.: Application of the Kolmogorov-Zurbenko filter
1369 and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model, Atmos. Environ., 80,
1370 58–69, 2013.
1371 Kim, S. W., Heckel, A., Frost, G. J., Richter, A., Gleason, J., Burrows, J. P., McKeen, S., Hsie, E. Y., Granier, C., and Trainer, M.:
1372 NO_2 columns in the western United States observed from space and simulated by a regional chemistry model and their
1373 implications for NO_x emissions, J. Geophys. Res.-Atmos., 114, D11301, doi:10.1029/2008jd011343, 2009.

1374 Kouznetsov, R. & Sofiev, M., 2012. A methodology for evaluation of vertical dispersion and dry deposition of atmospheric
1375 aerosols. Journal of Geophysical Research, 117(D01202).
1376 Kouznetsov, R. M. Sofiev, M. 2014: "Wet deposition scheme for SILAM chemical transport model" Proc. 16th Int. Conf. on
1377 Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 8-11 September 2014, Varna,
1378 Bulgaria, p 336
1379 Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: TNO-MACC_II emission inventory; a
1380 multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmos.
1381 Chem. Phys., 14, 10963-10976, doi:10.5194/acp-14-10963-2014, 2014.
1382 Lattuati, M.: Contribution `a l´´etude du bilan de l'ozone troposph ´erique `a l'interface de l'Europe et de l'Atlantique Nord:
1383 Lin, Y.L., Farley, R.D., Orville, H.D., 1993. Bulk parameterization of the snow field in a cloud model. J. Clim. Appl. Meteorol.
1384 22, 1065-1092.
1385 M. Schaap, F. Sauter, R.M.A. Timmermans, M. Roemer, G. Velders, J. Beck and P.J.H. Builtjes, The LOTOS-EUROS model:
1386 description, validation and latest developments, Int. J. Environment and Pollution, Vol. 32, No. 2, pp.270–290, 2008.
1387 M. Schaap, F. Sauter, R.M.A. Timmermans, M. Roemer, G. Velders, J. Beck and P.J.H. Builtjes, The LOTOS-EUROS model:
1388 description, validation and latest developments, Int. J. Environment and Pollution, Vol. 32, No. 2, pp.270–290, 2008.
1389 Martensson, E. M., E. D. Nilsson, G. de Leeuw, L. H. Cohen, and H.-C. Hansson (2003). "Laboratory simulations and
1390 parameterization of the primary marine aerosol production". Journal of Geophysical Research: Atmospheres 108.D9,
1391 n/a–n/a. DOI: 10.1029/2002JD002263.
1392 Mason, R., Zubrow, A., Eyth, A., 2012. Technical Support Document (TSD) Preparation of Emissions Inventories for the
1393 Version 5.0, 2007 Emissions Modeling Platform. Available at:
1394 http://epa.gov/ttn/chief/emch/2007v5/2007v5_2020base_EmisMod_TSD_13dec2012.pdf
1395 Mass, C. F. and Ovens, D.: Fixing WRF's high speed wind bias: a new subgrid scale drag parameterization and the role of
1396 detailed verification, 91st AMS Annual Meeting, Seattle, WA, available at:
1397 http://ams.confex.com/ams/91Annual/webprogram/Paper180011.html, 2011.
1398 Mellor, G. L., Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys.,
1399 20, 851–875.
1400 Menut L, B.Bessagnet, D.Khvorostyanov, M.Beekmann, N.Blond, A.Colette, I.Coll, G.Curci, G.Foret, A.Hodzic, S.Mailler,
1401 F.Meleux, J.L.Monge, I.Pison, G.Siour, S.Turquety, M.Valari, R.Vautard and M.G.Vivanco, 2013, CHIMERE 2013: a model
1402 for regional atmospheric composition modelling, Geoscientific Model Development, 6, 981-1028, doi:10.5194/gmd-6-
1403 981-2013
1404 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono and S. A. Clough, 1997: Radiative transfer for inhomogeneous
1405 atmospheres: RRTM, a validated correlated–k model for the longwave. J. Geophys. Res., 102, 16663–16682.
1406 Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous
1407 atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res.-Atmos., 102, 16663–16682,
1408 1997.
1409 Monahan, E.C., D.E. Spiel, and K.L Davidson (1986). "A model of marine aerosol generation via whitecaps and wave
1410 disruption". In: Oceanic Whitecaps and their role in air/sea exchange. Ed. by G. Monahan E.C.and Mac Niocaill. Reidel,
1411 Norwell, Mass.,USA, pp. 167–174.
1412 Morrison, H., G. Thompson, V. Tatarskii, 2009: Impact of Cloud Microphysics on the Development of Trailing Stratiform
1413 Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. Mon. Wea. Rev., 137, 991–
1414 1007. doi: http://dx.doi.org/10.1175/2008MWR2556.1
1415 Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform
1416 precipitation in a simulated squall line: comparison of one- and two-moment scheme, Mon. Weather Rev., 137, 991-
1417 1007, doi:10.1175/2008mwr2556.1, 2009.
1418 Nakanishi, M. and Niino, H.: An improved Mellor-Yamada Level-3 Model: its numerical stability and application to a
1419 regional prediction of advection fog, Bound.-Lay. Meteorol., 119, 397–407, doi:10.1007/s10546-005-9030-8, 2006.
1420 Nenes, A, C. Pilinis, and S.N. Pandis. 1999. Continued Development and Testing of a New Thermodynamic Aerosol Module
1421 for Urban and Regional Air Quality Models. Atmos. Environ. 33, 1553-1560.
1422 Otero, N., Sillmann, J., Schnell, J.L., Rust, H.W., Butler, T., 2016. Synoptic and meteorological drivers of extreme ozone
1423 concentrations over Europe. Environ. Res. Lett. 11 (2016) 024005. doi:10.1088/1748-9326/11/2/024005
1424 Otte, T. L. Pleim, J. E., 2010. The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system:
1425 updates through MCIPv3.4.1, Geosci. Model Dev., 3, 243–256, doi:10.5194/gmd-3-243-2010.
1426 Pleim, J. E. and Gilliam, R., 2009.: An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land
1427 surface model, J. Appl. Meteorol. Clim., 48, 1362–1376.
1428 Pleim, J. E. and Xiu, A., 2003. Development of a land surface model. Part II: data assimilation, J. Appl. Meteorol., 42, 1811–
1429 1822, 2003
1430 Pleim, J. E., 2006: A simple, efficient solution of flux-profile relationships in the atmospheric surface layer, J. Appl. Meteor.
1431 and Clim., 45, 341–347.
1432 Pleim, J. E., 2007b: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part II: application
1433 and evaluation in a mesoscale meteorological model, J. Appl. Meteorol. Clim., 46, 1396–1409
1434 Pleim, J., R. Gilliam, and J. Godowitch, 2010: Evaluation of PBL models compared to GABLS experiments and testing in
1435 meteorology and air quality models, presented at the 19th Symposium on Boundary Layers and Turbulence, American

Meteorological Society, August 2–6, 2010, Keystone, CO, recorded presentation available at http://ams.confex.com/ams/19Ag19BLT9Urban/recordingredirect.cgi/id/15728

Pleim, J., Ran, L., 2011. Surface Flux Modeling for Air Quality Applications. Atmosphere 2, 271-302

Pleim, J., Gilliam, R., Appel, W., Ran, L., 2016. Recent advances in modelling of the atmospheric boundary layer and land surface in the coupled WRF-CMAQ model. Air Pollution Modelling and Its Application XXIV, Springer Proceeding in Complexity, Switzerland, pp. 391-396

Pouliot, G., and et al. 2015. Analysis of the emission inventories and model-ready emission datasets of Europe and North America for phase 2 of the AQMEII project. Atmospheric Environment 115, 340-360

Poupkou, A., Giannaros, T., Markakis, K., Kioutsioukis, I., Curci, G., Melas, D., Zerefos, C., 2010. A model for European Biogenic Volatile Organic Compound emissions: Software development and first validation. Environ. Model. Softw. 25, 1845–1856. doi:10.1016/j.envsoft.2010.05.004

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., Sokhi, R., 2016. 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

Rao, S. T., Galmarini, S., and Puckett, K.: Air quality model evaluation international initiative (AQMEII), B. Am. Meteorol. Soc., 92, 23–30, doi:10.1175/2010BAMS3069.1, 2011.

Rao, S. T., Zurbenko, I. G., Neagu, R., Porter, P. S., Ku, J. Y., and Henry, R. F.: Space and time scales in ambient ozone data, B. Am. Meteorol. Soc., 78, 2153e2166, doi:10.1175/1520-0477(1997)078<2153:SATSIA>2.0.CO;2, 1997.

Ritter, B. and Geleyn, J. F., (1992). A comprehensive radiation scheme for numerical weather prediction models withpotential applications in climate simulations, Mon. Weather Rev., 120, 303–325, doi:10.1175/1520-0493 Rockel, B., Will, A., and Hense, A., (2008). The Regional Climate Model COSMO-CLM (CCLM). Meteorol.Z. 17, 347–248.

Rockel, B., Will, A., and Hense, A., (2008). The Regional Climate Model COSMO-CLM (CCLM). Meteorol. Z. 17, 347–248.

Sarwar, G., Appel, K. W., Carlton, A. G., Mathur, R., Schere, K., Zhang, R., and Majeed, M. A., 2011a: Impact of a new condensed toluene mechanism on air quality model predictions in the US, Geosci. Model Dev., 4, 183-193, doi:10.5194/gmd-4-183-2011

Sarwar, G., K. Fahey, S. Napelenok, S. Roselle, R. Mathur, 2011b. Examining the impact of CMAQ model updates on aerosol sulfate predictions, Presentation at the 10th Annual CMAS Models-3 User's Conference, October 2011, Chapel Hill, NC, available online at http://www.cmascenter.org/conference/2011/slides/sarwar_examining_impact_2011.pdf

Sarwar, G.; Luecken, D. & Yarwood, G. Borrego, C. & Renner, E., (2007). (Eds.) Chapter 2.9 Developing and implementing an updated chlorine chemistry into the community multiscale air quality model Air Pollution Modeling and Its Application XVIII, Elsevier, 2007, 6, 168 - 176, doi: 10.1016/S1474-8177(07)06029-9

Sauter, F., Swaluw, E. van der, Manders-Groot, A., Wichink Kruit, R, Segers, A., Eskes, H., (2012). LOTOS-EUROS v1.8 Reference Guide, TNO Report TNO-060-UT-2012-01451, Utrecht, The Netherlands,

Sauter, F., Van der Swaluw, E., Manders-Groot, A., Wichink Kruit, R., Segers, A., Eskes, H., 2012. TNO-060-UT-2012-01451LOTOS-EUROS v. 1.8 Reference Guide. TNO, Utrecht, The Netherlands

Schaettler, U., Doms, G., and Schraff, C., (2008). A Description of the Nonhydrostatic Regional COSMO-Model Part VII: User's Guide, Tech. rep., Deutscher Wetterdienst.

Schaettler, U., Doms, G., and Schraff, C., (2008). A Description of the Nonhydrostatic Regional COSMO-Model Part VII: User's Guide, Tech. rep., Deutscher Wetterdienst.

Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A., 2001. Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. J. Geophys. Res. 106 (D22), 28275-28293.

Schrodin, R. and Heise, E., (2001). The multi-layer-version of the DWD soil model TERRA/LM,Tech. Rep, Consortium for Small-Scale Modelling (COSMO), available at: http: //www.cosmo-model.org/content/model/documentation/techReports/docs/techReport02.pdf (last access: 8 April 2015)

Schrodin, R. and Heise, E., (2001). The multi-layer-version of the DWD soil model TERRA/LM,Tech. Rep, Consortium for Small-Scale Modelling (COSMO), available at: http: //www.cosmomodel. org/content/model/documentation/techReports/docs/techReport02.pdf (last access: 8 April 2015)

Schwede, D., Pouliot, G. A., and Pierce, T.: Changes to the Biogenic Emissions Inventory System Version 3 (BEIS3), In Proceedings of the 4th CMAS Models-3 Users' Conference, Chapel Hill, NC, 26–28 September 2005

Seifert, A. and Beheng, K. D., (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection, Atmos. Res., 59–60, 265–281, doi: 10.1016/S0169-8095(01)00126-0

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

Shrivastava, M., Fast, J., Easter, R., Gustafson, W.I., Zaveri, R. a., Jimenez, J.L., Saide, P., Hodzic, a., 2011. Modeling organic aerosols in a megacity: comparison of simple and complex representations of the volatility basis set approach. Atmos. Chem. Phys. 11, 6639–6662. doi:10.5194/acp-11-6639-2011

Simon, H., Bhave, P.V.; Simulating the degree of oxidation in atmospheric organic particles. Environ. Sci. Technol. 2012, 46, 331-339

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

1499 Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting
1500 applications, J. Comp. Phys., 227, 3465–3485, 2008.
1501 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.:
1502 A description of the Advanced Research WRF version 3, National Center for Atmospheric Research Tech. Note,
1503 NCAR/TN- 475+STR, 113 pp., 2008
1504 Soares, J., Sofiev, M., Hakkarainen, J., 2015. Uncertainties of wild-land fi res emission in AQMEII phase 2 case study. Atmos.
1505 Environ. doi:10.1016/j.atmosenv.2015.01.068
1506 Sofiev, M., 2000. A model for the evaluation of long-term airborne pollution transport at regional and continental scales.
1507 Atmos. Environ. 34, 2481–2493. doi:10.1016/S1352-2310(99)00415-X
1508 Sofiev, M., Soares, J., Prank, M., de Leeuw, G., Kukkonen, J., 2011. A regional-to-global model of emission and transport of
1509 sea salt particles in the atmosphere. J. Geophys. Res. 116, 25. doi:10.1029/2010JD014713
1510 Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., and Genikhovich, E.: Construction of the SILAM Eulerian
1511 atmospheric dispersion model based on the advection algorithm of Michael Galperin, Geosci. Model Dev., 8, 3497-
1512 3522, doi:10.5194/gmd-8-3497-2015, 2015.
1513 Solazzo, E., Bianconi, R., Pirovano, G., Moran, M. D., Vautard, R., Hogrefe, C., Appel, K. W., Matthias, V., Grossi, P.,
1514 Bessagnet, B., Brandt, J., Chemel, C., Christensen, J. H., Forkel, R., Francis, X. V., Hansen, A. B., McKeen, S., Nopmongcol,
1515 U., Prank, M., Sartelet, K. N., Segers, A., Silver, J. D., Yarwood, G., Werhahn, J., Zhang, J., Rao, S. T., and Galmarini, S.:
1516 Evaluating the capability of regional-scale air quality models to capture the vertical distribution of pollutants, Geosci.
1517 Model Dev., 6, 791–818, doi:10.5194/gmd-6-791-2013, 2013
1518 Solazzo, E., Galmarini, S., 2016. Error apportionment for atmospheric chemistry transport models - a new approach to
1519 model evaluation. Atmospheric Chemistry and Physics 16, 6263-6283
1520 Solazzo,E., Galmarini, S., 2015. Comparing apples with apples: using spatially distributed time series of monitoring data for
1521 model evaluation. Atmospheric Environment 112, 234-245
1522 Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.), Contribution of
1523 Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007,
1524 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
1525 Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordonez, C., Prevot, A. S. H., and
1526 Hueglin,C.: Nitrogen dioxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques,
1527 J. Geophys. Res.,112, D11307, doi:10.1029/2006JD007971, 2007
1528 Steppeler, J., Doms, G., Schattler, U., Bitzer, H. W., Gassmann, A., Damrath, U., and Gregoric, G., (2003). Meso-gamma scale
1529 forecasts using the nonhydrostatic model LM, Meteorol. Atmos. Phys., 82, 75–96, doi:10.1007/s00703-001-0592-9
1530 Stockwell, W. R., Kirchner, F. K., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling,
1531 J. Geophys. Res., 102, 25847–25879, doi:10.1029/97JD00849, 1997.
1532 Stockwell, W.R., Middleton, P., Chang, J.S., Tang, X., 1990. The second generation regional acid deposition model chemical
1533 mechanism for regional air quality modeling. J. Geophys. Res. 95 (D10),
1534 16343e16367. http://dx.doi.org/10.1029/ JD095iD10p16343.
1535 Strader, R., F. Lurmann, and S.N. Pandis. 1999. Evaluation of secondary organic aerosol formation in winter. Atmos.
1536 Environ., 33, 4849-4863.
1537 Tanaka, P. L.; Allen, D. T.; McDonald-Buller, E. C.; Chang, S.; Kimura, Y.; Mullins, C. B.; Yarwood, G. & Neece, J. D., (2003),
1538 Development of a chlorine mechanism for use in the carbon bond IV chemistry model J. Geophys. Res. Atmos., 2003,
1539 108, doi:10.1029/2002JD002432.
1540 Tiedtke, M., (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Weather
1541 Rev., 117, 1779–1800, doi:10.1175/1520-0493(1989)117
1542 Tong, D. Q., G.E., Bowker, S., He, D. W., Byun, R. Mathur, D. A. Gillette, 2011: Development of a Windblown Dust Module
1543 within the Community Multi-scale Air Quality (CMAQ) Model: Description and Preliminary Applications in the
1544 Continental United States, submitted, J. Geophys. Res., 2011.
1545 Tsyro, S.G. 2005. To what extent can aerosol water explain the discrepancy between model calculated and gravimetric
1546 PM10 and PM2.5? *Atmos. Chem. Phys*., 5, 515–532.
1547 Tuccella, P., Curci, G., Visconti, G., Bessagnet, B., Menut, L., Park, R. J. (2012a), Modelling of gas and aerosol with
1548 WRF/Chem over Europe: evaluation and sensitivity study, J. Geophys. Res., 117, D03303, doi:10.1029/2011JD016302
1549 Tuccella, P., Curci, G., Visconti, G., Bessagnet, B., Menut, L., and Park, R. J.: Modeling of gas and aerosol with WRF/Chem
1550 over Europe: Evaluation and sensitivity study, J. Geophys. Res., 117, D03303, doi:10.1029/2011JD016302, 2012b
1551 Tuccella, P., Curci, G., Grell, G. A., Visconti, G., Crumeyrolle, S., Schwarzenboeck, A., and Mensah, A. A.: A new chemistry
1552 option in WRF-Chem v. 3.4 for the simulation of direct and indirect aerosol effects using VBS: evaluation against
1553 IMPACT-EUCAARI data, Geosci. Model Dev., 8, 2749-2776, doi:10.5194/gmd-8-2749-2015, 2015.
1554 Valmartin, M., Heald, C. L., and Arnold, S. R., 2014. Coupling dry deposition to vegetation phenology in the Community
1555 Earth System Model: Implications for the simulation of surface O3, Geophys. Res. Lett., 41, 2988–2996.
1556 Vestreng, V., Ntziachristos, L., Semb, A., Reis, S., Isaksen, I. S. A., and Tarrasón, L., 2009. Evolution of NOx emissions in
1557 Europe with focus on road transport control measures, Atmos. Chem. Phys. 9, 1503–1520, doi:10.5194/acp-9-1503-
1558 2009.
1559 Vukovich, J. and Pierce, T.: The Implementation of BEIS3 within the SMOKE Modeling Framework, in: Proceedings of the
1560 11th International Emissions Inventory Conference, Atlanta, Georgia, available at:
1561 www.epa.gov/ttn/chief/conference/ei11/modeling/vukovich.pdf, 15–18 April 2002

1562 Walcek, C. J. and Taylor, G. R.: A theoretical method for computing vertical distributions of acidity and sulfate production
1563 within cumulus clouds, J. Atmos Sci., 43, 339-355, doi: http://dx.doi.org/10.1175/1520-
1564 0469(1986)043<0339:ATMFCV>2.0.CO;2, 1986.
1565 Wesely M. L. (1989) Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models.
1566 Atmospheric Environment, 23, 1293-1304
1567 Wesely, M. L., and Hicks, B. B.: A review of the current status of knowledge on dry deposition, Atmos. Environ., 34(12-14),
1568 2261-2281, doi:10.1016/S1352-2310(99)00467-7, 2000.
1569 Wesely, M.L., Hicks, B.B., 2000. Atmospheric Environment 34, 2261-2282.
1570 Whitten, G.Z., Heo,G., Kimura, Y., McDonald-Buller, E. Allen, D.T., Carter, W.P.L., Yarwood, G. (2010). A new condensed
1571 toluene mechanism for Carbon Bond: CB05-TU. Atmospheric Environment, 44, 5346-5355.
1572 Wild, O., X. Zhu, and M. J. Prather (2000), Fast-J: Accurate simulation of in- and below cloud photolysis in tropospheric
1573 chemical models, J. Atmos. Chem., 37, 245–282, doi:10.1023/A:1006415919030.
1574 Wise, E. K. and Comrie, A. C.: Extending the KZ filter: application to ozone, particulate matter, and meteorological trends, J.
1575 Air Waste Manage., 55, 1208e1216, doi:10.1080/10473289.2005.10464718, 2005.
1576 Xiu, A. and Pleim, J. E.: Development of a land surface model. Part I: application in a mesoscale meteorological model, J.
1577 Appl. Meteorol., 40, 192–209, 2001
1578 Yarwood, G., Rao, S., Yocke, M., and Whitten, G.: Updates to the Carbon Bond Chemical Mechanism: CB05, Final Report to
1579 the US EPA, RT-0400675, available at: http://www.camx.com/publ/pdfs/cb05_final_report_120805.pdf (last access: 16
1580 December 2014), 2005.
1581 Zanten, M.C. van, F.J. Sauter, R.J. Wichink Kruit, J.A. van Jaarsveld, W.A.J. van Pul, Wichink Kruit, R.J., 2010. Description of
1582 the DEPAC module. Dry deposition modelling with DEPAC GCN2010. Tech. rep. Bilthoven, The Nehterlands:
1583 Rijksinstituut voor volksgezondheid en Milieu, RIVM report 680180001. URL: http : / / www . rivm . nl
1584 /Documenten_en_publicaties/Wetenschappelijk/Rapporten/2010/oktober/Description_of_the_DEPAC_module_Dry_d
1585 eposition_modelling_with_DEPAC_GCN2010
1586 Zare, A., J. H. Christensen, A. Gross, P. Irannejad, M. Glasius and J. Brandt, 2014: Quantifying the contributions of natural
1587 emissions to ozone and total fine PM concentrations in the Northern Hemisphere. Atmospheric Chemistry and Physics,
1588 Vol. 14, pp. 2735-2756, 2014, www.atmos-chem-phys.net/14/2735/2014/. doi:10.5194/acp-14-2735-2014.
1589 Zhang, L., J. R. Brook, and R. Vet. 2003. A revised parameterization for gaseous dry deposition in air-quality models. Atmos.
1590 Chem. Phys., 3, 2067–2082.
1591 Zhang, L., S. Gong, J. Padro, L. Barrie. 2001. A size-segregated particle dry deposition scheme for an atmospheric aerosol
1592 module. Atmos. Environ., 35, 549-560
1593 Zurbenko, I. G.: The Spectral Analysis of Time Series, North-Holland, Amsterdam, 236 pp., 1986.
1594

1595

1596

1597 # TABLES

1598

1599

1600

1601

1602

1603

1604

1605

1606

1607

1608

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

1609 **TABLE 1. PARTICIPATING MODELLING SYSTEMS AND KEY FEATURES**

Operated by	Modelling system	Horizontal grid	Vertical grid	Deposition scheme	Global meteo data provider	NO_x emission share of NO and NO_2	Gaseous chemistry module
EUROPEAN DOMAIN							
Finnish Meteorological Institute	ECMWF-SILAM_H, SILAM_M	0.25 x 0.25 deg Lat x Lon	12 uneven layers up to 13km. First layer ~30m	Dry: Kouznetsov and Sofiev (2012) Wet: Kouznetsov and Sofiev (2014)	ECMWF (nudging within the PBL)	90/10	CBM-IV
Netherlands Organization for Applied Scientific Research	ECMWF-L.-EUROS	0.5 x 0.25 deg Lat x Lon	Surface layer (~25m depth), mixing layer, 2 reservoir layers up to 3.5km.	Wet: below-cloud scavening Dry: Zhang et al. (2001) for particles, Depac (Zanten et al., 2012) for gases	Direct interpolation from ECMWF	97/3	CBM-IV
University of L'Aquila	WRF-WRF/Chem1	270x225 cells, 23 km	33 levels up to 50hPa. 12 layers below 1km. First layer ~12m	Dry: Wesely (1989) Wet: Grell and Freitas (2014)	ECMWF (nudging above the PBL)	95/5	RACM-ESRL
University of Murcia	WRF-WRF/Chem2	270 x 225 cells,t 23 km x 23 km	33 levels, from ~24m to 50hPa	Dry: Wesley resistance approach, (Wesley, 1989) Wet: Grid scale wet deposition (Easter et al, 2004) and convective wet deposition	ECMWF (nudging above the PBL)	90/10	RADM2
Ricerca Sistema Energetico	WRF-CAMx	265x220 cells, 23 km x 23 km	14 layers up to 8km. First layer ~25m.	Dry: Resistance model for gases (Zhang et al., 2003) and aerosols (Zhang et al., 2001) Wet: Scavenging model for gases and aerosols (Seinfeld and Pandis, 1998)	ECMWF (nudging within the PBL)	95/5	CB05
University of Aarhus	WRF-DEHM	50 km x 50 km	29 layers up to 100hPa	Wet and dry as in Simpson et al. (2003)	ECMWF (no nudging within the PBL)	90/10	Brandt et al. (2012)
Istanbul Technical University	WRF-CMAQ1	184 x 156 cells, 30 km x 30 km	24 layers up to 10hPa	Wet and Dry as in Foley et al. (2010)	NCEP (nudging within PBL)	95/5	CB05
Kings College	WRF-CMAQ4	15 km x 15 km	23 layers up to 100hPa, 7 layer below 1km. First layer ~14m	Wet: Taken from the RADM (Chang et al., 1987) Dry: Electrical resistance analog model	NCEP (Nudging within the PBL)	90/10	CB05
Ricardo E&E	WRF-CMAQ2	30 km x 30 km	23 layers up to 100hPa, 7 layers below 1km. First layer ~15m	Wet: Byun and Schere (2006) Dry: Pleim and Ran (2011)	NCEP (nudging above the PBL)	Road transport: 86/14; non-road: 95/5	CB05-TUCL

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

Helmholtz-Zentrum Geesthacht	CCLM-CMAQ	24 km x 24 km	30 vertical layers from ~40m to 50hPa	Wet: Byun and Schere (2006) Dry: Pleim and Ran (2011)	NCEP (spectral nudging above free troposhere)	90/10	CB05-TUCL
University of Hertfordshire	WRF-CMAQ3	18 km x 18 km	35 vertical layers from ~20m to ~16km	Dry: resistance analogy model (Wesley, 1989). Wet: Asymmetric Convective model algorithm in CMAQ cloud module	ECMWF (nudging above PBL)	90/10	CB05-TUCL
INERIS/CIEMAT	ECMWF-Chimere_H Chimere_M	0.25 x 0.25 deg Lat x Lon	9 layers up to 500hPa. First layer ~20m	Wet: in-cloud and sub-cloud scavenging for gases and aerosols (Menut et al. 2013) Dry: resistance approach as Emberson (2000a,b)	Direct interpolation from ECMWF	95% NO 4.5% NO_2 0.5% HONO	MELCHIOR2
NORTH AMERICAN DOMAIN							
Helmholtz-Zentrum Geesthacht	CCLM-CMAQ	24 km x 24 km	30 vertical layers from ~40m to 50hPa.	Wet: Byun and Schere (2006) Dry: Pleim and Ran (2011)	NCEP (spectral nudging above free troposhere)	90/10	CB05-TUCL
Environmental Protection Agency of the USA	WRF-CMAQ	459x299 cells 12 km x 12 km	35 layers, up to 50hPa. First layer ~19m	Wet: Byun and Schere (2006) Dry: Pleim and Ran (2011)	NCEP (nudging above the PBL)	90/10 Calculated by MOVES for transport	CB05-TUCL
RAMBOLL Environ	WRF-CAMx	459x299 cells, 12 Km x 12 km	26 layers up to 97.5hPa	Dry: Resistance model for gases (Zhang et al., 2003) Wet: Scavenging model for gases and aerosols (Seinfeld and Pandis, 1998)	NCEP (nudging above the PBL)	90/10	CB05
University of Aarhus	WRF-DEHM	50 km x 50 km	29 layers up to 100hPa	Wet and dry as in Simpson et al. (2003)	Direct interpolation from ECMWF	90/10	Brandt et al. (2012)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

1614 **TABLE 2. EXTENSION OF THE SUB-REGIONS AND NUMBER OF RECEPTORS USED IN THE ANALYSIS**

	EU1/NA1 42–57.2N; -9–1.3W / 40–49.5; -83– -66W	EU2/NA2 47.5–56N; 1.3–18W / 30–38N; -91–-75W	EU3/NA3 43.5–46N; 7–14W / 33.5–43; -124–-118.5W	EU/NA 30–65N; -10–33W / 26–51N; -125–-55W
Ozone	134/165	352/63	120/93	972/667
CO	32/29	91/8	70/12	418/103
NO (EU)	27	367	161	836
NO_2	149/97	529/21	176/54	1390/340
SO_2	96/69	296/3	55/3	865/141
PM_{10} (EU) **$PM_{2.5}$ (NA)**	47 89	347 9	2 22	619 226
WS	168/229	305/245	5/59	827/1721
Temp	168/232	305/243	5/46	830/1546

1615

1616 **TABLE 3. SUMMARY OF OZONDESONDES DATA FOR OZONE**

EU			
Station	**O_3 Records**	**Period**	**Local time**
316	52	Year(4-5 launches per month)	11-12
308	52	Year(4-5 launches per month)	10-11
318	37	Year(3-4 launches per month, mostly winter and autumn)	11-12
242	46	January-April(10-12 launches per month)	11-12
156	144	Year(12 launches per month)	10-12
099	66	Year(5-6 launches per month)	Mostly early morning 4-6
053	149	Year(11-13 launches per month)	11-12
043	51	Year(4-5 launches per month)	11-12
NA			
021	44	Year(3-4 launches per month)	11-12
107	54	Year(4-5 launches per month)	16-20
338	50	Year(2-4 per month; 17 in July; none in September)	14-15 July-August 17-18 other months
456	57	2-5 per month; 25 in July	17-18
457	75	Year(2-5 per month; 18-20 in May-June)	23-00
458	71	Year(3-8 per month; 20 in July)	23-00

1617

1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

FIGURES

Figure 1. Sub-regions of the two continental domains (a) EU; b) NA). Overlaid are the ozone monitoring stations classified based on the network

Figure 2. RMSE for a) Temp and b) WS in Europe

FIGURE 3 RMSE for a) Temp and b) WS in North America

Figure 4. Mean Bias (mod – obs) for the vertical profiles of Wind Speed measured by ozonesondes launched from the European locations indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel

Figure 5. Mean Bias (mod – obs) for the vertical profiles of Temperature measured by ozonesondes launched from the European locations indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel

Figure 6. Mean Bias (mod – obs) for the vertical profiles of Wind Speed measured by ozonesondes launched from the North American locations indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel

Figure 7. Mean Bias (mod – obs) for the vertical profiles of Temperature measured by ozonesondes launched from the North American locations indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel

Figure 8. Cumulated modelled deposition per unit area over the continental regions of a) EU and b) NA for the full year of 2010. The boxes extend between the minimum and the 5^{th} percentile, while the maximum is reported by the number at the top of each box. results are displayed for the models and species for which data have been made available

Figure 9. RMSE (ppb) for CO by spectral component and season (panel *a* for Europe and *b* for North America). FT is the full (un-filtered) time series, LT, SY, DU, are the Long Term, Synoptic and diurnal components, respectively.

Figure 10. MSE (ppb^2) breakdown into bias squared, variance and *mMSE* for the spectral components of the spatial average time series of CO during the months of December, January, and February (DJF), based on EQ.6. The bias is entirely accounted for by the LT component. The signs within the bias and variance portion of the bars indicate model overestimation (+) or underestimation (-) of the bias and variance. The colour of the *mMSE* share of the error is coded based on the values of *r*, the correlation coefficient, according to the colour scale at the bottom of each plot. Top panel: EU; lower panel: NA. Similar plots for the other two sub-regions are reported in the supplementary material.

Figure 11. RMSE variation between the 's20%' scenario (anthropogenic emission and boundary condition reduced by 20%) and the base case for CO in EU2

Figure 12. Top panel: as in Figure 9 for NO (EU only). Lower panel: as in Figure 10 for NO (EU only)

Figure 13. RMSE variation between the 's20%' scenario (anthropogenic emission and boundary condition reduced by 20%) and the base case for anthropogenic NO (aNO) in eu2

Figure 14. As in Figure 9 for NO_2.

Figure 15. As in Figure 10 for NO_2 in EU2. Upper panel: Urban sites only (223 stations); lower panel: Rural sites only (159 stations)

Figure 16. As in Figure 10 for NO_2 in NA1. Upper panel: Urban sites only (39 stations); Lower panel: Rural sites only (10 stations).

Figure 17. As in Figure 9 for ozone

Figure 18. As in Figure 10 for ozone during the months from May to September

1677 Figure 19. Ozone mixing-ratio profiles measured by ozonesondes launched from the European location indicated on the
1678 inset map (lower-right corner) of each panel. The profiles are time-averaged over the number of hourly records reported in
1679 the parenthesis at the top of each panel. Legend as in the first panel.

1680 Figure 20. As in Figure 19 for North America

1681 Figure 21. Ozone vs NO modelled mean bias for the EU2 sub-region, color-coded by temperature bias and symbols
1682 according to the NO_x emission fraction of NO and NO_2. Each point represents a model. *a)* winter months and *b)* summer
1683 months.

1684 Figure 22. As in Figure 9 for SO_2

1685 Figure 23. As in Figure 10 for SO_2

1686 Figure 24. As in Figure 9 for PM_{10} in Europe (error units in $\mu g/m^3$)

1687 Figure 25. As in Figure 9 for $PM_{2.5}$ in North America (error units in $\mu g/m^3$)

1688 Figure 26. As in Figure 10 for PM_{10} in Europe (error units in $\mu g/m^3$)

1689 Figure 27. As in Figure 10 for $PM_{2.5}$ in North America (error units in $\mu g/m^3$)

1690 Figure 28. Spatial map of the ozone monitoring stations coloured based on the 'delta hour' values, i.e. the difference in
1691 hours between the zero of the autocorrelation function (acf) for the modelled ozone minus the zero of the acf of the
1692 observed one. The acf is calculated on the long term component for the months of May to September. Negative values
1693 indicate too short memory and excess of removal (vice-versa for positive values). The box on the right summarises the
1694 delta hour percentile distribution.

1695 Figure 29. As in Figure 28 for North America.

1696

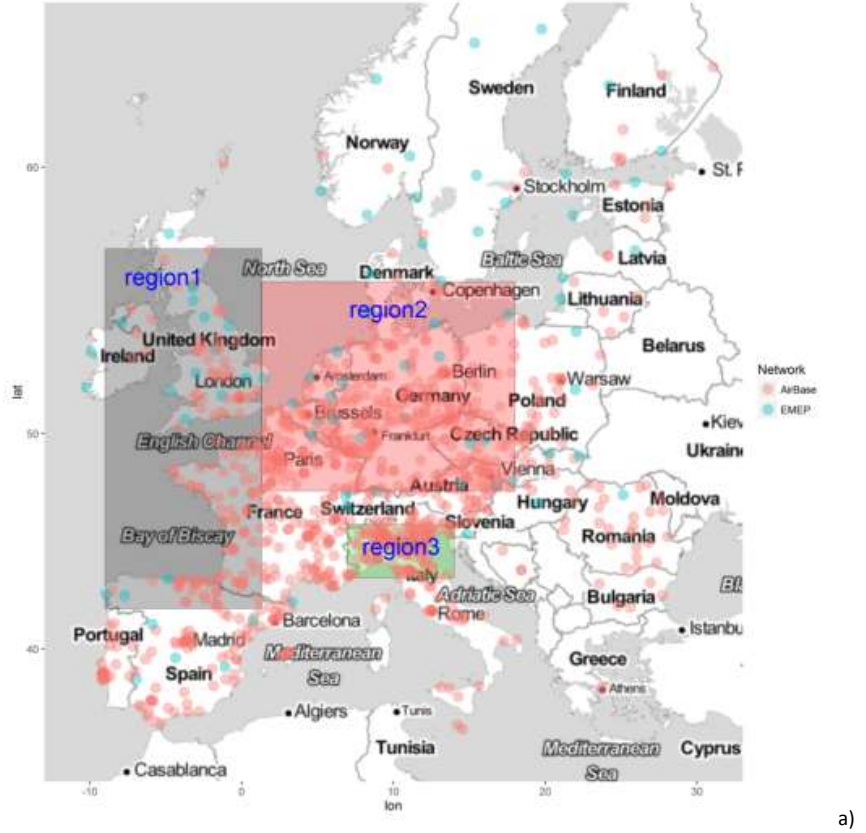
1697

1698

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

FIGURES



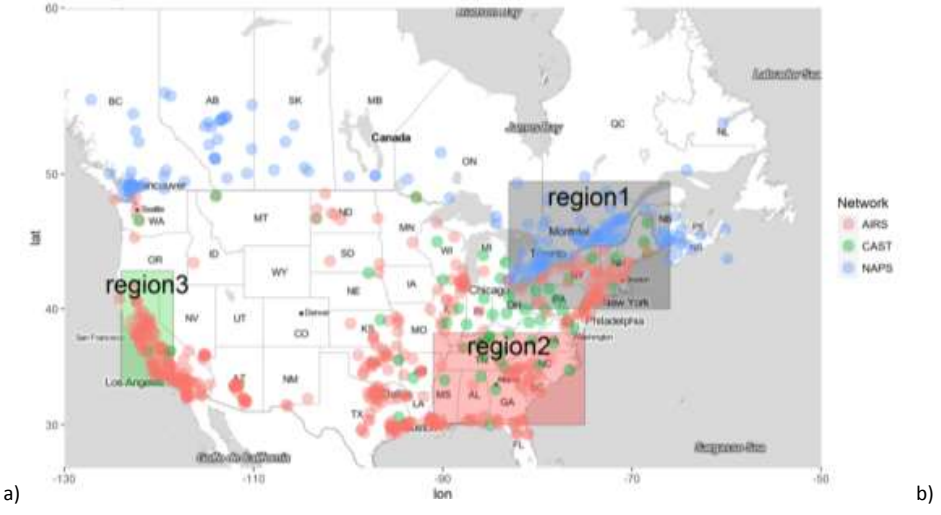
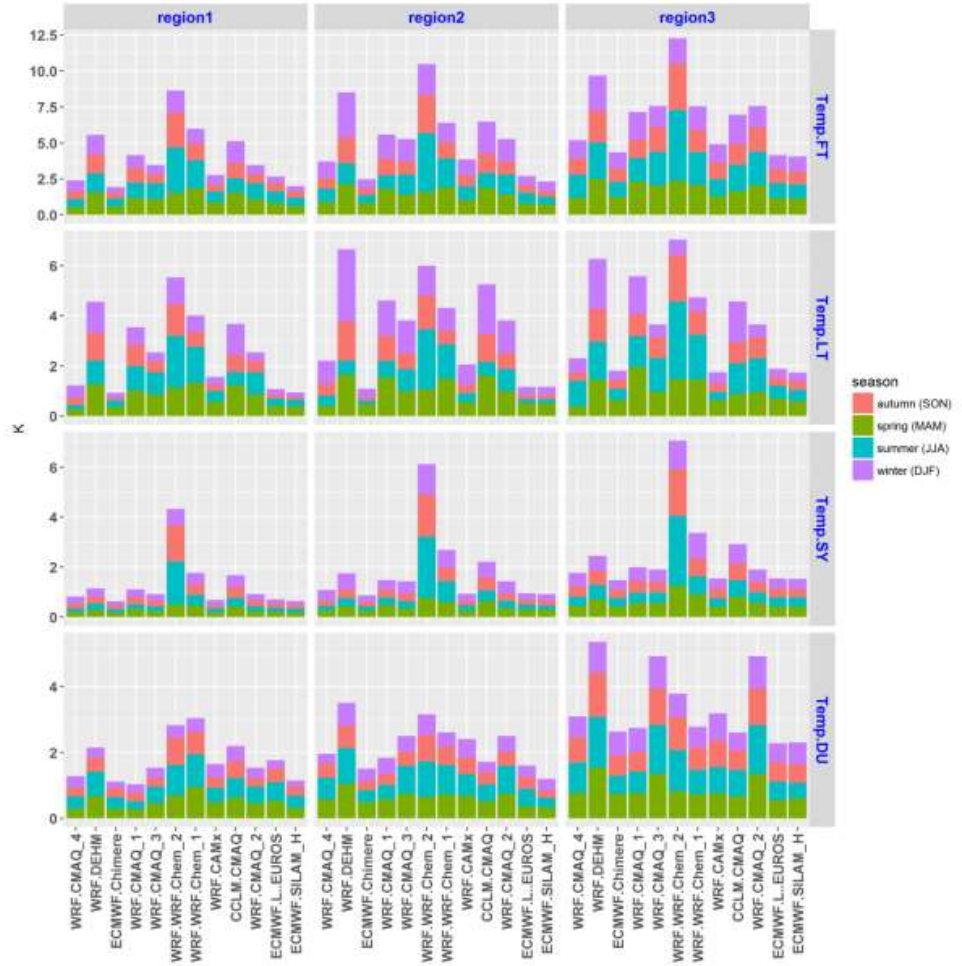


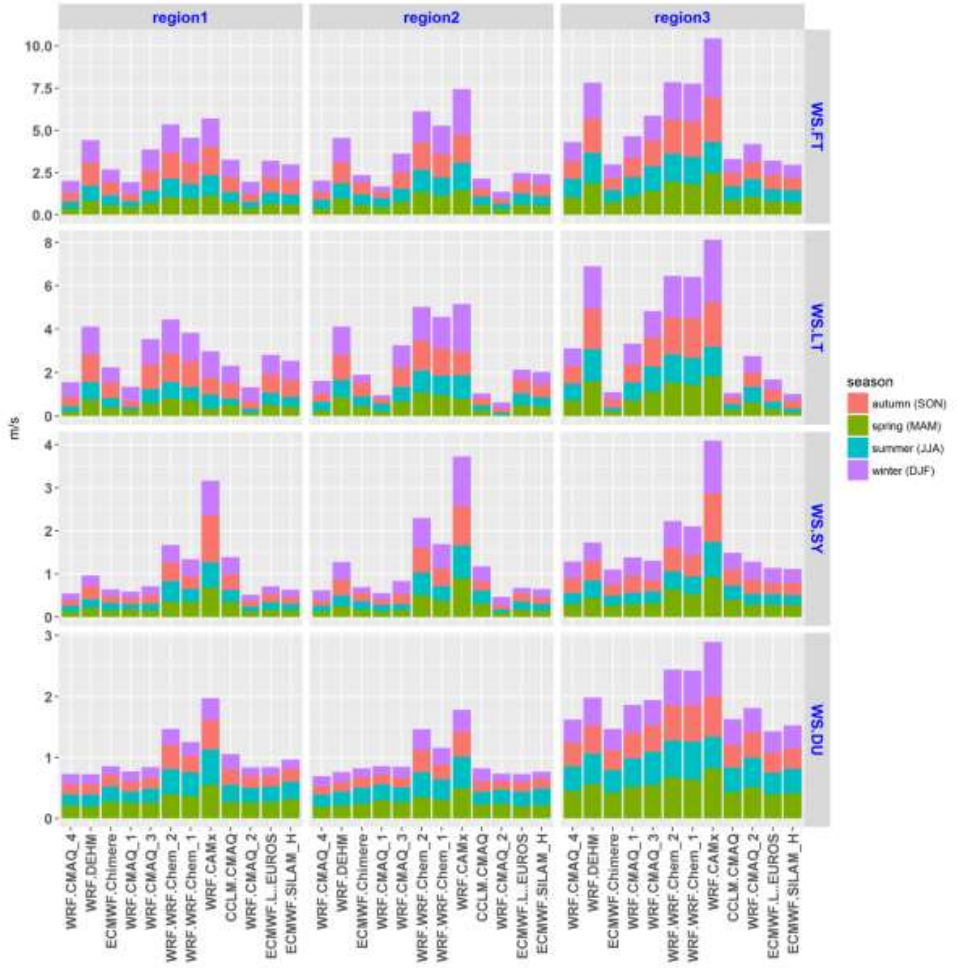
FIGURE 1. Sub-regions of the two continental domains (a) EU; b) NA). Overlaid are the ozone monitoring stations classified based on the network

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)

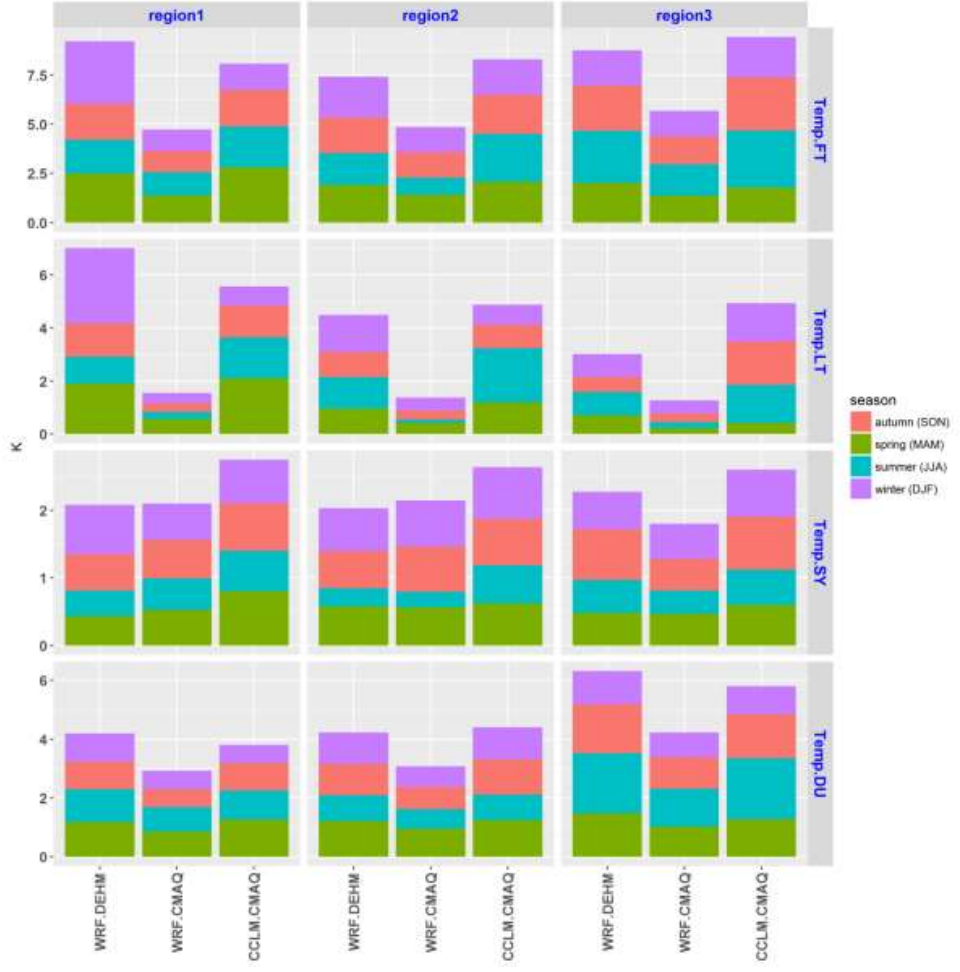
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



b)

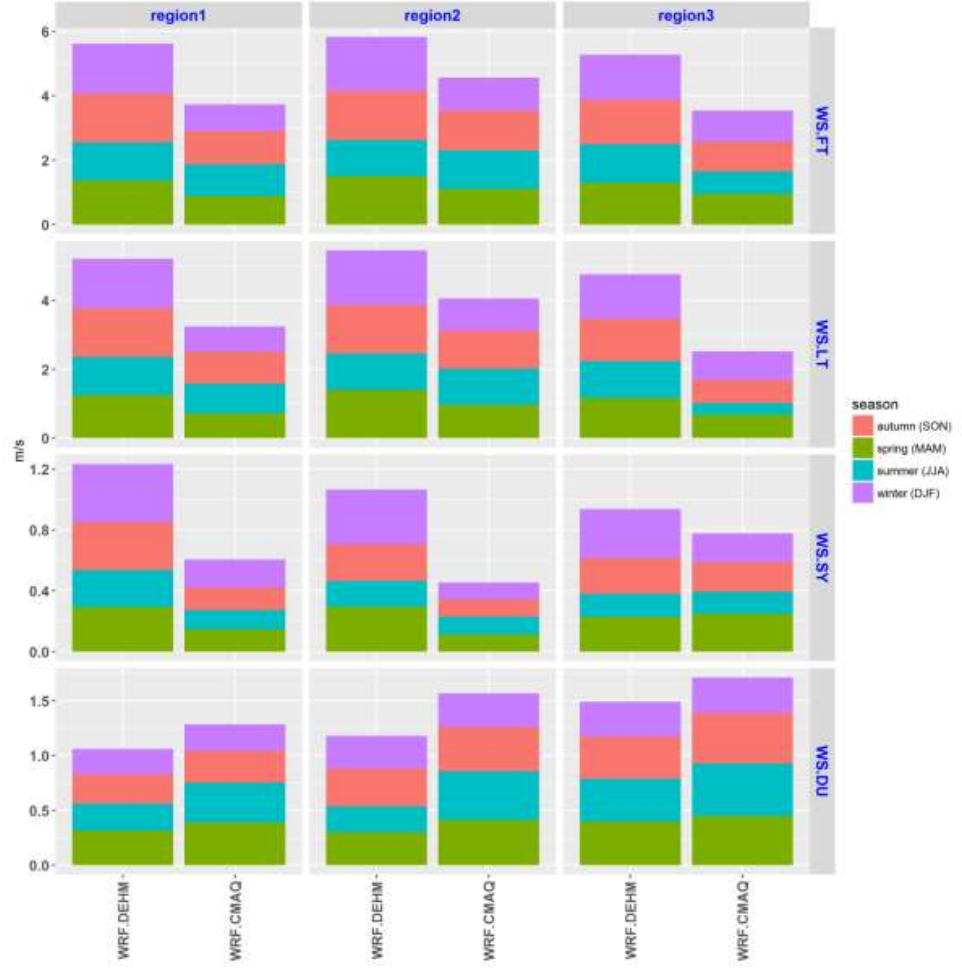
FIGURE 2. RMSE FOR A) TEMP AND B) WS IN EUROPE

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



b)

FIGURE 3. RMSE FOR A) TEMP AND B) WS IN NORTH AMERICA

Ozonesondes. Time averaged vertical profiles for WIND SPEED BIAS -EUROPE

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

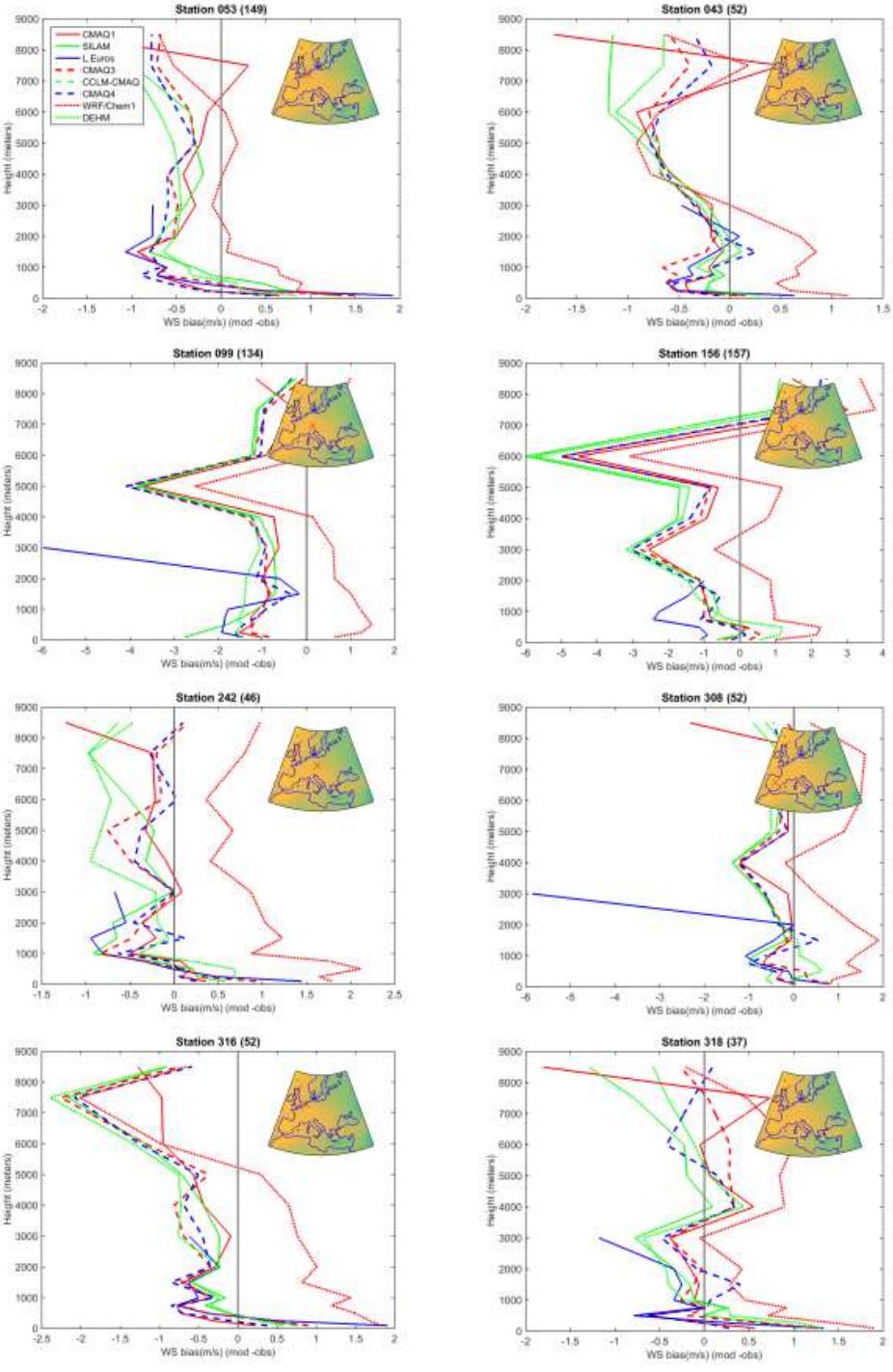
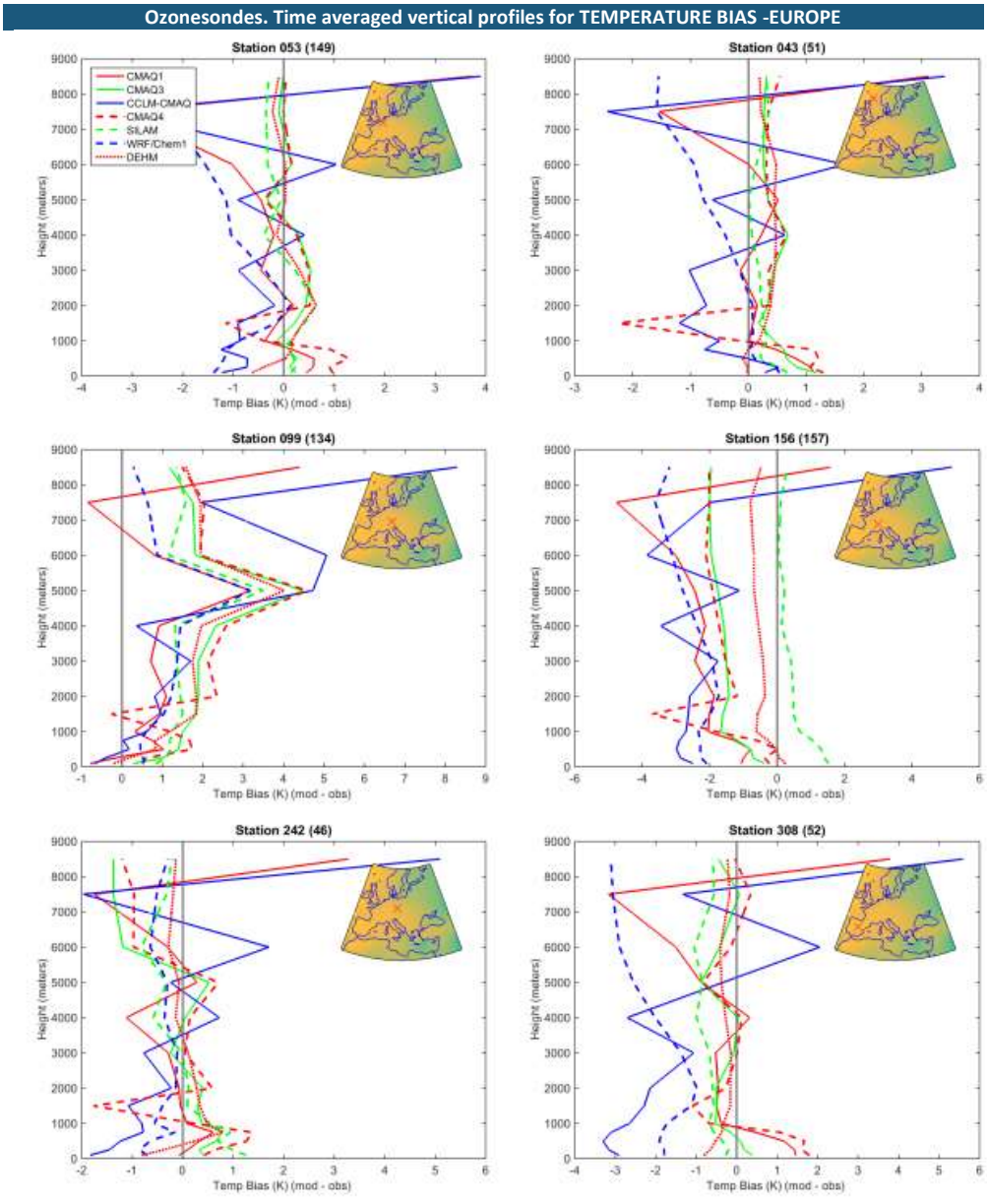


Figure 4. Mean Bias (mod – obs) for the vertical profiles of wind speed measured by ozonesondes launched from the European location indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

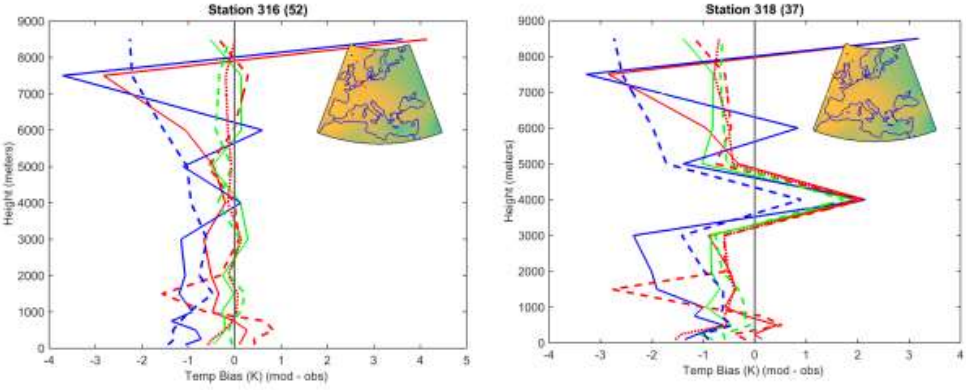
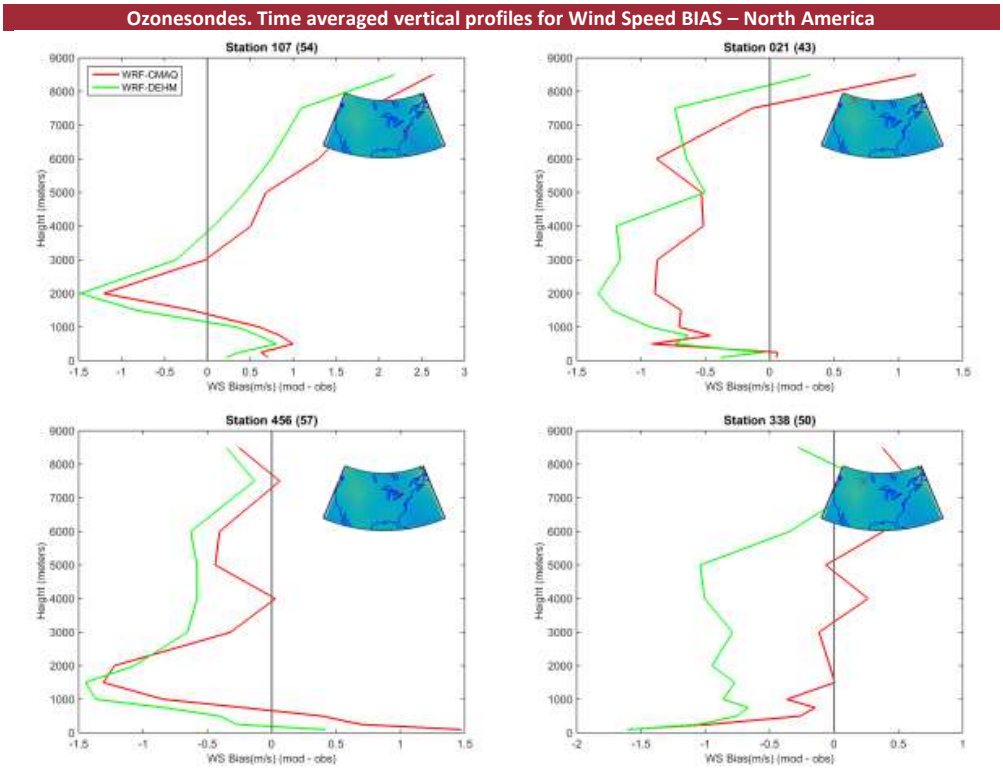


Figure 5. Mean Bias (mod – obs) for the vertical profiles of temperature measured by ozonesondes launched from the European location indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel



Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

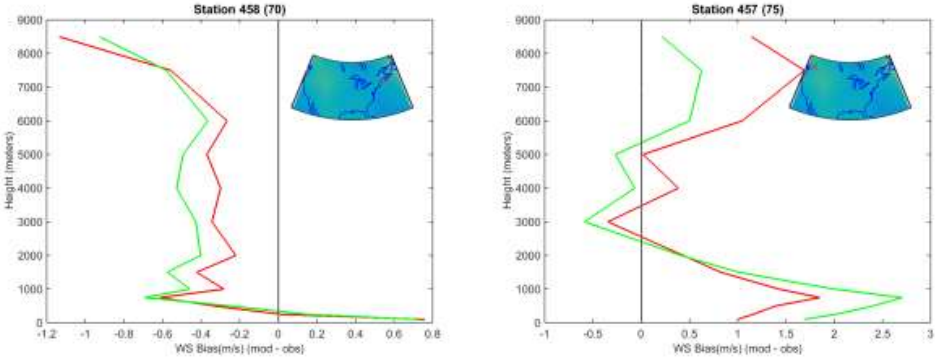
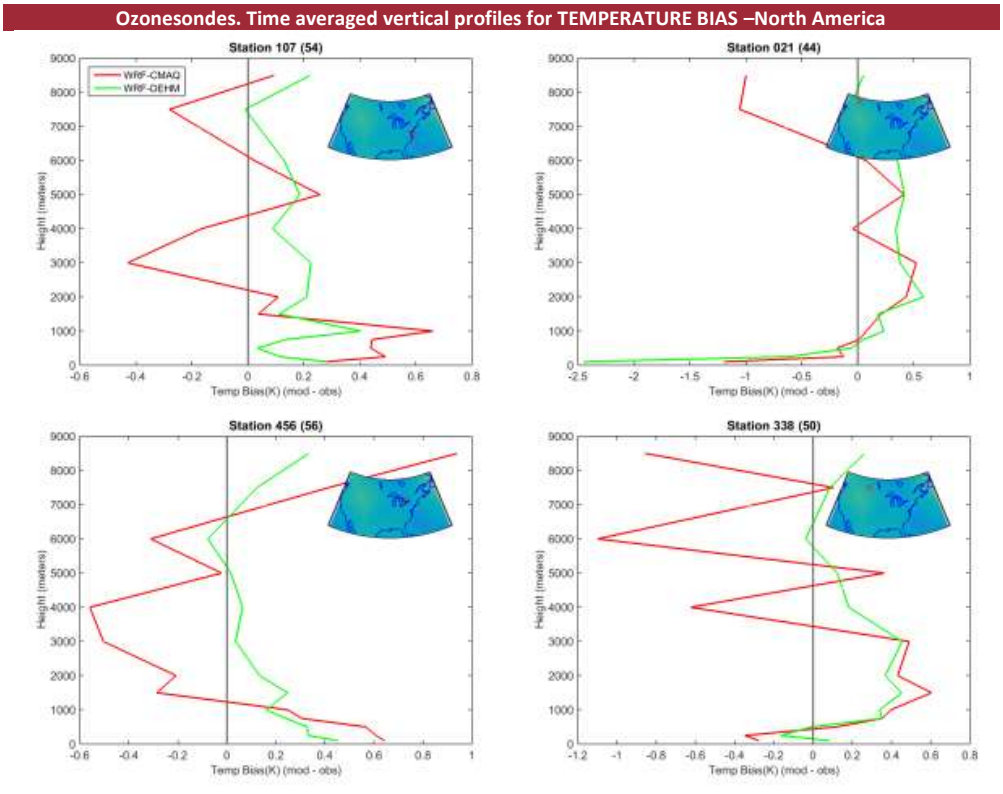


FIGURE 6. Mean Bias (mod – obs) for the vertical profiles of wind speed measured by ozonesondes launched from the North American locations indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel



Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

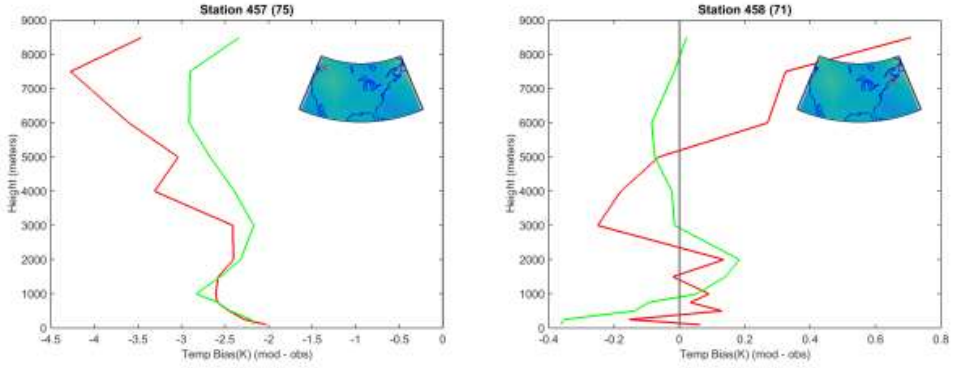
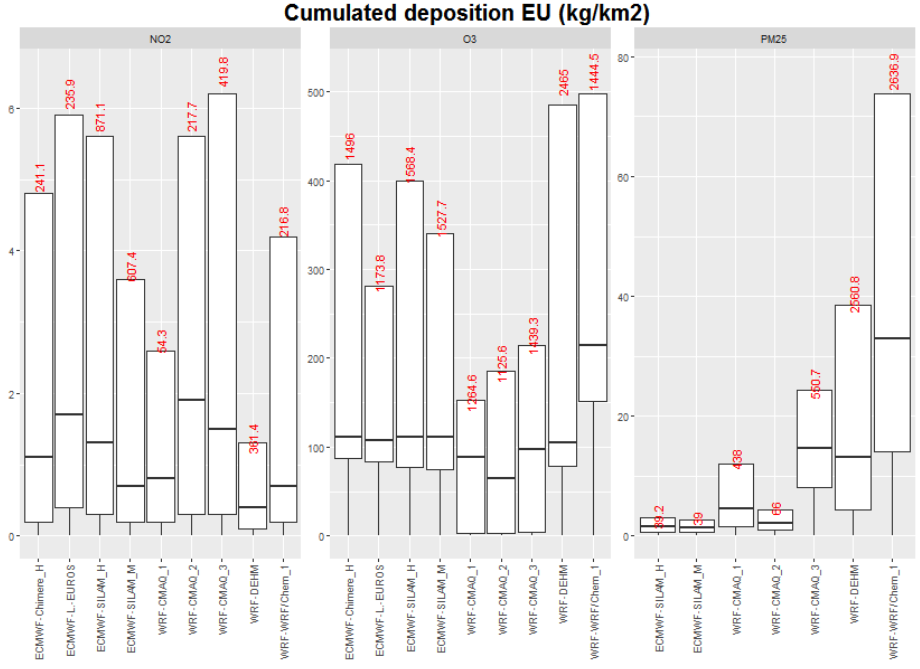


FIGURE 7. Mean Bias (mod – obs) for the vertical profiles of Temperature measured by ozonesondes launched from the North American location indicated on the inset map of each panel. The number of hourly profiles available for each site is reported in the parenthesis at the top of each panel



a)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

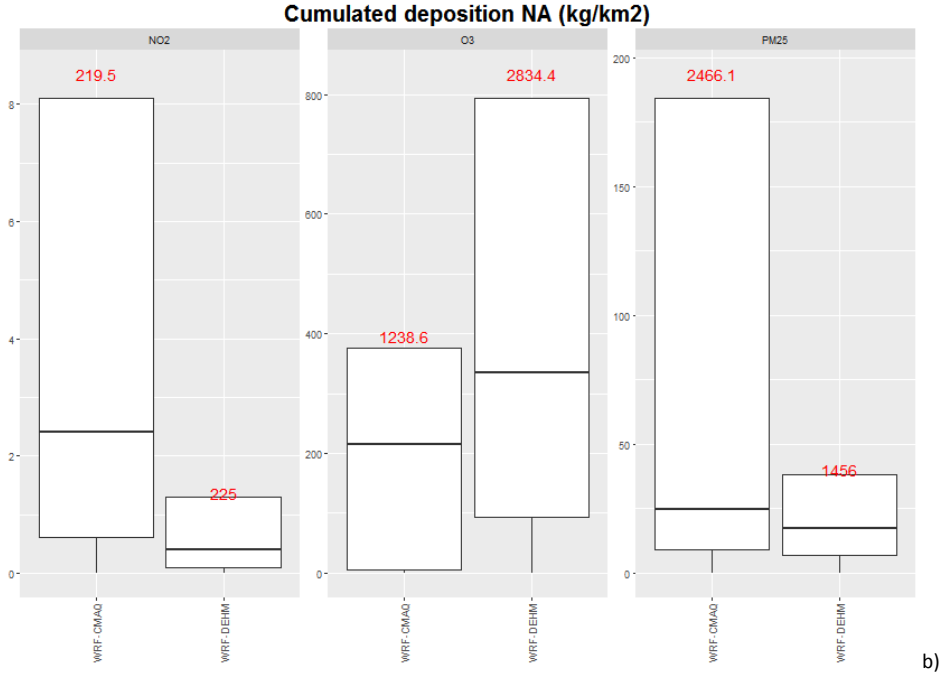
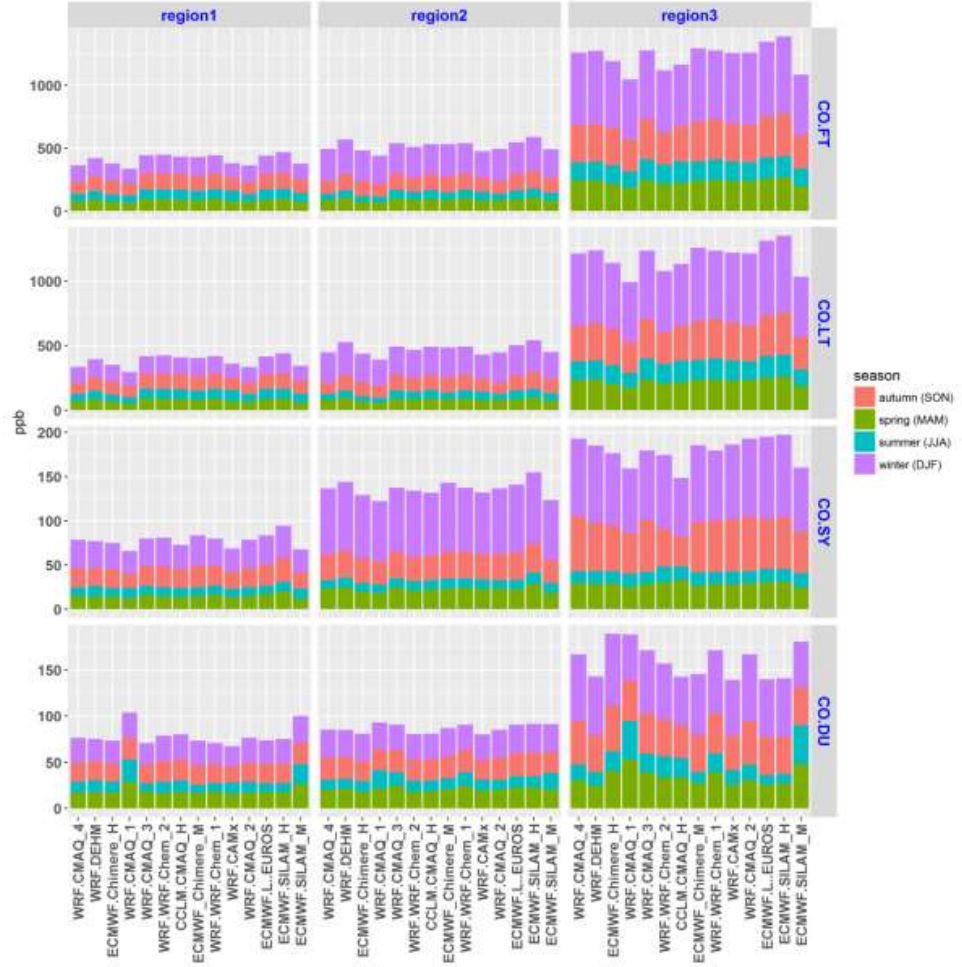


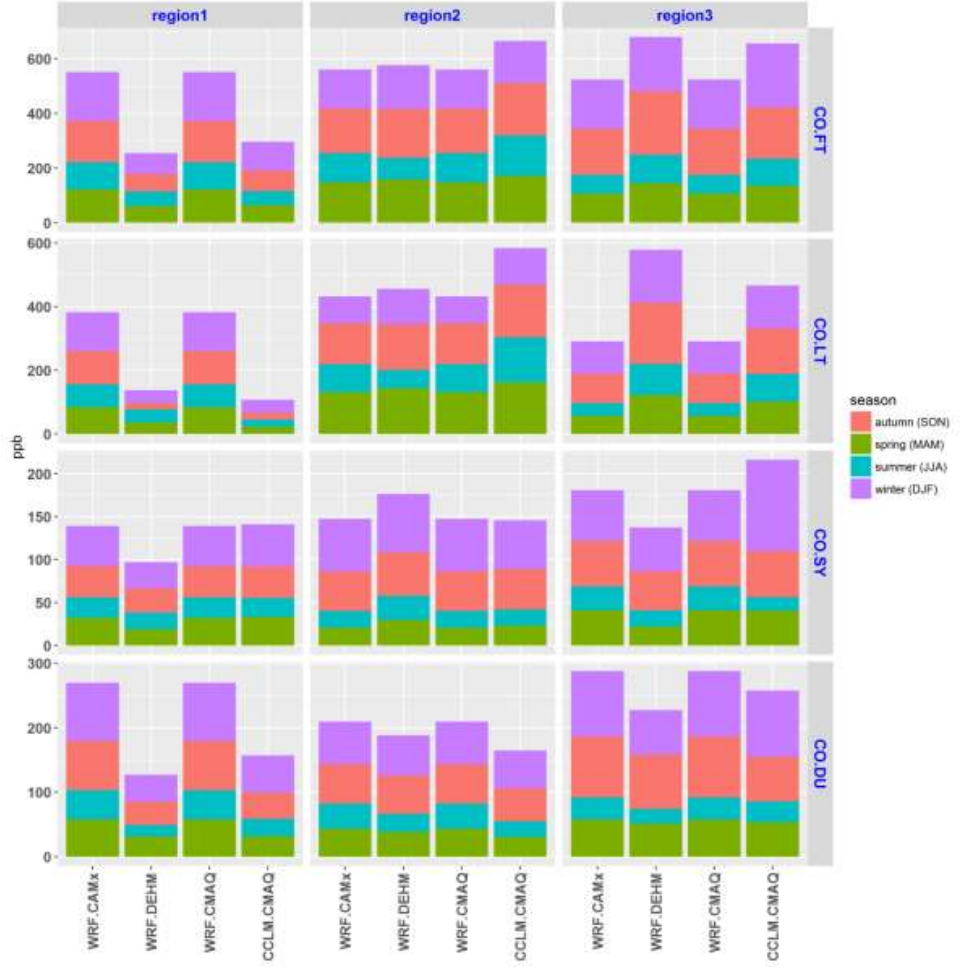
Figure 8. Cumulated modelled deposition per unit area over the continental regions of a) EU and b) NA for the full year of 2010. The boxes extend between the minimum and the 5th percentile, while the maximum is reported by the number at the top of each box. results are displayed for the models and species for which data have been made available.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)

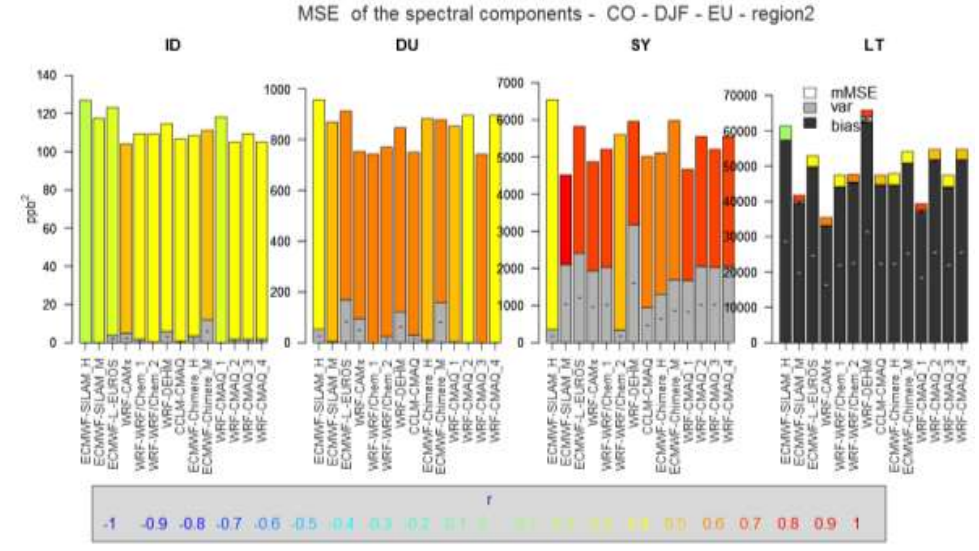
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



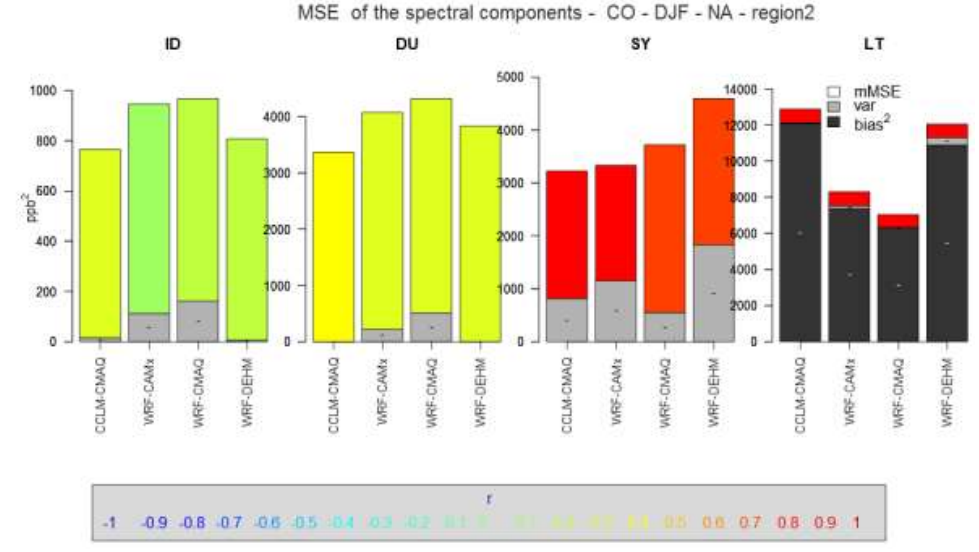
b)

FIGURE 9. RMSE (ppb) for CO by spectral component and season (panel *a* for Europe and *b* for North America). FT is the full (un-filtered) time series, LT, SY, DU, are the Long Term, Synoptic and diurnal components, respectively.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)



b)

FIGURE 10. MSE (ppb^2) breakdown into bias squared, variance and *mMSE* for the spectral components of the spatial average time series of CO during the months of December, January, and February (DJF), based on EQ.6. The bias is entirely accounted for by the LT component. The signs within the bias and variance portion of the bars indicate model overestimation (+) or underestimation (-) of the bias and variance. The colour of the *mMSE* share of the error is coded based on the values of *r*, the correlation coefficient, according to the colour scale at the bottom of each plot. Top panel: EU; lower panel: NA. Similar plots for the other two sub-regions are reported in the supplementary material.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

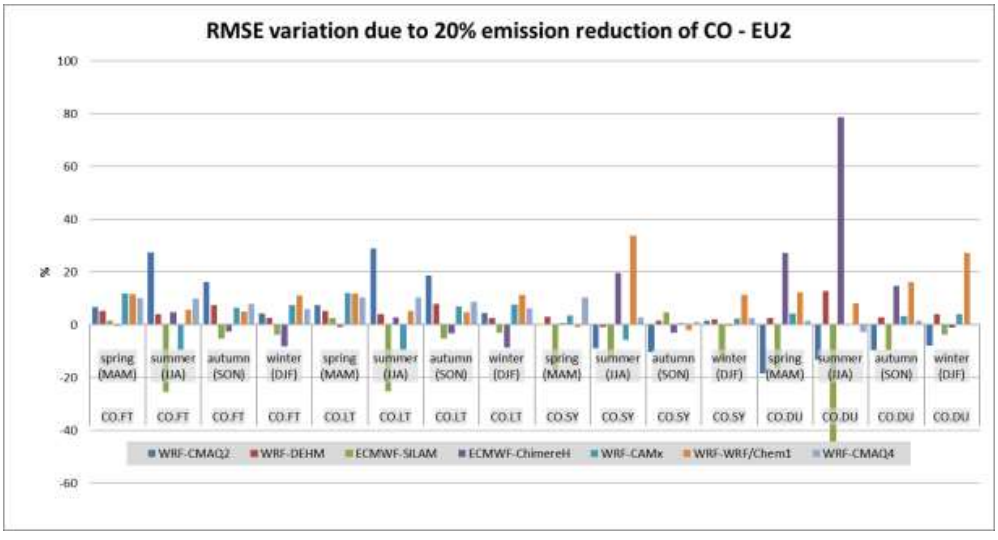
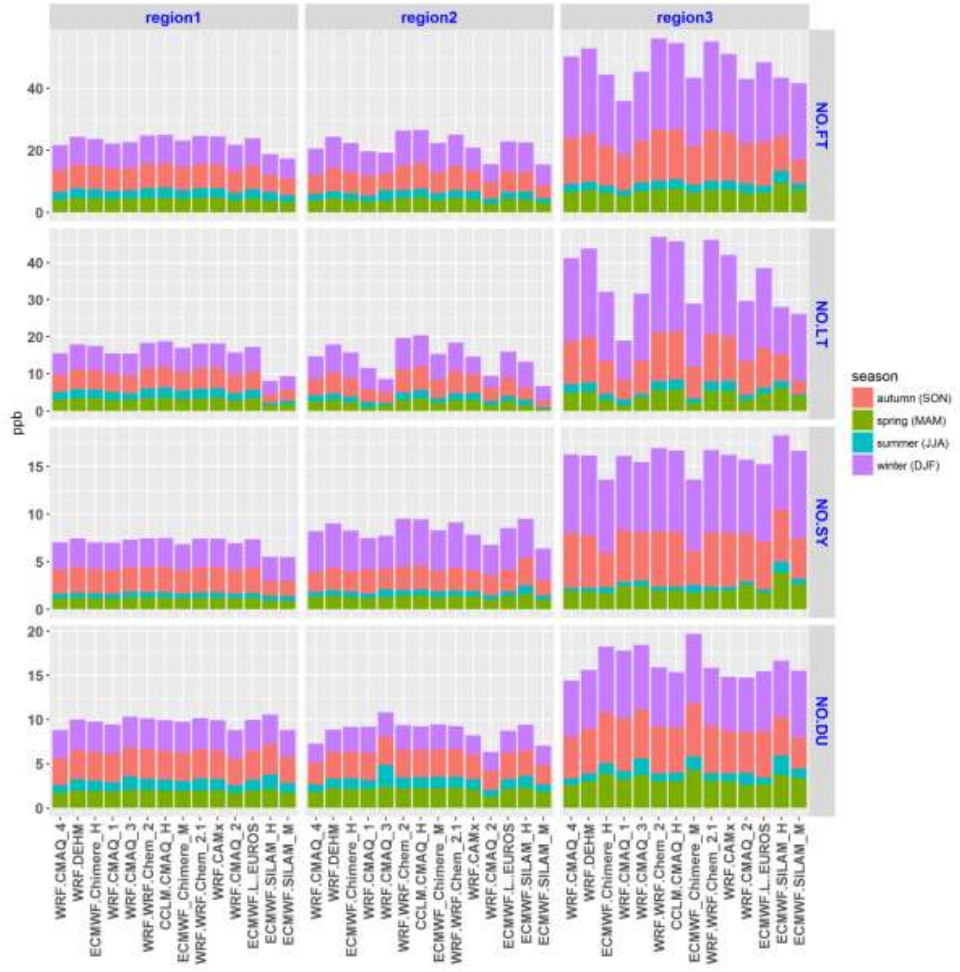


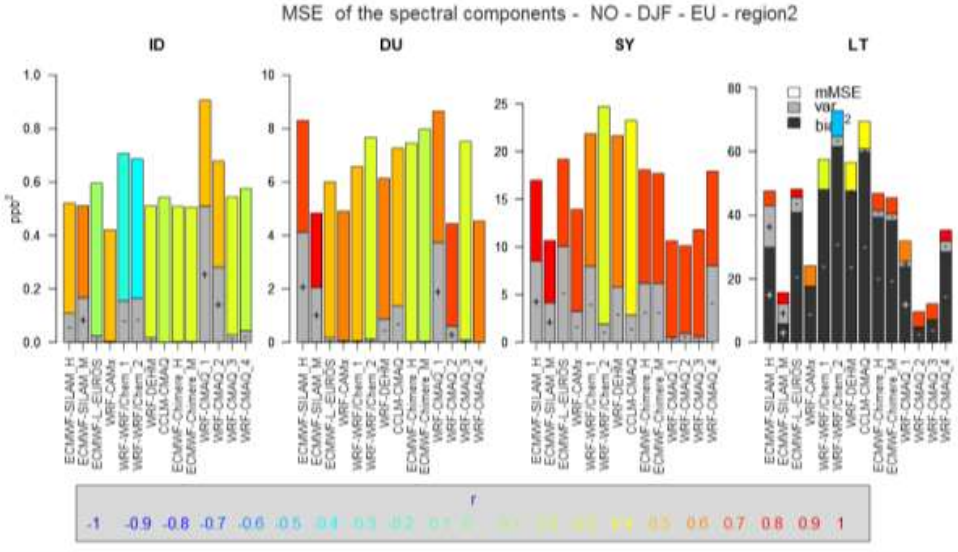
FIGURE 11. RMSE variation between the 's20%' scenario (anthropogenic emission and boundary condition reduced by 20%) and the base case for CO in EU2

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



b)

FIGURE 12. Top panel: as in **FIGURE 9** for NO (EU only). Lower panel: as in **FIGURE 10** for NO (EU only)

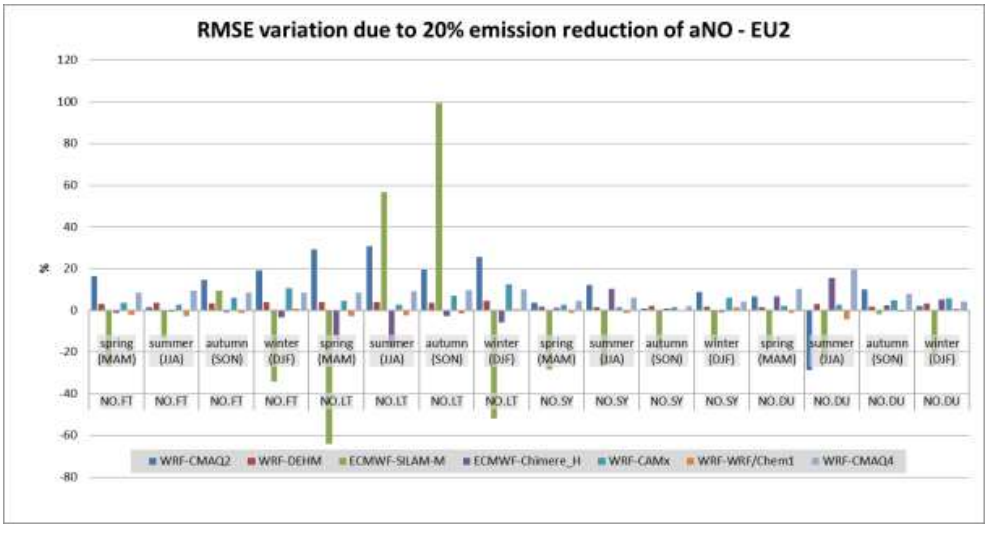
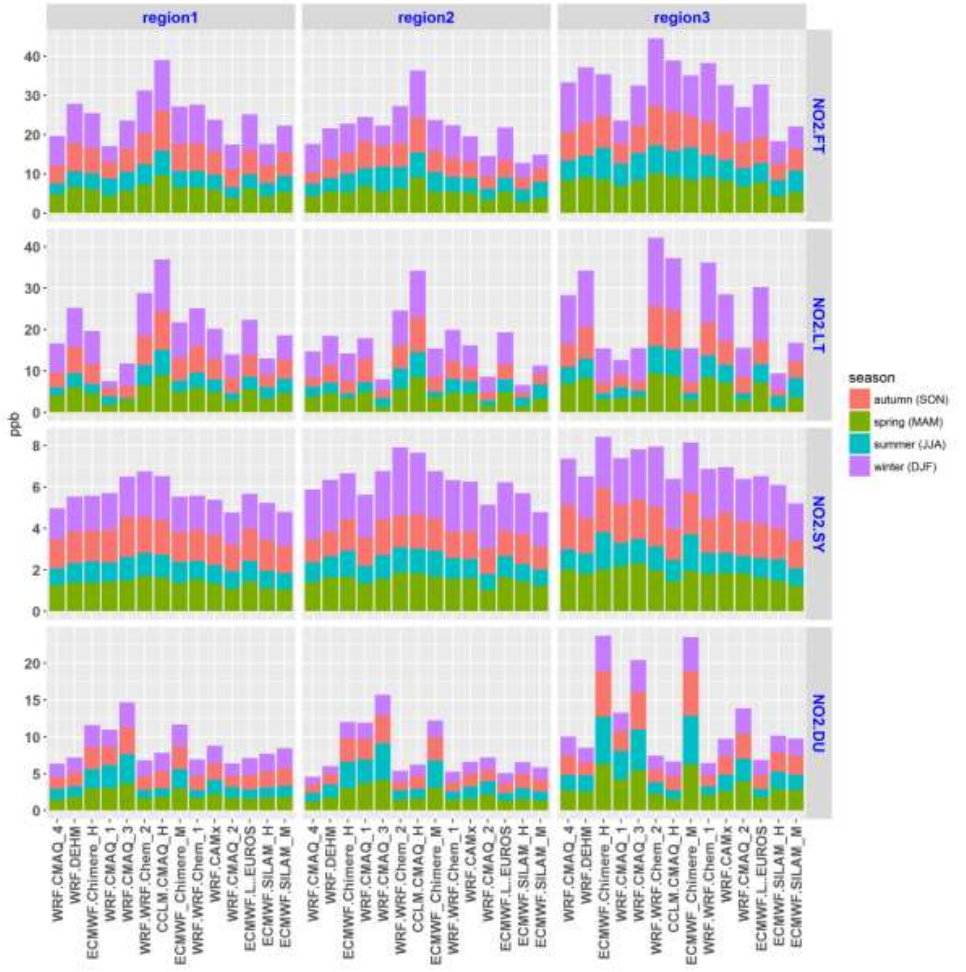


FIGURE 13. RMSE variation between the 's20%' scenario (anthropogenic emission and boundary condition reduced by 20%) and the base case for anthropogenic NO (aNO) in EU2

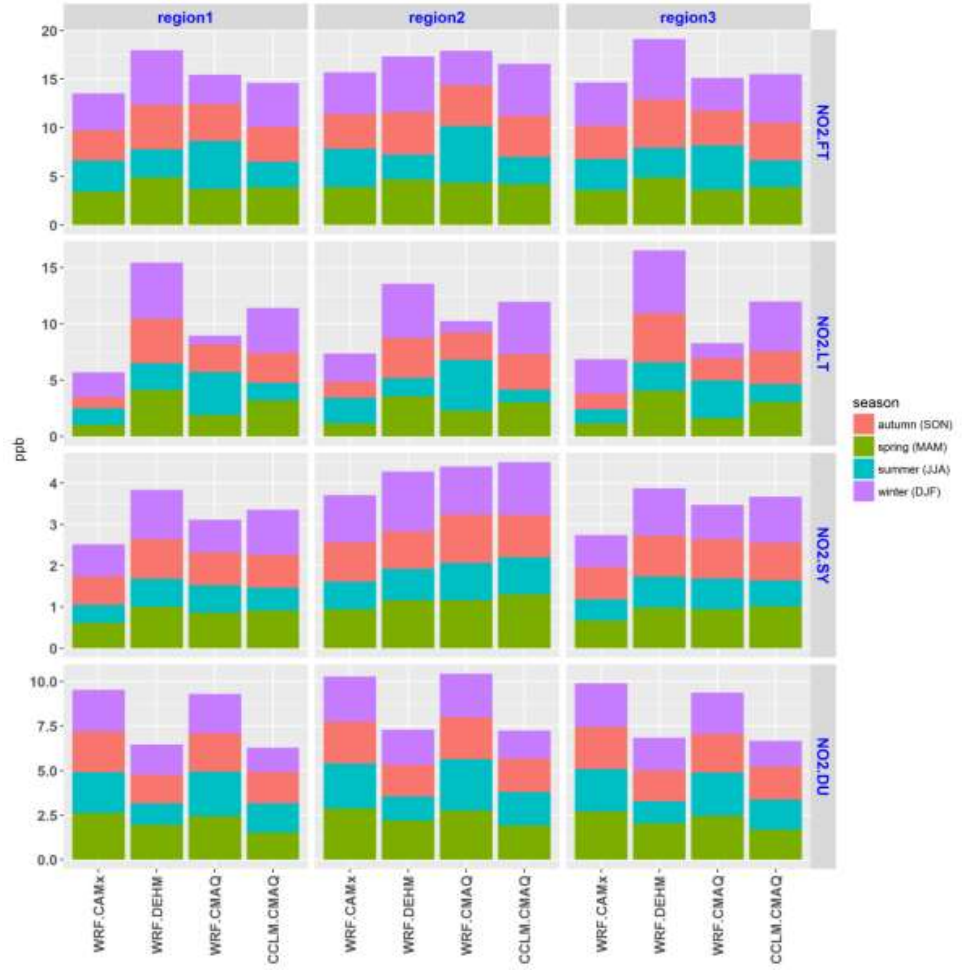
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



a)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

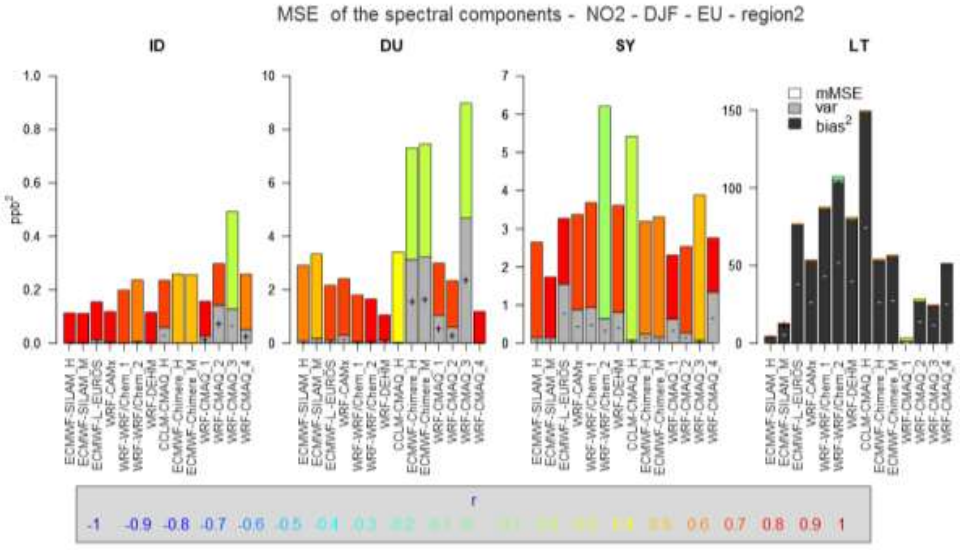


b)

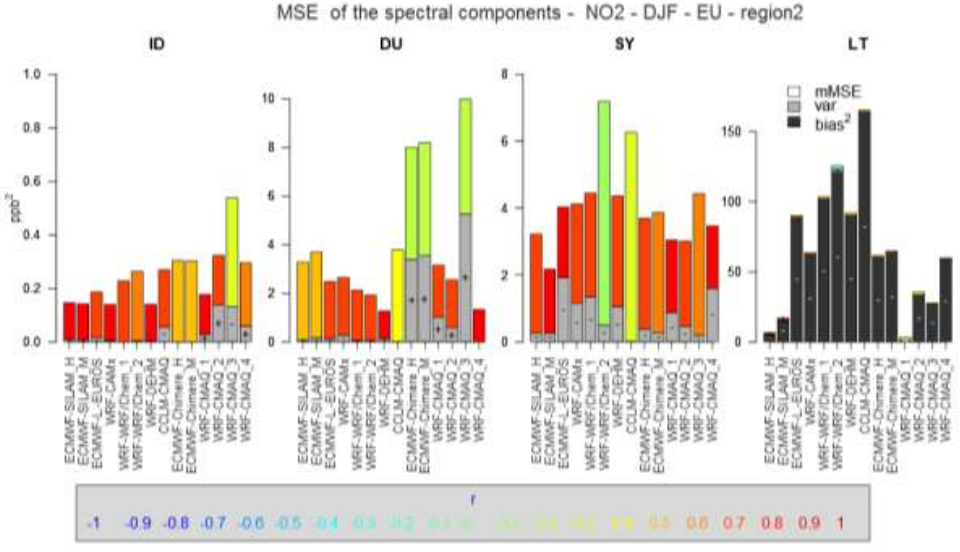
FIGURE 14. As in **FIGURE 9** for NO_2

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



a) Urban NO_2 in EU2 sub-region (223 stations)

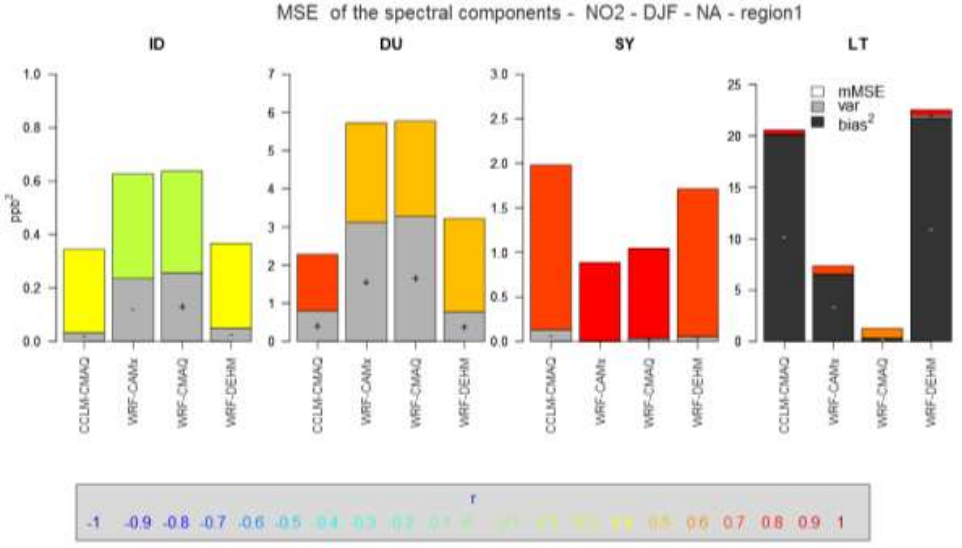


b) Rural NO_2 in EU2 sub-region(159 stations)

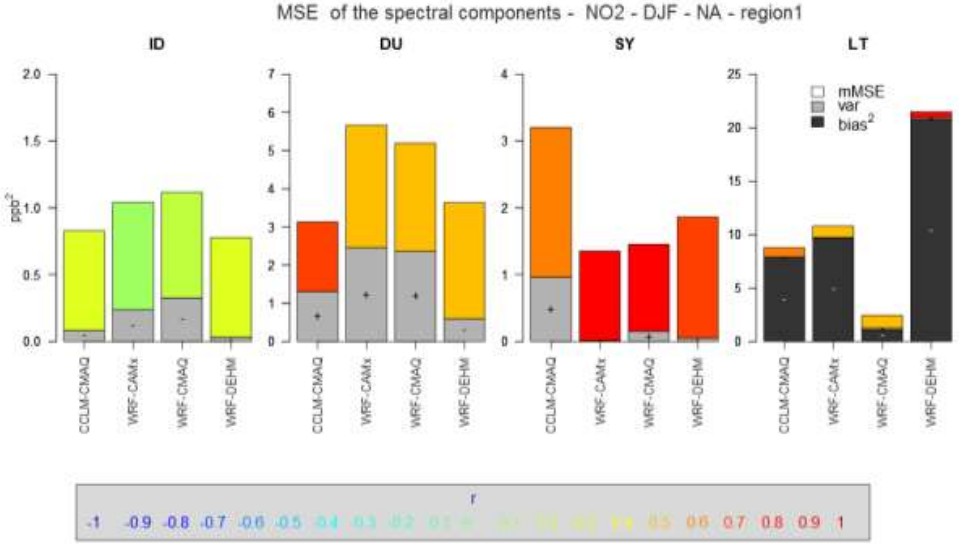
FIGURE 15. AS IN FIGURE 10 FOR NO_2 IN EU2. UPPER PANEL: URBAN SITES ONLY (223 STATIONS); LOWER PANEL: RURAL SITES ONLY (159 STATIONS)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



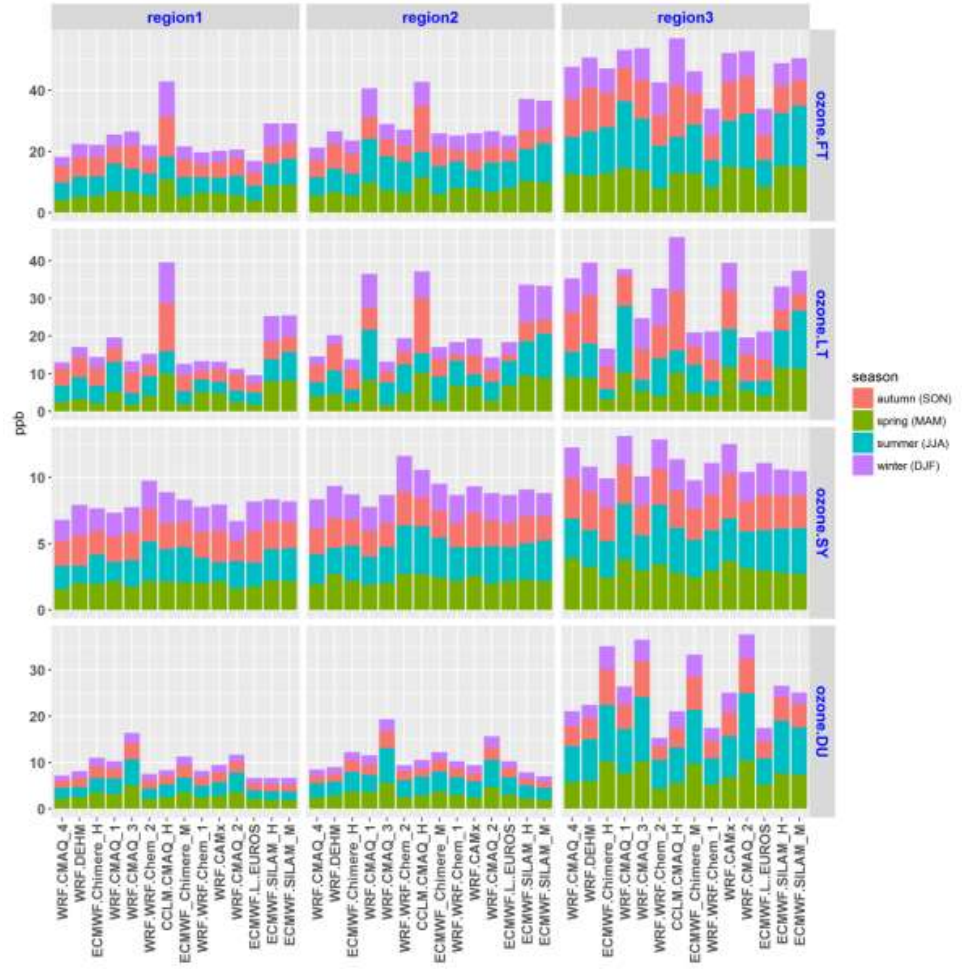
a) NA1 urban (39 stations)



b) NA1 rural (10 stations)

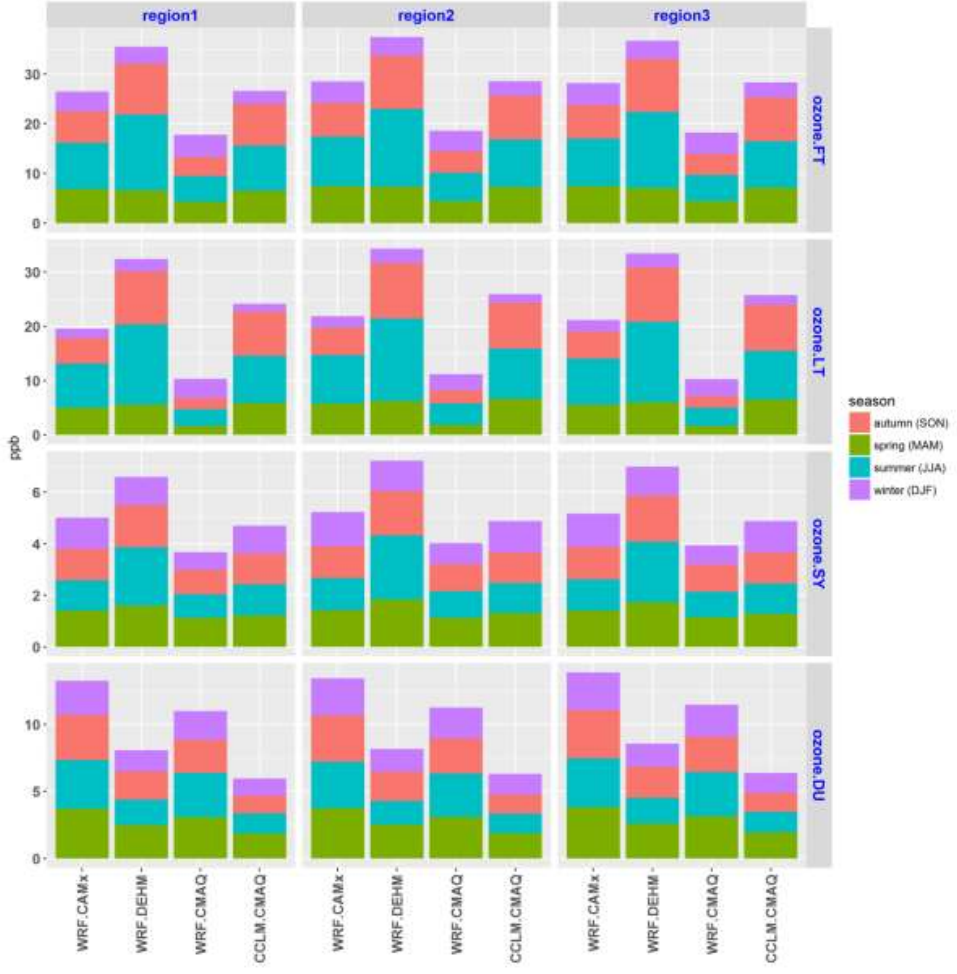
FIGURE 16. AS IN FIGURE 10 FOR NO_2 IN NA1. UPPER PANEL: URBAN SITES ONLY (39 STATIONS); LOWER PANEL: RURAL SITES ONLY (10 STATIONS).

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



A)

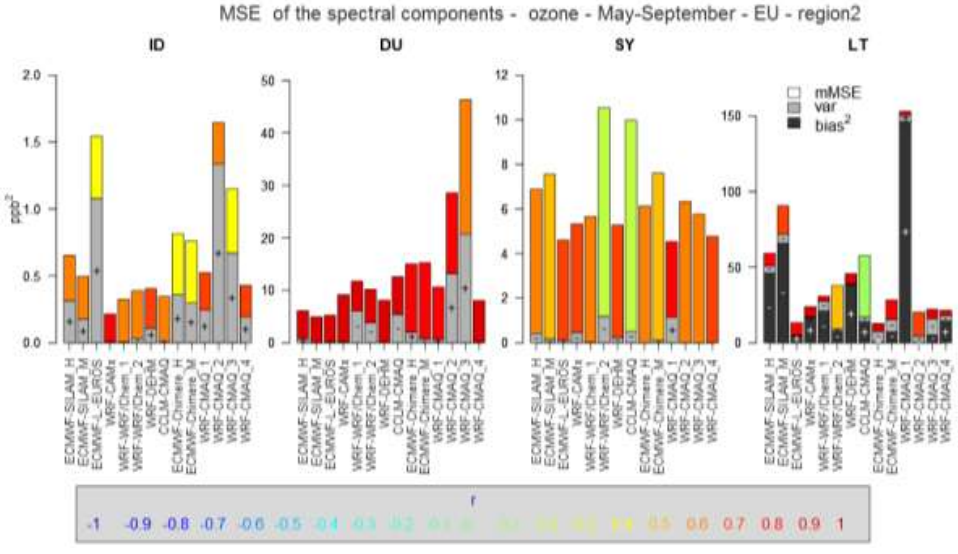
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



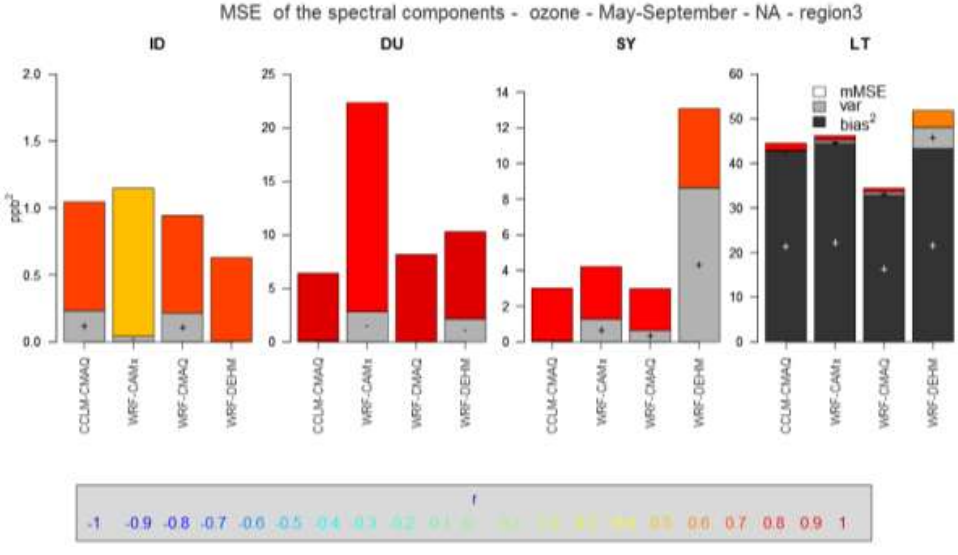
B)

FIGURE 17. AS IN FIGURE 9 FOR OZONE

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)

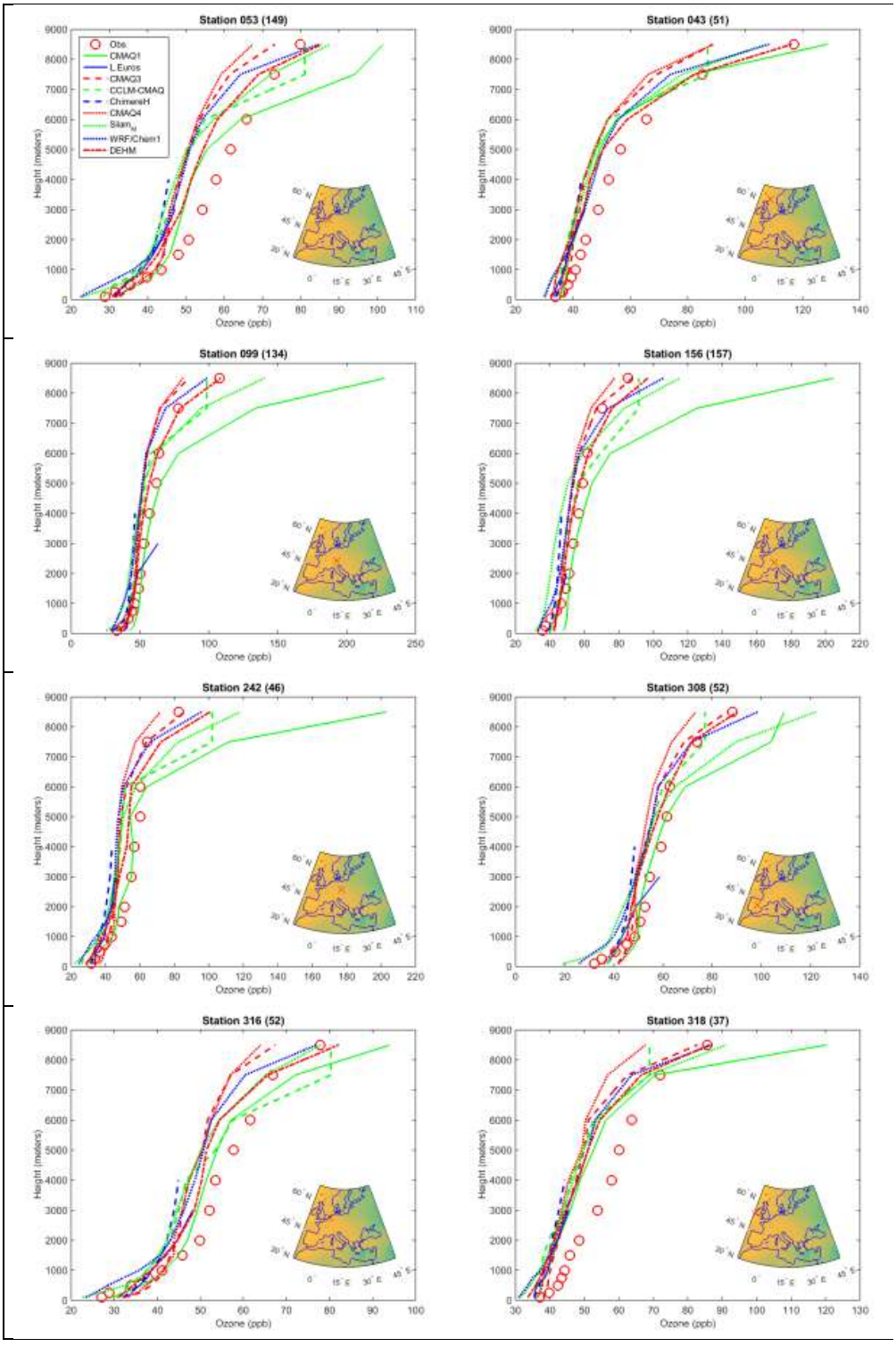


b)

FIGURE 18. AS IN FIGURE 10 FOR OZONE DURING THE MONTHS FROM MAY TO SEPTEMBER

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

Figure 19. Ozone mixing-ratio profiles measured by ozonesondes launched from the European location indicated on the inset map (lower-right corner) of each panel. The profiles are time-averaged over the number of hourly records reported in the parenthesis at the top of each panel. Legend as in the first panel**.**

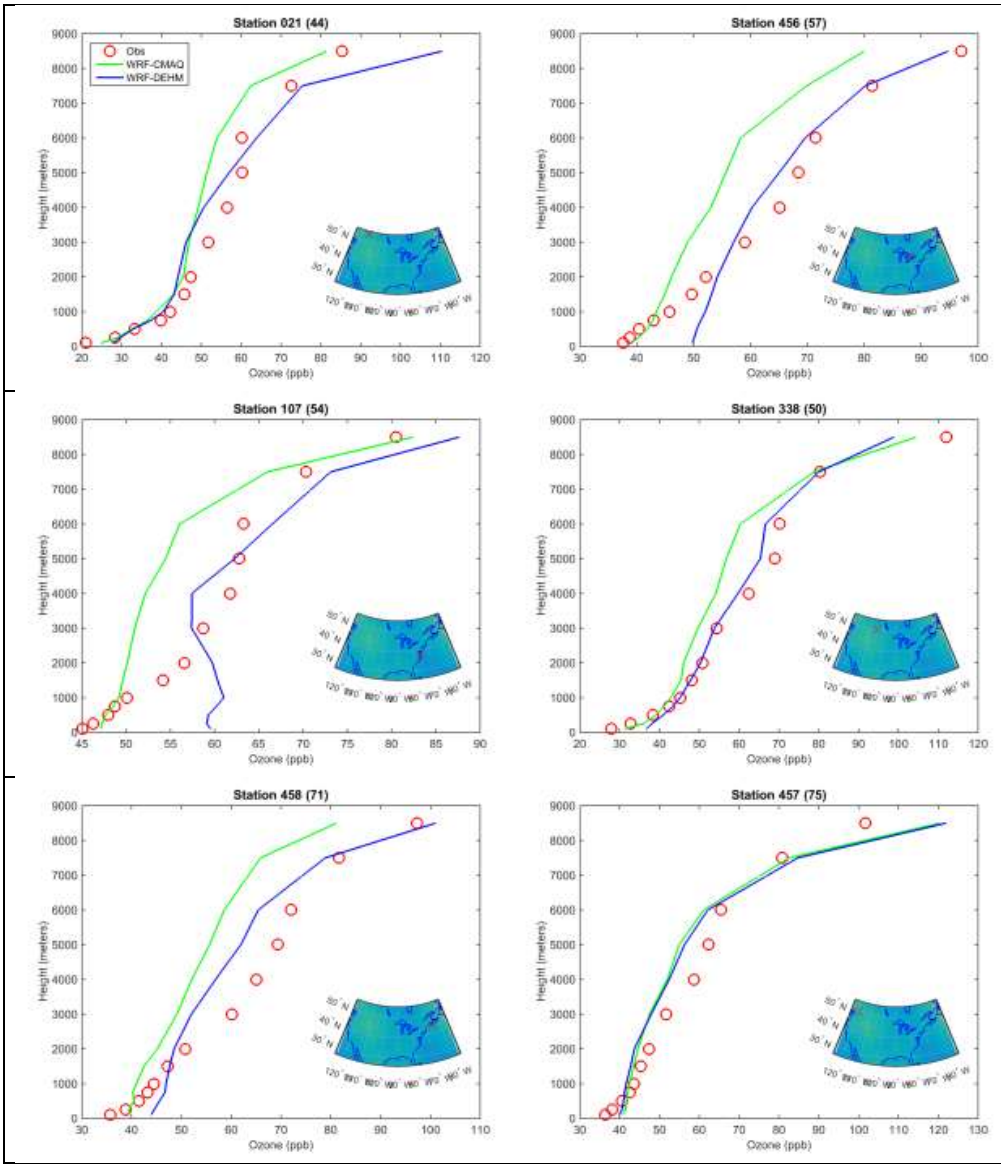


FIGURE 20 . As in **FIGURE** 19 for North America

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

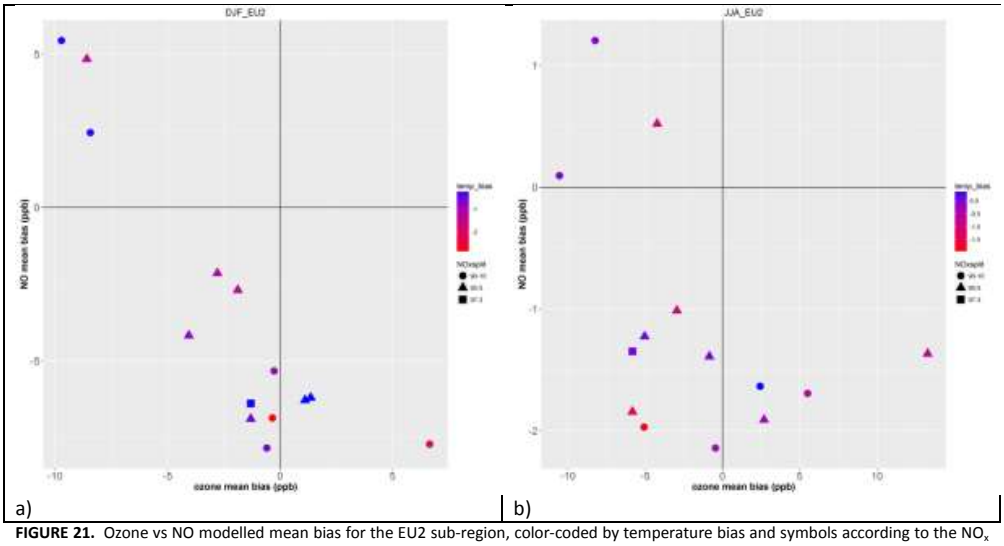
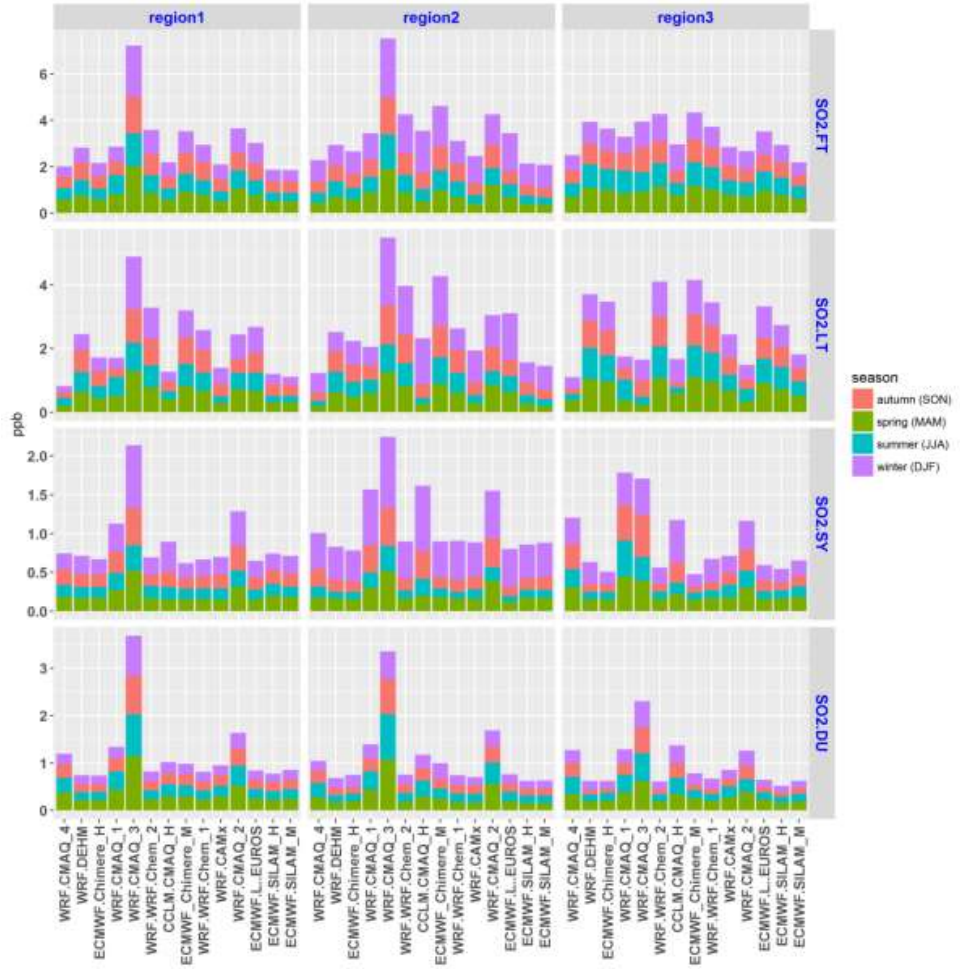


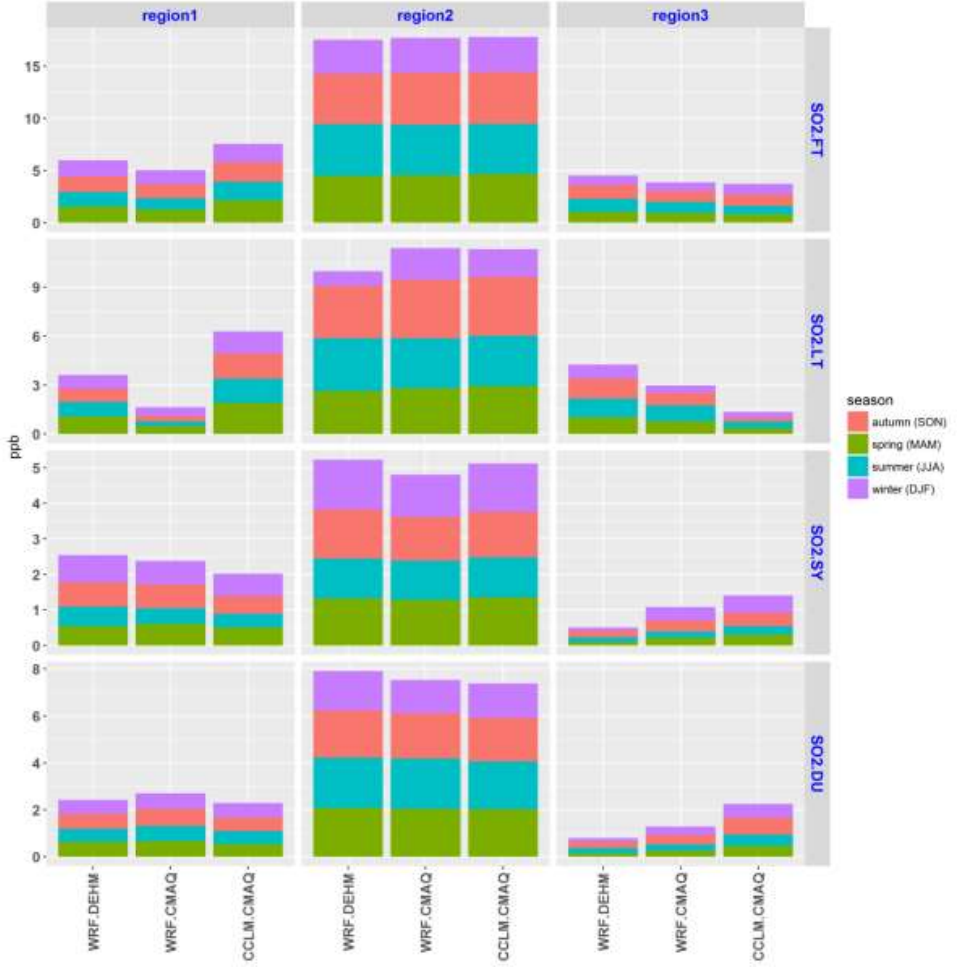
FIGURE 21. Ozone vs NO modelled mean bias for the EU2 sub-region, color-coded by temperature bias and symbols according to the NO_x emission fraction of NO and NO_2. Each point represents a model. *a)* winter months and *b)* summer months.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)

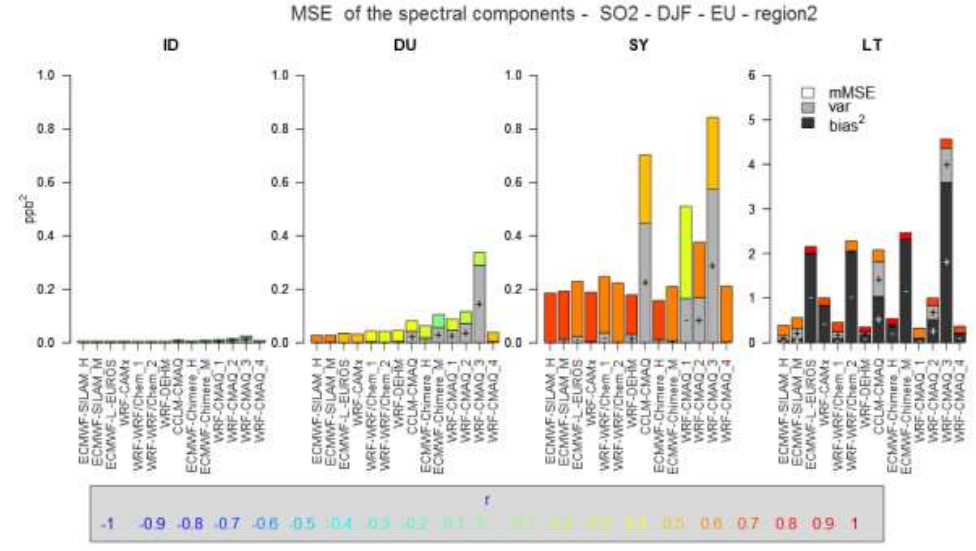
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



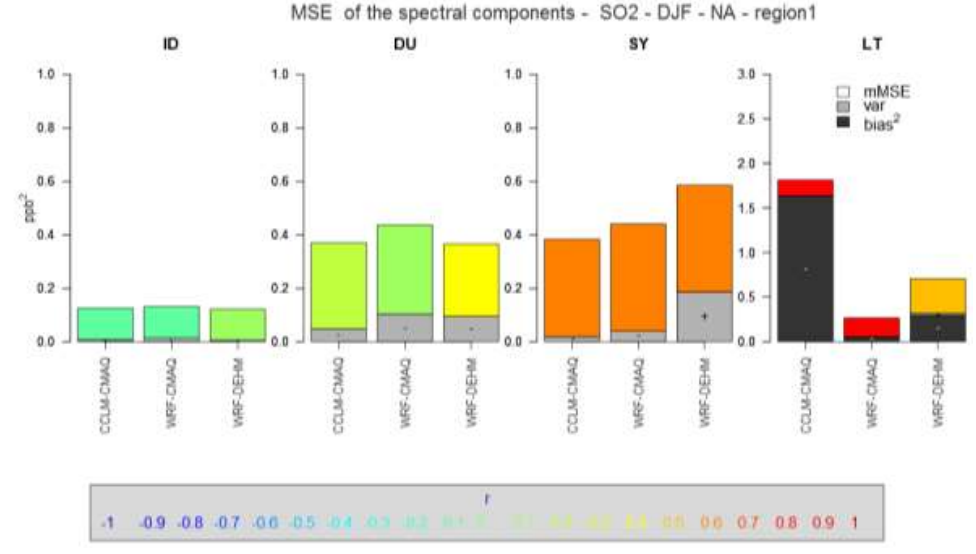
b)

FIGURE 22. AS IN FIGURE 9 FOR SO_2

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016



a)



b)

FIGURE 23. AS IN FIGURE 10 FOR SO_2

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

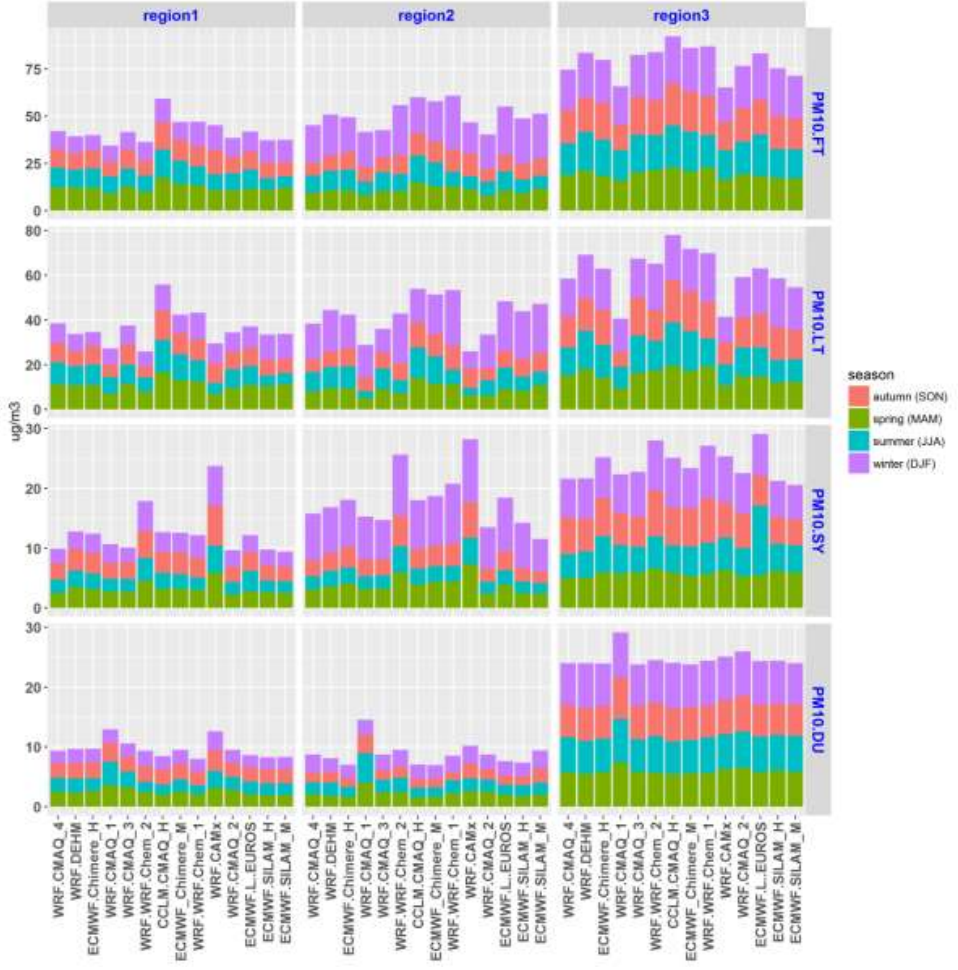


FIGURE 24. As in Figure 9 for PM_{10} in Europe (error units in $\mu g/m^3$)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

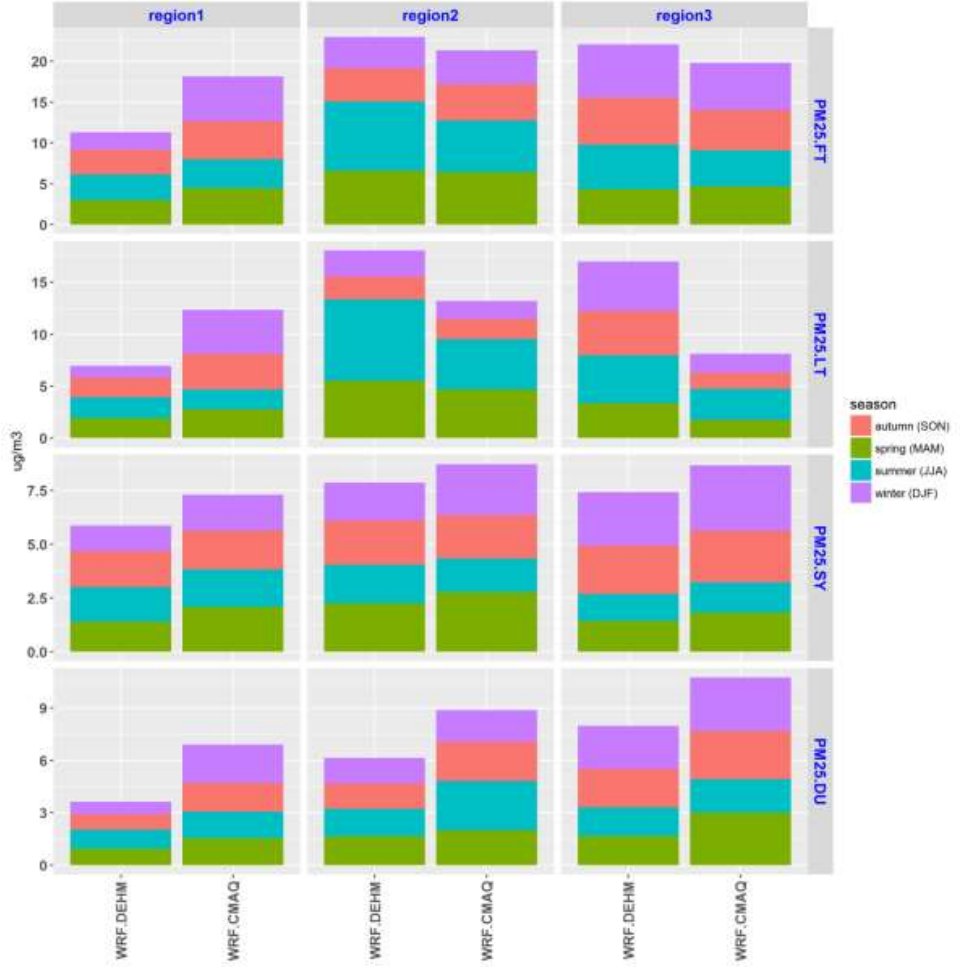


FIGURE 25. As in Figure 9 for $PM_{2.5}$ in North America (error units in $\mu g/m^3$)

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions

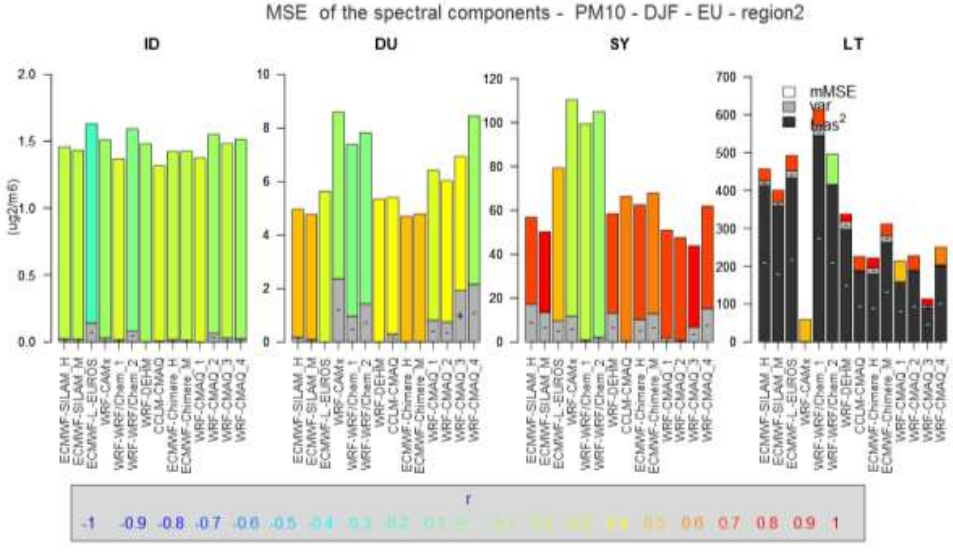


FIGURE 26. As in Figure 10 for PM_{10} in Europe (error units in $\mu g/m^3$)

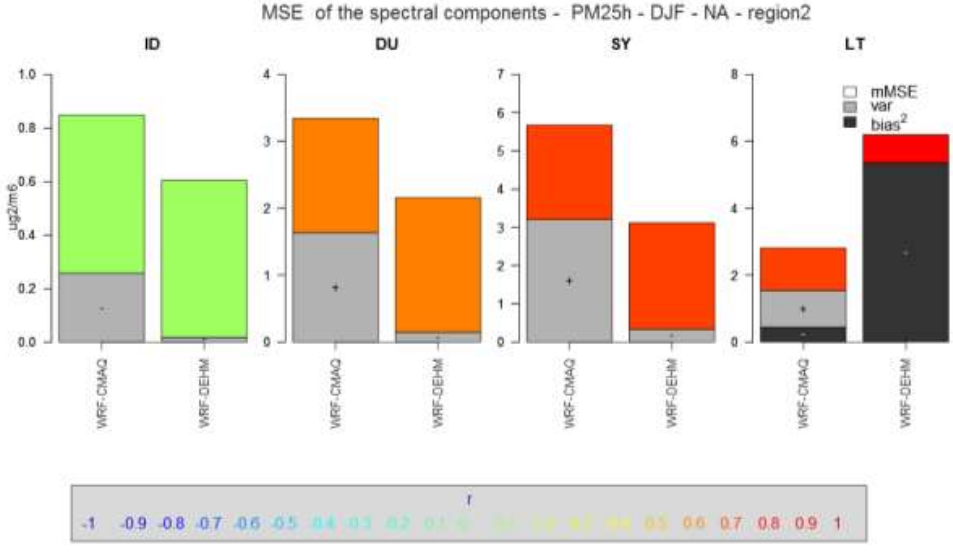
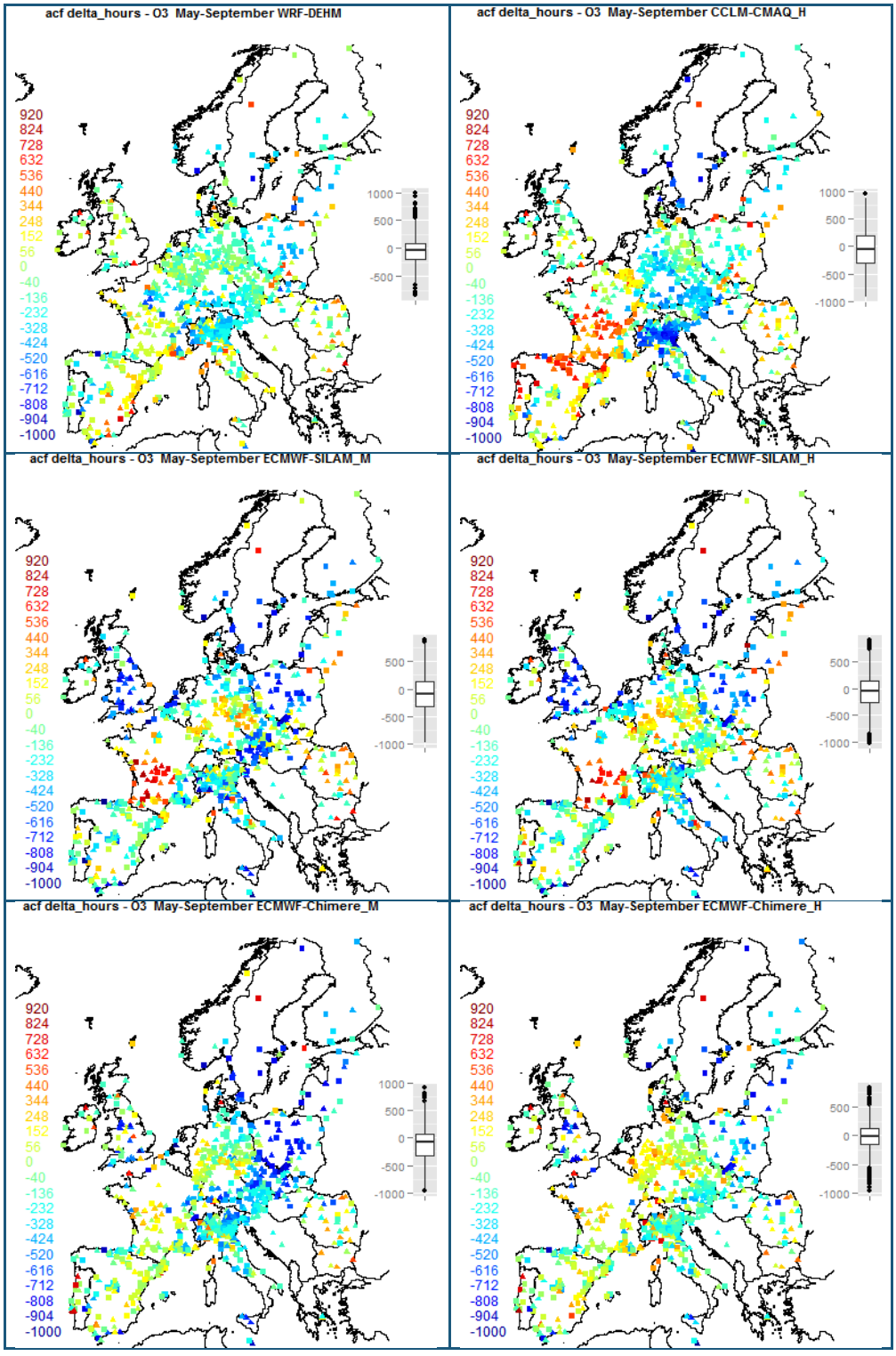


FIGURE 27. As in Figure 10 for hourly $PM_{2.5}$ in North America (error units in $\mu g/m^3$)

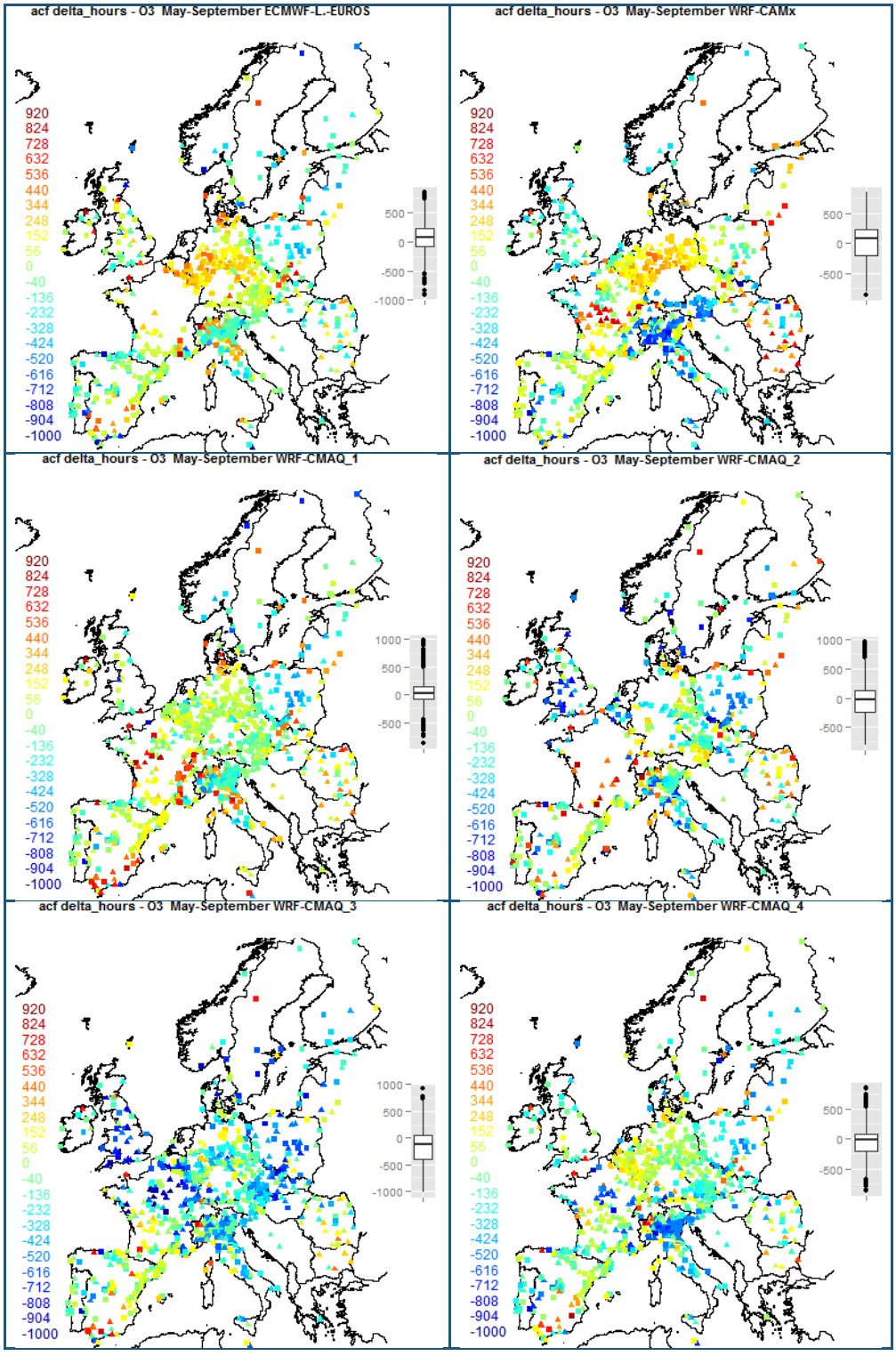
Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

Atmospheric Chemistry and Physics Discussions



Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-682, 2016
Manuscript under review for journal Atmos. Chem. Phys.
Published: 7 September 2016

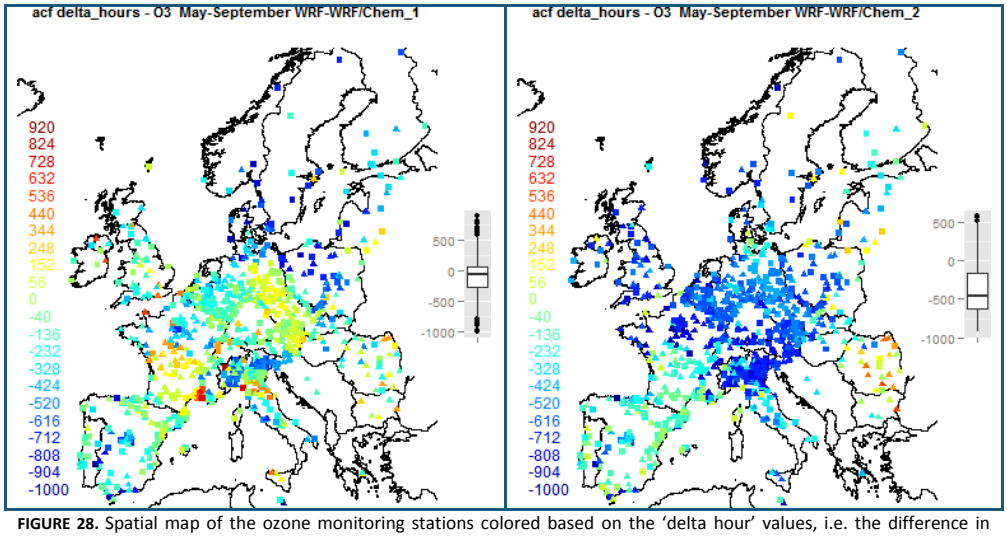


FIGURE 28. Spatial map of the ozone monitoring stations colored based on the 'delta hour' values, i.e. the difference in hours between the zero of the autocorrelation function (acf) for the modelled ozone minus the zero of the acf of the observed one. The acf is calculated on the long term component for the months of May to September. Negative values indicate an excess of removal (viceversa for positive values). The box on the right summarises the delta hour percentile distribution.

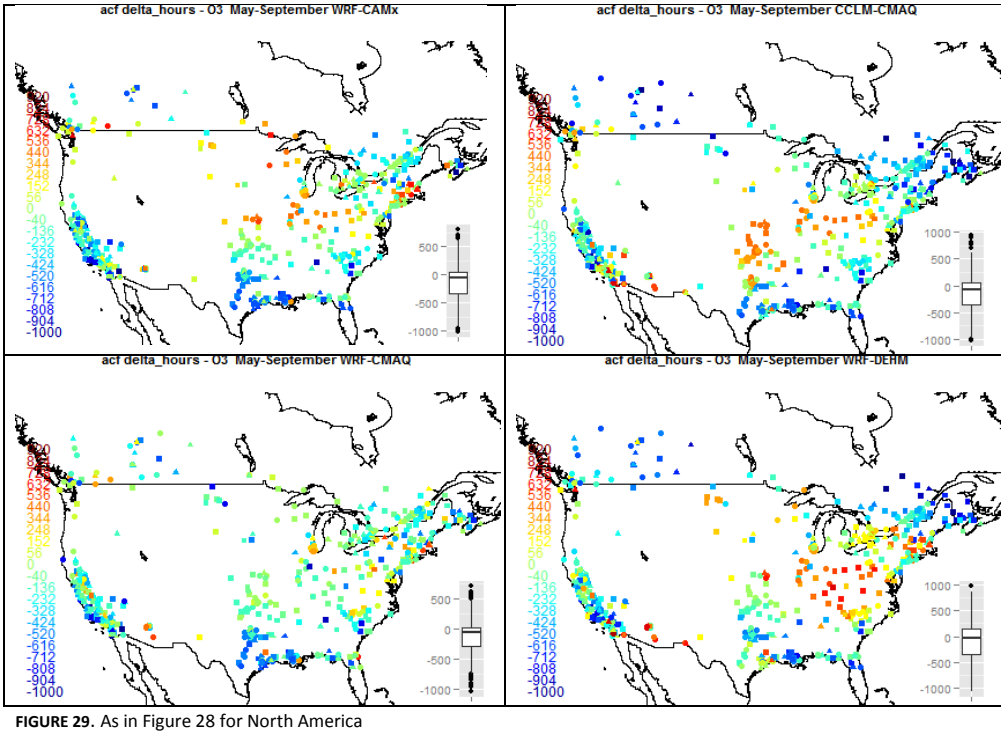


FIGURE 29. As in Figure 28 for North America