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], Paolo Tuccella[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], 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.

46 The application of the error apportionment method to the AQMEII Phase 3 simulations provides several key
47 insights. In addition to reaffirming the strong impact of model inputs (emissions and boundary conditions) and
48 poor representation of the stable boundary layer on model bias, results also highlighted the high inter-
49 dependencies among meteorological and chemical variables, as well as among their errors. This indicates that the
50 evaluation of air quality model performance for individual pollutants needs to be supported by complementary
51 analysis of meteorological fields and chemical precursors to provide results that are more insightful from a model
52 development perspective. This will require evaluation methods that are able to frame the impact on error of
53 processes, conditions, and fluxes at the surface. For example, error due to emission and boundary conditions is
54 dominant for primary species (CO, PM), while errors due to meteorology and chemistry are most relevant to
55 secondary species, such as ozone. Some further aspects emerged whose interpretation requires additional
56 consideration, such as, among others, the uniformity of the synoptic error being region and model-independent,
57 observed for several pollutants; the source of unexplained variance for the diurnal component; and the type of
58 error caused by deposition and at which scale.

1. INTRODUCTION

60 The Air Quality Model Evaluation International Initiative (AQMEII, Rao et al., 2010) has been active since 2008 with
61 the aim of promoting the research on regional air quality model evaluation across the modelling communities of
62 Europe and North America. It is coordinated by the European Joint Research Centre (JRC) and the U.S.
63 Environmental Protection Agency (EPA) and it has now reached its third phase, referred to as AQMEII3 hereafter.
64 The experience gathered in the first two phases consisted of important advancement in the model evaluation
65 research as well as establishing a large community of participating regional modeling groups, and have made
66 AQMEII a natural candidate to collaborate with the Hemispheric Transport of Air Pollution (HTAP) initiative. HTAP,
67 a taskforce of the Long Range Transport of Air Pollution program (LTRAP) acting within the UNECE program, relies
68 on a community of global scale chemical transport models to investigate the fate of air pollutants emitted in the
69 Northern hemisphere and determine the contribution of remote sources as well as their impacts to the
70 background concentration in different parts of the globe. HTAP is in its second phase and the activities undertaken
71 during this second phase include coordinating simulations by both global and regional scale models. The regions of
72 interest in the Northern hemisphere are North America, Europe and South East Asia. The regional-scale modelling
73 component of this activity for Europe and North America is being coordinated by AQMEII while the Asian
74 component is being coordinated by MICs-ASIA (Model Intercomparison Study-Asia). Global-scale models
75 participating in HTAP are used by the AQMEII regional models as boundary conditions and special attention has
76 been given to the emission inventory to ensure that it is consistent between the global and regional-scale
77 simulations as described in Janssens-Maenhout et al. (2015). The activity described here relates to the evaluation
78 of the base case scenario set up within the context of HTAP and AQMEII (Galmarini et al., 2017).

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

90 accommodate the data and to pair modelled to observational values in time and space to provide direct
91 comparison and statistical analysis.

92 The model evaluation approach proposed and applied in this study combines aspects of operational and diagnostic
93 evaluation as defined by Dennis et al. (2010). It makes use of the classical statistical indicators typically employed
94 for operational evaluation based on the direct comparison with observations, but also provides more indications
95 on the processes contributing to model errors, which is the focus of diagnostic model evaluation (Solazzo and
96 Galmarini, 2016). The data used in the analysis are not process specific but are ordinary time series of modelled
97 and monitoring data which are decomposed into four spectral components: ID (intra-day), DU (diurnal), SY
98 (synoptic), and LT (long-term), each determined by different physical and chemical processes (Rao et al., 1997).
99 The error apportionment applied to each spectral component can provide indications on the possible sources of
100 error. The scope of the diagnostic evaluation, as also highlighted by Gupta et al. (2009), is to move beyond the
101 usual aggregate metrics that only offer a statistical interpretation, towards the use of measures selected for the
102 quality of the information they can provide to model developers and users.

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

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

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

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

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

2. METHODOLOGY

125 The first step of the analysis is the spectral decomposition of the time series of modelled and observed species, as
126 outlined in the methodology proposed in Solazzo and Galmarini (2016). Because each spectral component
127 represents a range of processes in a specific spectral range, the deviation of the modelled from the observed
128 spectral component is informative about the process(es) causing the error. The second step is to separate the
129 mean square error (MSE) of each spectral component into its constituent parts: the bias, variance and covariance.

130 These time-scale specific errors, expressed in terms of bias, variance, and covariance then allow a more precise
131 diagnosis of their cause.

2.1 ERROR BREAK DOWN

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

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

EQ 1

134

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

$$bias = E(mod - obs)$$

EQ 2

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

$$MSE = var(mod - obs) + bias^2$$

EQ 3

137

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

$$var(mod - obs) = var(mod) + var(obs) - 2cov(mod, obs)$$

EQ 4

139

140 the MSE can be expressed as:

$$MSE = bias^2 + var(mod) + var(obs) - 2cov(mod, obs),$$

EQ 5

141

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

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

EQ 6

145 where

$$mMSE = \sigma_{obs}^2(1 - r^2)$$

EQ 7

146 is the minimum error achievable by an accurate (unbiased, $\overline{mod} = \overline{obs}$) and precise ($\sigma_{mod} = \sigma_{obs}$) modelling
147 system (r is the linear correlation coefficient). The first term on right hand side of Eq 6 is the mean unconditional
148 bias (how much the time averaged modelled concentration is shifted with respect to the averaged observation);
149 the second term includes variance and covariance types of error (due to differences in the amplitude and timing
150 between the modelled and observed signals), and the *MSE* is the 'unexplained' portion of the error, reflecting the
151 amount of observed variance not accounted for by a linear model (Murphy, 1995). The *mMSE* type of error is
152 caused by the variability of the observation not reproduced by the models, which includes incommensurability,
153 noise, timing of the signal, and linearization of non-linear processes, summarised by the coefficient of
154 determination (Solazzo and Galmarini, 2016).

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

2.2. SPECTRAL DECOMPOSITION AND ERROR ATTRIBUTION

159 Spectral filtering has been applied to the measured and modelled hourly-averaged time series at the monitoring
160 sites using the Kolmogorov-Zurbenko (kz) low-pass filter (Zurbenko, 1986). This allows to separate different
161 phenomena having distinct signals, such as long-term and short-term fluctuations in the observed and modelled
162 time series (Rao et al., 1997). Applications of the *kz* filter to ozone have been described in a number of previous
163 studies (Rao et al., 1997; Wise and Comrie, 2005; Hogrefe et al., 2000; 2003; 2014; Galmarini et al., 2013; Kang et
164 al., 2013; Solazzo and Galmarini, 2015 and 2016; Kioutsioukis et al., 2016).

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

$$FT(S) = LT(S) + SY(S) + DU(S) + ID(S)$$ **EQ 8**

170

171 where *S* is the time series of the species being analysed and FT is the full (un-decomposed) time series. Another
172 possibility, not explored here, is to avoid the use of the band-pass property but rather use the kz filter to filter out
173 the unwanted fluctuations directly from the FT time series.

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

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

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

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

193   1. The non-orthogonality of the spectral components is one of the major drawbacks of the signal decomposition. The relationship among the spectral components of Eq 8 is non-linear in *m* and *k* and thus an orthogonal separation is not achievable (Rao et al., 1997; Kang et al., 2013). 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 do 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), or are more convoluted (e.g. kz-Fourier Transform), or simply have not been applied as frequently as the kz filter to air quality data (e.g. Bowdalo et al., 2016). Hogrefe et al. (2003) provided an exhaustive comparison among four techniques for separating different time scales in atmospheric variables (kz, kz-Fourier Transform, wavelet transform and elliptic filter) and concluded that they all gave qualitatively similar results in terms of the variance distribution among components and that no single filter outperformed the others for all applications.

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.

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 TNO-MACC-II (Netherlands Organization for Applied Scientific Research, Monitoring Atmospheric Composition and Climate) inventory of anthropogenic emissions for the year 2009 was used, while biogenic emissions (meteorology-dependent) were specifically calculated for the year of 2010 by several groups. Five modelling systems have used the EDGAR-HTAPv2.2 emission inventory (Janssens-Maenhout et al., 2015), which complements the standard MACC inventory in regions outside EU (Table 1). The two inventories (MACC and HTAP) are approximately the same over the common part of EU (the standard MACC inventory does not cover North Africa, while it does cover eastern Europe, including Russia and Turkey) and only differ for regions outside the EU borders but within the domain boundaries, such as North Africa. Some discrepancies might exist among the two inventories (e.g in the emissions from ships). Two EU modelling systems (CHIMERE and SILAM) made results available with both the MACC and the HTAP inventories. 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 Sparse Operating Operator Kernel Emissions (SMOKE) processing system (Mason et al., 2012). Specific updates for the year of 2010 were made for several sectors, including mobile

235 sources, power plants, wildfires, and biogenic emissions. Details are given in Im at al. (2015a,b) and Pouliot et al.
236 (2015).

237 Typically, emission processors use annual emission total, while AQ models require hourly input values. Therefore,
238 proxies variables and surrogate fields are used to spatially disaggregate the annual total and to allocate them
239 temporally. The overall model accuracy heavily depends on the degree of similarity between the disaggregation of
240 total emission and the true spatial and temporal distribution (Makar et al., 2014). Furthermore, the emissions for
241 EU, being compiled on a country-wise basis, are affected by gaps and inconsistency across borders which require
242 further processing and manipulation (Pouliot et al., 2015).

243 Emissions from lightning and volcanic sources are not contained in the EU and NA emissions inventories, since not
244 all participating models include robust methods for estimating these emissions.

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

2.3.1 MODEL FEATURES

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

251 Three models (CHIMERE, SILAM, Lotos-Euros) have used the meteorological inputs extracted by the ECMWF
252 (European Centre for Medium-Range Weather Forecasts) operational archive, the Cosmo-CLM (CCLM from now
253 on) model has driven the CMAQ simulations provided by the HZG (Helmholtz-Zentrum Geesthacht) institute, and
254 all remaining models have been driven by the meteorological fields generated by the WRF (Weather Research and
255 Forecasting, Grell et al., 2005) model.

256 Bearing in mind that small changes in model configuration can produce significantly different outcomes (e.g.
257 Herwehe et al., 2011), Table 2 summarises the configuration of the WRF runs, detailing difference and
258 commonalities. Without entering in the detail of each parameterisation, the differences among the PBL
259 formulations (detailed review provided by Cohen et al., 2015) have a profound impact on the discussion of the
260 error, especially (but not exclusively) on the diurnal scale. One of the main differences is the local vs. non-local
261 closure of the PBL equations, indicating the depth over which the PBL variables influence the prediction at a given
262 point. Non-local scheme offer more advantages with respect to local ones, as the latter may not fully account for
263 deeper vertical mixing associated with larger eddies, while non-local schemes are overall more accurate in
264 simulating deeper vertical mixing in buoyancy-driven PBLs (Cohen et al., 2015). With reference to Table 2, the
265 MYNN, MYJ (Janjic, 1994) are local schemes, the YSU (Hong et al., 2006) is a non-local scheme, while the ACM2
266 (Pleim, 2007) can be regarded as hybrid one in that it incorporates local and non-local closures for potential
267 temperature and velocity, resulting in more accurate vertical mixing.

268 The land-surface processes are used to calculate the surface heat and moisture fluxes and strongly affect the
269 prediction of temperature and humidity. RUC and NOAH land surface models have shown to behave similarly over
270 US (Jin at al., 2010), while Mooney et al. (2013) found the NOAH surface scheme yielding more accurate surface
271 temperature results compared to RUC.

272

273 Six groups have operated the CMAQ (Community Multiscale Air Quality) model. The main differences among the
274 CMAQ runs reside in the number of vertical levels (minimum of 23 for CMAQ4 up to 35 for CMAQ3 and WRF-
275 CMAQ in NA) and horizontal spacing (from 12 km by WRF-CMAQ in NA down to 30 km by CMAQ3 and CMAQ4) and
276 in the estimation of biogenic emissions. CMAQ4, CCLM-CMAQ and WRF-CMAQ calculated biogenic emissions
277 using the BEIS (Biogenic Emission Inventory System version 3) either as implemented in SMOKE v2.6
278 (https://www.cmascenter.org/smoke) or as implemented directly into CMAQ while CMAQ1, CMAQ2 and CMAQ3
279 calculated biogenic emissions through the MEGAN model (Guenther et al., 2012). Moreover, the CCLM-CMAQ
280 model does not include the dust module, while the other CMAQ instances use the inline calculation (Appel et al.,
281 2013) and CMAQ1 uses the dust calculation previously calculated for AQMEII phase 2. Finally, all runs have been
282 carried out by using CMAQ version 5.0.2 except for CMAQ1, which is based on the 4.7.1 version. A series of known
283 shortcomings of CMAQ v.5.0.2 are discussed in Appel et al. (2016) (and partially addressed in the new version 5.1
284 of the model), among which is the tendency to underestimate the vertical mixing during transition periods, with
285 the net results of increasing the concentration of primary pollutants and reduce that of ozone as consequence of
286 more availability of NO_x.

287 Hereafter, more detailed information is provided for each modelling system.

288 The FMI (Finnish Meteorological Institute) has taken part with the ECMWF-SILAM system (ECMWF-SILAM_M and
289 ECMWF-SILAM_H of Table 1, indicating the instances of the SILAM model using the MACC and the HTAP emission
290 inventory, respectively). SILAM v5.4 (Sofiev et al., 2015) has been used, with meteorological input extracted from
291 the ECMWF operational archives. The simulation included sea-salt emissions as in Sofiev et al. (2011) (but not from
292 the boundaries), biogenic VOC (volatile organic compounds) emissions as in Poupkou et al. (2010) and wild-land
293 fire emissions as in Soares et al. (2015). The wind-blown dust is only included from the lateral boundary conditions.
294 The volatility distribution of anthropogenic OC was taken from Shrivastava et al. (2011). The gas phase chemistry
295 was simulated with CBM-IV, with reaction rates updated according to the recommendations of IUPAC
296 (http://iupac.pole-ether.fr) and JPL (http://jpldataeval.jpl.nasa.gov). The secondary inorganic aerosol formation
297 was computed with updated DMAT scheme (Sofiev 2000) and secondary organic aerosol formation with the
298 Volatility Basis Set (VBS, Ahmadov et al., 2012). Pressure and latitude dependent photolysis rates of the FinROSE
299 model (Damski et al., 2007) are used and reduced proportionally to cloud cover below the clouds down to half the
300 original value at full cloud cover. The SILAM model does not account for extra plume rise in addition to that
301 prescribed by the emission profiles. A known deficiency of the SILAM version used in this study is the
302 overestimation of ozone dry deposition.

303 The LOTOS-EUROS modelling system (Schaap et al. 2008, Sauter et al. 2012) has been applied by TNO (the
304 Netherlands Organization for Applied Scientific Research), using version v1.10.1. The meteorological inputs have
305 been extracted from the ECMWF operational archives. For biogenic emissions the approach as described in
306 Beltman et al. (2013) has been used. Gas-phase chemistry is based on CBM-IV (modified reaction rates, see Sauter
307 et al., 2012), secondary inorganic aerosol (SIA) formation on Isorropia II (Fountoukis and Nenes, 2009) and for
308 semivolatile species the VBS approach was used (Donahue et al. 2006, Bergström et al. 2012), with 100% of the
309 emitted OC mass in the 4 lowest volatility classes that are predominantly solid and an additional 150% in the five
310 higher volatility bins. Modelled terpene emissions were reduced by 50% to limit their contribution to SOA
311 (secondary organic aerosol) formation which was found to be too high otherwise (Bergström et al., 2012). No NO_x
312 emissions from soil were taken into account. The model includes pH dependent conversion rates for SO_2 (Banzhaf
313 et al., 2012), while only below-cloud scavenging is used for wet deposition. Mineral dust emissions were calculated
314 on-line, including emissions from road resuspension and agricultural activities, according to Schaap et al. (2009).
315 For sea spray the parameterizations by Monahan et al. (1986) and Martensson et al. (2003) were used. Photolysis
316 rates are based on clear-sky photolysis rate by Roeths flux algorithm (function of solar zenith angle) (Poppe et al.,

317 1996) and multiplied by an attenuation factor in case of clouds. The LOTOS-EUROS model does not account for
318 extra plume rise in addition to that prescribed by the emission profiles. A specific feature of LOTOS-EUROS is that it
319 only covers the lower 3.5 km of the atmosphere, with a static 25 m surface layer, a dynamic mixing layer and two
320 dynamic reservoir layers. This makes the model relatively fast in terms of computation time but has implications
321 for the vertical mixing of species for instances where the mixing layer rapidly changes in height.

322 The INERIS and CIEMAT institutes jointly applied the ECMWF-CHIMERE system. CHIMERE (version CHIMERE 2013)
323 has been run with meteorology provided by ECMWF IFS (Integrated Forecast System). Biogenic VOC emissions
324 from vegetation and soil NO emissions have been calculated with the MEGAN model (version 2.04; Guenther et al.,
325 2006, 2012). Sea salt emissions inside the domain have been calculated according to Monahan (1986). The wind-
326 blown dust is only included from the lateral boundary conditions. CHIMERE uses the MELCHIOR2 chemical
327 mechanism (Lattuati, 1997) and ammonium nitrate equilibrium was calculated with ISORROPIA (Nenes et al.,
328 1999). Dry deposition is based on the resistance approach (Emberson 2000a,b) and both in-cloud and sub-cloud
329 scavenging have been considered for wet deposition.

330 WRF-WRF/Chem1 is applied by the University of L'Aquila (Italy). The version 3.6 of the Weather Research and
331 Forecasting model with Chemistry model (WRF/Chem) has been used, modified to include the new chemistry
332 option implemented by Tuccella et al. (2015) that includes in the simulation of direct and indirect aerosol effects a
333 better representation of the secondary organic aerosol mass, calculated as in Ahmadov et al. (2012). Here only
334 direct effects have been included in the simulation, for computational expediency. The model uses RACM-ESRL gas
335 phase chemical mechanism (Kim et al., 2009), an updated version of the Regional Atmospheric Chemistry
336 Mechanism (RACM) (Stockwell et al., 1997). The inorganic aerosols are treated with the Modal Aerosol Dynamics
337 Model for Europe (MADE) (Ackermann et al., 1998). The parameterization for SOA production is based on the VBS
338 approach. The aerosol direct and semi direct effects are taken in account following Fast et al. (2006). Cloud
339 chemistry in the convective updraft is modelled using the scheme of Walcek and Taylor (1986), while the aqueous
340 phase oxidation of SO_2 by H_2O_2 in the grid-resolved clouds is parameterized with the scheme used in GOCART
341 (Goddard Chemistry Aerosol Radiation and Transport). Wet deposition from convective and resolved precipitation
342 is included following Grell and Freitas (2014). The photolysis frequencies are calculated with the Fast-J scheme
343 (Fast et al., 2006). Dry deposition and photolysis schemes were modified to take in account the effects of the soil
344 snow coverage following Ahmadov et al. (2015). The anthropogenic emissions are taken from TNO-MACC
345 inventory for 2009 (Kuenen et al., 2014) and have been adapted to the chemical mechanism used following the
346 method of Tuccella et al. (2012). .

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

354 Simulations of WRF-CAMx over EU have been performed by RSE (Italy) using CAMx version 6.10 (Environ, 2014)
355 with Carbon Bond 2005 (CB05) gas phase chemistry (Yarwood et al., 2005) and the Coarse-Fine (CF) aerosol
356 module. Input meteorological data were generated by WRF model version 3.4.1 (Skamarock et al., 2008a,b), driven
357 by ECMWF analysis fields. Grid nudging of wind speed, temperature and water vapour mixing ratio has been
358 employed within the PBL, with a nudging coefficient of 0.0003 sec^{-1}. WRF-Chem has been adopted to predict
359 GOCART dust emissions (Ginoux et al., 2001) along with the meteorology. The WRFCAMx pre-processor (version

360 4.2; ENVIRON, 2014) was used to create CAMx ready input files collapsing the 33 vertical layers used by WRF to 14
361 layers in CAMx but keeping identical the layers up to 230 m above ground level. Biogenic VOC emissions were
362 computed by applying the MEGAN emission model v2.04. Sea salt emissions were computed using published
363 algorithms (de Leeuw et al., 2000; Gong, 2003).

364 Aarhus University (Denmark) applied the WRF-DEHM modelling system over EU and NA. The DEHM model used
365 anthropogenic emissions from the EDGAR-HTAP database and biogenic emissions are calculated using the MEGAN
366 model. The gas-phase chemistry module includes 58 chemical species, 9 primary particles and 122 chemical
367 reactions (Brandt et al., 2012). Secondary organic aerosols (SOA) are calculated following the two-product
368 approach assuming that hydrocarbons undergo oxidation through O_3, OH and NO_3 and for only two semi-volatile
369 gas products (Zare et al., 2014). However, the module is simple as it does not include aging processes and further
370 reactions in the gas and particulate phase (Zare et al., 2014).

371   WRF-CMAQ1 has been applied by the ITU (Istanbul Technical University) over EU. The MCIP version 3.6 (Otte and
372 Pleim, 2010) has been used to process WRF output for CMAQ. The MEGANv2.1 (Guenther et al., 2012) model has
373 been used to calculate the biogenic VOC emissions from vegetation, using surface temperature and radiation from
374 MCIP output. CMAQv4.7.1 (Foley et al., 2010) was configured with the CB05 chemical mechanism and the AERO5
375 module (Foley et al., 2010) for the simulation of gas-phase chemistry and aerosol and aqueous chemistry,
376 respectively.

377 The WRF-CMAQ2 system has been applied by Ricardo Energy & Environment (Ricardo-E&E) over EU. It has been
378 configured using WRFv3.5.1 and CMAQ v5.0.2. The CMAQ model adopted the CB05-TUCL chemical mechanism
379 (Whitten et al., 2010; Sarwar et al., 2011a), the AERO6 three mode aerosol module (Appel et al., 2013). The MCIP
380 version 4.2 has been used to process WRF output for CMAQ. The MEGANv2.0.4 model has been used to calculate
381 the biogenic VOC emissions from vegetation, using surface temperature and radiation from MCIP output.

382 The WRF-CMAQ3 modelling system has been applied by the University of Hertfordshire and utilized the uncoupled
383 version of the WRF-v3.4.1 model and CMAQ v5.0.2. The results from WRF simulations were pre-processed for
384 CMAQ using Meteorology-Chemistry Interface Process (MCIP) version 3.6 (Otte et al., 2005). In CMAQ model, the
385 gas phase chemical mechanism was based on carbon bond chemical mechanism version 5 (Foley et al., 2010) with
386 updated toluene and chlorine chemistry (CB05-TUCL) and the aerosol chemical reaction were treated with AERO6
387 module. The biogenic emissions were derived from MEGAN.

388 The WRF-CMAQ4 simulation has been performed by the Kings College (UK) using CMAQ v5.0.2 (Byun and Schere,
389 2006) with CB05 chemical mechanism that includes aqueous and aerosol chemistry. The CMAQ model is driven by
390 meteorological fields from the WRF v3.4.1. The anthropogenic emissions for most part of the model domain are
391 from MACC and the missing information have been filled with the emissions provided by EDGAR/HTAP. The
392 biogenic emissions were estimated using the BEIS3 model. The dust (Tong, et al, 2011) and sea-salt (Gantt et al.,
393 2015) emissions are generated using CMAQ inline modules.

394 HZG has used the COSMO-CLM meteorological model to drive the CMAQ model. For AQMEII3 the CMAQ version
395 5.0.1 was used, with the CB05-TUCL scheme and the multi-pollutant aerosol module AERO6. CMAQ was run using
396 the optional in-line calculation of dry deposition velocities. Wet deposition processes include in-cloud and sub-
397 cloud scavenging processes. All atmospheric parameters were taken from regional atmospheric simulations with
398 the COSMO-CLM (CCLM) mesoscale meteorological model (version 4.8) for the year 2010 (Geyer, 2014) using NCEP
399 forcing data employing a spectral nudging method for large-scale effects (Kalnay et al., 1996). CCLM is the climate
400 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. Biogenic emissions and NO emissions from soil were calculated using the BEIS3 model. 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. 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 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. No imputation on missing values has been performed.

In addition, we also make use of vertical profiles of ozone, temperature and wind speed data measured by ozonesondes and 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 over the sub-regions shown in Figure 1, and prior to the spectral decomposition (the original time series

443 have been spatially averaged first and then this spatial average time series has been spectrally decomposed). As
444 noted in the introduction, unsupervised hierarchical clustering was used to determine sub-regions where the LT
445 and SY components showed similar characteristics – spatial averaging within these sub-regions was carried out due
446 to the similarity of the observation data within these regions implying they will experience common physical and
447 chemical characteristics. Errors due to the heterogeneity induced by country-specific emission profiles (in EU) are
448 therefore included in the DU component. As a consequence of the spatial averaging, the relative importance of the
449 ID component is likely reduced, since the ID fluctuations are highly variable in space (Hogrefe et al., 2014). Further,
450 no land-use type filtering has been applied to the stations used for evaluation. While this choice has limited impact
451 on the SY and LT components (Solazzo and Galmarini, 2015; Galmarini et al., 2013), the DU components of some
452 species (such as ozone, PM, NO_x) might be strongly influenced by the vicinity of urban stations to emissions
453 sources.

454 Details of the modelled regions and number of receptors are reported in Table 3.

3. RESULTS

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

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

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

3.1 METEOROLOGICAL DRIVERS: TEMPERATURE, WIND SPEED, AND WIND DIRECTION

3.1.1 WIND SPEED AND TEMPERATURE
472 The RMSE for surface temperature and wind speed is reported in Figure 2 (EU) and Figure 3 (NA). For EU (Figure
473 2a), the RMSE of the full (i.e. not spectrally decomposed and denoted as "FT" in the plots) time series of
474 temperature for the entire year is, on a seasonal average, on the order of ~0.5-2K (but often exceeding 3K in EU3),
475 with higher values typically occurring in spring and winter. The CHIMERE and SILAM models (both directly driven
476 by the global meteorological fields provided by ECMWF) report the smallest error in EU1 and EU2, while the
477 WRF/Chem2 model has the largest error in all sub-regions (up to ~5K for EU3 in summer) which is largely caused
478 by the unusually large error in the SY component when compared to other models. The RMSE of the LT component
479 resembles the behaviour of the full time series, with the highest error in spring and winter (on average). The RMSE
480 of the SY component is below ~2K (slightly higher in EU3) except for WRF/Chem2, whereas the DU component
481 shows a more marked regional dependence, with the EU3 sub-region reporting, on average, approximately 50%

482   higher seasonal error than the other two sub-regions, more pronounced in summer. The correlation coefficient is
483   higher than 0.90 for the majority of models and spectral components (Table S2).

484   The bias for temperature is predominantly negative (model underestimation) for all EU models and sub-regions,
485   except for WRF-CMAQ4 in EU3, where the model overestimates the measured temperature in summer and winter.
486   According to Katragkou et al. (2015), cold bias during summer by WRF is typically related to the CAM radiation
487   scheme, and, in general the land surface model is pivotal in determining the sign and amount of bias (Mooney et
488   al., 2013), and in particular the combination of NOAH surface scheme and CAM radiation model seems more prone
489   to cold bias.

490   For NA (Figure 3a) the temperature RMSE of the WRF-DEHM and CCLM-CMAQ models (peaking in winter and
491   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
492   DU component is significantly higher (between 0. 5K and 2K). The WRF-CMAQ model has a small bias (LT error
493   small) so that the overall error is dominated by the error in the DU component. The bias is negative for the WRF-
494   DEHM model in all sub-regions and has the same sign for CCLM-CMAQ and WRF-CMAQ, i.e. negative in spring and
495   positive in the other seasons (although for NA2 and NA3 WRF-CMAQ reports a slightly negative bias also in winter)
496   (Table S2).

497   The RMSE of the surface WS for EU shows large model-to-model variability, more markedly for the LT and SY
498   components (all sub-regions, Figure 2b), whereas the error of the DU component is more evenly distributed across
499   models (and significantly higher in EU3, where low-wind speed conditions are predominant). Although the
500   meteorological fields are assimilated within the models (either from NCEP or from ECMWF, see Table 2), there are
501   profound differences in the way these fields are ingested and interpolated to the model grid, as well as differences
502   in the parameterisation of the boundary and surface layer which impact the modelled wind speed and
503   temperature. For example, the two instances of WRF/Chem applied the assimilation of the meteorological fields
504   (wind speed, temperature, and relative humidity) of global meteorological fields only above the PBL, whereas
505   other models (e.g. WRF-CAMx) assimilated the global data also within the PBL. For the models directly driven by
506   the global fields, (e.g SILAM, CHIMERE) the seasonal error for WS (~0.5-1 ms^{-1}) and temperature (0.4-1.2K) (Figure
507   2a,b) can be considered as the uppermost limit the accuracy of the models can achieve. Thus, the assimilation and
508   interpolation methods errors (which are specific to the configuration of the meteorological model) can add up
509   more than 1.5K and 2ms^{-1} to the total error.

510   The full WS time series of the WRF-DEHM, WRF/Chem1 and WRF/Chem2 models report the largest error (in excess
511   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
512   remaining models have an error of 0.5-0.7m/s. Most of the error is apportioned to the LT component, with the SY
513   and DU below 0.3 m/s (except for WRF-CAMx and the other models mentioned above).

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

519   For NA (Figure 3b), the WRF-DEHM model reports an error of ~1-1.2 m/s during all seasons and sub-regions, while
520   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
521   the SY and DU components is small (below 0.3m/s for each season) for both models. Both models are biased high
522   (all instances) and the correlation coefficient is in the order of ~0.9 or above (Table S3).

3.1.2 VERTICAL PROFILES OF WIND SPEED AND TEMPERATURE

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 4. The launches in EU predominately occurred during daylight hours, whereas for NA measurements are also available 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 responsible 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 nudging of meteorological fields only above the PBL, and only during the first 12 hours of meteorological spin-up, while for the other WRF instances the nudging is active during the entire run). 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 WIND DIRECTION

The spatial and temporal distributions of wind direction (WD) are reported in Figure 8. The boxes summarize the temporal and spatial variability of the WD values at the receptors of each sub-region (no averages have been applied). For EU1 (Figure 8a), the median of all models but WRF-CAMx is within ±5 degrees that of the observation, and similarly for EU2. Also the modelled 22th and 75th percentiles are in line with the observations in these two sub-regions (the CCLM model predicts slightly larger variability).

The EU3 sub-region is topographically more complex, and the analysis is based on four stations with only 55% data validity over the entire period. Southern winds are predominant (based on the observation) while the models show large variability and, even the several instances of WRF (but not all) and the ECMWF data tend to under-predict the median value. The only two models over-predicting the median observed value are WRF-CMAx and WRF-CMAQ1, both apply grid nudging also within the PBL along with WRF-CAMQ4 which, however, shows a slight under-estimation.

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 ~140-220 ppb in EU1 and EU2, and up to 600 ppb 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 overestimation of WS discussed in section 3.1 also contributes to further dilute the concentration of primary species such as CO (for example corr($bias_{CO}$, $bias_{WS}$) = 0.60 for the CMAQ4 model in EU2 during winter).

604 The error due to variance in EU (under-estimated by the models) and *mMSE* are significant in the DU and SY
605 components in winter (Figure 10a). In particular, the variance error of winter DU is small compared to the *mMSE*,
606 which accounts for almost the entire DU error, up to over 30 ppb. For SY, the model SILAM_H shows an *mMSE*
607 error of over 75 ppb, the variance part being approximately null. On average, the DU and SY errors are
608 approximately similar for all EU models (~45 ppb for DU and ~65 ppb for SY), indicating some common error
609 leading to poor associativity, which typically corresponds to lagged timing of the observed and modelled signals.
610 An example of such might be the poor representation of the diurnal variation of the emissions (e.g. Makar et al.,
611 2014). A further reason could stem from the lack of temperature dependent emissions (the current emission
612 inventory processing approach employs constant temporal emission profiles, and therefore cold/warm episodes
613 are not incorporated in the modelled emissions while these episodes do affect real-world emissions). The lack of
614 temperature-dependant emission is likely to have a strong effect for CO, as about 50% CO emissions comes from
615 residential heating (at least in mid/north European countries). A test to this hypothesis is currently under
616 investigation by running the CCLM-CMAQ model with a set of emissions using temperature data for the temporal
617 disaggregation for residential heating emissions.

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

3.3.1.A SENSITIVITY SIMULATIONS WITH REDUCED EMISSIONS AND BOUNDARY CONDITIONS

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

627 $$\%RMSE = 100 * \frac{RMSE_{CO}^{s20\%} - RMSE_{CO}^{base}}{RMSE_{CO}^{base}}$$

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

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

3.3.2 NO

642 NO is emitted by both natural and anthropogenic sources and its chemistry patterns are closely connected to those
643 of NO_2 and ozone. Due to the fast ozone-NO titration reaction, the uncertainty in emissions, transport, and vertical

644 mixing dominates the uncertainty in chemistry. As no observational data was available for NA, the discussion is
645 limited to EU. The European Environment Agency (EEA) reports an estimated uncertainty for NO_x emission of ~20%
646 (EEA, 2011); Vestreng et al. (2009) found ±8-25% uncertainties in EU NO_x emissions, in line with other similar
647 bottom-up uncertainty studies (see Pouliot et al., 2015). A further source of uncertainty and model to model
648 difference is the vertical emission profiles adopted and how this is interpolated to the vertical grids used by the
649 models. Within the SILAM model, for example, the vehicular traffic emissions are released largely at the bottom of
650 the first layer and this sub-grid information about the vertical location of the plume used in the vertical transport
651 scheme further supresses the mixing to the upper layers, thus keeping the surface concentrations higher.

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

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

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

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

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

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

684 3.3.2.A SENSITIVITY SIMULATIONS WITH REDUCED EMISSIONS AND BOUNDARY CONDITIONS
685 The analysis discussed in Section 3.3.1.A is repeated here for NO and results are presented in Figure 13. A decrease
686 by 20% of the amount of NO in the domain produces a variation of RMSE of ~8% (averaged over models and
687 spectral components). A naïve projection indicates that a reduction of 100% (thus removing the production of NO
688 from emissions and boundary conditions) would produce a variation of the error of ~35%. Such an amount is less
689 than that found for CO (~50%, section 3.3.1.A), which is consistent with the photochemical processes involving NO
690 but not CO.

691 The LT component is the most sensitive to changes for NO, with an average of ~17% error variation (and up to 20%
692 in autumn, both positive and negative). Again, the SILAM model is the most sensitive to changes in the amount of
693 pollutants entering the domain. Remarkable differences between the 's20%' scenario and the base case are
694 detected for summer and autumn (LT error variation of 100%) (Figure 13). The improvement of the error of SILAM
695 (and of the other models) for the 's20%' scenario is due to the overestimation of NO mean concentration in the
696 base case (positive bias, Table S6).

697 ### 3.3.3 NO_2
698 Primary NO_2 is emitted by a variety of combustion sources and plays a major role in atmospheric reactions that
699 produce ground-level ozone. NO_2 is also a precursor to nitrates, which contribute to PM formation. As for NO, only
700 a small portion of the total error is expected to stem from the boundary conditions. The AQMEII3 modelling
701 systems attribute a fraction of NO_2 emission ranging between 3% and 10% of the total NO_x emissions (some
702 models treat the NO_2 emission from the transport sector differently, see Table 1). The results of the error analysis
703 discussed hereafter do not reveal, though, grouping of model behaviour consistent with the choice of the NO_2 to
704 NO_x emissions ratio, also in light of the fast chemistry between NO and NO_2.

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

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

714 The bias is the main contributor to the NO_2 error and stems from a model under-prediction of the mean observed
715 concentration during the entire year (but, with the exception of the winter season, is positive for WRF-CMAQ in NA
716 and WRF-CMAQ1 in EU) (Table S7). The bias is probably caused by a combination of factors, including emissions
717 estimate (e.g. underestimation of residential combustion), PBL height and vertical mixing at night (when wood
718 combustion emissions tend to be maximum, e.g. Denier Van Der Gon et al., 2015), and missing processes acting as
719 systematic errors, such as shading effects of forested canopies (e.g. Makar et al., 2016). However, the tendency of
720 NO_2 measurements to be likely overestimated by the majority of commercially available instruments for detecting
721 NO_x (Steinbacher et al., 2007) needs to be taken in to account. The magnitude of the bias is higher in EU (from
722 ~1.3pbb of WRF-CMAQ1 in EU1 to ~-12.5 ppb of CCLM-CMAQ in EU3) than in NA (the maximum being ~5.5 ppb in
723 NA3 by the WRF-DEHM model), with the mean observed values being of 11.5 ppb and 10.5 ppb for EU and NA,
724 respectively.

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

729 The median value of the modelled accumulated deposition per unit area (Fig. S11) for NO_2 ranges from ~0.4 to
730 ~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
731 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
732 uniform across the EU models, while the deviation between the two NA models for deposition is not negligible,
733 also in light of the different native grid sizes of 50km and 12km (WRF-DEHM and WRF-CMAQ, respectively).
734 Therefore, for the majority of the EU models model-to-model differences in the error are unlikely due to significant
735 difference in the deposition, while it remains a possibility for NA.

736 The magnitude of the error for NO_2 resembles that of NO and ozone, although the apportionment reveals
737 significant differences. In fact, while NO includes variance error and a uniform share of *mMSE*, the LT error of NO_2
738 for winter is almost completely determined by the bias, for both continents (Figure 15 and Figure 16). The other
739 NO_2 spectral components (ID, DU, SY) reveal more profound difference with respect to NO, both in terms of bias
740 and of error apportionment. The ID error for NO_2 is even smaller than that of NO (less than 1 ppb) and can be
741 regarded as noise. Also the DU (~1.5 ppb) and SY (~1 ppb) errors are considerably smaller than for NO (both
742 continents), although the DU error presents some excess of variance for WRF-CMAQ3 and the two instances of the
743 CHIMERE model (Figure 15).

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

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

753 3.3.4 OZONE
754 Due to the adverse effects on human health and to the impact on climate, tropospheric ozone is regulated in EU
755 and NA and substantial efforts are made to improve the models' predictive skill for this pollutant. Tropospheric
756 ozone can be either transported from regions outside the modelled domain, be the result of
757 stratosphere/troposphere exchange, or be produced locally by photochemistry through oxidation of VOCs (volatile
758 organic compounds) and CO in the presence of NO_x and sunlight. Due to its photochemical nature, ozone
759 production is directly influenced by temperature through speeding up the rates of the chemical reactions and
760 increasing the emissions of VOCs (e.g. isoprene) from vegetation (Jacob and Winner, 2009). Along with dry
761 deposition, chemistry can act as local sink to ozone depending on the photochemical regime.

762 Results of the AQMEII3 modes for ozone are reported in Figure 17 and Figure 18, and in Table S4. Overall, the
763 correlation between modelled and observed ozone time series is higher for the winter and fall seasons than the
764 spring and summer seasons in EU, while the opposite holds true in NA where the maximum correlation is observed
765 in summer (all sub-regions) (Table S4). In EU, the RMSE is generally lower in winter than in the warm seasons

766   (summer and spring) (RMSE in summer ranges between 4.3 ppb of WRF/Chem1 in EU1 and 21 ppb of WRF-CAMQ1
767   in EU3), with the exception of the CCLM-CMAQ model for which the RMSE peaks in autumn (all sub-regions).

768   Due to the strong and well defined diurnal cycle characterized by ozone formation and loss, the correlation
769   coefficient is generally higher for the DU component, while it tends to be lowest for the SY component (Table S4
770   and Figure 18). The SY component often exhibits the lowest correlation among all components, especially in
771   summer (EU) and spring (NA), possibly due to the combined effect of transport of precursors, deposition and
772   chemistry (formation/depletion of ozone from precursor emission in the regions where the ozone is transported)
773   (Bowdalo et al., 2016). However, the SY error is generally small (~2-3 ppb, although higher for EU3, where the SY
774   error is double that of the other sub-regions) and is mostly due to *mMSE*. It is thus characterised by poor
775   coefficients of determination and underestimated variability (Eq 6). Therefore, the SY component suffers from low
776   precision (for some models $r < 0.3$) meaning that the variability of the synoptic mechanisms needs further
777   attention, especially in the meteorological conditions leading to high ozone level episodes and in relation to
778   temperature, cloudiness, and radiation. The WRF/Chem2 model (having the highest error for temperature, Figure
779   2b) reports the largest SY error for ozone (especially the variance part). For this model, the correlation between
780   the ozone and the Temp error for the SY component $corr(err_{O3}, err_{Temp})_{SY}$ is 0.44 for the summer months in EU2
781   (not shown), among the highest, which helps to explain part of the SY error for ozone. Further possible causes
782   could be associated to tropopause folding events, especially downwind of mountain areas (e.g. Bonasoni et al.,
783   2000; Makar et al., 2010), which would also be in line with the larger synoptic error of ozone in EU3 (Figure S4b),
784   comparable for all models in the range of 3-4 ppb. In order to characterise better the *mMSE* part of the error for
785   the periodic components, such as DU and SY, analysis of the phase and amplitude are ongoing.

786   The error of the DU component is largely due to the *mMSE* term (Figure 18a) which is comparable for all models in
787   the range of 2-5 ppb, with some significant excess of variance for WRF-CMAQ2 and WRF-CMAQ3 in EU2 (~5 ppb).
788   One possible reason is the dynamics of the nocturnal PBL as well as the timing of the ozone cycle, with an either
789   too fast or too slow modelled ozone peak (e.g. Pirovano et al., 2012). Limitations of the models to reproduce the
790   amplitude and phase of the daily ozone cycle were already highlighted in the first and second phase of AQMEII,
791   mostly related to the representation of night-time and stable conditions. Further, the variance error for WRF-
792   CMAQ2 and WRF-CMAQ3 can be induced by the bias of temperature and/or concentration of ozone precursors.
793   For WRF-CMAQ2 (WRF-CMAQ3), $corr(err_{O3}, err_{Temp})_{DU}$ is 0.88 (0.94) and $corr(err_{O3}, err_{NO2})_{DU}$ is 0.86 (0.83) (summer
794   months, EU2) (not shown), which indicates that the error of the Temp and NO_2 fields are strongly associated with
795   the error of ozone at the DU scale. PBL representation during transitions is a long standing issue of AQ models.

796   The error in the LT component is dominated by the bias error (Figure 18) (almost completely for NA) although with
797   significant exceptions in EU (for CCLM-CMAQ the *mMSE* error of the LT component is larger than the bias portion).
798   The May-September ozone LT bias for EU2 peaks at 12-13 ppb (WRF-CMAQ1), while it is ~6 ppb in NA3 (but in
799   excess of 20ppb in NA2 by the WRF-DEHM model) (the yearly average measured ozone mixing ratio is 26.5 and 29
800   ppb for EU and NA, respectively). The bias of the precursors and of the meteorological fields is typically highly
801   correlated with the bias of ozone. For instance, in EU2 for the WRF-CMAQ1 model $corr(bias_{O3}, bias_{Temp})$ is 0.65 and
802     $corr(bias_{O3}, bias_{WS})$ is 0.81 (summer months). The almost null NO_2 bias for CMAQ1 (among the lowest), combined
803   with the positive bias for NO suggest that chemistry also affects the ozone bias of CMAQ1. Furthermore, the excess
804   of ozone intrusion for the troposphere (discussed next) may also factor in determining the high positive bias at the
805   surface for this model.

806   According to Bowdalo et al. (2016) the bias of the ozone amplitude cycle linearly evolves with NO_x emissions,
807   suggesting that correction of the error for ozone needs to start from NO_x emissions. Otero et al. (2016) have
808   shown that meteorological drivers account for most of the explained variance of ozone, especially over central and

809 northwest Europe. One of the main drivers of ozone is the daily maximum temperature, in relation to the effect of
810 temperature on emissions of VOCs. Therefore, while part of the bias error is possibly due to NO_x emissions, the
811 *mMSE* and variance error are also likely induced by error in meteorology. Other documented of biases are
812 transcontinental transport in winter (Hogrefe et al., 2011) and missing processes during spring and summer, such
813 as the large scale effect of the absence of forest shading in the models (Makar et al., 2016), a too rapid production
814 of ozone form available precursors together with an underestimation of ozone deposition (Herwehe et al., 2011).
815 Im et al. (2015b) also indicated a range of factors determining the difference in performance among models, such
816 as the chemical mechanism, biogenic module and VOC pre-processing and difference in microphysics affecting the
817 photolysis, temperature and radiation acting on the production of ozone.

818 Although the concentration peaks are associated with the ID and DU components, the contribution to the total
819 error of the ID component is small (< 2 ppb) due to the flattening of the spikes operated by the spatial averaging
820 carried out prior of the spectral decomposition. The noise of the ID component is reflected by the correlation
821 coefficient being lower than the correlation of the DU component.

822 3.3.4.A OZONE VERTICAL PROFILES
823 Several studies have demonstrated the importance of extending the evaluation of air quality models to the
824 troposphere (e.g. Solazzo et al., 2013; Makar et al., 2010; Herwehe et al., 2011), not only because of the vertical
825 turbulent transport, but also for the key role played by coupling of the PBL and the free troposphere aloft in
826 determining the ozone intrusion to the surface. In this section profiles of modelled ozone are compared against
827 ozonesondes measurements.

828 A summary of the records provided by the ozonesondes for ozone are reported in Table 4. Plots of the simulated
829 and observed ozone levels at fixed heights (through the ENSEMBLE system models and measurements are paired
830 at the heights of 0, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000 m) are reported in Figure 19 and Figure
831 20. The ozonesonde data are mainly available during daylight, although two stations with night-time data are
832 available for NA (Table 4).

833 Overall, the general tendency of the models in both continents is to underestimate the ozone levels above the PBL,
834 suggesting that not enough ozone enters the continental domains through the inflow boundaries. The most
835 significant underestimation (~10 ppb) is observed at the two stations closer to the west boundary for EU (stations
836 318 and 043). The boundary layer deficit of ozone is a long standing issue, as similar conclusions were derived for
837 the first (Solazzo et al., 2013) and second (Im et al., 2015b; Giordano et al., 2015) phase of AQMEII, as well as in
838 other studies (Katragkou et al., 2015), emphasizing the strong dependence of regional models on the lateral
839 boundary, whose effects propagate far into the interior of the domain.

840 Towards the interior of the EU domain (stations 134, 157, 242) the profiles are in closer agreement with the
841 observations, with the WRF-CMAQ1 model performing the best throughout the troposphere, possibly due to the
842 overestimation of the entrainment of upper tropospheric ozone, as revealed by the strong gradient of WRF-
843 CMAQ1 at 6000m (Figure 19). With respect to the other models (and SILAM in particular), the CMAQ runs show
844 larger ozone availability in the residual layer above the PBL, which act as a reservoir of ozone that becomes
845 depleted the next day, increasing the concentration at the surface. Possibly, the PBL and vertical mixing within
846 these models is too weak (Appel et al., 2016). Further analyses restricted to specific season and time of the day are
847 required to validate this hypothesis.

848 For NA (Figure 20), the general tendency is of slight to consistent (stations 71 and 75) over-estimation within the
849 PBL and underestimation aloft for the WRF-CMAQ model and of overestimation (stations 107, 456, and 458 –
850 afternoon/night launches) at the surface and mild underestimation above the PBL for the WRF-DEHM model.

851 3.3.4.B RELATIONSHIP BETWEEN THE BIAS OF OZONE, NO_x AND TEMPERATURE
852 The relationship between the bias of NO and the bias of ozone is reported in Figure 21 for the EU2 region (similar
853 plots including the bias of NO_2 for EU and NA are provided in the supplementary material). A linear relationship
854 between the biases of the two species is detectable, more evident in winter. Large, positive ozone bias is driven by
855 underestimation of NO (a primary species) whereas the largest negative ozone bias correspond to the largest
856 overestimation of NO. The role of the temperature bias is less clear, but the NO_2 and ozone relationship (Figure S7)
857 suggests that large NO_2 bias is associated with temperature under-prediction. The partition of NO_x emission into
858 primary NO and NO_2 seems to suggest that the models adopting a 95%-5% ratio suffer lower ozone bias (at least in
859 winter), although in general the clustering of models based on the NO/NO_2 share of total NO_x emission is far from
860 robust. A simple linear regression between NO bias and ozone bias (based on the yearly time series) among the EU
861 models suggests that the NO_x and temperature biases can explain, on average, ~35% and ~16% of the variability of
862 the ozone bias, respectively.

863 3.3.5 SO_2
864 SO_2 is another primary regulated pollutant which, in EU and NA, is mainly emitted from coal power plants and also
865 from the residential heating and waste disposal sector. SO_2 acts as a precursor to sulfates, which are one of the
866 main components of PM in the atmosphere. Any error in SO_2 is likely inherited by these secondary species. The
867 majority of models employed the prescribed vertical distribution by EMEP (Vestreng and Støren, 2000), while
868 CMAQ4 in EU and WRF-CMAQ in NA adopted the Briggs plume rise algorithm (Briggs, 1971; 1972) accounting for
869 the effects of modelled meteorology, and SILAM, CHIMERE, and CCLM-CMAQ adopted the sector dependent
870 vertical emission profiles as in Bieser et al. (2011b). The EEA reports an estimated uncertainty for SO_2 emission of
871 ~10% (EEA, 2011), therefore SO_2 emissions are expected to be more accurate than NO_x emissions. This is reflected
872 in the low bias in both continents (~1-2 ppb in winter, mostly due to model underestimation) (Figure 22 and Figure
873 23). The averaged observed concentration of SO_2 is of 1.92 ppb and 2.7 ppb in EU and NA, respectively.

874 The seasonal modelled error for SO_2 ranges, on average, between 0.65 and 1.3 ppb in EU and between in excess of
875 ~1 and 5 ppb in NA (the maximum error in NA2), peaking in autumn.

876 In EU and NA1, the error of ID, DU and SY components is comparable for all seasons and, on average, below 0.6
877 ppb. There are some exceptions, most notably the WRF-CMAQ3 model, which is the only one significantly biased
878 high (Figure 23a) and shows an excess of variance significantly larger than the other models. By contrast,
879 WRF/Chem2, CHIMERE and L.-Euros show significant low bias (the latter two models have the smallest number of
880 vertical layers). Overall, though, the bias error does not group consistently by PBL scheme and/or vertical
881 resolution. For example, CMAQ2, CMAQ3, CMAQ4 employ the same PBL scheme based on ACM2 and have
882 comparable number of vertical levels (CMAQ3 has even more), but the bias of CMAQ3 is much larger than that of
883 CMAQ4 and CMAQ2 which, in turn, have comparable bias but opposite in sign. The two instances of WRF/Chem
884 show significantly different bias, which might be due to the different PBL and cloud scheme, influencing the SO_2
885 oxidation (Table 2).

886 The large variability of the model-to-model error (especially in EU) and correlation coefficient in both continents is
887 an indication that the mechanisms governing the initial mixing and subsequent transport and chemical
888 transformation suffer from different sources of error, at all scales. In no instance the correlation coefficient is
889 consistently above 0.5 for all seasons and spectral components while there are several instances of negative

890   correlation between the spectral components of observed and modelled SO_2 (e.g. CCLM-CMAQ model in EU and
891   several others). The poor correlation coefficient of, especially, the ID and DU components for both continents,
892   indicates that the peaks of the SO_2 concentration are not caught by the models, leading to low precision. Although
893   the mean fluctuations are, generally, well reproduced (low variance error in both continents), it remains a
894   significant portion of unexplained variance (*mMSE*) error, which can derive from meteorology and chemistry.
895   Bieser et al. (2011b) showed that the height of the release and vertical distribution of the SO_2 emission influence
896   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
897   are the major source of SO_2 further analysis should investigate the impact of differences in the vertical emission
898   distribution between models.

3.3.6 PARTICULATE MATTER

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

902   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
903   dictated by the availability of hourly measurements in the two continents. The RMSE distribution is reported in
904   Figure 24 (PM_{10} for EU) and Figure 25 ($PM_{2.5}$ for NA). The error distribution for EU reveals that, despite the large
905   numbers of modelling options and parameters characterising the chemistry and physics of particles, the error
906   distribution for DU and SY is homogeneous among the EU models. For these components the error is
907   approximately uniform over seasons, although with some exceptions (significantly higher in EU3, although based
908   on two receptors only). EU3 is a small area compared to EU1 and EU2, but is densely populated, intensively
909   farmed, with a large amount of wood burning in winter, and agricultural area in summer. It is surrounded by
910   mountains and stagnant flow conditions are predominant. It is, thus, a challenging area for current modelling
911   systems, especially for primary species such as PM.

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

915   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
916   (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
917   EU3) with the largest share in autumn, spring, and winter and slightly lower in summer. The error of the LT
918   component ranges between ~11-15 µg m^{-3} in EU1 and EU2 and up to 25 µg m^{-3} during winter in EU3.

919   The analysis of the correlation coefficient reveals that the model to model differences in the correlation coefficient
920   with the observed component time series tend to be most pronounced for the DU and ID components, indicating
921   that these two components are pivotal in determining the overall model skill in terms of capturing observed
922   fluctuations in PM_{10} concentration. In more detail, the correlation is poor for the DU component (especially in EU2
923   and EU3, Table S9), possibly due to PBL dynamics and emission profiles (as discussed above for the RMSE at the DU
924   scale). The LT component has correlation values highly varying among models and, for the same model, among
925   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).

926   In winter the LT and SY error is more severe likely due to the larger uncertainties in PM_{10} emissions of combustion
927   processes (wood burning, residential heating) (Van der Gon et al., 2014), as well as due to the current limitations in
928   modelling the vertical mixing during stable conditions, as mentioned for the gaseous species (especially CO, being
929   another primary species). The majority of the EU models show an LT error in winter between 12 and 16 µg m^{-3},
930   eight models above 16 µg m^{-3} and only one (WRF-CAMx) below 10 µg m^{-3}. The absence of background sea-salt for

931 all EU models (see end of section 2.3) can also be responsible for low bias of the LT component for PM_{10}, especially
932 in the vicinity of the coastline.

933 The SY winter error exceeds 5 μg m^{-3} for all models (all sub-regions) and three instances (WRF-CAMx, WRF/Chem1
934 and WRF/Chem2, this latter showing the highest accumulated deposition for $PM_{2.5}$, Fig. S11) report an error above
935 7.5 μg m^{-3}, possibly due to the low nitrate concentration and high sulphate concentration during winter months,
936 resulting from the GOCART parameterization of the aqueous cloud chemistry. All the remaining models have
937 comparable *mMSE* and variance errors (Figure 26), and are biased low (model under-prediction), possibly due to
938 missing PM source and overestimated surface wind speed. As for the WRF-CAMx model, the low bias on LT
939 component and the relatively high *mMSE* error in the SY fraction suggest that the model was able to capture the
940 mean magnitude of PM concentration over the entire year, but failed in reconstructing the correct variability of
941 the different episodes, whose timing is generally driven by the synoptic time scale.

942 The analysis of *corr*(bias$_{Temp}$, bias$_{PM10}$)$_{LT}$ shows that the error of these two variables are related, especially during
943 the spring months and more consistently in EU3 (up to 0.74 for the WRF/Chem1 model) and during autumn in EU1
944 (the bias of Temp and the bias of PM_{10} are anti-correlated up to -0.67 for CMAQ1). Other models (e.g. the CAMx
945 model), on the other hand, do not show any significant correlation.

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

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

956 As inferred for the species discussed above, the uniformity of model behaviour is indicative of errors stemming
957 from external fields, likely emissions, where missing sources of PM can affect the error within certain time scales
958 for all models. Further common causes of error are intrinsic to the model-observation comparison as modelled
959 PMs is commonly dry while this is not always the conditions for the measurements. For instance, the filter-based
960 gravimetric measurements as recommended by the European Committee for Standardization (CEN) are likely to
961 retain part of the particle-bound water after the filter conditioning at a constant temperature of 20° C and relative
962 humidity of 50%. Recent findings by Prank et al. (2016) report the aerosol water content from the gravimetric
963 measurements to range between 5 and 20% for PM2.5 and between 10 and 25% for PM10. The particle-bound
964 water was found to be associated with hygroscopic particles such as sulphate, nitrate, and organic compounds.
965 This remaining water content can be up to approximately 10-35% depending on the chemical composition of
966 aerosols being measured (Tsyro, 2005, Kajino, et al., 2006, Jones and Harrison, 2006). The water aerosols should
967 therefore be accounted when compared with these measurements. Part of the problem lies in secondary organic
968 aerosol. In winter, in particular for wood burning part of the emissions are condensable gases that rapidly change
969 to the aerosol phase (Van der Gon et al 2014), but are missed since they are not part of the presently used PM
970 emission inventory. In summer, biogenic emissions that contribute to SOA formation and their yields are quite
971 uncertain. A good representation of SOA is still a problem for all models. In spring, the application of manure and
972 fertilizer leads to peaks of NH_3 emissions and subsequent NH_4 aerosol formation, contributing to PM_{10} and $PM_{2.5}$.

973 The timing of these emissions is parameterized based on long-time averages, whereas in practice they are strongly
974 related to meteorology. This can explain part of the discrepancy on the diurnal to synoptic time scale (Hendriks et
975 al 2015).

976 ## 4. MEMORY OF THE SIGNAL AND REMOVAL PROCESSES: THE CASE OF OZONE

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

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

983 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
984 observed LT component falls to zero, and finally

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

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

987

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

996 In the supplementary material Figure S9 and Figure S10, the *acf* for the network-wide spatial average and for the
997 full year is reported. The *acf* is calculated for the LT component of the observed (first panel) and modelled ozone
998 time series. The zero of the *acf* and the slope of the decay of *acf* of the observations (approximately a straight line
999 from 1 to 0 in 2000 hours) are replicated by the models with various degree of success (Figure S10). Our intent is to
1000 apply this analysis to the seasonal ozone time series at each receptor, and derive useful information about the
1001 modelled removal/production processes. The spatial analysis is proposed for ozone, for the months of May to
1002 September (Figure 28 and Figure 29) and for the full year (supplementary material Figure S9 and Figure S10).

1003 The average life time of ozone in the troposphere is of approximately 20-30 days (Solomon et al., 2007). By
1004 analysing the LT component (processes > ~21 days) we therefore screen out the daily removal/transformation due
1005 to chemistry and can focus on seasonal transport, deposition of the free tropospheric ozone, long term chemistry
1006 (seasonal changes in vegetation that affect biogenic VOCs emissions and ozone deposition, and also the monthly
1007 variations applied to the anthropogenic emission) and influence of boundary conditions. The structure of the *acf*
1008 also benefits from the removal of short time scale processes as it is less affected by noise and the results are easier
1009 to interpret.

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

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

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

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

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 direction, 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.

1048 This work is primarily meant to provide a wide overview of the performance of current regional AQ modelling
1049 systems and to set the basis for additional diagnostic analysis that is currently in progress.

1050 The model evaluation is carried out by quantifying the components of the error (bias, variance, *mMSE*) at four
1051 time-scales (ID, DU, SY, LT) each describing physical processes in a specific time range. The bias and variance
1052 measure the departure from the first and second moment of the observed distribution (mean and standard
1053 deviation), while the *mMSE* accounts for the unexplained observed variability. The apportionment of the error to
1054 the relevant time-scales and the analysis of the quality of the error have revealed that the LT bias is, by far, the
1055 first cause of error, followed by the variance error (fluctuations about the mean value) of the DU component and
1056 the unexplained variance of the DU and SY components, depending on the species and season. In more detail:

1057 • The mean concentration of the primary species (NO, CO, PM_{10}, SO_2) is underestimated by the vast majority of
1058 the models in both continents, more markedly during the winter and autumn seasons. The largest share of
1059 error for these species is the bias of the LT components, most probably due to error of the fluxes at the
1060 boundaries (emission, deposition, and boundary conditions) and to the effects of comparing point
1061 measurements to volume averaged concentrations.
1062 • The bias is, by far, the primary source of error and the most important from a model evaluation/development
1063 point of view. Because it is essentially a shift of the mean concentration, the causes of it need to be sought in
1064 processes and conditions at the boundaries that have a systematic effect of displacing the concentration values
1065 while approximately preserving the shape of the distribution. Thus, processes like emission timing, chemistry
1066 transformation, autocorrelation structures, stratospheric intrusion, atmospheric stability are unlikely
1067 responsible for systematic bias-type error (while they can be source of casual inaccuracy for limited periods).
1068 On the other hand deposition fluxes, magnitude of emission, input from the lateral boundaries are more
1069 probable sources of bias error. The effect of meteorology is more complex, as errors in synoptic circulation can
1070 induce surface wind velocity and direction to be inaccurate, and thus negatively impacting on the long term
1071 modelled concentrations causing bias error.
1072 • The meteorological fields of temperature and wind speed are consistently biased low and high, respectively.
1073 Based on the results of the European models directly driven by the global fields for meteorology (e.g SILAM,
1074 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
1075 considered as the uppermost limit the accuracy of the models can currently achieve. The use of nudging and
1076 interpolation methods (specific to the configuration of the meteorological model) can add more than 1.5K and
1077 2ms^{-1} to the total error. The analysis of the available vertical profiles suggests that the models overestimate the
1078 wind speed within the PBL and vice versa above the PBL, possibly inducing a net outward flux of pollutants at
1079 the PBL interface.
1080 • Modelled CO is affected by high errors, uniformly across models and components, more pronounced in winter
1081 and predominantly driven by the negative bias of the LT component, followed by variance error of the SY
1082 component. Modelled NO and NO_2 also report negative bias but, in contrast to CO, there is significant model-
1083 to-model difference in error variability, possibly due to the chemistry of NO_x. The SY and DU errors of NO are
1084 comparable in magnitude (3-5 ppb) and mostly due to *mMSE* error. Preliminary sensitivity investigations for CO
1085 and NO seem to suggest that at most ~50% and ~35% of the total error, respectively, could be due to
1086 emissions. Finally, based on spatially averaged analysis, the error for NO/NO_2 is the same for urban and rural
1087 stations (i.e. the error is insensitive to the area-type of the stations).
1088 • The error analysis for ozone shows large model-to-model variability for all errors and spectral components,
1089 with the exception of the SY component for which the error is similar among models and possibly driven by the
1090 error in temperature and in the boundary conditions, as modelled vertical ozone profiles near the domain's
1091 boundaries are typically underestimated in both continents by all models. The bias is prevalently positive, while

1092   the variance error is generally small. While the bias error for ozone is likely driven by error in NO_x emissions,
1093   the error in meteorology may factor in determining the *mMSE* and variance error. In fact, there are several
1094   models for which the bias of temperature and the bias of NO_2 are strongly associated with the DU error of
1095   ozone. A simple linear regression between NO_x bias and ozone bias (based on the yearly time series) among the
1096   EU models suggests that the NO_x and temperature biases can explain, on average, ~35% and ~16% of the
1097   variability of the ozone bias, respectively. Ongoing analyses are focusing on explaining the origin of the *mMSE*
1098   error by investigating the phase shift between the modelled and observed DU and SY components as well as on
1099   looking at maximum daily values rather than to the full time series.

1100 •   PM analysis (PM_{10} for Europe and $PM_{2.5}$ for North America) reveals that, for Europe, the error distribution for
1101   DU and SY is homogeneous and season independent among the models, despite the large numbers of
1102   modelling options and parameters characterising the chemistry and physics of particles. A common source of
1103   model bias (model underestimation, especially in winter) for PM_{10} likely lies in the emissions (missing sources)
1104   and in the overestimation of surface wind speed, whereas variance error may stem from PBL dynamics under
1105   stable conditions and missing processes in the model (SOA formation is a known issue for all models). The
1106   analysis of $PM_{2.5}$ (based on two models only) shows an excess of variance and low correlation coefficient in the
1107   DU component, possibly due to the timing of the PM cycle. Further analyses dealing with the PM components
1108   are needed.

1109 •   The analysis of the memory of the ozone signal has revealed a strong model deficit in continental Europe,
1110   where the seasonal modulation of ozone is overestimated by the majority of the models. The opposite holds
1111   true in the continental US.

1112 Although remarkable progress has been made since the first phase of AQMEII, both in terms of model
1113 performance and also in terms of developing a more versatile and robust evaluation procedure, results of AQ
1114 model evaluation and inter-comparison remain generic as they fail to associate errors with processes, or at least to
1115 narrow down the list of processes responsible for model error. AQ models are meant to be applicable to a variety
1116 of geographic (and topographic) scenarios, under almost any type of weather, season, and emission conditions. For
1117 such a wide range of conditions the inherent non-linearity among processes are difficult to disentangle and
1118 specifically designed sensitivity runs seems the only viable alternative. A model evaluation strategy relying solely
1119 on the comparison of modelled vs. observed time series would never be able to quantify exactly the error induced
1120 e.g. by biogenic emissions, vertical emission profiles and their dependence on temperature, deposition, vertical
1121 mixing, chemistry, and the analysis approach presented in this work is no exception. In fact, the methodology
1122 devised to carry out the evaluation activity in this study has not succeeded in determining the 'actual' causes of
1123 model error, although providing much clearer indications of the processes responsible for the error with respect to
1124 conventional operational model evaluation.

1125 The highly non-linear nature of current AQ models requires the study of the relationships among error fields, those
1126 of the meteorological drivers and those of the precursors. When the seasonal and spectral structures of these
1127 relationships is analysed together with the error of the input fields (emissions and boundary conditions), then it
1128 would be possible to diagnose and explain accurately the processes responsible for the error. Future AQ model
1129 evaluation activities should envision sensitivity simulations and process specific analyses. The 'theory of
1130 evaluation' based on information theory currently being developed by the hydrology modelling community
1131 (Nearing et al., 2016 and references therein) is a promising way forward and the AQ community should be
1132 prepared to catch those developments.

1133 Ongoing work (Solazzo et al., 2017) is being devoted to deepen the investigation of causes of model errors by
1134 focusing on two models (CMAQ for NA and CHIMERE for EU), for which additional model runs have been carried
1135 out to frame the effect of fluxes (emissions, boundary conditions and deposition) on modelled ozone.

1136 APPENDIX 1.

1137 Following Hogrefe et al. (2000) and Galmarini et al. (2013) the time windows (m) and the smoothing parameter (k)

1138 have been selected as follow:

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

1139 where $\mathbf{x}(t)$ is the time series vector. The additive property of the components whose summation returns the
1140 original time series might be questioned. In the original work by Rao et al. (1997) it is highlighted the importance of
1141 log-transform the components to stabilize the variance. In the case of log-transformation the original time series is
1142 obtained by the product of exponential functions whose exponents are the spectral components. For the purposes
1143 of the error apportionment analysis presented here, the results of using additive time series component of log-
1144 transformed did not produce substantial differences.

1145 A clear-cut separation of the components of Eq. S1 is not achievable, since the separation is a non-linear function
1146 of the parameters m and k (Rao et al., 1997). It follows that the components of Eq. S1 are not completely
1147 orthogonal and that there is some level of overlapping energy (Kang et al., 2013). Galmarini et al. (2013) found that
1148 the explained variance by the spectral components account for 75 to 80% of the total, the remaining portion being
1149 on account of the interactions between the components.

1150 APPENDIX 2.

1151 Statistical indicators:

1152 Root Mean Square Error

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

1153 Mean Bias (MB)

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

1154 Pearson correlation coefficient (r)

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

1155 Where M and O are the n-element modelled and observed time series, respectively, σ is the standard deviation
1156 and the overbar indicates temporal averaging.

1157 ACKNOWLEDGEMENTS
1158 We gratefully acknowledge the contribution of various groups to the third air Quality Model Evaluation
1159 international Initiative (AQMEII) activity. The following agencies have prepared the data sets used in this study:
1160 U.S. EPA (North American emissions processing and gridded meteorology); U.S. EPA, Environment Canada,
1161 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., Napelenok, S. L., Foley, K. M., Pye, H. O. T., Hogrefe, C., Luecken, D. J., Young, J. O., 2016. Overview and evaluation of the Community Multiscale Air Quality (CMAQ) model version 5.1. Geosci.Model Dev. Discuss. doi:10.5194/gmd-2016-226

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

1212 Bieser, J., A. Aulinger, V. Matthias, M. Quante and P. Builtjes, 2011a. SMOKE for Europe adaptation, modification and evaluation
1213   of a comprehensive emission model for Europe, Geosci. Model Dev., 4, 4768.
1214 Bieser, J., Aulinger, A., Matthias, V., Quante, M., and Denier van der Gon, H. A. C., 2011b. Vertical emission profiles for Europe
1215   based on plume rise calculations, Environ. Pollut., 159, 2935–2946, doi:10.1016/j.envpol.2011.04.030
1216 Binkowski, F. S., and S. J. Roselle, Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component, 1, Model
1217   description, J. Geophys. Res., 108(D6), 4183, doi:10.1029/2001JD001409, 2003
1218 Binkowski, F.S., S. Arunachalam, Z. Adelman, and J.P. Pinto, 2007: Examining Photolysis Rates with a Prototype Online
1219   Photolysis Module in CMAQ. J. Appl. Meteor. Climatol., 46, 1252–1256.
1220 Bonansoni, P., Evangelisti, F., Bonafe, U., Ravegnani, F., Calzolari, F., Stohl, A., Tositti, L., Tubertini, O., Colombo, T., 2000.
1221   Stratospheric ozone intrusion episodes recorded at Mt. Cimone during the VOLTALP project: case studies. Atmospheric
1222   Environment, 34, 1355-1365.
1223 Bowdalo, D.R., Evans, M.J., Sofen, E.D., 2016. Spectral analysis of atmospheric composition: application to surface ozone model-
1224   measurement comparison. Atmospheric Chemistry and Physics Discussion 172. doi:10.5194/acp-2016-172
1225 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, A.
1226   Zare, and J. H. Christensen, 2012. An integrated model study for Europe and North America using the Danish Eulerian
1227   Hemispheric Model with focus on intercontinental transport. Atmospheric Environment 53, 156-176
1228 Briggs, G. A., 1971, Some Recent Analyses of Plume Rise Observation, pp. 1029-1032 in *Proceedings of the Second International*
1229   *Clean Air Congress*, edited by H. M. Englun and W. T. Beery. Academic Press, New York.
1230 Briggs, Gary A., 1972, Discussion on Chimney Plumes in Neutral and Stable Surroundings, *Atmos.Environ.* 6, 507-510.
1231 Byun, D.W., Ching, J.K.S.: Science Algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ)
1232   Modeling System. EPA/600/R-99/030, US EPA National Exposure Research Laboratory, Research Triangle Park, NC, 1999.
1233 Byun, D.W., Schere, (2006). Review of the governing equations, computational algorithms, and other components of the
1234   Models-3 community Multiscale Air Quality (CMAQ) modeling system. Applied Mechanics Reviews. v59 i2, 51-77
1235 Carlton, A. G., P. Bhave, S. Napelenok, E. O. Edney, G. Sarwar, R. W. Pinder, G. Pouliot, and M. Houyoux. Model Representation
1236   of Secondary Organic Aerosol in CMAQ v4.7. Environmental Science & Technology, American Chemical Society, Washington,
1237   DC, 44(22):8553-8560, (2010)
1238 Carslaw, D.C., Ropkins, K., 2012. openair – an R package for air quality data analysis. Environmental Modelling and Software 27-
1239   28, 52-61.
1240 Chang, J.S., Brost, R.A., Isaksen, I.S.A., Madronich, S., Middleton, P., Stockwell, W.R., Walcek, C.J., (1987) A three dimensional
1241   Eulerian acid deposition model: physical concepts and formulation, Geophys. Res. 92, 14681-14700.
1242 Chapman, E. G., Gustafson Jr., W. I., Easter, R. C., Barnard, J. C., Ghan, S. J., Pekour, M. S., and Fast, J. D.: Coupling aerosolcloud-
1243   radiative processes in the WRF-Chem model: Investigating the radiative impact of elevated point sources, Atmos. Chem.
1244   Phys., 9, 945–964, doi:10.5194/acp-9-945-2009, 2009.
1245 Chen, F. and Dudhia, J., 2001 Coupling an advanced landsurface/hydrology model with the Penn State/ NCAR MM5 modeling
1246   system. Part I: model description and implementation, Mon. Weather Rev., 129, 569–585.
1247 Cohen, A.E., Cavallo, S.M., Coniglio, M.C., Brooks, H.E, 2015. A Review of Planetary Boundary Layer Parameterization Schemes
1248   and Their Sensitivity in Simulating Southeastern U.S. Cold Season Severe Weather Environments. Weather And Forecasting
1249   30, 591-612.
1250 Christensen, J. H., J. Brandt, L. M. Frohn and H. Skov, 2004: "Modelling of mercury in the Arctic with the Danish Eulerian
1251   Hemispheric Model. Atmospheric Chemistry and Physics. Vol. 4, pp 2251-2257.
1252 Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., and Brown, P. D.,
1253   Atmospheric radiative transfer modeling: a summary of the AER codes, J. Quant. Spectrosc. Ra., 91, 233–244, 2005
1254 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.
1255   Zhang, and Y. Dai, 2004. Description of the NCAR community atmosphere model (cam 3.0), Technical report, National
1256   Center for Atmospheric Research. NCAR/TN–464+STR, NCAR TECHNICAL NOTE (http://www.cesm.ucar.edu/models/atm-
1257   cam/docs/description/description.pdf)
1258 Damski, J., Thölix, L., Backman, L., Taalas, P., Kulmala, M., 2007. FinROSE: middle atmospheric chemistry transport model.
1259   Boreal Environ. Res. 12, 535–550
1260 de Leeuw, G., Neele, F.P., Hill, M., Smith, M.H., Vignati, E., 2000. Production of sea spray aerosol in the surf zone. J. Geophys.
1261   Res. 105 (D24), 29397e29409.
1262 Denier van der Gon, H. a. C., Bergström, R., Fountoukis, C., Johansson, C., Pandis, S. N., Simpson, D. and Visschedijk, a.:
1263   Particulate emissions from residential wood combustion in Europe – revised estimates and an evaluation, Atmos. Chem.
1264   Phys. Discuss., 14(23), 31719–31765, doi:10.5194/acpd-14-31719-2014, 2014
1265 Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S. T., Scheffe, R., Schere, K., Steyn, D., and
1266   Venkatram, A.: A framework for evaluating regional-scale numerical photochemical modeling systems, Environ. Fluid Mech.,
1267   10, 471–489, doi:10.1007/s10652-009-9163-2, 2010

Doms, G., (2011). A Description of the Nonhydrostatic Regional COSMO model. Part I: Dynamics and Numerics., Tech. rep., Deutscher Wetterdienst, available at: http://www.cosmo-model.org/content/model/documentation/core/cosmoDyncsNumcs.pdf (last access: 8 April 2015)

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mrionow, D., Raschendorfer, M., Reinhart, T., Ritter, B., Schrodin, R., Schulz, J.-P., and Vogel, G., (2011). A Description of the Nonhydrostatic Regional COSMO Model. Part II: Physical Parameterization, Tech. rep., Deutscher Wetterdienst, available at: http://www.cosmomodel.org/content/model/documentation/core/cosmoPhysParamtr.pdf (last access: 8 April 2015)

Donahue, N. M., A. L. Robinson, C. O. Stanier, and S. N. Pandis (2006). "Coupled partitioning, dilution, and chemical aging of semivolatile organics," Environ. Sci. Technol., 40.8, pp. 635–2643.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two–dimensional model. J. Atmos. Sci., 46, 3077–3107.

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., and Zaveri, R. A.: MIRAGE: Model Description and Evaluation of Aerosols and Trace Gases, J. Geophys. Res., 109, D20210, doi:10.1029/2004JD004571, 2004.

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

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

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

Environ, 2014. CAMx (Comprehensive Air Quality Model with Extensions) User's Guide Version 6.1 ENVIRON International Corporation, Novato, CA.

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 meteorological data. Bull. Amer. Meteor. Soc. 78, 1473-1483.

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. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. J. Geophys. Res. 111, D21305. http://dx.doi.org/10.1029/2005JD006721.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A., 2015 Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 8, 975-1003, doi:10.5194/gmd-8-975-2015.

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. C., Nolte, C. G., Kelly, J. T., Gilliland, A. B., and Bash, J. O.: Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7, Geosci. Model Dev., 3, 205–226, doi:10.5194/gmd-3-205-2010, 2010

Fountoukis, C., Nenes, A., 2007. ISORROPIA II: a computationally efficient thermo-dynamic equilibrium model for K^+–Ca^{2+}–Mg^{2+}–NH_4^+–Na^+–SO_4^{2-}–NO_3^-–Cl^-–H_2O aerosols. Atmos. Chem.Phys. 7, 4639-4659.

Galmarini, S., Bianconi, R., Appel, W., Solazzo, E., et al., 2012. ENSEMBLE and AMET: two systems and approaches to a harmonised, simplified and efficient assistance to air quality model developments and evaluation. Atmos. Environ. 53, 51-59.

Galmarini, S., Kioutsioukis, I., and Solazzo, E.: E pluribus unum: ensemble air quality predictions, Atmos. Chem. Phys., 13, 7153–7182, doi:10.5194/acp-13-7153-2013, 2013

Galmarini, S., Koffi, B., Solazzo, E., Keating, T., Hogrefe, C., Schulz, M., Benedictow, A., Griesfeller, J.J., Janssens-Maenhout, G., Carmichael, G., Fu, J., Dentener, F., 2017. Harmonization of the multi-scale multi-model activities HTAP, AQMEII and MICS-Asia: simulations, emission inventories, boundary conditions and output formats. Atmospheric Chemistry and Physics, in press

Gantt, B, J. T. Kelly and J. O. Bash, 2015: Updating sea spray aerosol emissions in the Community Multiscale Air Quality (CMAQ) model version 5.0.2, Geosci. Model Dev. Discuss., 8, 39053939, doi:10.5194/gmdd-8-3905-2015.

Geels, C., Christensen, J.H., Frohn, L.M., Brandt, J., 2002. Simulating spatiotemporal variations of atmospheric CO2 using a nested hemispheric model. Physics and Chemistry of the Earth, Parts A/B/C 27 (35), 1495e1505.

Geyer, B. (2014). High-resolution atmospheric reconstruction for Europe 1948–2012: coastDat2, Earth Syst. Sci. Data, 6, 147–164, doi:10.5194/essd-6-147-2014

Gilliam, R.C., J. M. Godowitch, and S. T. Rao, Improving the horizontal transport in the lower troposphere with four dimensional data assimilation, Atmospheric Environment, Volume 53, June 2012, Pages 186-201, ISSN 1352-2310, http://dx.doi.org/10.1016/j.atmosenv.2011.10.064

Gilliam, R.C., Pleim, J.E., 2010. Performance assessment of new land surface and planetary boundary layer physics in the WRF-ARW. J. Appl. Meteor. Climatol. 49, 760-774

Gilliam, R. C., C. Hogrefe, J. M. Godowitch, S. Napelenok, R. Mathur, and S. T. Rao (2015), Impact of inherent meteorology uncertainty on air quality model predictions, J. Geophys. Res. Atmos., 120, 12,259–12,280, doi:10.1002/2015JD023674.

Ginoux, P., et al. (2001), Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106(D17), 20,255–20,273.

Giordano, L., Brunner, D., Flemming, J., Hogrefe, C., Im, U., Bianconi, R., and et al., 2015. Assessment of the MACC reanalysis and its influence as chemical boundary conditions for regional air quality modelling in AQMEII-2. Atmospheric Environment 115, 371-388.

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

Grell, G. A., Peckham, S. E., McKeen, S., Schmitz, R., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled ''online'' chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, doi:10.1016/j.atmosenv.2005.04.027, 2005.

Grell, G.A., Freitas, S.R., 2014. A scale and aerosol aware stochastic convective parameterization for weather and air quality modelling. Atmos. Chem. Phys., 14, 5233–5250, doi:10.5194/acp-14-5233-2014

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

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

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, doi:10.5194/acp-6-3181-2006, 2006.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of Hydrology 377, 80-91.

Hansen, K.M., Christensen, J.H., Brandt, J., Frohn, L.M., Geels, C., 2004. Modelling atmospheric transport of a-hexachlorocyclohexane in the Northern Hemisphere with a 3-D dynamic model: DEHM-POP. Atmospheric Chemistry and Physics 4, 1125e1137.

Hendriks, C., Kranenburg, R., Kuenen, J.J.P. ,Van den Bril, B., Verguts, V., Schaap, M. Ammonia emission time profiles based on manure transport data improve ammonia modelling across north western Europe, Atmos. Environment., Volume 131, Pages 83-96 (April 2016)

Hogrefe, C., Vempaty, S., Rao, S.T., Porter, S.P., 2003. A comparison of four techniques for separating different time scales in atmospheric variables, Atmospheric Environment 37(3), 313-325

Hogrefe, C., Rao, S. T., Zurbenko, I. G., Porter, P. S., 2000. Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the Eastern United States, B. Am. Meteorol. Soc., 81, 2083-2106, doi:0.1175/1520-0477(2000)0812.3.CO;2.

Hogrefe, C., Hao, W., Zalewsky, E. E., Ku, J.-Y., Lynn, B., Rosenzweig, C., Schultz, M. G., Rast, S., Newchurch, M. J., Wang, L., Kinney, P. L., and Sistla, G. 2011. An analysis of long-term regional-scale ozone simulations over the Northeastern United States: Variability and trends. Atmospheric Chemistry and Physics 11, 567-582

Hogrefe, C., Roselle, S., Mathur, R., Rao, S. T., Galmarini, S., 2014. Space-time analysis of the Air Quality Model Evaluation International Initiative (AQMEII) phase 1 air quality simulation, J.Air Waste Manage., 64, 388–405.

Hong, S.–Y., Lim, J.–O. J. ,2006. The WRF single–moment 6–class microphysics scheme (WSM6). J. Korean Meteor. Soc., 42, 129–151.

Hong, S.-Y., Dudhia, J., and Chen, S.-H., 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, Mon. Weather Rev., 132, 103–120.

Hong, S.Y., Pan, H.L., 1996. Nonlocal boundary layer vertical diffusion in a medium- range forecast model. Mon. Weather Rev. 124 (10), 2322-2339.

Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A., Collins, W.D., 2008. Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. J. Geophys. Res. 113, D13103. http://dx.doi.org/ 10.1029/2008JD009944.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A.,Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C.,Curci, G., Denier van der Gon, H., Flemming, J., Forkel, R.,Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak,L., Jorba, O., Knote, C., Makar, P. A., Manders-Groot, A.,Neal, L., Pérez, J. L., Pirovano, G., Pouliot, G., San Jose, R., Savage,N., Schroder,W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella,P., Wang, K., Werhahn, J., Wolke, R., Zabkar, R., Zhang,Y., Zhang, J., Hogrefe, C., and Galmarini, S.: Evaluation of operational online coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: particulate matter, Atmos. Environ., 115, 421–441, 2015a.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A.,Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C.,Curci, G., Flemming, J., Forkel, R., Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote,C., Kuenen, J. J. P., Makar, P. A., Manders-Groot, A., Neal, L.,Pérez, J. L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N.,Schroder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P.,Werhahn, J.,Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang,J., Hogrefe, C., and Galmarini, S.: Evaluation of operational online-coupled regional air quality models over Europe and NorthAmerica in the context of AQMEII phase 2. Part I: ozone, Atmos.Environ., 115, 404–420, 2015b.

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

Jacob, D. J. and Winner, D. A.: Effect of climate change on air quality, Atmospheric Environment, 43, 51 – 63, 2009

Janjić Z.I., 1994. The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. Monthly Weather Review 122 (5), 927-945

Janjic, Z.I., 2002: Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso model, NCEP Office Note, No. 437, 61 pp

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411-11432, doi:10.5194/acp-15-11411-2015, 2015.

Jimenez, P. A. and Dudhia, J.: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model, J. Appl. Meteor. Climatol., 51, 300–316, 2012

Jin,J., Miller, N.L, Schlegel,N., 2010. Sensitivity Study of Four Land Surface Schemes in the WRF Model. Advances in Meteorology, Volume 2010, Article ID 167436, 11 pages, doi:10.1155/2010/167436

Johnson, R.: Assessment of Bias with Emphasis on Method Comparison, Clin. Biochem., 29, S37–S42, 2008.

Jones, A. M. and Harrison, R.M. 2006. Assessment of natural components of PM10 at UK urban and rural sites, *Atmospheric Environment*, 40, 7733–7741

Kain, J. S., 2004. The Kain-Fritsch convective parameterization: An update, J. Appl. Meteorol., 43, 170–181

Kajino, M., Winiwarter, W and Ueda, H. 2006. Modeling retained water content in measured aerosol mass, *Atmospheric Environment*, 40, 5202–5213.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J.,Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A.,Reynolds, R., Jenne, R., and Joseph, D., (1996). The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471

Kang, D., Hogrefe, C., Foley, K. L., Napelenok, S. L., Mathur, R., and Rao, S. T.: Application of the Kolmogorov-Zurbenko filter and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model, Atmos. Environ., 80, 58–69, 2013.

Katragkou, E., García-Díez, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R. M., Colette, A., Fernandez, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P. M. M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., and Jacob, D.: Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble, Geosci. Model Dev., 8, 603-618, doi:10.5194/gmd-8-603-2015, 2015

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.: NO_2 columns in the western United States observed from space and simulated by a regional chemistry model and their implications for NO_x emissions, J. Geophys. Res.-Atmos., 114, D11301, doi:10.1029/2008jd011343, 2009.

Kioutsioukis, I., Im, U., Solazzo, E., Bianconi, R., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Denier van der Gon, H., Flemming, J., Forkel, R., Giordano, L., Jiménez-Guerrero, P., Hirtl, M., Jorba, O., Manders-Groot, A., Neal, L., Pérez, J. L., Pirovano, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Tuccella, P., Werhahn, J., Wolke, R., Hogrefe, C., and Galmarini, S., 2016. Improving the deterministic skill of air quality ensembles, Atmos. Chem. Phys. 16, 15629-15652, doi: 10.5194/acp-16-15629-2016

Kouznetsov, R. & Sofiev, M., 2012. A methodology for evaluation of vertical dispersion and dry deposition of atmospheric aerosols. Journal of Geophysical Research, 117(D01202).

Kouznetsov, R. M. Sofiev, M. 2014: "Wet deposition scheme for SILAM chemical transport model" Proc. 16th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 8-11 September 2014, Varna, Bulgaria, p 336

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmos. Chem. Phys., 14, 10963-10976, doi:10.5194/acp-14-10963-2014, 2014.

Lattuati M., 1997: Impact des emissions européennes sur le bilan d'ozone troposphérique à l'interface de l'Europe et de l'Atlantique Nord : apport de la modélisation lagrangienne et des mesures en altitude. Ph.D. Thesis, Université Pierre et Marie Curie, Paris,France

Lin, Y.L., Farley, R.D., Orville, H.D., 1993. Bulk parameterization of the snow field in a cloud model. J. Clim. Appl. Meteorol. 22, 1065-1092.

Lim, K.–S. S., and S.–Y. Hong, 2010: Development of an effective double–moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. Mon. Wea. Rev., 138, 1587–1612.

Makar, P., Nissen, R., Teakles, A., Zhang, J., Zheng, J., Moran, M.D., Yau, H., diCenzo, C., 2014. Turbulent transport, emissions and the role of compensating errors in chemical transport models. Geosci. Model Dev. 7, 1001-1024.

Makar, P., Staebler, R., Akingunola, A., Zhang, J., McLinden, C., Kharol, S., Moran, M., Robichaud, A., Zhang, L., Stroud, C., Pabla, B., Cheung, P., 2016. Forest Canopy Processes in a Regional Chemical Transport Model. Geophysical Research Abstracts Vol. 18, EGU2016-1763, 2016, EGU General Assembly 2016, Vienna.

Martensson, E. M., E. D. Nilsson, G. de Leeuw, L. H. Cohen, and H.-C. Hansson (2003). "Laboratory simulations and parameterization of the primary marine aerosol production". Journal of Geophysical Research: Atmospheres 108.D9, n/a–n/a. DOI: 10.1029/2002JD002263.

Mason, R., Zubrow, A., Eyth, A., 2012. Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 5.0, 2007 Emissions Modeling Platform. Available at: http://epa.gov/ttn/chief/emch/2007v5/2007v5_2020base_EmisMod_TSD_13dec2012.pdf

Mass, C. F. and Ovens, D.: Fixing WRF's high speed wind bias: a new subgrid scale drag parameterization and the role of detailed verification, 91st AMS Annual Meeting, Seattle, WA, available at: http://ams.confex.com/ams/91Annual/webprogram/Paper180011.html, 2011.

Mellor, G. L., Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys., 20, 851–875.

Menut L, B.Bessagnet, D.Khvorostyanov, M.Beekmann, N.Blond, A.Colette, I.Coll, G.Curci, G.Foret, A.Hodzic, S.Mailler, F.Meleux, J.L.Monge, I.Pison, G.Siour, S.Turquety, M.Valari, R.Vautard and M.G.Vivanco, 2013, CHIMERE 2013: a model for regional atmospheric composition modelling, Geoscientific Model Development, 6, 981-1028, doi:10.5194/gmd-6-981-2013

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated–k model for the longwave. J. Geophys. Res., 102, 16663–16682.

Monahan, E.C., D.E. Spiel, and K.L Davidson (1986). "A model of marine aerosol generation via whitecaps and wave disruption". In: Oceanic Whitecaps and their role in air/sea exchange. Ed. by G. Monahan E.C.and Mac Niocaill. Reidel, Norwell, Mass.,USA, pp. 167–174.

Mooney, P. A., Mulligan, F. J., and Fealy, R.: Evaluation of the sensitivity of the weather research and forecasting model to parameterization schemes for regional climates of europe over the period 1990–95, J. Climate, 26 1002–1017, 2013.

Morrison, H., G. Thompson, V. Tatarskii, 2009: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. Mon. Wea. Rev., 137, 991–1007. doi: http://dx.doi.org/10.1175/2008MWR2556.1

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment scheme, Mon. Weather Rev., 137, 991-1007, doi:10.1175/2008mwr2556.1, 2009.

Murphy, A.H., 1995. The coefficients of correlation and determination as measures of performance in forecast verification. Weather and Forecasting 10, 681-688.

Nakanishi, M. and Niino, H.: An improved Mellor-Yamada Level-3 Model: its numerical stability and application to a regional prediction of advection fog, Bound.-Lay. Meteorol., 119, 397–407, doi:10.1007/s10546-005-9030-8, 2006.

Nearing, G.S., Tian, Y., Gupta, H.V., Clark,M.P., Harrison, K.W., Wejs,V., 2016. A philosophical basis for hydrological uncertainty. Hydrological Sciences Journal 6(9), 1666-1678.

Nenes, A, C. Pilinis, and S.N. Pandis. 1999. Continued Development and Testing of a New Thermodynamic Aerosol Module for Urban and Regional Air Quality Models. Atmos. Environ. 33, 1553-1560.

Otero, N., Sillmann, J., Schnell, J.L., Rust, H.W., Butler, T., 2016. Synoptic and meteorological drivers of extreme ozone concentrations over Europe. Environ. Res. Lett. 11 (2016) 024005. doi:10.1088/1748-9326/11/2/024005

Otte, T. L. Pleim, J. E., 2010. The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1, Geosci. Model Dev., 3, 243–256, doi:10.5194/gmd-3-243-2010.

Pirovano, G., Balzarini,A., Bessagnet, B., Emery, C., Kallos, G., Meleux, F., Mitsakou, C., Nopmongcol,U., Riva, G.M., Yarwood, G. 2012. Investigating impacts of chemistry and transport model formulation on model performance at European scale, Atmospheric Environment, Volume 53, June 2012, Pages 93-109, ISSN 1352-2310, http://dx.doi.org/10.1016/j.atmosenv.2011.12.052.

Pleim, J. E. and Gilliam, R., 2009.: An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land surface model, J. Appl. Meteorol. Clim., 48, 1362–1376.

Pleim, J. E. and Xiu, A., 2003. Development of a land surface model. Part II: data assimilation, J. Appl. Meteorol., 42, 1811–1822, 2003

Pleim, J. E., 2006: A simple, efficient solution of flux-profile relationships in the atmospheric surface layer, J. Appl. Meteor. and Clim., 45, 341–347.

Pleim, J. E., 2007b: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part II: application and evaluation in a mesoscale meteorological model, J. Appl. Meteorol. Clim., 46, 1396–1409

Pleim, J., R. Gilliam, and J. Godowitch, 2010: Evaluation of PBL models compared to GABLS experiments and testing in meteorology and air quality models, presented at the 19th Symposium on Boundary Layers and Turbulence, American

1499   Meteorological Society, August 2–6, 2010, Keystone, CO, recorded presentation available at
1500   http://ams.confex.com/ams/19Ag19BLT9Urban/recordingredirect.cgi/id/15728

1501   Pleim, J., Ran, L., 2011. Surface Flux Modeling for Air Quality Applications. Atmosphere 2, 271-302

1502   Pleim, J., Gilliam, R., Appel, W., Ran, L., 2016. Recent advances in modelling of the atmospheric boundary layer and land surface
1503   in the coupled WRF-CMAQ model. Air Pollution Modelling and Its Application XXIV, Springer Proceeding in Complexity,
1504   Switzerland, pp. 391-396

1505   Poppe, D., Y. Andersson-Sköld, A. Baart, P.J.H. Builtjes, M. Das, F. Fiedler, O. Hov, F. Korchner, M. Kuhn, P.A. Makar, J.B.Milford,
1506   M.G.M. Roemer, R. Runhke, D Simpson, W.R. Stockwell, Gas-Phase reactions in atmospheric chemistry and transport
1507   models, Tech Rep. Garmisch-Partenkirchen, Germany: Eurotrac report.

1508   Pouliot, G., and et al. 2015. Analysis of the emission inventories and model-ready emission datasets of Europe and North
1509   America for phase 2 of the AQMEII project. Atmospheric Environment 115, 340-360

1510   Poupkou, A., Giannaros, T., Markakis, K., Kioutsioukis, I., Curci, G., Melas, D., Zerefos, C., 2010. A model for European Biogenic
1511   Volatile Organic Compound emissions: Software development and first validation. Environ. Model. Softw. 25, 1845–1856.
1512   doi:10.1016/j.envsoft.2010.05.004

1513   Prank, M., Sofiev, M., Tsyro, S., Hendriks, C., Semeena, V., Vazhappilly Francis, X., Butler, T., Denier van der Gon, H., Friedrich,
1514   R., Hendricks, J., Kong, X., Lawrence, M., Righi, M., Samaras, Z., Sausen, R., Kukkonen, J., Sokhi, R., 2016. Evaluation of the
1515   performance of four chemical transport models in predicting the aerosol chemical composition in Europe in 2005. Atmos.
1516   Chem. Phys. 16, 6041–6070. doi:10.5194/acp-16-6041-2016

1517   Rao, S. T., Galmarini, S., and Puckett, K.: Air quality model evaluation international initiative (AQMEII), B. Am. Meteorol. Soc.,
1518   92, 23–30, doi:10.1175/2010BAMS3069.1, 2011.

1519   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.
1520   Meteorol. Soc., 78, 2153e2166, doi:10.1175/1520-0477(1997)078<2153:SATSIA>2.0.CO;2, 1997.

1521   Ritter, B. and Geleyn, J. F., (1992). A comprehensive radiation scheme for numerical weather prediction models withpotential
1522   applications in climate simulations, Mon. Weather Rev., 120, 303–325, doi:10.1175/1520-0493 Rockel, B., Will, A., and
1523   Hense, A., (2008). The Regional Climate Model COSMO-CLM (CCLM). Meteorol.Z. 17, 347–248.

1524   Rockel, B., Will, A., and Hense, A., (2008). The Regional Climate Model COSMO-CLM (CCLM). Meteorol. Z. 17, 347–
1525   248.

1526   Sarwar, G., Appel, K. W., Carlton, A. G., Mathur, R., Schere, K., Zhang, R., and Majeed, M. A., 2011a: Impact of a new condensed
1527   toluene mechanism on air quality model predictions in the US, Geosci. Model Dev., 4, 183-193, doi:10.5194/gmd-4-183-
1528   2011

1529   Sarwar, G., K. Fahey, S. Napelenok, S. Roselle, R. Mathur, 2011b. Examining the impact of CMAQ model updates on aerosol
1530   sulfate predictions, Presentation at the 10th Annual CMAS Models-3 User's Conference, October 2011, Chapel Hill, NC,
1531   available online at http://www.cmascenter.org/conference/2011/slides/sarwar_examining_impact_2011.pdf

1532   Sarwar, G.; Luecken, D. & Yarwood, G. Borrego. C. & Renner, E., (2007). (Eds.) Chapter 2.9 Developing and implementing an
1533   updated chlorine chemistry into the community multiscale air quality model Air Pollution Modeling and Its Application XVIII,
1534   Elsevier, 2007, 6, 168 - 176, doi: 10.1016/S1474-8177(07)06029-9

1535   Sauter, F., Swaluw, E. van der, Manders-Groot, A., Wichink Kruit, R, Segers, A., Eskes, H., (2012). LOTOS-EUROS v1.8 Reference
1536   Guide, TNO Report TNO-060-UT-2012-01451, Utrecht, The Netherlands,

1537   Schaettler, U., Doms, G., and Schraff, C., (2008). A Description of the Nonhydrostatic Regional COSMO-Model Part VII: User's
1538   Guide, Tech. rep., Deutscher Wetterdienst.

1539   Schaap, M., Sauter,F., Timmermans, R.M.A., Roemer, M., Velders, G., Beck, J., Builtjes,P., 2008. The LOTOS-EUROS model:
1540   description, validation and latest developments, Int. J. Environment and Pollution, Vol. 32, No. 2, pp.270–290.

1541   Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A., 2001. Modeling the formation of secondary organic aerosol within
1542   a comprehensive air quality model system. J. Geophys. Res. 106 (D22), 28275-28293.

1543   Schrodin, R. and Heise, E., (2001). The multi-layer-version of the DWD soil model TERRA/LM,Tech. Rep, Consortium for Small-
1544   Scale Modelling (COSMO), available at: http: //www.cosmomodel.
1545   org/content/model/documentation/techReports/docs/techReport02.pdf (last access: 8 April 2015)

1546   Schwede, D., Pouliot, G. A., and Pierce, T.: Changes to the Biogenic Emissions Inventory System Version 3 (BEIS3), In
1547   Proceedings of the 4th CMAS Models-3 Users' Conference, Chapel Hill, NC, 26–28 September 2005

1548   Seifert, A. and Beheng, K. D., (2001). A double-moment parameterization for simulating autoconversion, accretion and
1549   selfcollection, Atmos. Res., 59–60, 265–281, doi: 10.1016/S0169-8095(01)00126-0

1550   Seinfeld, J.H., and S.N. Pandis. 1998. Atmospheric Chemistry and Physics, From Air Pollution to Climate Change. John Wiley and
1551   Sons, Inc., NY.

1552   Shrivastava, M., Fast, J., Easter, R., Gustafson, W.I., Zaveri, R. a., Jimenez, J.L., Saide, P., Hodzic, a., 2011. Modeling organic
1553   aerosols in a megacity: comparison of simple and complex representations of the volatility basis set approach. Atmos.
1554   Chem. Phys. 11, 6639–6662. doi:10.5194/acp-11-6639-2011

1555   Simon, H., Bhave, P.V.; Simulating the degree of oxidation in atmospheric organic particles. Environ. Sci. Technol. 2012, 46, 331-
1556   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.

Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comp. Phys., 227, 3465–3485, 2008.

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.: A description of the Advanced Research WRF version 3, National Center for Atmospheric Research Tech. Note, NCAR/TN-475+STR, 113 pp., 2008

Smirnova, T.G., Brown, J.M., Benjamin, S.G., Kim, D., 2000. Parameterization of cold-season processes in the MAPS land-surface scheme, Journal of Geophysical Research D, vol. 105, no. 3, pp. 4077–4086

Soares, J., Sofiev, M., Hakkarainen, J., 2015. Uncertainties of wild-land fi res emission in AQMEII phase 2 case study. Atmos. Environ. doi:10.1016/j.atmosenv.2015.01.068

Sofiev, M., 2000. A model for the evaluation of long-term airborne pollution transport at regional and continental scales. Atmos. Environ. 34, 2481–2493. doi:10.1016/S1352-2310(99)00415-X

Sofiev, M., Soares, J., Prank, M., de Leeuw, G., Kukkonen, J., 2011. A regional-to-global model of emission and transport of sea salt particles in the atmosphere. J. Geophys. Res. 116, 25. doi:10.1029/2010JD014713

Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., and Genikhovich, E.: Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin, Geosci. Model Dev., 8, 3497-3522, doi:10.5194/gmd-8-3497-2015, 2015.

Solazzo, E., Bianconi, R., Pirovano, G., Moran, M. D., Vautard, R., Hogrefe, C., Appel, K. W., Matthias, V., Grossi, P., Bessagnet, B., Brandt, J., Chemel, C., Christensen, J. H., Forkel, R., Francis, X. V., Hansen, A. B., McKeen, S., Nopmongcol, U., Prank, M., Sartelet, K. N., Segers, A., Silver, J. D., Yarwood, G., Werhahn, J., Zhang, J., Rao, S. T., and Galmarini, S.: Evaluating the capability of regional-scale air quality models to capture the vertical distribution of pollutants, Geosci. Model Dev., 6, 791–818, doi:10.5194/gmd-6-791-2013, 2013

Solazzo, E., Galmarini, S., 2016. Error apportionment for atmospheric chemistry transport models - a new approach to model evaluation. Atmospheric Chemistry and Physics 16, 6263-6283

Solazzo,E., Galmarini, S., 2015. Comparing apples with apples: using spatially distributed time series of monitoring data for model evaluation. Atmospheric Environment 112, 234-245

Solazzo, E., Hogrefe, C., Colette, A., Vivanco, M., Galmarini, S., 2017. Diagnostic error analysis of the CMAQ and CHIMERE modelling systems. In preparation for the Atmospheric Chemistry and Physics journal.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.), Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

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

Steppeler, J., Doms, G., Schattler, U., Bitzer, H. W., Gassmann, A., Damrath, U., and Gregoric, G., (2003). Meso-gamma scale forecasts using the nonhydrostatic model LM, Meteorol. Atmos. Phys., 82, 75–96, doi:10.1007/s00703-001-0592-9

Stockwell, W. R., Kirchner, F. K., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling, J. Geophys. Res., 102, 25847–25879, doi:10.1029/97JD00849, 1997.

Stockwell, W.R., Middleton, P., Chang, J.S., Tang, X., 1990. The second generation regional acid deposition model chemical mechanism for regional air quality modeling. J. Geophys. Res. 95 (D10), 16343e16367. http://dx.doi.org/10.1029/ JD095iD10p16343.

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

Tanaka, P. L.; Allen, D. T.; McDonald-Buller, E. C.; Chang, S.; Kimura, Y.; Mullins, C. B.; Yarwood, G. & Neece, J. D., (2003), Development of a chlorine mechanism for use in the carbon bond IV chemistry model J. Geophys. Res. Atmos., 2003, 108, doi:10.1029/2002JD002432.

Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca, 2004: Implementation and verification of the unified NOAH land surface model in the WRF model. 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, pp. 11–15

Tiedtke, M., (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Mon. Weather Rev., 117, 1779–1800, doi:10.1175/1520-0493(1989)117

Tong, D. Q., G.E., Bowker, S., He, D. W., Byun, R. Mathur, D. A. Gillette, 2011: Development of a Windblown Dust Module within the Community Multi-scale Air Quality (CMAQ) Model: Description and Preliminary Applications in the Continental United States, submitted, J. Geophys. Res., 2011.

Tsyro, S.G. 2005. To what extent can aerosol water explain the discrepancy between model calculated and gravimetric PM_{10} and $PM_{2.5}$? *Atmos. Chem. Phys.*, 5, 515–532.

1614 Tuccella, P., Curci, G., Visconti, G., Bessagnet, B., Menut, L., Park, R. J. (2012a), Modelling of gas and aerosol with WRF/Chem
1615   over Europe: evaluation and sensitivity study, J. Geophys. Res., 117, D03303, doi:10.1029/2011JD016302

1616 Tuccella, P., Curci, G., Visconti, G., Bessagnet, B., Menut, L., and Park, R. J.: Modeling of gas and aerosol with WRF/Chem over
1617   Europe: Evaluation and sensitivity study, J. Geophys. Res., 117, D03303, doi:10.1029/2011JD016302, 2012b

1618 Tuccella, P., Curci, G., Grell, G. A., Visconti, G., Crumeyrolle, S., Schwarzenboeck, A., and Mensah, A. A.: A new chemistry option
1619   in WRF-Chem v. 3.4 for the simulation of direct and indirect aerosol effects using VBS: evaluation against IMPACT-EUCAARI
1620   data, Geosci. Model Dev., 8, 2749-2776, doi:10.5194/gmd-8-2749-2015, 2015.

1621 Van der Gone, H., Bergström, R., Fountoukis, C., Johansson, C., Pandis, S.N., Simpson, D., Visschedijk , A., 2015. Particulate
1622   emissions from residential wood combustion in Europe. Atmospheric Chemistry and Physics 2015, 15, 6503-6519.

1623 Valmartin, M., Heald, C. L., and Arnold, S. R., 2014. Coupling dry deposition to vegetation phenology in the Community Earth
1624   System Model: Implications for the simulation of surface O3, Geophys. Res. Lett., 41, 2988–2996.

1625 Vautard, R., Moran, M., Solazzo, E., Gilliam, R., Volker, M., Bianconi, R., and et al., 2012. Evaluation of the meteorological
1626   forcing used for the Air Quality Model Evaluation International Initiative (AQMEII) air quality simulations. Atmospheric
1627   Environment 53, 15-37

1628 Vestreng, V., Ntziachristos, L., Semb, A., Reis, S., Isaksen, I. S. A., and Tarrasón, L., 2009. Evolution of NOx emissions in Europe
1629   with focus on road transport control measures, Atmos. Chem. Phys. 9, 1503–1520, doi:10.5194/acp-9-1503-2009.

1630 Vestreng, V., Støren, E., 2000. Analysis of the UNECE/EMEP Emission Data. MSC-W Status Report 2000. Norwegian
1631   Meteorological Institute, Blindern, Oslo

1632 Vukovich, J. and Pierce, T.: The Implementation of BEIS3 within the SMOKE Modeling Framework, in: Proceedings of the 11th
1633   International Emissions Inventory Conference, Atlanta, Georgia, available at:
1634   www.epa.gov/ttn/chief/conference/ei11/modeling/vukovich.pdf, 15–18 April 2002

1635 Walcek, C. J. and Taylor, G. R.: A theoretical method for computing vertical distributions of acidity and sulfate production within
1636   cumulus clouds, J. Atmos Sci., 43, 339-355, doi: http://dx.doi.org/10.1175/1520-0469(1986)043<0339:ATMFCV>2.0.CO;2,
1637   1986.

1638 Wesely M. L. (1989) Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models.
1639   Atmospheric Environment, 23, 1293-1304

1640 Wesely, M. L., and Hicks, B. B.: A review of the current status of knowledge on dry deposition, Atmos. Environ., 34(12-14), 2261-
1641   2281, doi:10.1016/S1352-2310(99)00467-7, 2000.

1642 Wesely, M.L., Hicks, B.B., 2000. Atmospheric Environment 34, 2261-2282.

1643 Whitten, G.Z., Heo,G., Kimura, Y., McDonald-Buller, E. Allen, D.T., Carter, W.P.L., Yarwood, G. (2010). A new condensed toluene
1644   mechanism for Carbon Bond: CB05-TU. Atmospheric Environment, 44, 5346-5355.

1645 Wild, O., X. Zhu, and M. J. Prather (2000), Fast-J: Accurate simulation of in- and below cloud photolysis in tropospheric
1646   chemical models, J. Atmos. Chem., 37, 245–282, doi:10.1023/A:1006415919030.

1647 Wise, E. K. and Comrie, A. C.: Extending the KZ filter: application to ozone, particulate matter, and meteorological trends, J. Air
1648   Waste Manage., 55, 1208e1216, doi:10.1080/10473289.2005.10464718, 2005.

1649 Xiu, A. and Pleim, J. E.: Development of a land surface model. Part I: application in a mesoscale meteorological model, J. Appl.
1650   Meteorol., 40, 192–209, 2001

1651 Yarwood, G., Rao, S., Yocke, M., and Whitten, G.: Updates to the Carbon Bond Chemical Mechanism: CB05, Final Report to the
1652   US EPA, RT-0400675, available at: http://www.camx.com/publ/pdfs/cb05_final_report_120805.pdf (last access: 16
1653   December 2014), 2005.

1654 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 the
1655   DEPAC module. Dry deposition modelling with DEPAC GCN2010. Tech. rep. Bilthoven, The Nehterlands: Rijksinstituut voor
1656   volksgezondheid en Milieu, RIVM report 680180001. URL: http : / / www . rivm . nl
1657   /Documenten_en_publicaties/Wetenschappelijk/Rapporten/2010/oktober/Description_of_the_DEPAC_module_Dry_depos
1658   ition_modelling_with_DEPAC_GCN2010

1659 Zare, A., J. H. Christensen, A. Gross, P. Irannejad, M. Glasius and J. Brandt, 2014: Quantifying the contributions of natural
1660   emissions to ozone and total fine PM concentrations in the Northern Hemisphere. Atmospheric Chemistry and Physics, Vol.
1661   14, pp. 2735-2756, 2014, www.atmos-chem-phys.net/14/2735/2014/. doi:10.5194/acp-14-2735-2014.

1662 Zhang, D.–L., and R.A. Anthes, 1982: A high–resolution model of the planetary boundary layer– sensitivity tests and
1663   comparisons with SESAME–79 data. J. Appl. Meteor., 21, 1594–1609.

1664 Zhang, L., J. R. Brook, and R. Vet. 2003. A revised parameterization for gaseous dry deposition in air-quality models. Atmos.
1665   Chem. Phys., 3, 2067–2082.

1666 Zhang, L., S. Gong, J. Padro, L. Barrie. 2001. A size-segregated particle dry deposition scheme for an atmospheric aerosol
1667   module. Atmos. Environ., 35, 549-560

1668 Zurbenko, I. G.: The Spectral Analysis of Time Series, North-Holland, Amsterdam, 236 pp., 1986.

1669

1670   TABLES

1671

1672

1673

1674

1675

1676

1677

1678

1679

1680

1681

1682   **TABLE 1. PARTICIPATING MODELLING SYSTEMS AND KEY FEATURES**

Operated by	Modelling system	Emission	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	EDGAR-HTAP; TNO-MACC	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	TNO-MACC	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
INERIS/CIEMAT	ECMWF-CHIMERE_H CHIMERE_M	EDGAR-HTAP; TNO-MACC	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
University of L'Aquila	WRF-WRF/Chem1	TNO-MACC	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	TNO-MACC	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	TNO-MACC	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	EDGAR-HTAP	16.7 km x 16.7 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	TNO-MACC	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	TNO-MACC	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	TNO-MACC	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

Helmholtz-Zentrum Geesthacht	CCLM-CMAQ	EDGAR-HTAP	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	TNO-MACC	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
NORTH AMERICAN DOMAIN								
Helmholtz-Zentrum Geesthacht	CCLM-CMAQ	SMOKE	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
U.S. Environmental Protection Agency	WRF-CMAQ	SMOKE	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	SMOKE	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	EDGAR-HTAP	16.7 km x 16.7 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)

1683

1684   **TABLE 2. CONFIGURATION OF THE WRF MODEL BY MODELLING GROUP**

Operated by	Input data	Number of Vertical levels	1th Layer Height	PBL model	Surface Layer	Land Surface	Cloud Microphysics	Cumulus Convection	SW/LW Radiation	Data Assimilation
University of L'Aquila	ECMWF	33	10m	MYNN	MM5 Similarity	NOAH	Morrison	Grell-Freitas	RRTMG	Grid analysis nudging nudging above PBL
University of Murcia	ECMWF	33	21m	YSU	Eta Similarity	NOAH	Lin	Kain-Fritsch 2	RRTMG	Grid analysis nudging nudging above PBL
Ricerca Sistema Energetico	ECMWF	33	25m	YSU	Eta Similarity	NOAH	Morrison	Grell-Freitas	RRTMG	Grid Analysis nudging also within the PBL
University of Aarhus	ECMWF	29	20m	MYJ	Eta Similarity	NOAH	WSM5	Kain-Fritsch2	CAM	Grid analysis nudging nudging above PBL
Istanbul Technical University	NCEP FNL	30	10m	YSU	Eta Similarity	NOAH	WSM3	Kain-Fritsch2	Dudhia/RRTM	Grid Analysis nudging also within the PBL
Kings College	NCEP GFS	23	14m	ACM2	Pleim-Xiu	RUC	WSM6	Kain-Fritsch 2	Dudhia/RRTM	Grid Analysis nudging also within the PBL

Ricardo E&E	NCEP GFS	23	15m	ACM2	Pleim-Xiu	RUC	WSM6	Kain-Fritsch 2	Dudhia/RRTM	Grid analysis nudging nudging above PBL
University of Hertfordshire	ECMWF	36	25m	ACM2	Pleim-Xiu	5-layer thermal diffusion	Morrison	Kain-Fritsch2	RRTMG	Grid analysis nudging nudging above PBL
U.S. Environmental Protection Agency	NCEP NAM analysis	35	20m	ACM2	Pleim-Xiu	Pleim-Xiu	Morrison	Kain-Fritsch2	RRTMG	Grid analysis nudging above PBL;
RAMBOLL Environ	NCEP NAM analysis	35	20m	ACM2	Pleim-Xiu	Pleim-Xiu	Morrison	Kain-Fritsch2	RRTMG	Grid analysis nudging above PBL

1685 RRTMG: Rapid Radiative Transfer Method for Global for solar and infrared radiation (Iacono et al. 2008);
1686 RRTM: Rapid Radiative Transfer Method for infrared radiation (Mlawer et al., 1997)
1687 Dudhia shortwave radiation scheme (Dudhia, 1989)
1688 YNN: Mellor-Yamada Nakanishi-Niino (PBL) scheme (Nakanishi-Niino, 2006)
1689 MYJ: Mellor-Yamada-Janjic (Janjic, 1994)
1690 YSU: Yonsei University PBL scheme (Hong and Lim, 2006)
1691 Grell-Freitas scheme for cumulus clouds (Grell and Freitas, 2014)
1692 Eta similarity surface layer (Janjic, 2002)
1693 KF2: Kain-Fritsch (Kain, 2004) scheme for cumulus parameterisation
1694 CAM scheme for long and short radiation (Collins et al., 2004)
1695 Morrison microphysics from Morrison et al. (2009)
1696 WSM3 microphysics scheme (Hong et al., 2004)
1697 WSM5: Double Moment 5–class Scheme (Lim and Hong, 2010)
1698 WSM6: Double Moment 6–class Scheme (Lim and Hong, 2010)
1699 MM5 Similairity surface layer scheme (Zhang and Anthes, 1982)
1700 NCEP (National Centers for Environmental Prediction) FNL Operational Model Global Tropospheric Analyses
1701 GFS: Global Forecasting System
1702 FNL: Final (same as GFS but FNLs are prepared about an hour or so after the GFS is initialized so that more observational data can be used)
1703 NAM: North American Model
1704 RUC (Rapid Update Cycle, Smirnova et al., 2000)
1705 NOAH land-surface model (Tewari et al., 2004))
1706 ACM2: Asymmetric Convective Model with non-local upward mixing and local downward mixing (Pleim, 2007)
1707 5-layer thermal diffusion (Dudhia, 1996)
1708 Pleim-Xiu: Plein and Xiu (2003)
1709

1710

1711

1712 **TABLE 3. 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

1713

1714 **TABLE 4. 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

1715

1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734

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. Spatial and temporal variability of the wind direction for a) EU and b) NA for the full year 2010. The boxes extend between the 25th and 75th percentile of the total distribution. The whiskers extend from the minimum to the maxium values.

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

1775   Figure 19. Ozone mixing-ratio profiles measured by ozonesondes launched from the European location indicated on the
1776   inset map (lower-right corner) of each panel. The profiles are time-averaged over the number of hourly records reported in
1777   the parenthesis at the top of each panel. Legend as in the first panel.

1778   Figure 20. As in Figure 19 for North America

1779   Figure 21. Ozone vs NO modelled mean bias for the EU2 sub-region, color-coded by temperature bias and symbols
1780   according to the NO_x emission fraction of NO and NO_2. Each point represents a model. *a)* winter months and *b)* summer
1781   months.

1782   Figure 22. As in Figure 9 for SO_2

1783   Figure 23. As in Figure 10 for SO_2

1784   Figure 24. As in Figure 9 for PM_{10} in Europe (error units in $\mu g/m^3$)

1785   Figure 25. As in Figure 9 for $PM_{2.5}$ in North America (error units in $\mu g/m^3$)

1786   Figure 26. As in Figure 10 for PM_{10} in Europe (error units in $\mu g/m^3$)

1787   Figure 27. As in Figure 10 for $PM_{2.5}$ in North America (error units in $\mu g/m^3$)

1788   Figure 28. Spatial map of the ozone monitoring stations coloured based on the 'delta hour' values, i.e. the difference in
1789   hours between the zero of the autocorrelation function (acf) for the modelled ozone minus the zero of the acf of the
1790   observed one. The acf is calculated on the long term component for the months of May to September. Negative values
1791   indicate too short memory and excess of removal (vice-versa for positive values). The box on the right summarises the
1792   delta hour percentile distribution.

1793   Figure 29. As in Figure 28 for North America.

1794

1795

1796

1797

1798

1799

1800

1801

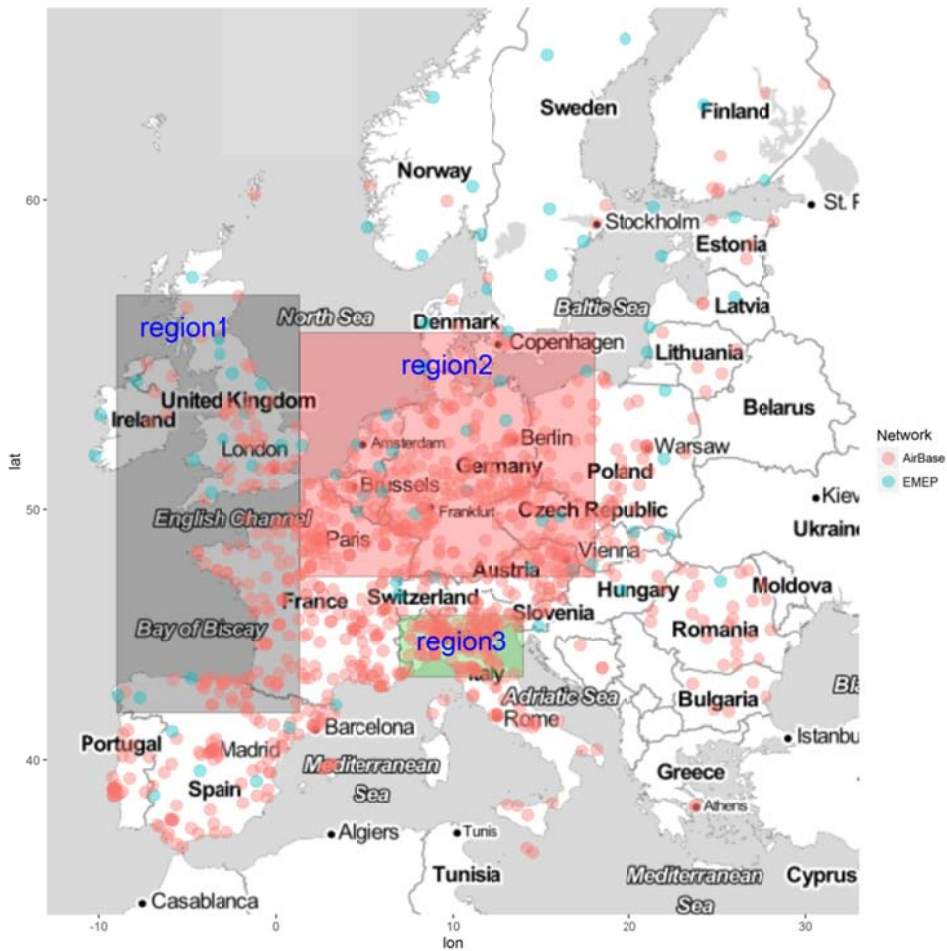
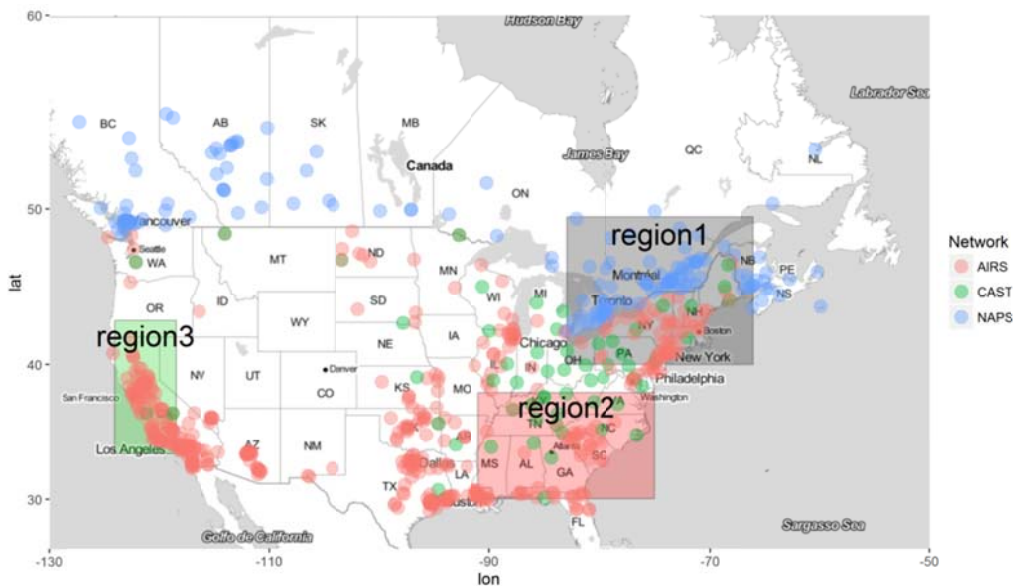
1802

1803

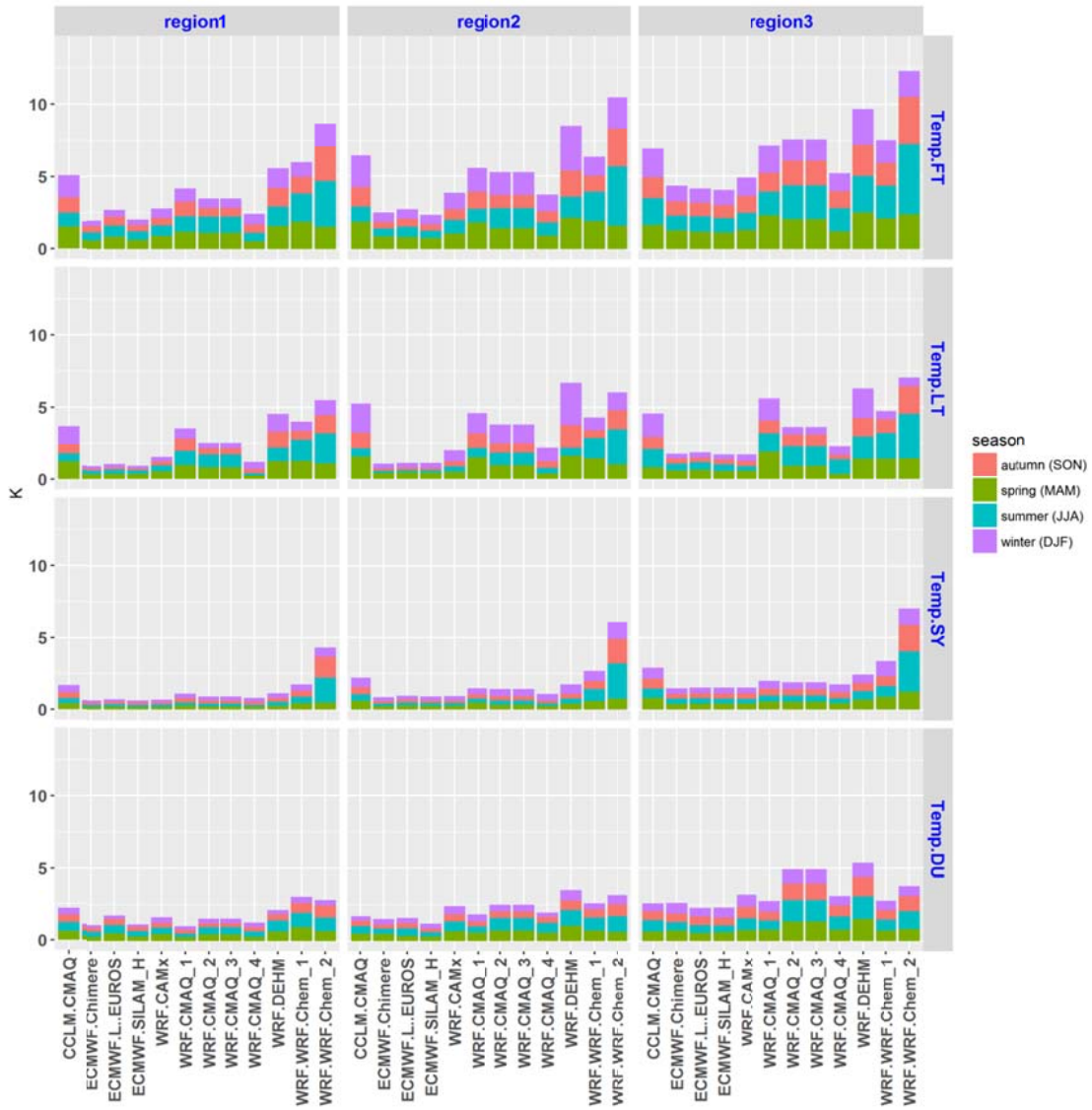
1804

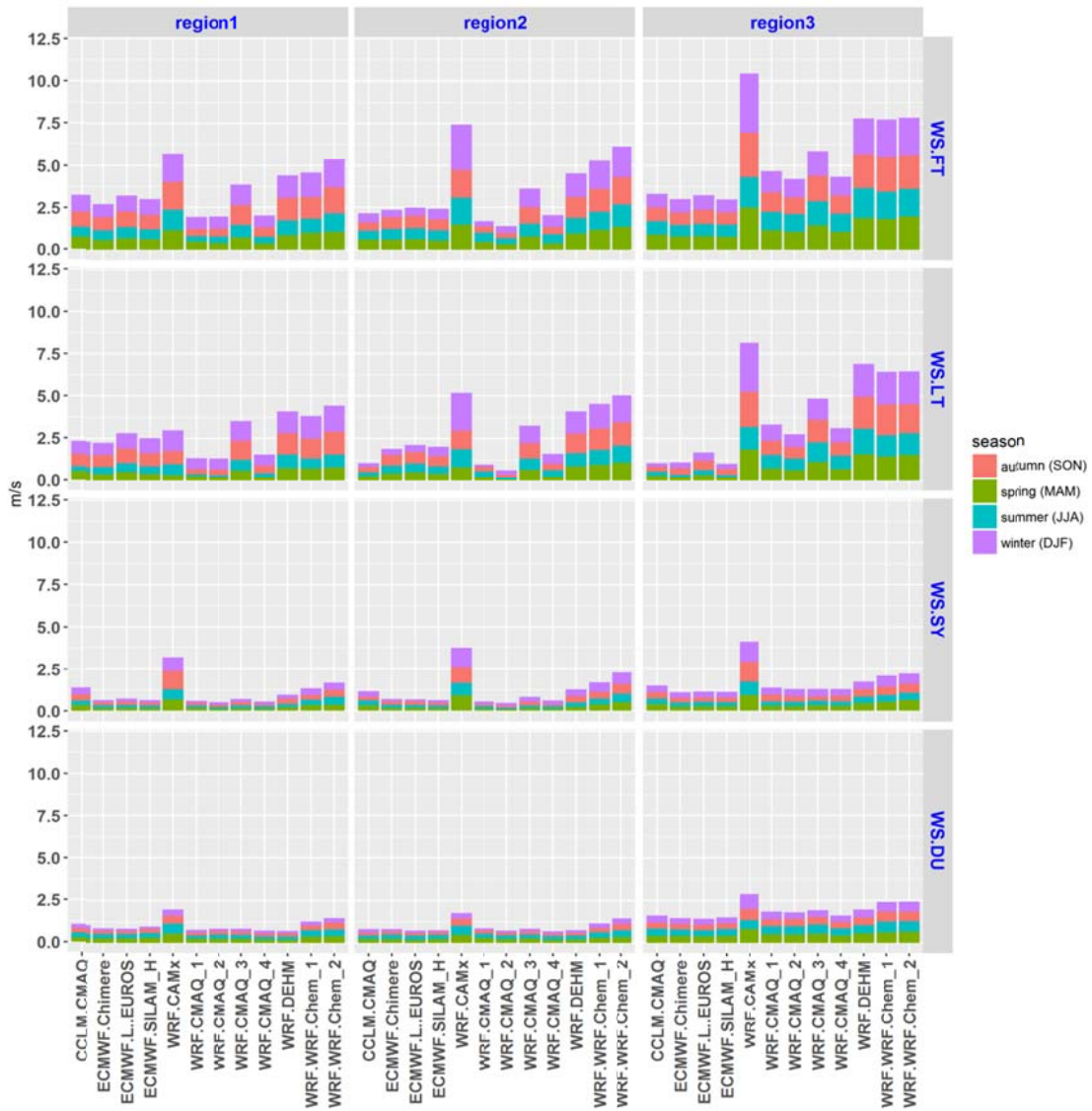
1805

1806

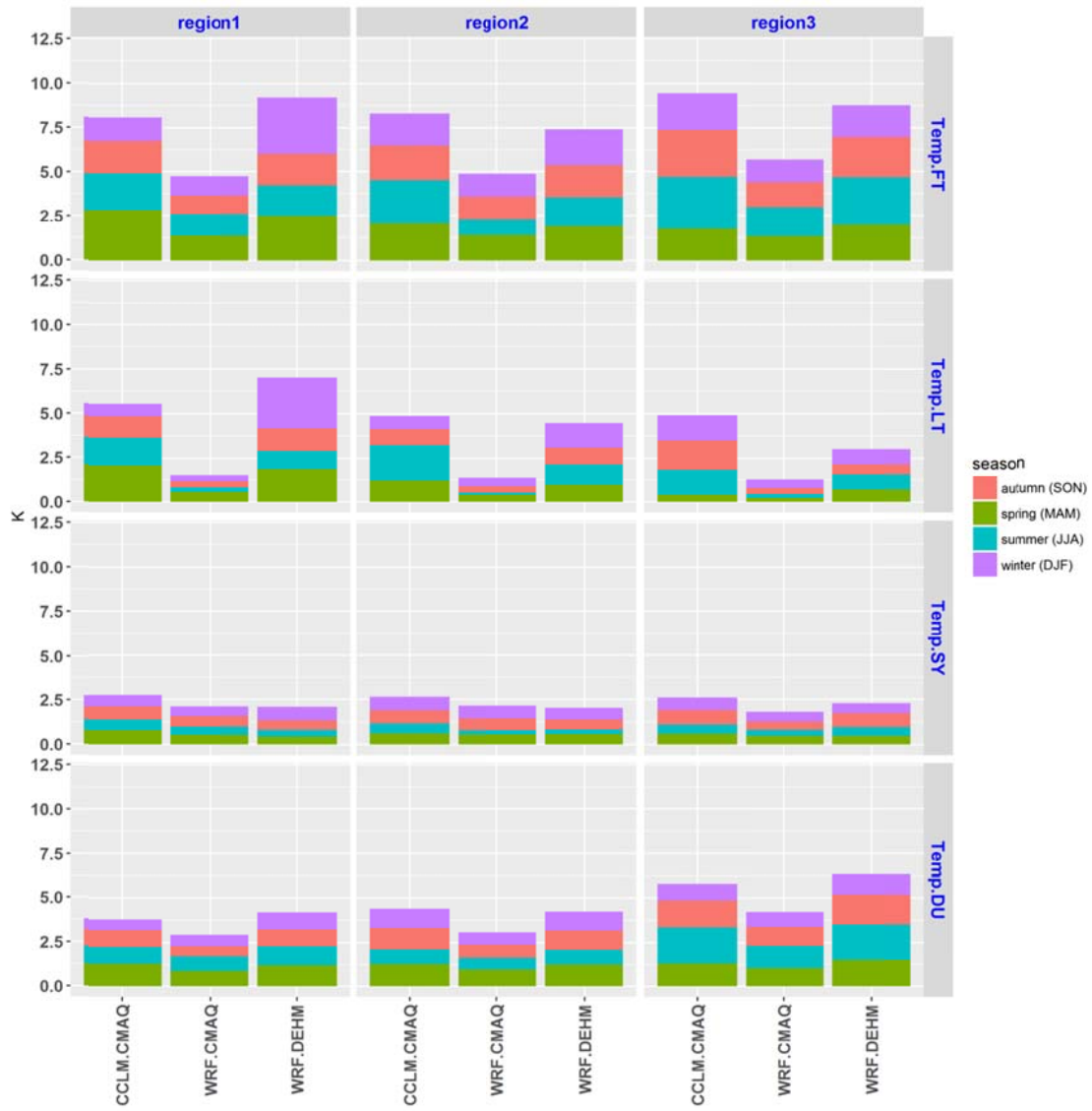
1807

1809   a)

1810   a)   b)

1811   **FIGURE 1. SUB-REGIONS OF THE TWO CONTINENTAL DOMAINS (A) EU; B) NA). OVERLAID ARE THE OZONE MONITORING STATIONS**
1812   **CLASSIFIED BASED ON THE NETWORK**

1813

1814

1815   a)

1816

1817   b)

1818   **FIGURE 2. RMSE FOR A) TEMP AND B) WS IN EUROPE**

1819

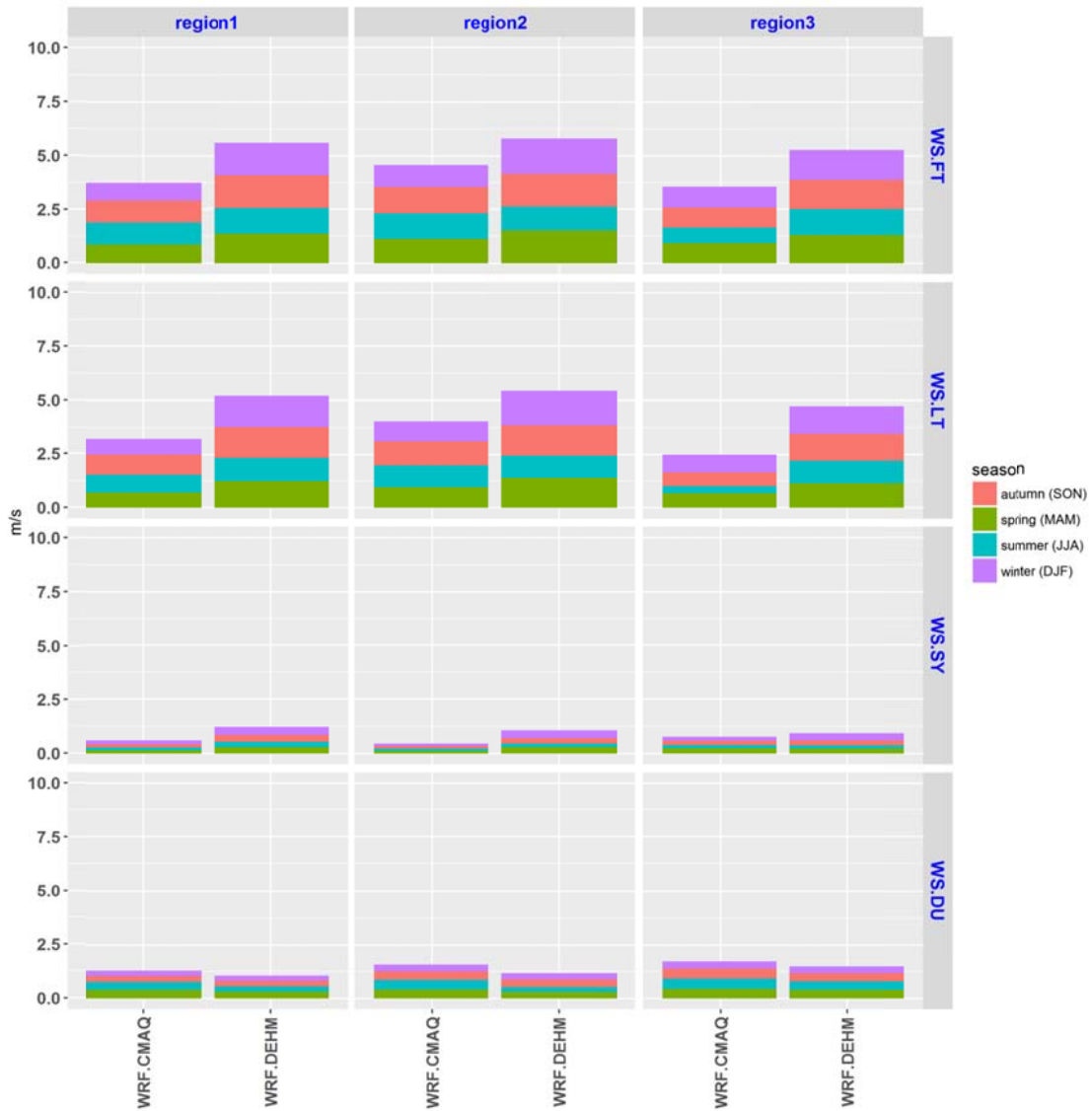
1820   a)

b)

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

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

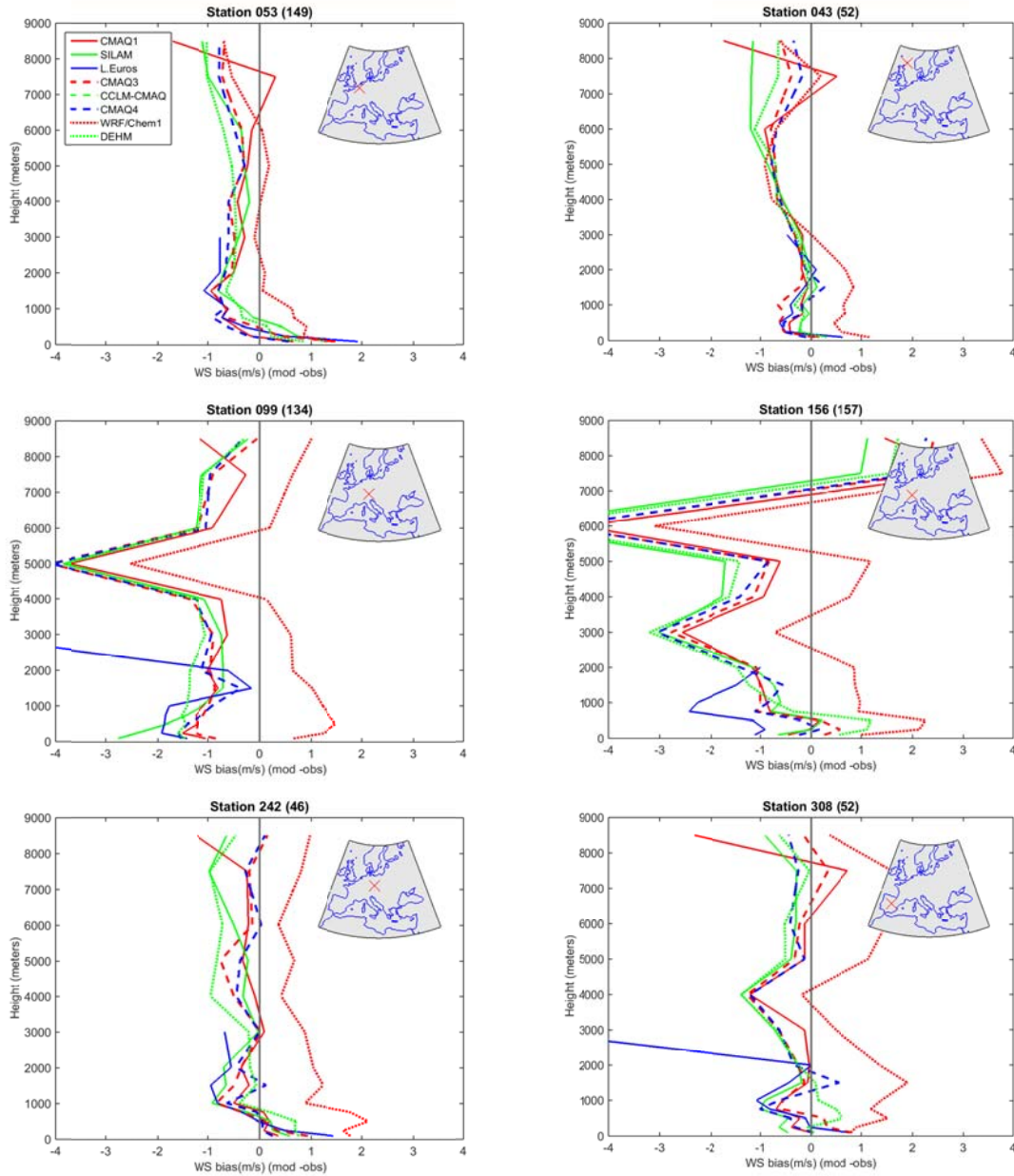
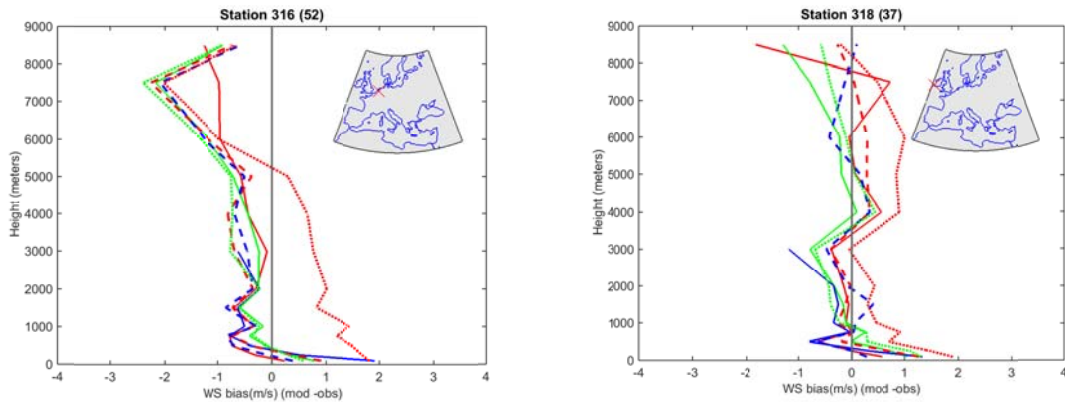



FIGURE 4. MEAN BIAS (MOD – OBS) FOR THE VERTICAL PROFILES OF WIND SPEED MEASURED BY OZONESONDES LAUNCHED FROM THE EUROPEAN LOCATION INDICATED ON THE INSET MAP OF EACH PANEL. THE NUMBER OF HOURLY PROFILES AVAILABLE FOR EACH SITE IS REPORTED IN THE PARENTHESIS AT THE TOP OF EACH PANEL

1833

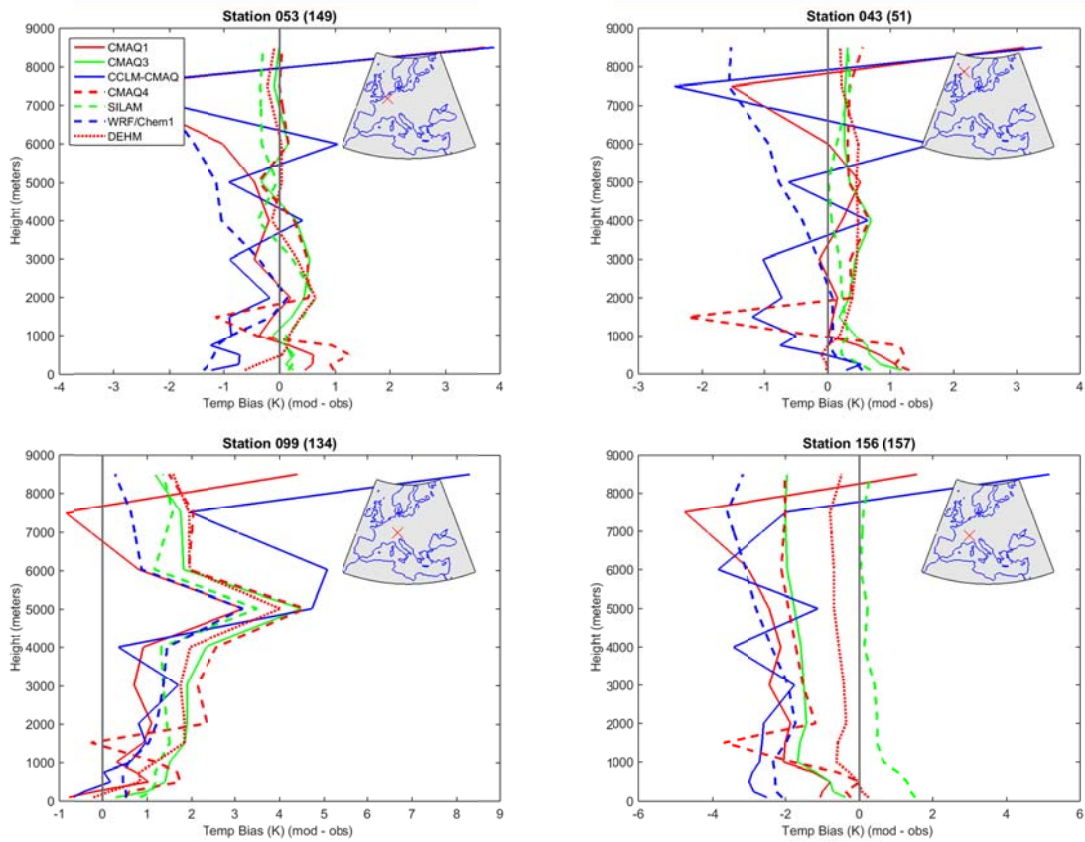
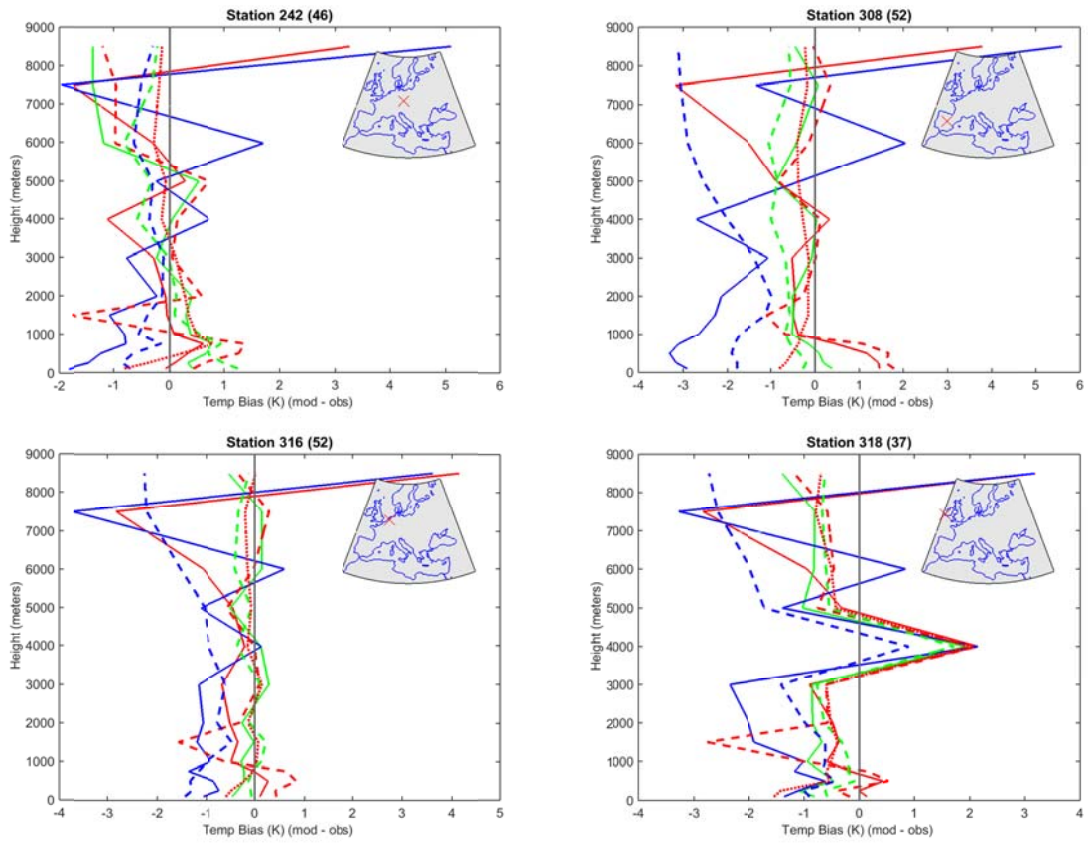



FIGURE 5. MEAN BIAS (MOD – OBS) FOR THE VERTICAL PROFILES OF TEMPERATURE MEASURED BY OZONESONDES LAUNCHED FROM THE EUROPEAN LOCATION INDICATED ON THE INSET MAP OF EACH PANEL. THE NUMBER OF HOURLY PROFILES AVAILABLE FOR EACH SITE IS REPORTED IN THE PARENTHESIS AT THE TOP OF EACH PANEL

1834

1835

1836

1837

1838

1839

1840

1841

1842

1843

1844

1845

1846

Ozonesondes. Time averaged vertical profiles for Wind Speed BIAS – North America

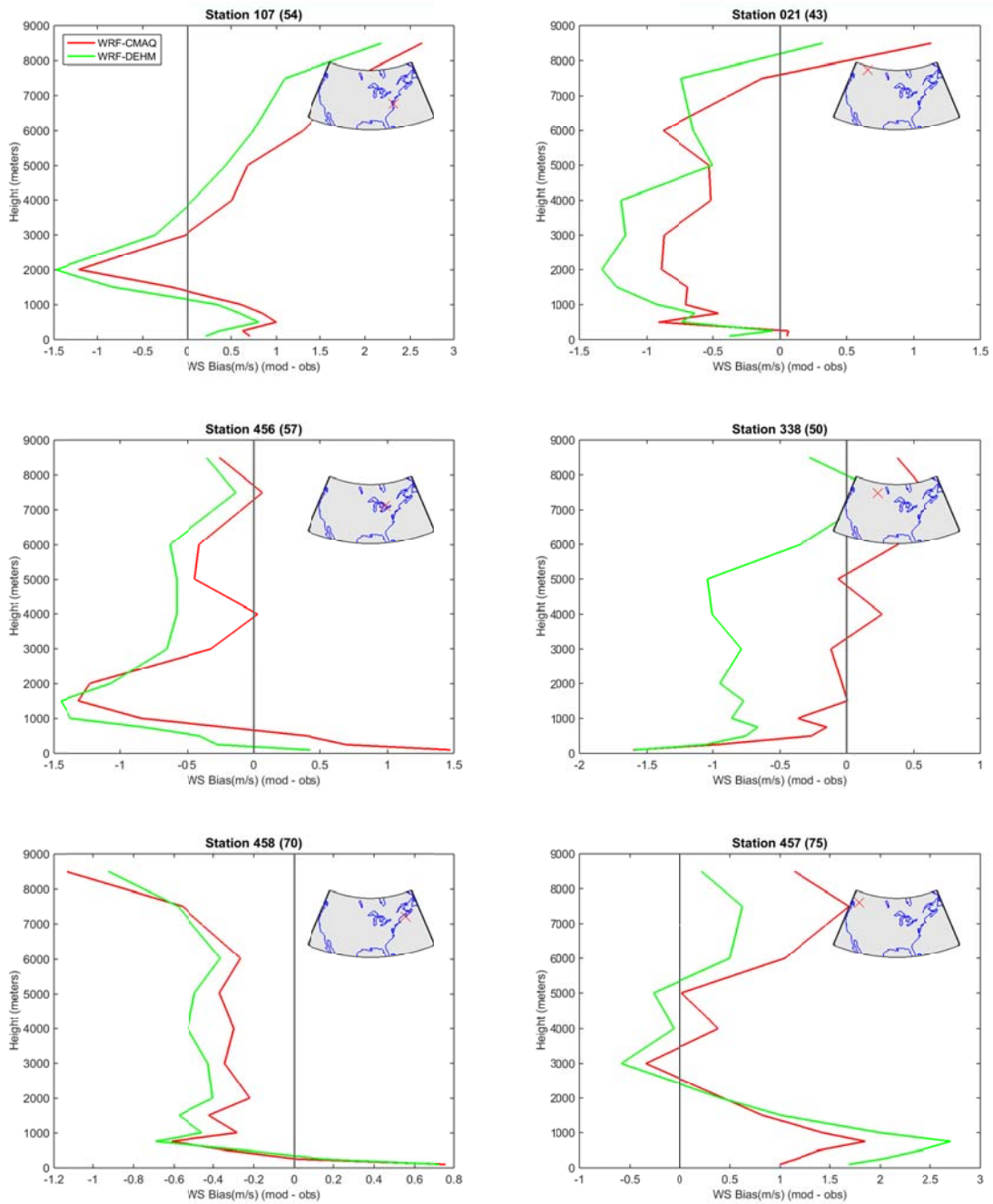



FIGURE 6. MEAN BIAS (MOD – OBS) FOR THE VERTICAL PROFILES OF WIND SPEED MEASURED BY OZONESONDES LAUNCHED FROM THE NORTH AMERICAN LOCATIONS INDICATED ON THE INSET MAP OF EACH PANEL. THE NUMBER OF HOURLY PROFILES AVAILABLE FOR EACH SITE IS REPORTED IN THE PARENTHESIS AT THE TOP OF EACH PANEL

1847

1848

1849

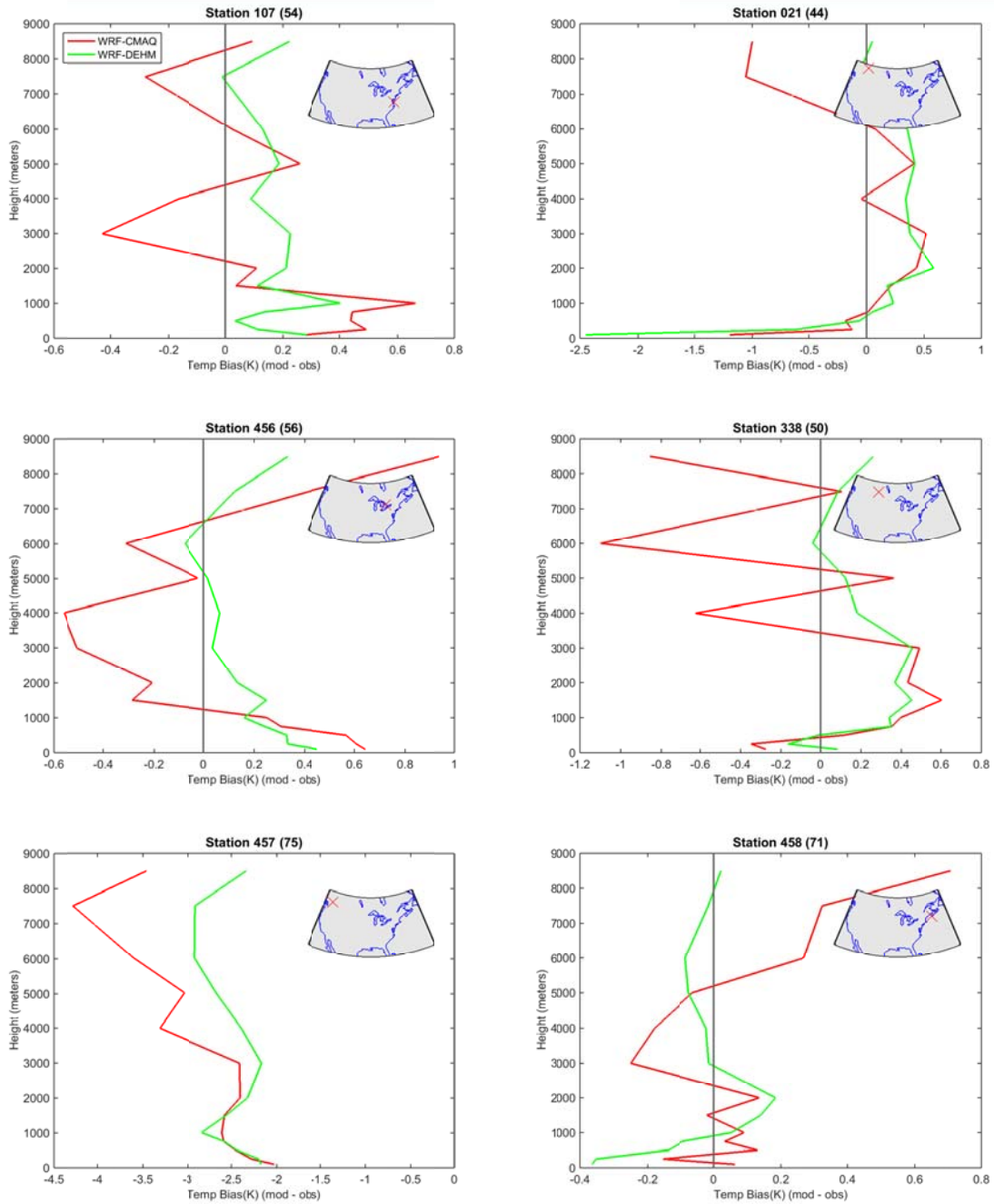
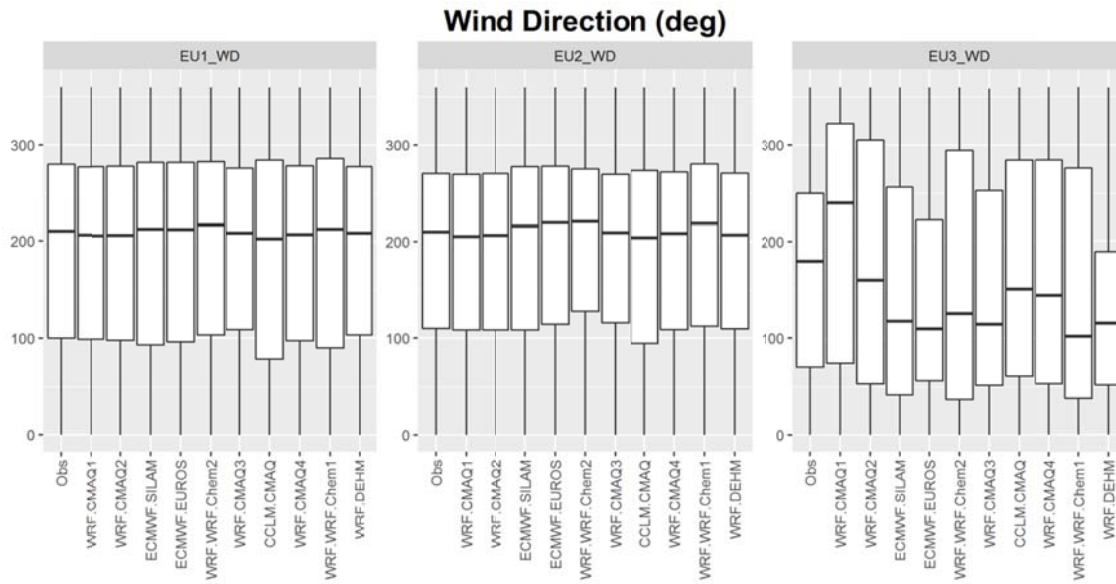
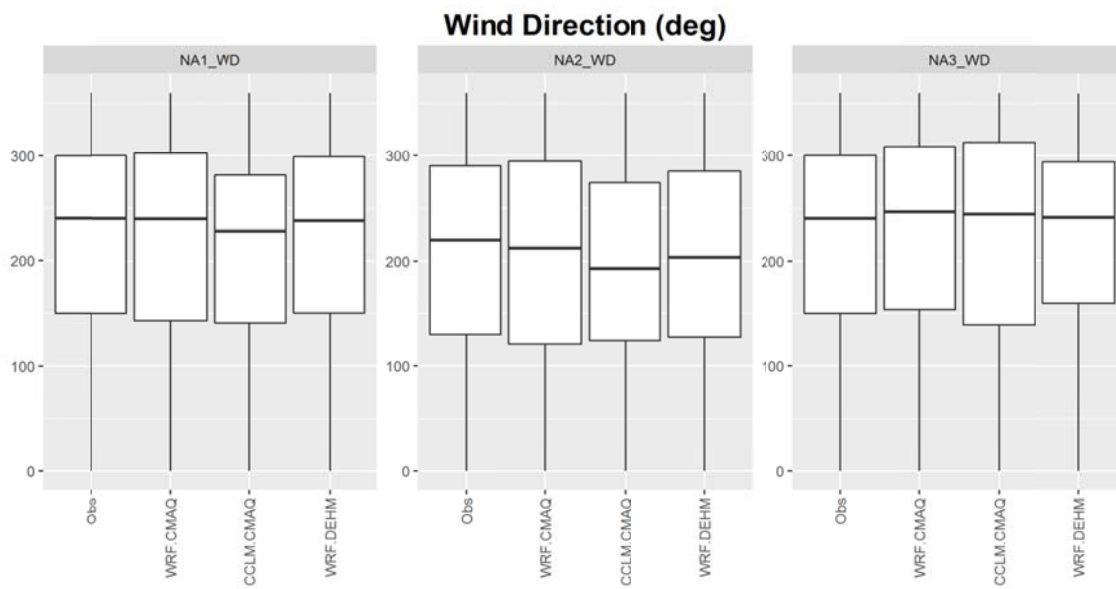
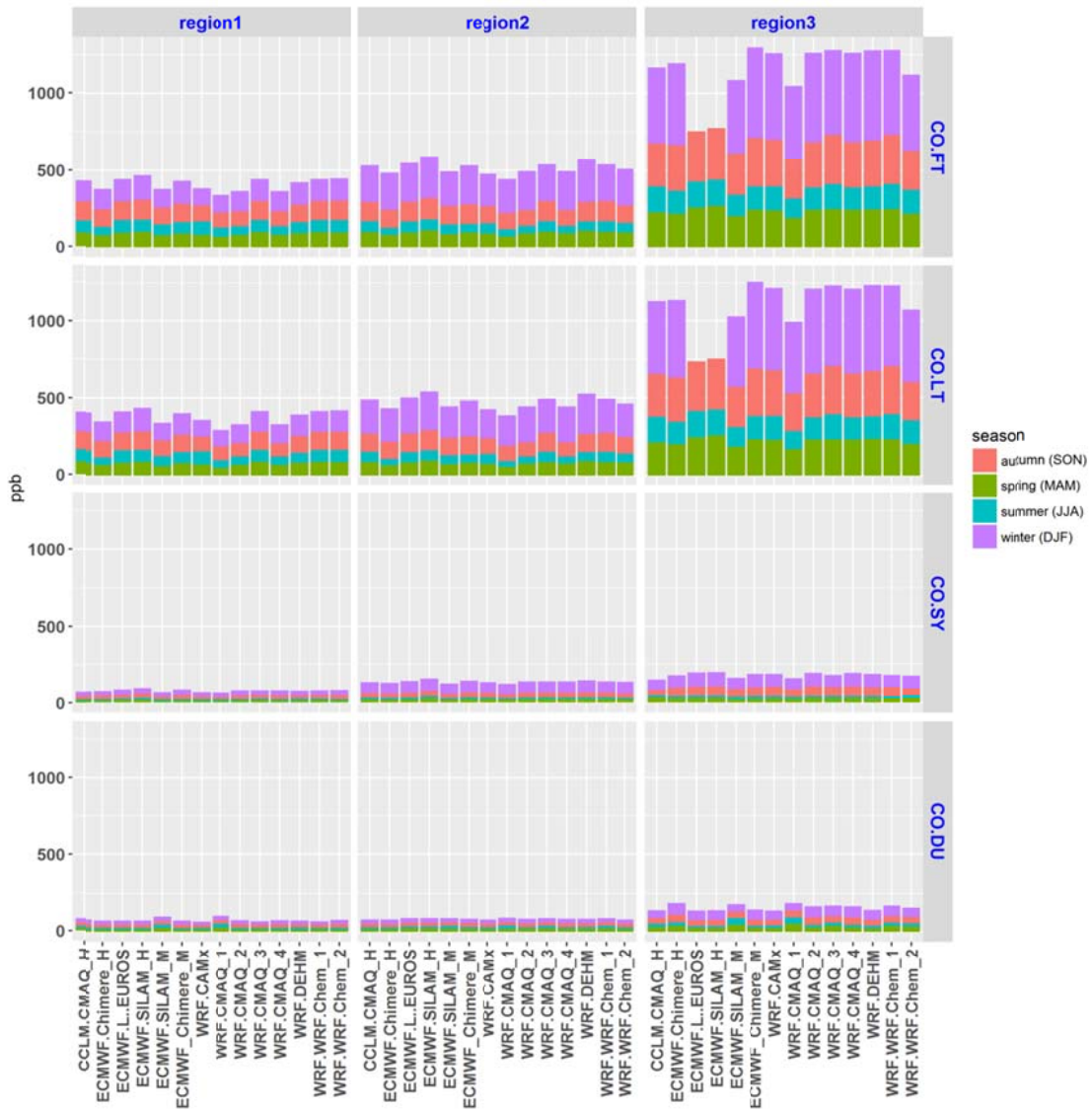
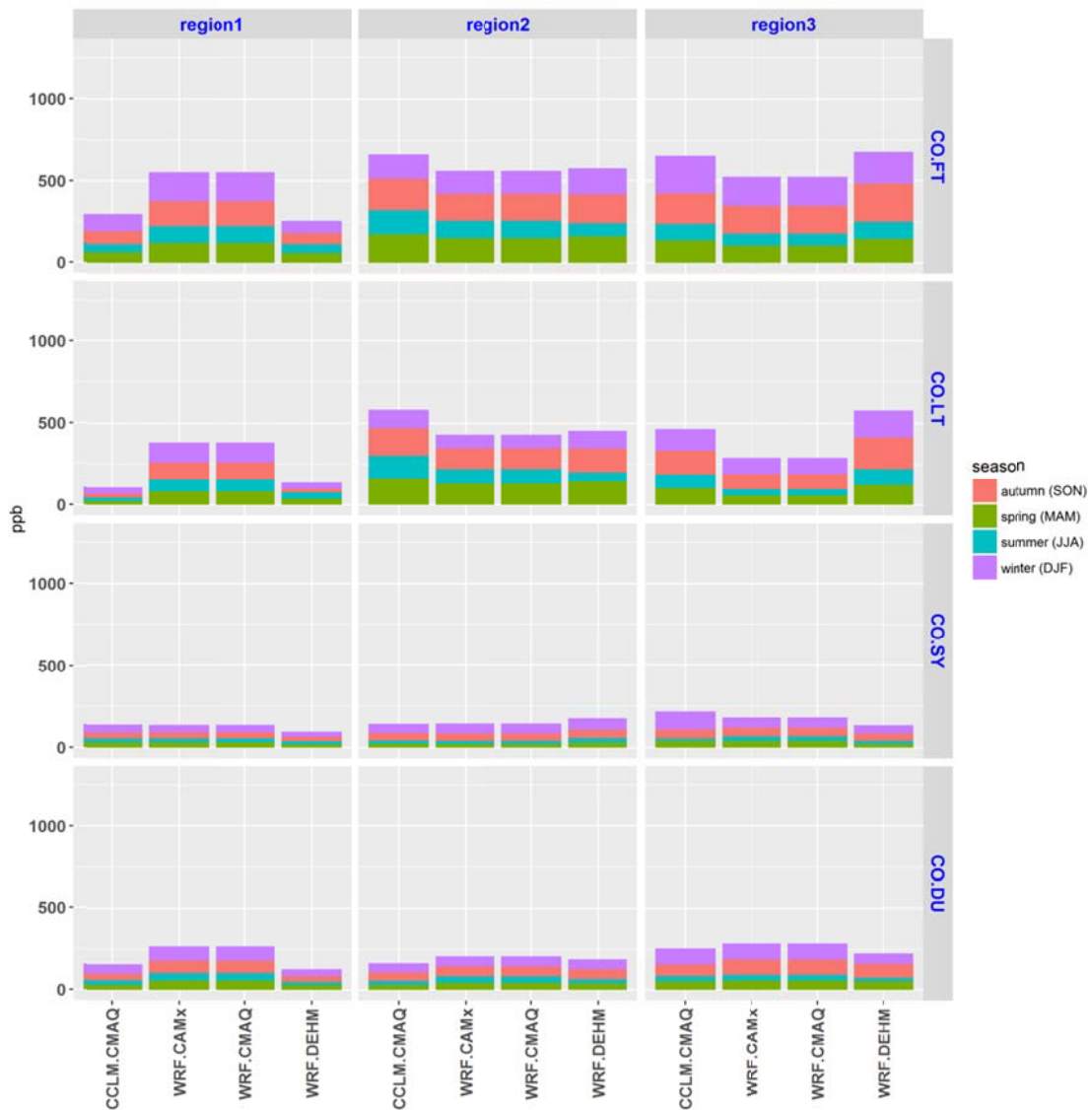



FIGURE 7. MEAN BIAS (MOD – OBS) FOR THE VERTICAL PROFILES OF TEMPERATURE MEASURED BY OZONESONDES LAUNCHED FROM THE NORTH AMERICAN LOCATION INDICATED ON THE INSET MAP OF EACH PANEL. THE NUMBER OF HOURLY PROFILES AVAILABLE FOR EACH SITE IS REPORTED IN THE PARENTHESIS AT THE TOP OF EACH PANEL

1850

1851   **a)**

1852

1853   **b)**

1854   FIGURE 8. SPATIAL AND TEMPORAL VARIABILITY OF THE WIND DIRECTION FOR A) EU AND B) NA FOR THE FULL YEAR 2010. THE BOXES
1855   EXTEND BETWEEN THE 25TH AND 75TH PERCENTILE OF THE TOTAL DISTRIBUTION. THE WHISKERS EXTEND FROM THE MINIMUM TO
1856   THE MAXIUM VALUES

1857

1858   a)

1859

1860   b)

1861   FIGURE 9. RMSE (PPB) FOR CO BY SPECTRAL COMPONENT AND SEASON (PANEL A FOR EUROPE AND B FOR NORTH AMERICA). FT IS THE
1862   FULL (UN-FILTERED) TIME SERIES, LT, SY, DU, ARE THE LONG TERM, SYNOPTIC AND DIURNAL COMPONENTS, RESPECTIVELY.

1863

1864

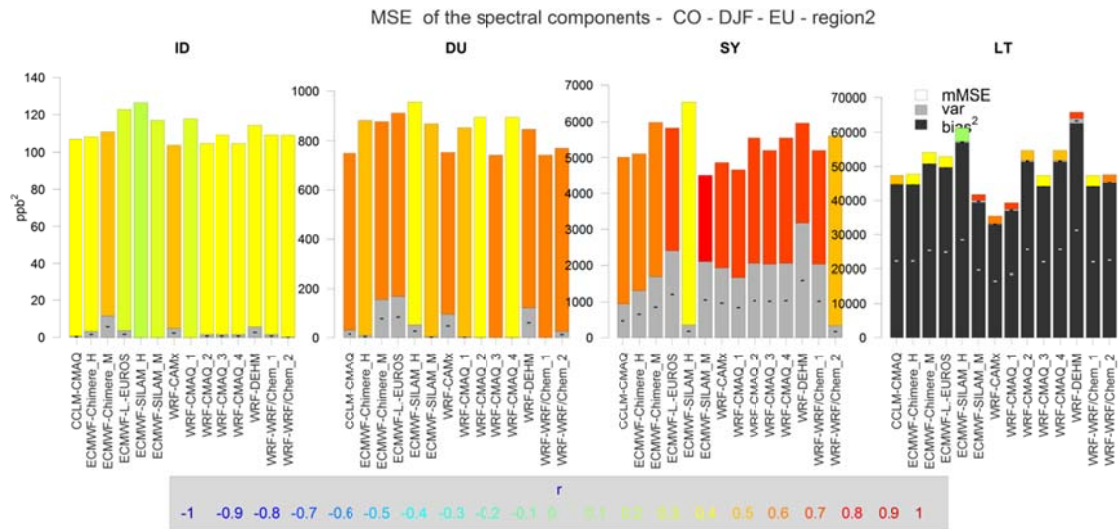
1865

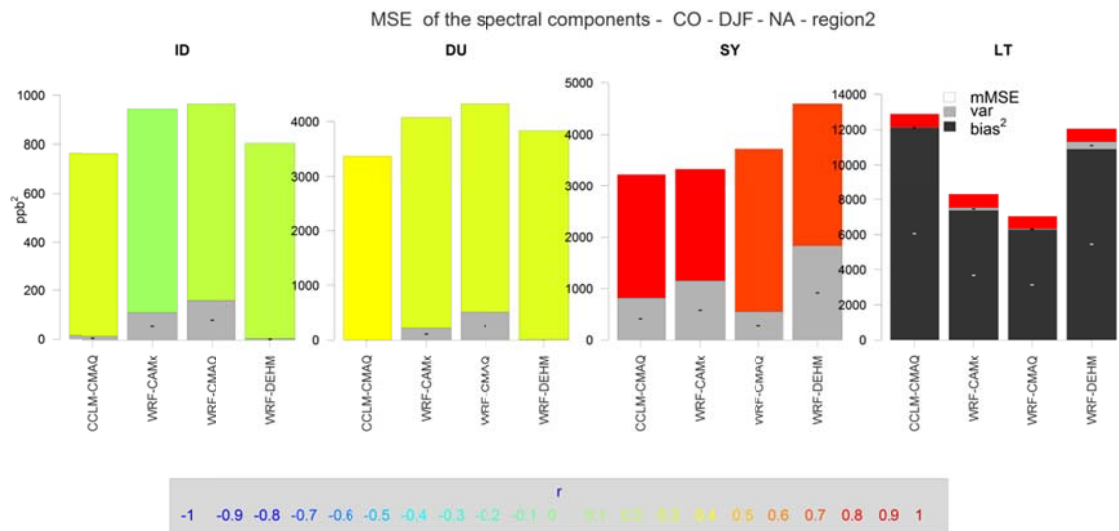
1866

1867

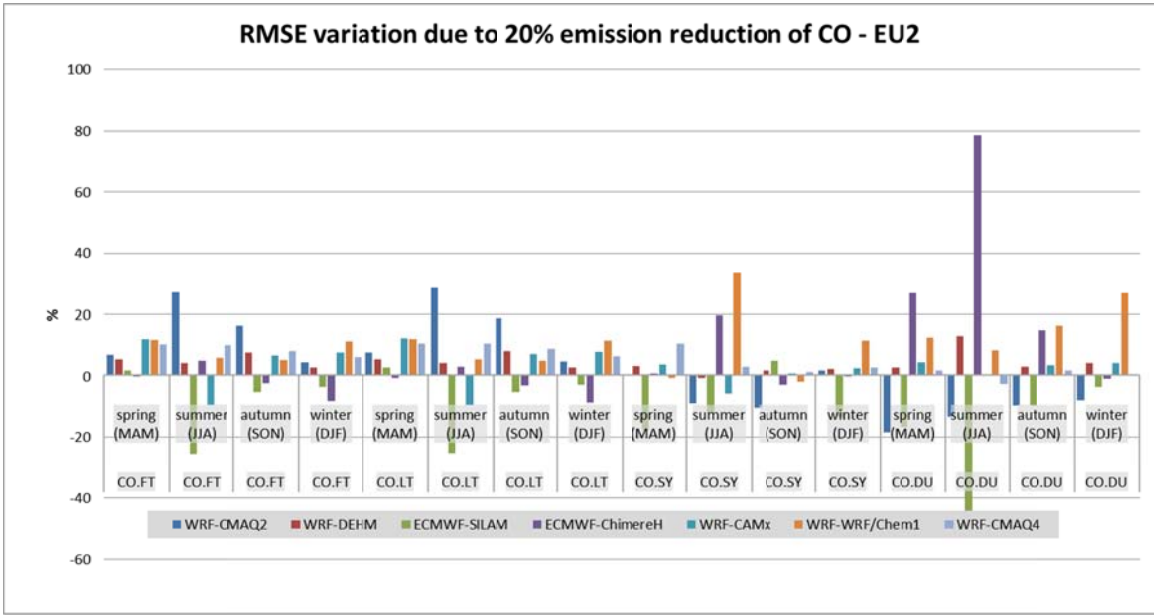
1868

1869

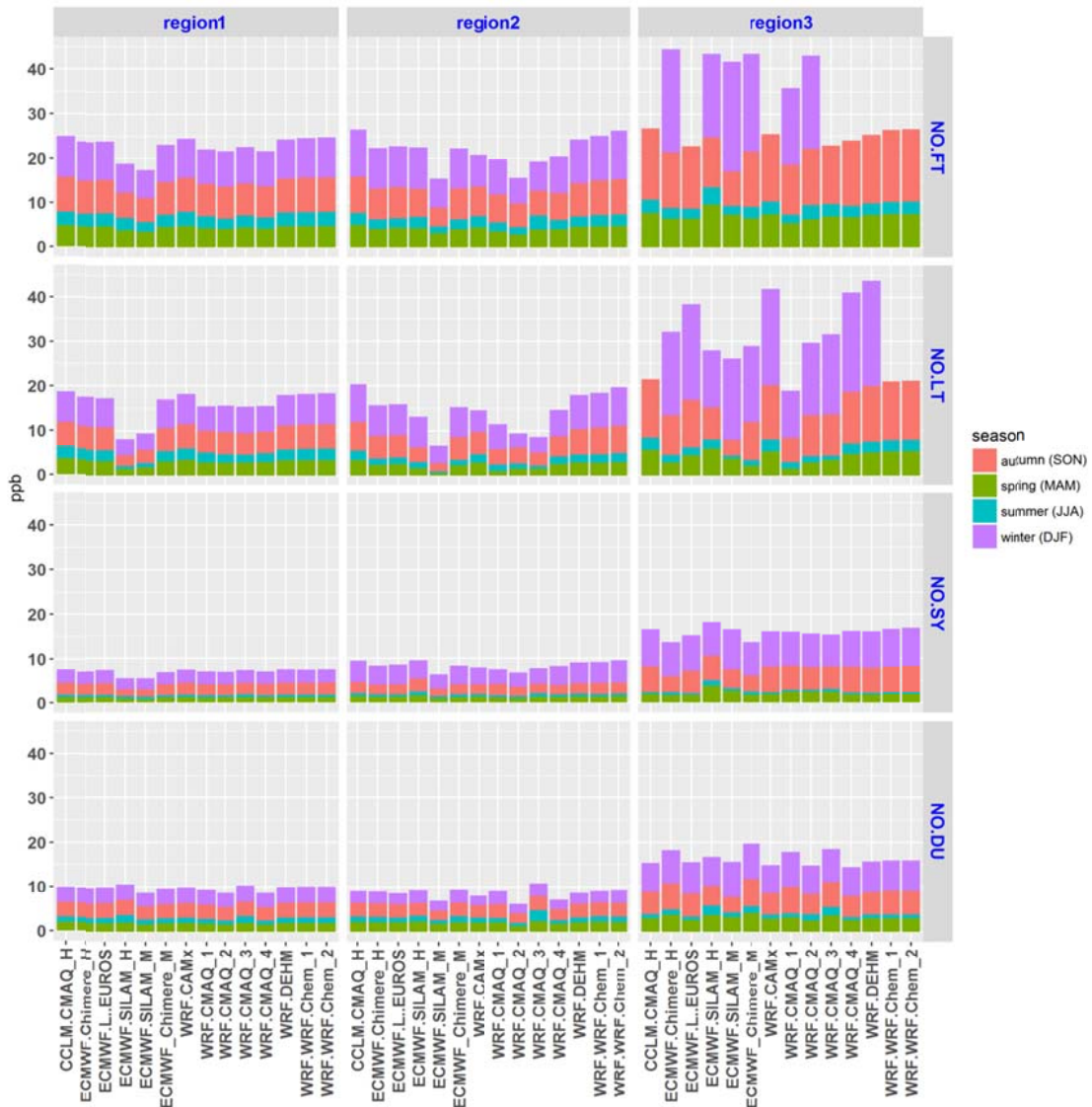
1870

1871

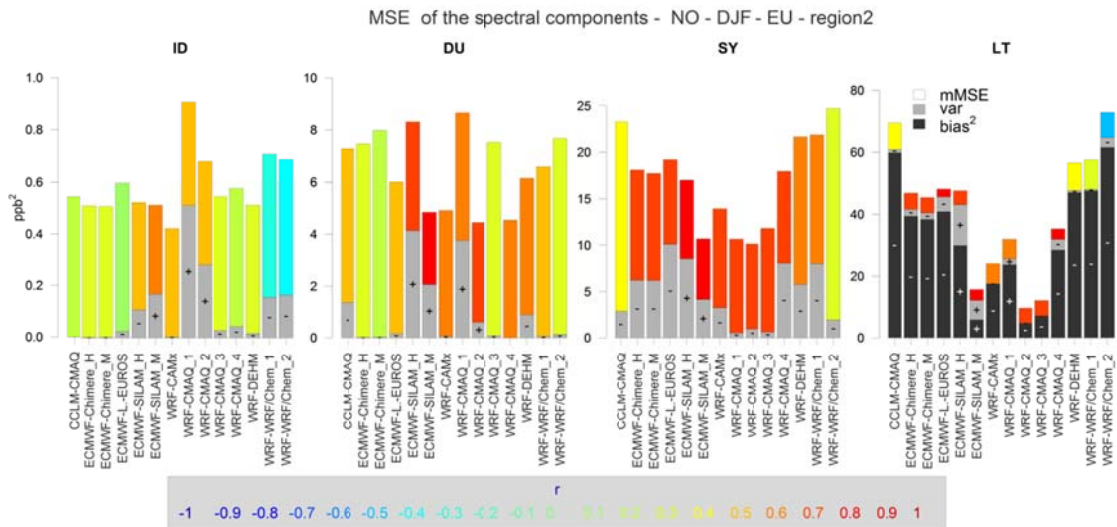
1872   a)

1873

1874   b)

1875   FIGURE 10. MSE (PPB2) BREAKDOWN INTO BIAS SQUARED, VARIANCE AND MMSE FOR THE SPECTRAL COMPONENTS OF THE SPATIAL
1876   AVERAGE TIME SERIES OF CO DURING THE MONTHS OF DECEMBER, JANUARY, AND FEBRUARY (DJF), BASED ON EQ.6. THE BIAS IS
1877   ENTIRELY ACCOUNTED FOR BY THE LT COMPONENT. THE SIGNS WITHIN THE BIAS AND VARIANCE PORTION OF THE BARS INDICATE
1878   MODEL OVERESTIMATION (+) OR UNDERESTIMATION (-) OF THE BIAS AND VARIANCE. THE COLOUR OF THE MMSE SHARE OF THE
1879   ERROR IS CODED BASED ON THE VALUES OF R, THE CORRELATION COEFFICIENT, ACCORDING TO THE COLOUR SCALE AT THE BOTTOM
1880   OF EACH PLOT. TOP PANEL: EU; LOWER PANEL: NA. SIMILAR PLOTS FOR THE OTHER TWO SUB-REGIONS ARE REPORTED IN THE
1881   SUPPLEMENTARY MATERIAL.

1882

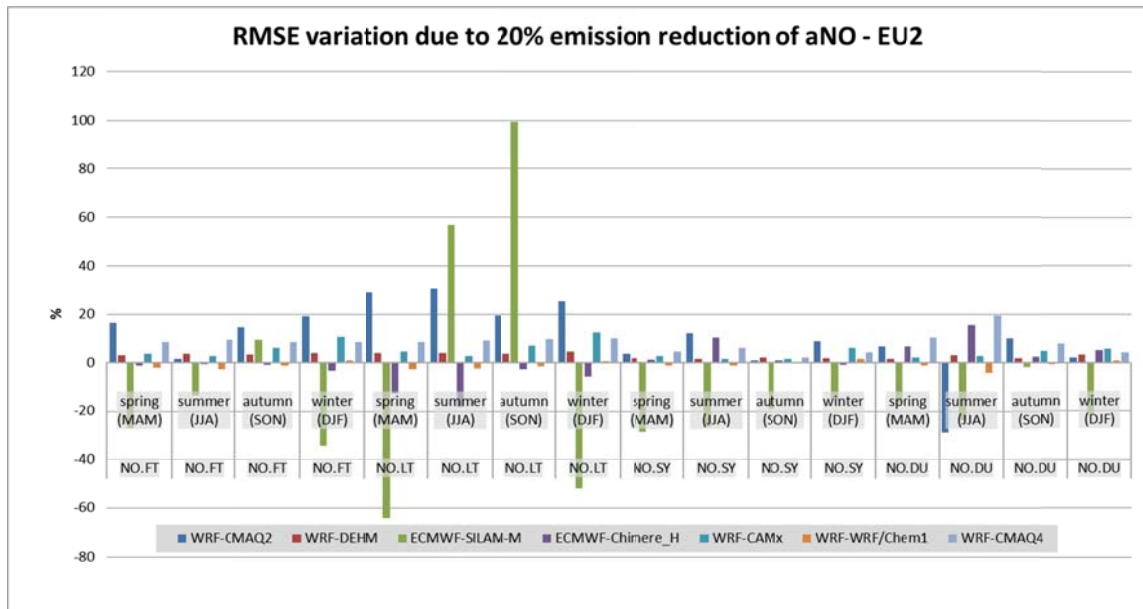
1883 **FIGURE 11. RMSE VARIATION BETWEEN THE 'S20%' SCENARIO (ANTHROPOGENIC EMISSION AND BOUNDARY CONDITION REDUCED BY**
1884 **20%) AND THE BASE CASE FOR CO IN EU2**

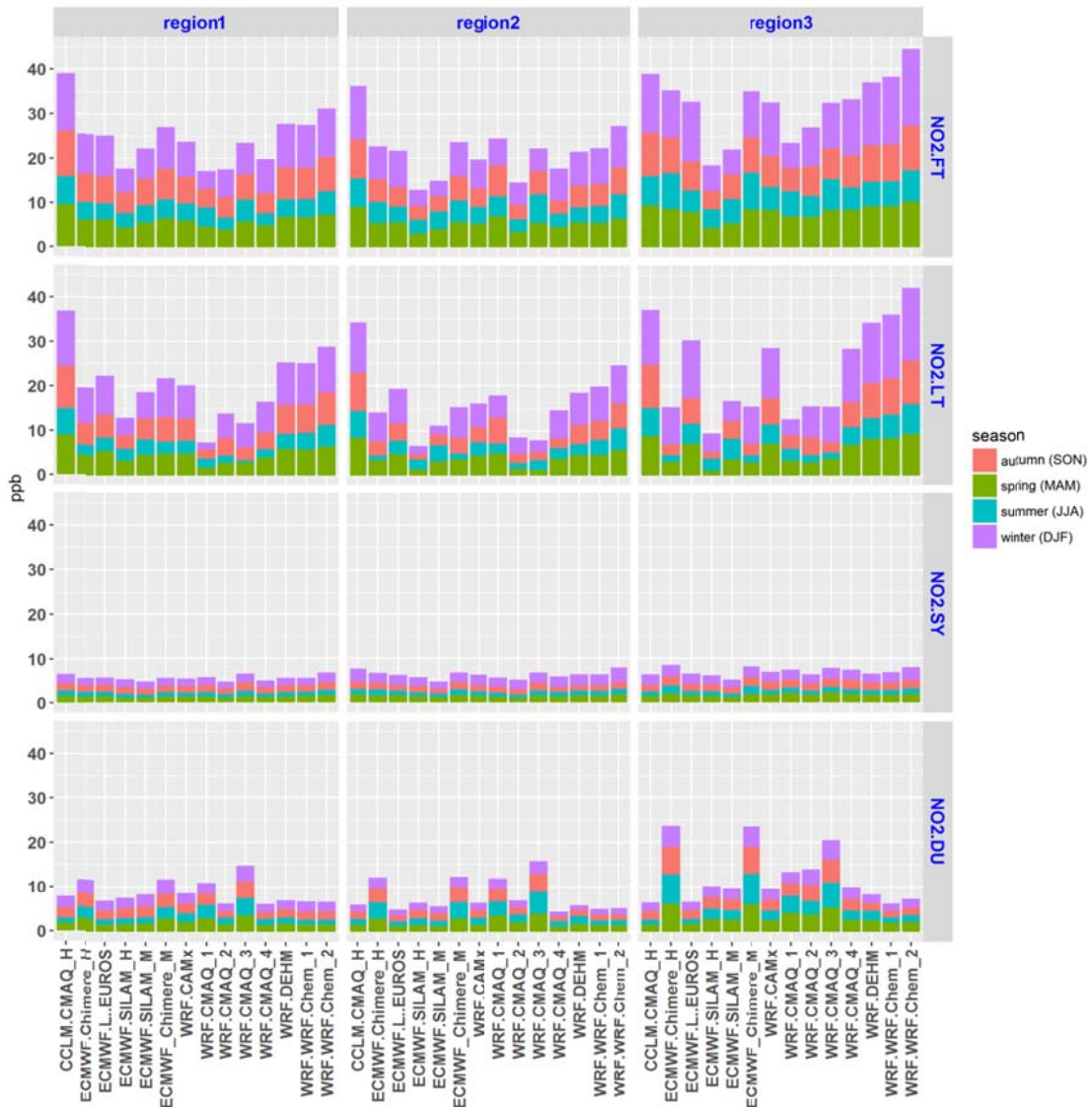
1885

1886

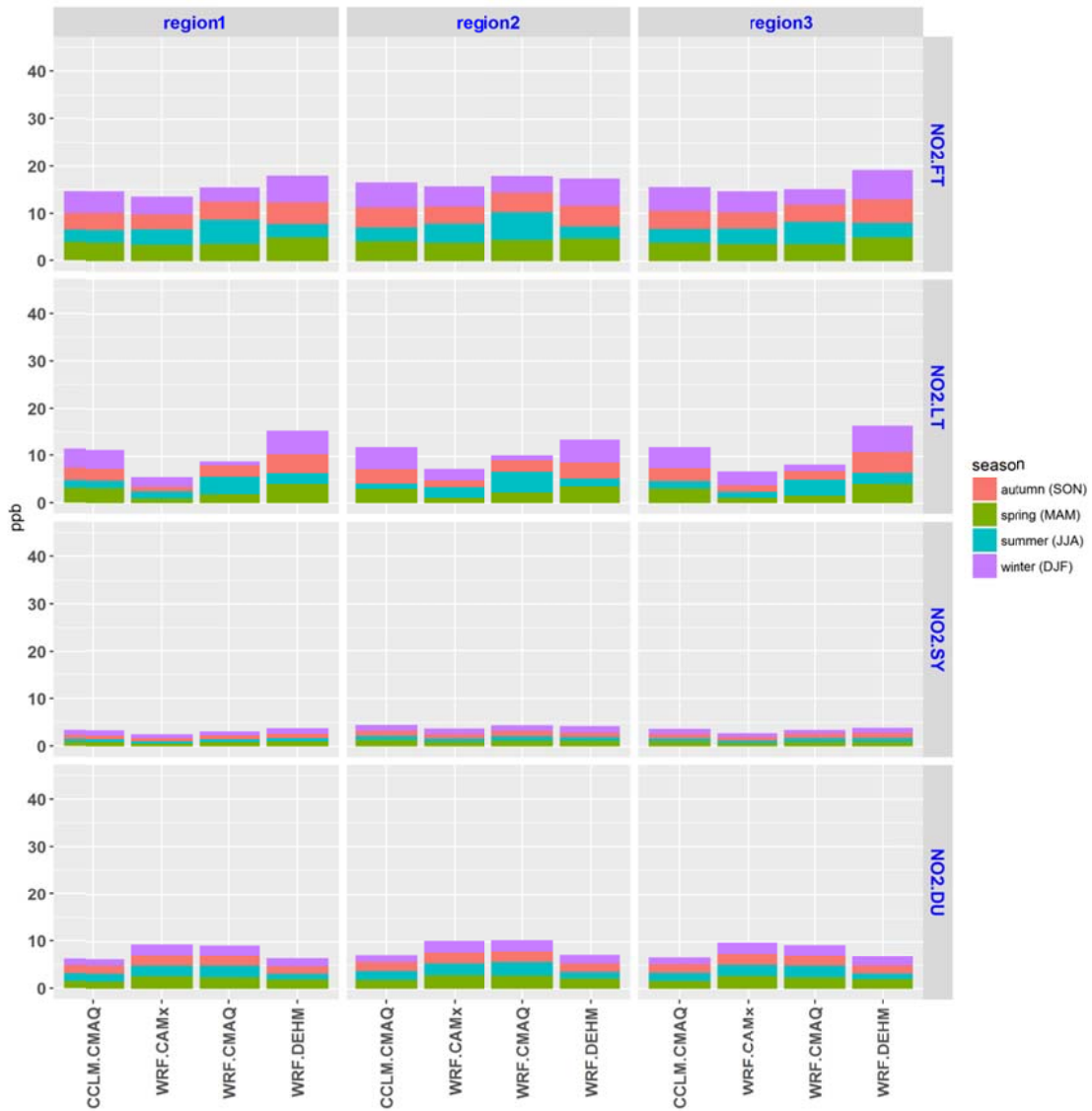
1887   a)

1888

1889 b)

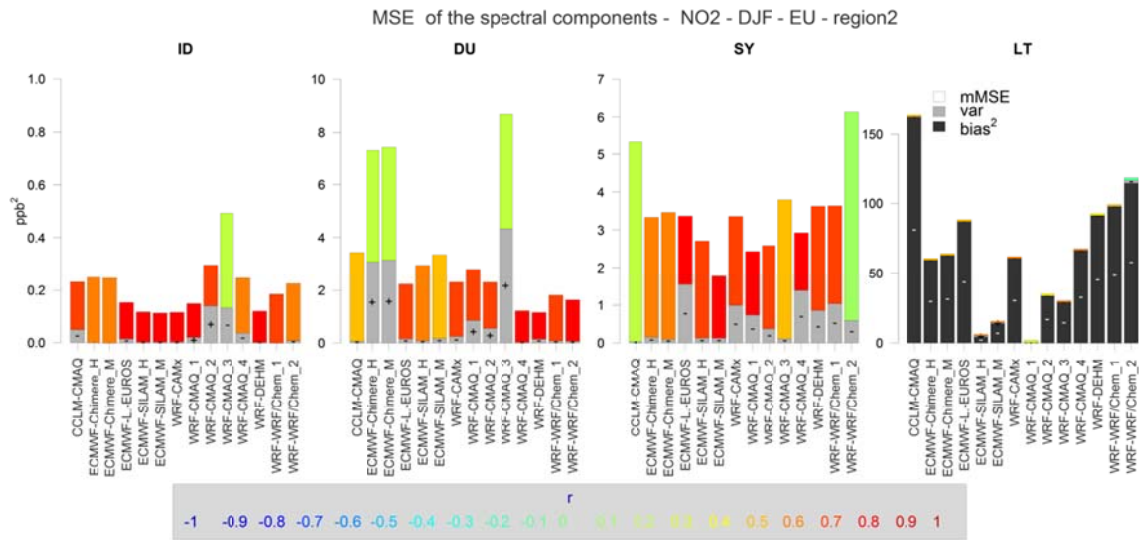
1891

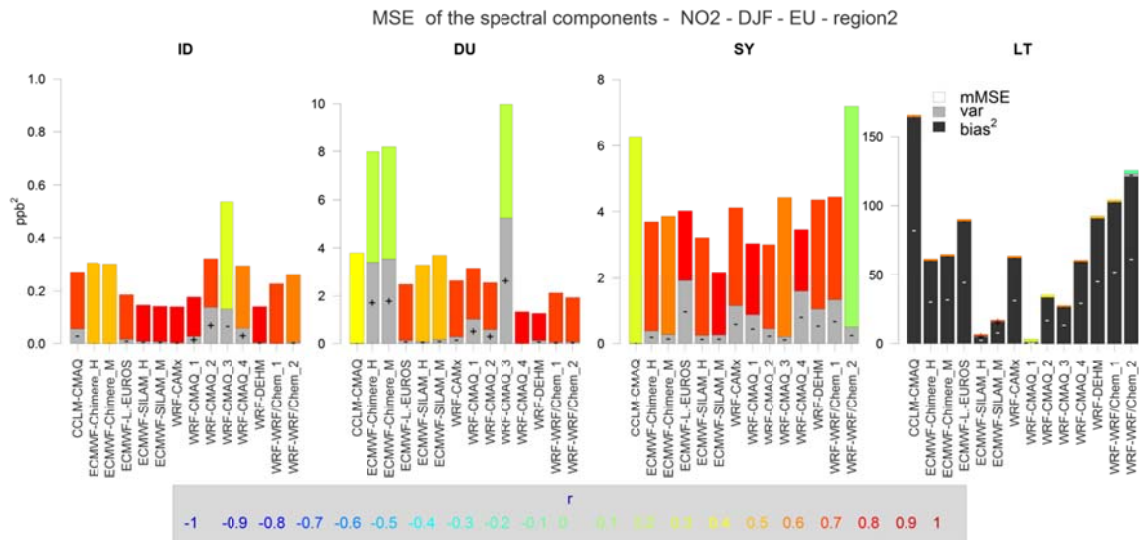
1892

1893

1894 FIGURE 13. RMSE VARIATION BETWEEN THE 'S20%' SCENARIO (ANTHROPOGENIC EMISSION AND BOUNDARY CONDITION REDUCED BY
1895 20%) AND THE BASE CASE FOR ANTHROPOGENIC NO (ANO) IN EU2

1896

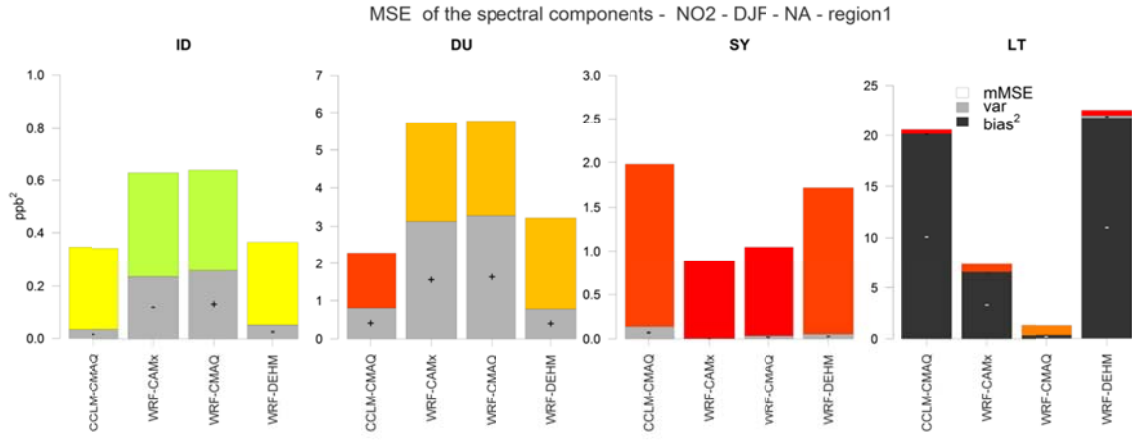
1897   a)

1898

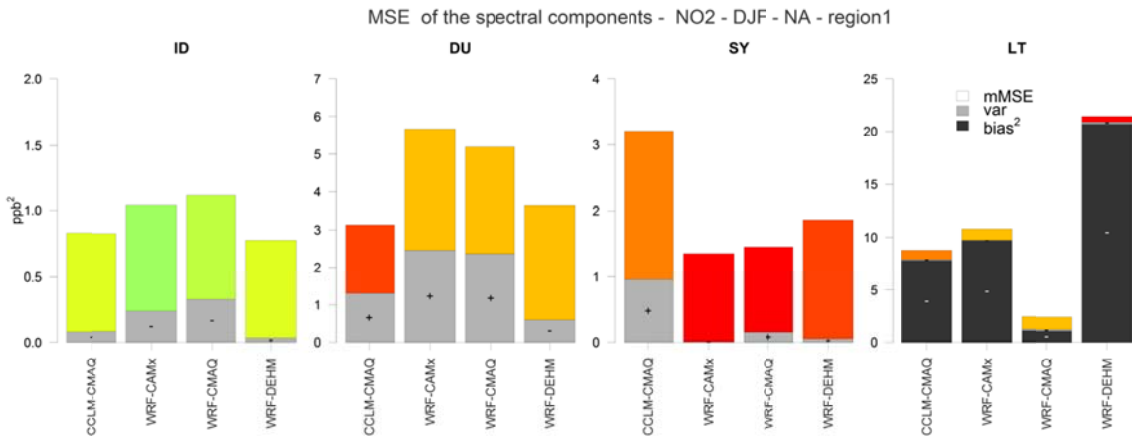
1899   b)

1900   **FIGURE 14. AS IN FIGURE 9 FOR NO₂**

1901

1902 a) Urban NO$_2$ in EU2 sub-region (223 stations)

1903

1904 b) Rural NO$_2$ in EU2 sub-region(159 stations)

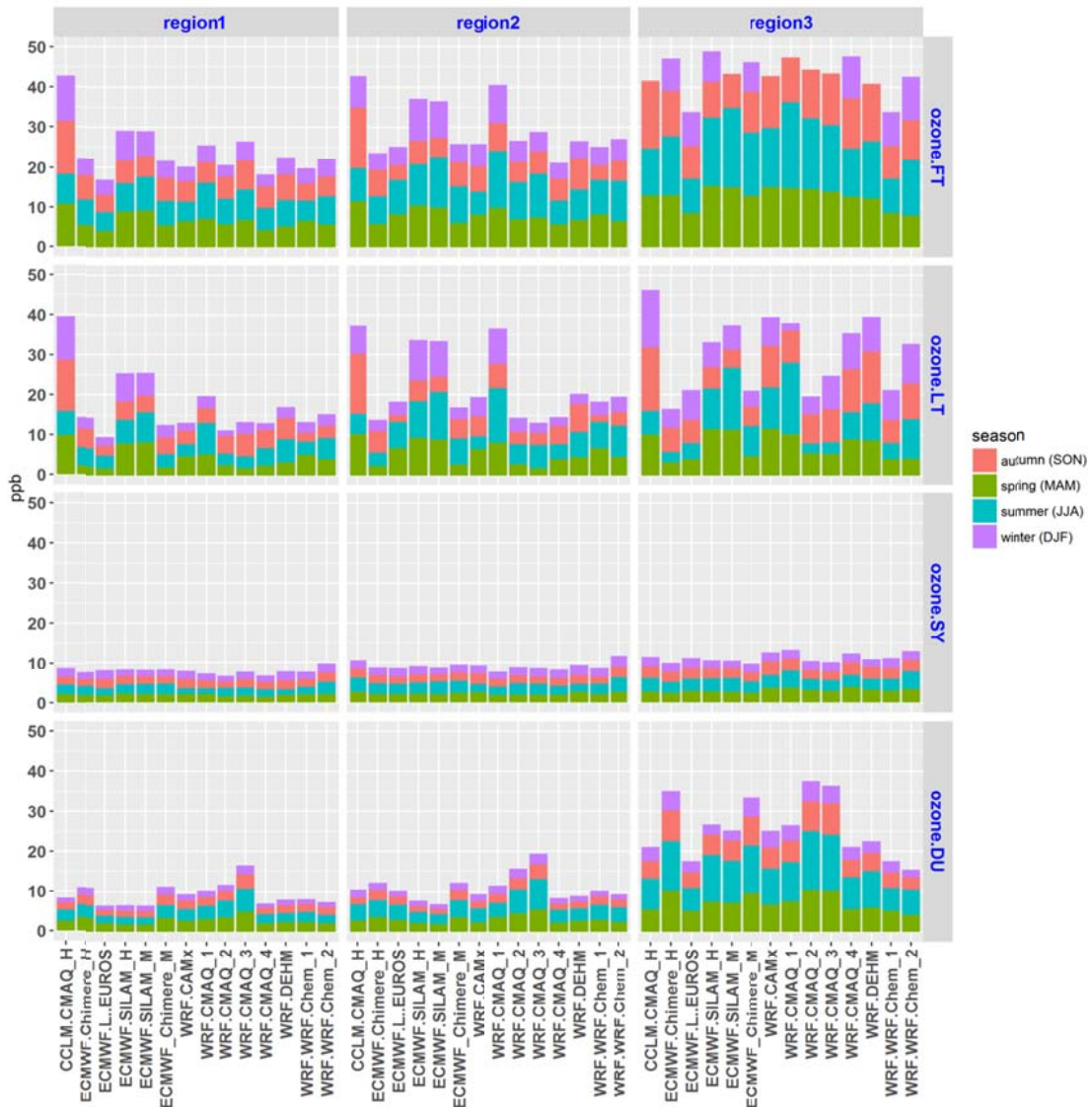
1905 FIGURE 15. AS IN FIGURE 10 FOR NO$_2$ IN EU2. UPPER PANEL: URBAN SITES ONLY (223 STATIONS); LOWER PANEL: RURAL SITES ONLY
1906 (159 STATIONS)

1907

1908   a) NA1 urban (39 stations)

1909

1910   b) NA1 rural (10 stations)

1911   FIGURE 16. AS IN FIGURE 10 FOR NO$_2$ IN NA1. UPPER PANEL: URBAN SITES ONLY (39 STATIONS); LOWER PANEL: RURAL
1912   SITES ONLY (10 STATIONS).

1913

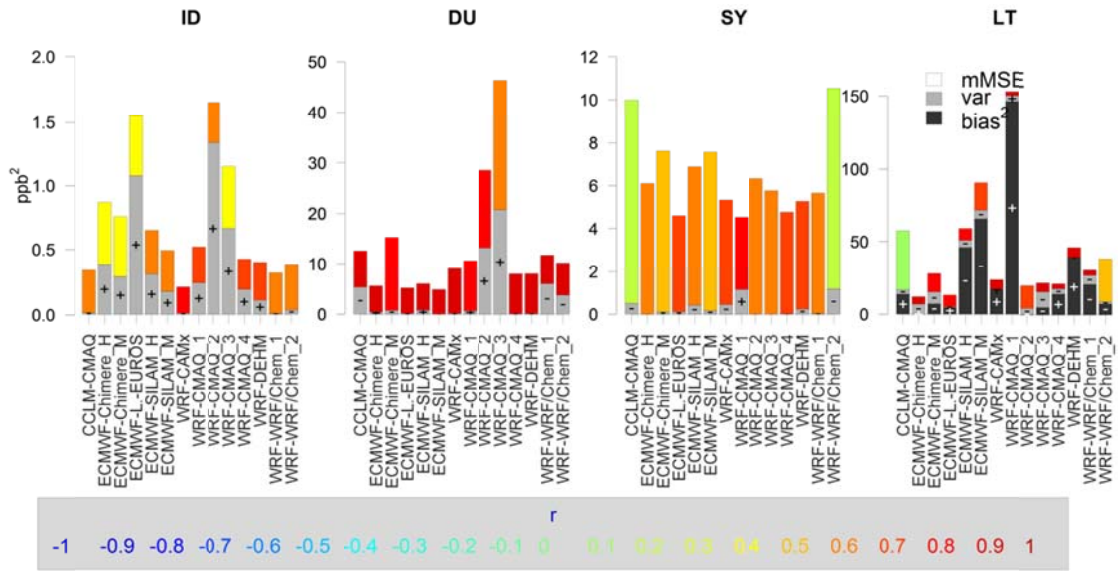
1914

1915

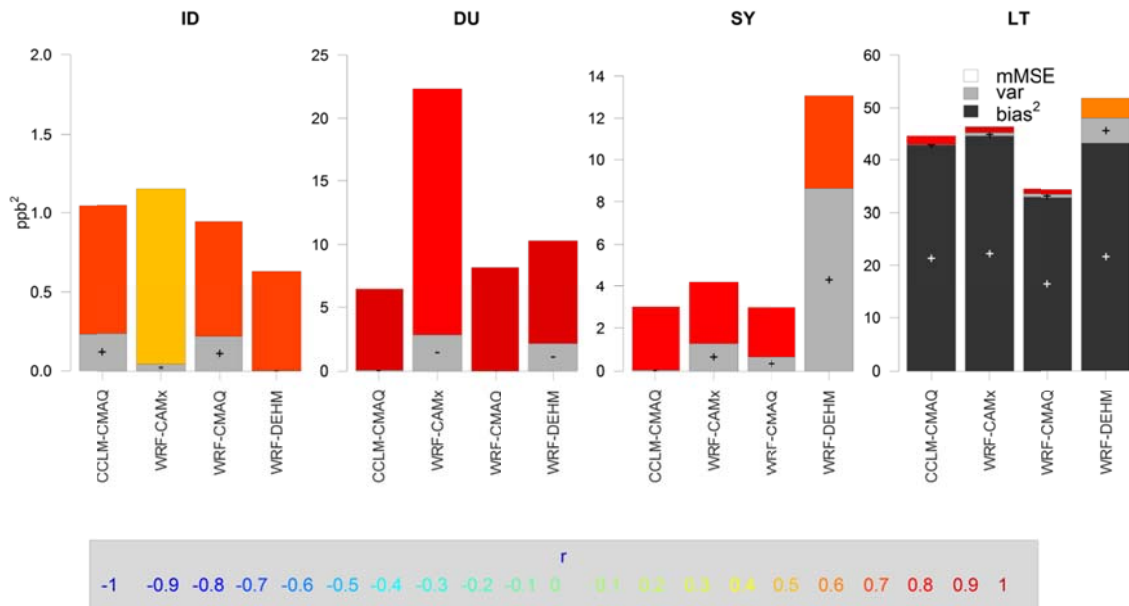
1916

1917   a)

1918

1919   b)

1920   **FIGURE 17. AS IN FIGURE 9 FOR OZONE**

1921

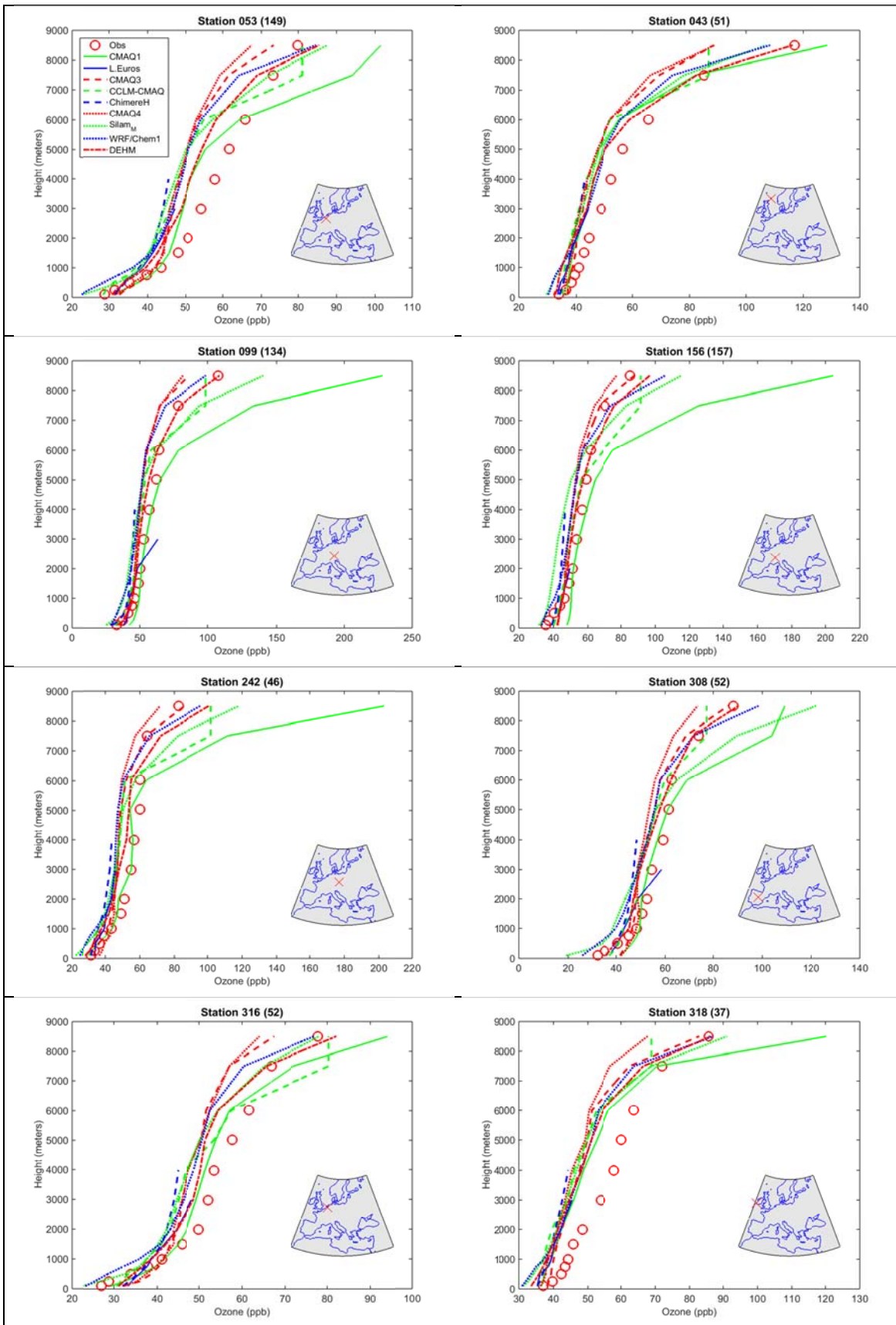
1922 a)

1923

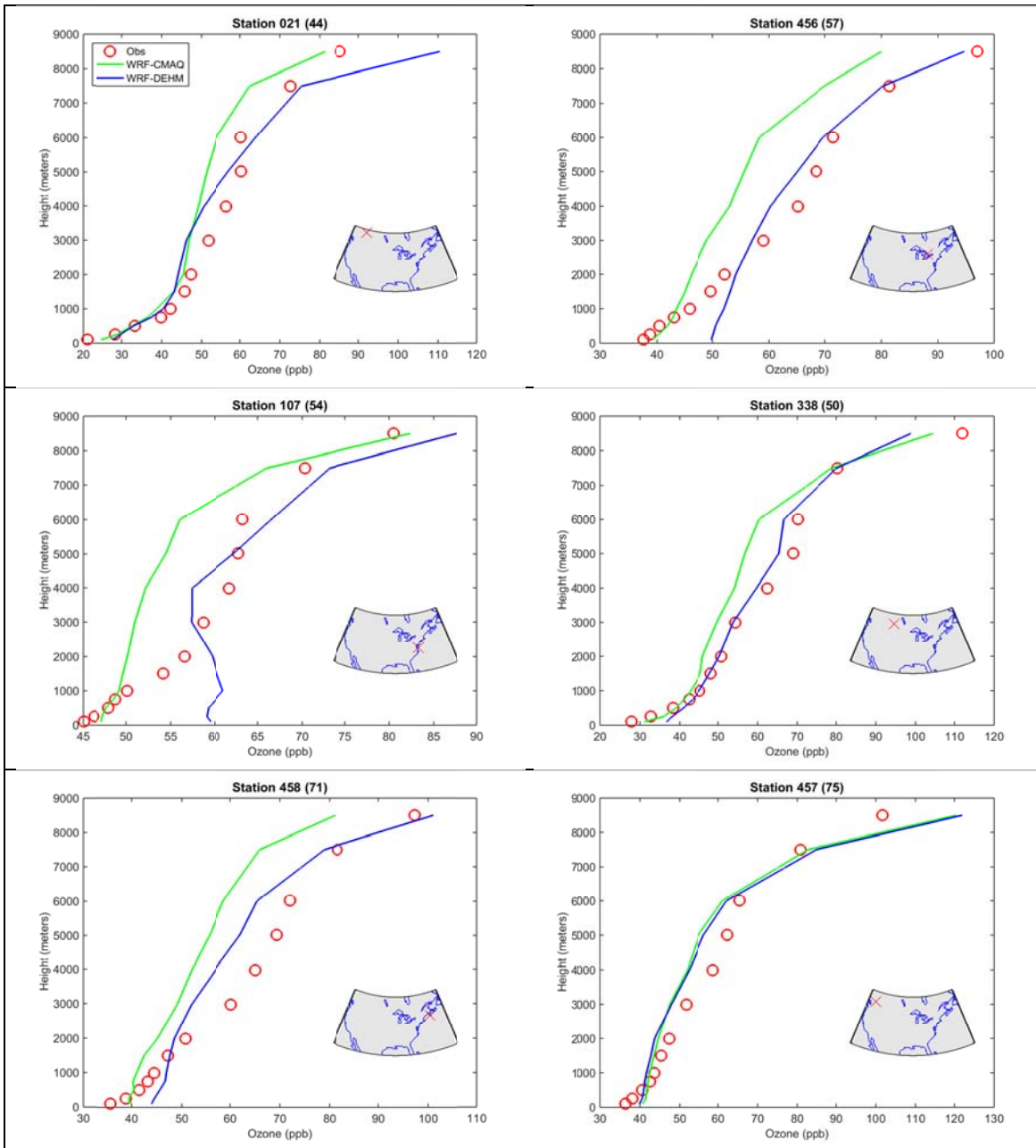
1924 b)

1925 FIGURE 18. AS IN FIGURE 10 FOR OZONE DURING THE MONTHS FROM MAY TO SEPTEMBER

1926

1927

1928

1929 FIGURE 19. OZONE MIXING-RATIO PROFILES MEASURED BY OZONESONDES LAUNCHED FROM THE EUROPEAN LOCATION INDICATED
1930 ON THE INSET MAP (LOWER-RIGHT CORNER) OF EACH PANEL. THE PROFILES ARE TIME-AVERAGED OVER THE NUMBER OF HOURLY
1931 RECORDS REPORTED IN THE PARENTHESIS AT THE TOP OF EACH PANEL. LEGEND AS IN THE FIRST PANEL.

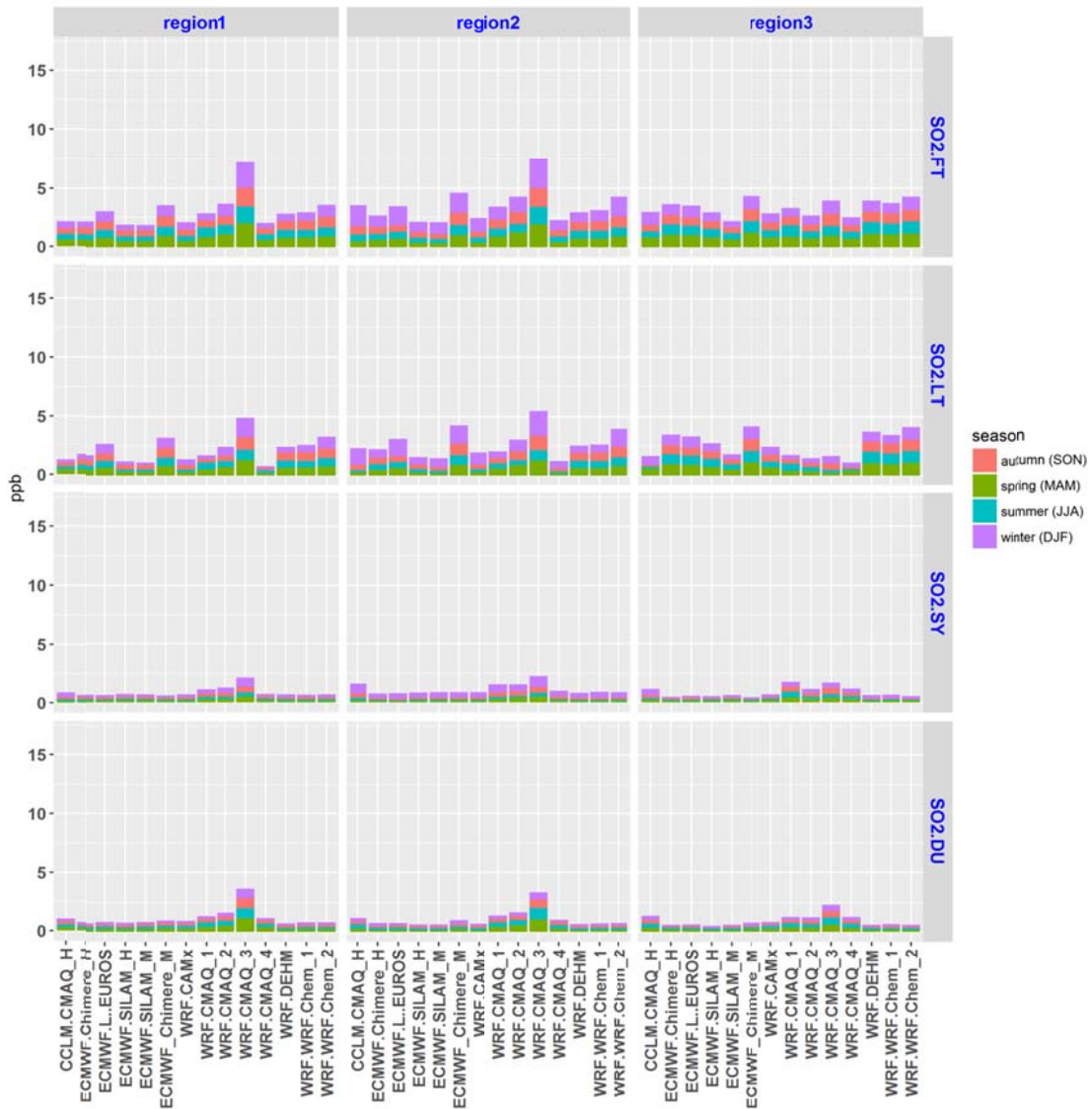
1932

1933

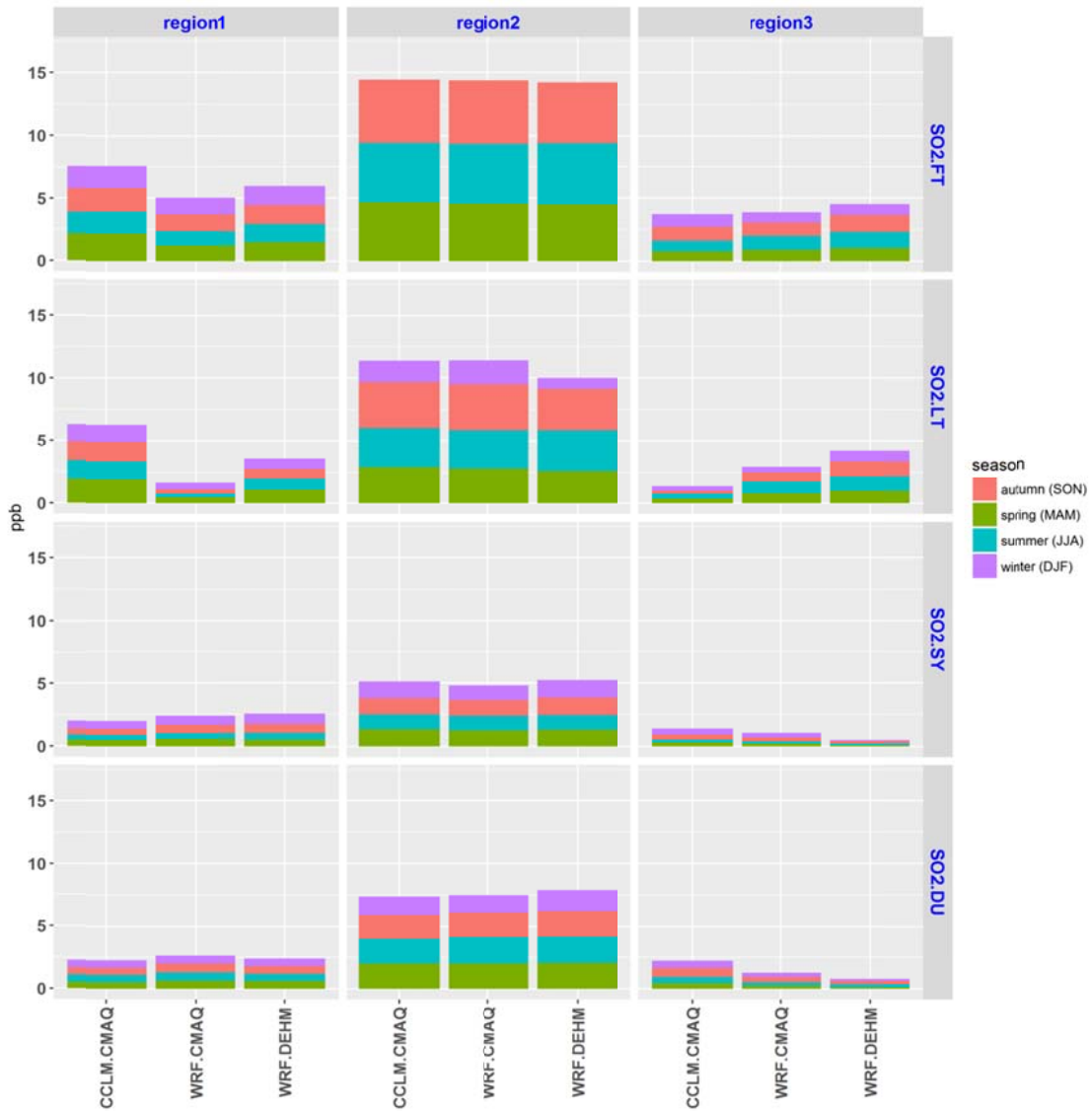
1934 FIGURE 20. AS IN FIGURE 19 FOR NORTH AMERICA

1935

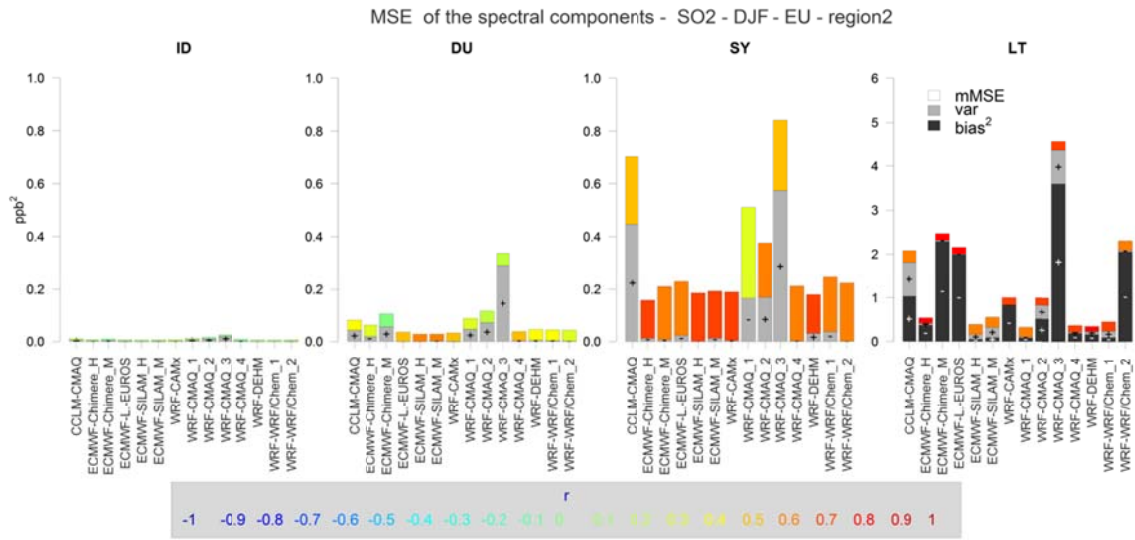
a)   b)

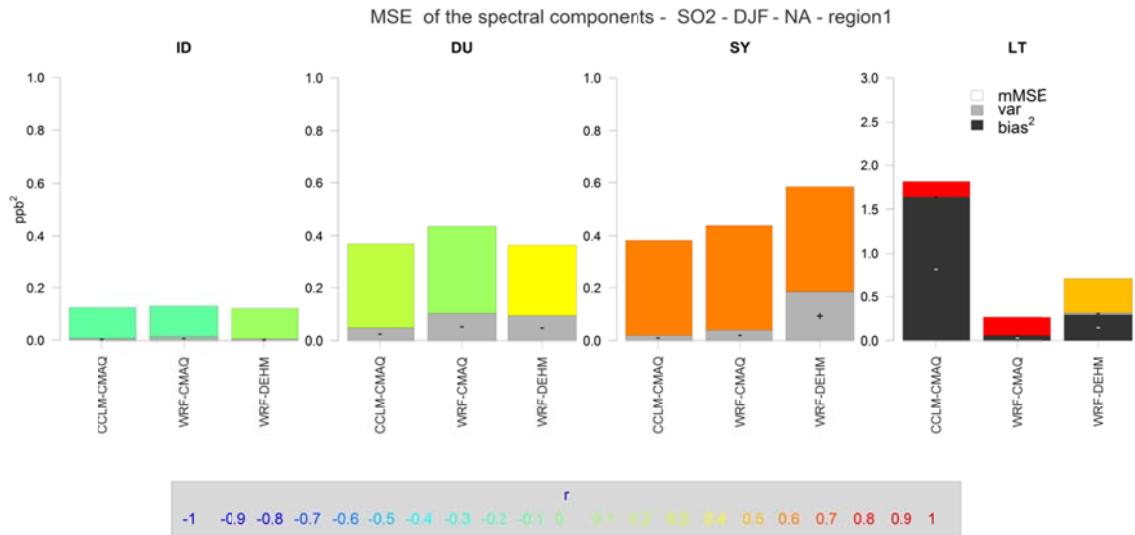
1936   FIGURE 21. OZONE VS NO MODELLED MEAN BIAS FOR THE EU2 SUB-REGION, COLOR-CODED BY TEMPERATURE BIAS AND SYMBOLS
1937   ACCORDING TO THE NO$_X$ EMISSION FRACTION OF NO AND NO$_2$. EACH POINT REPRESENTS A MODEL. A) WINTER MONTHS AND B)
1938   SUMMER MONTHS.

1939

1940

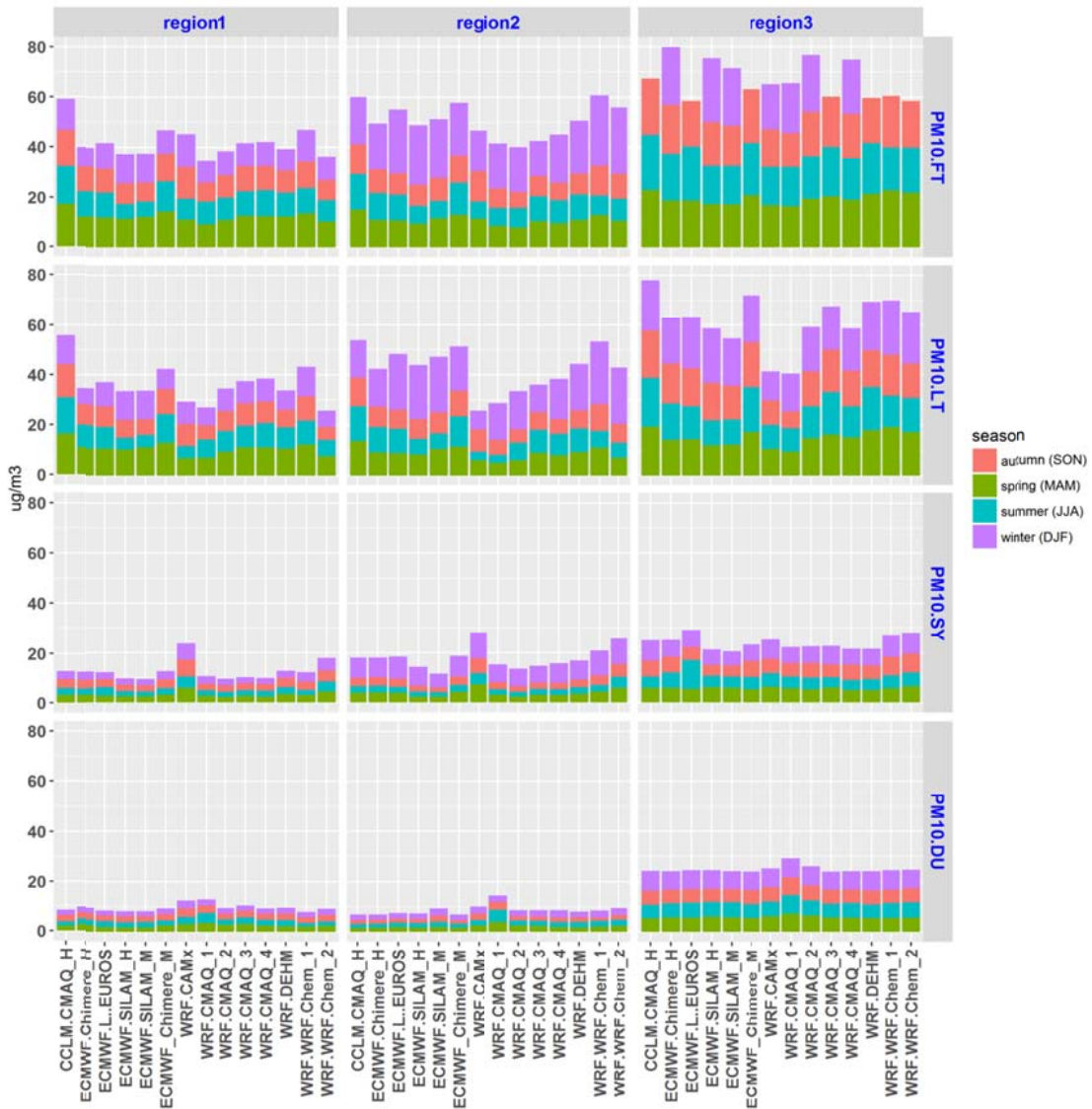
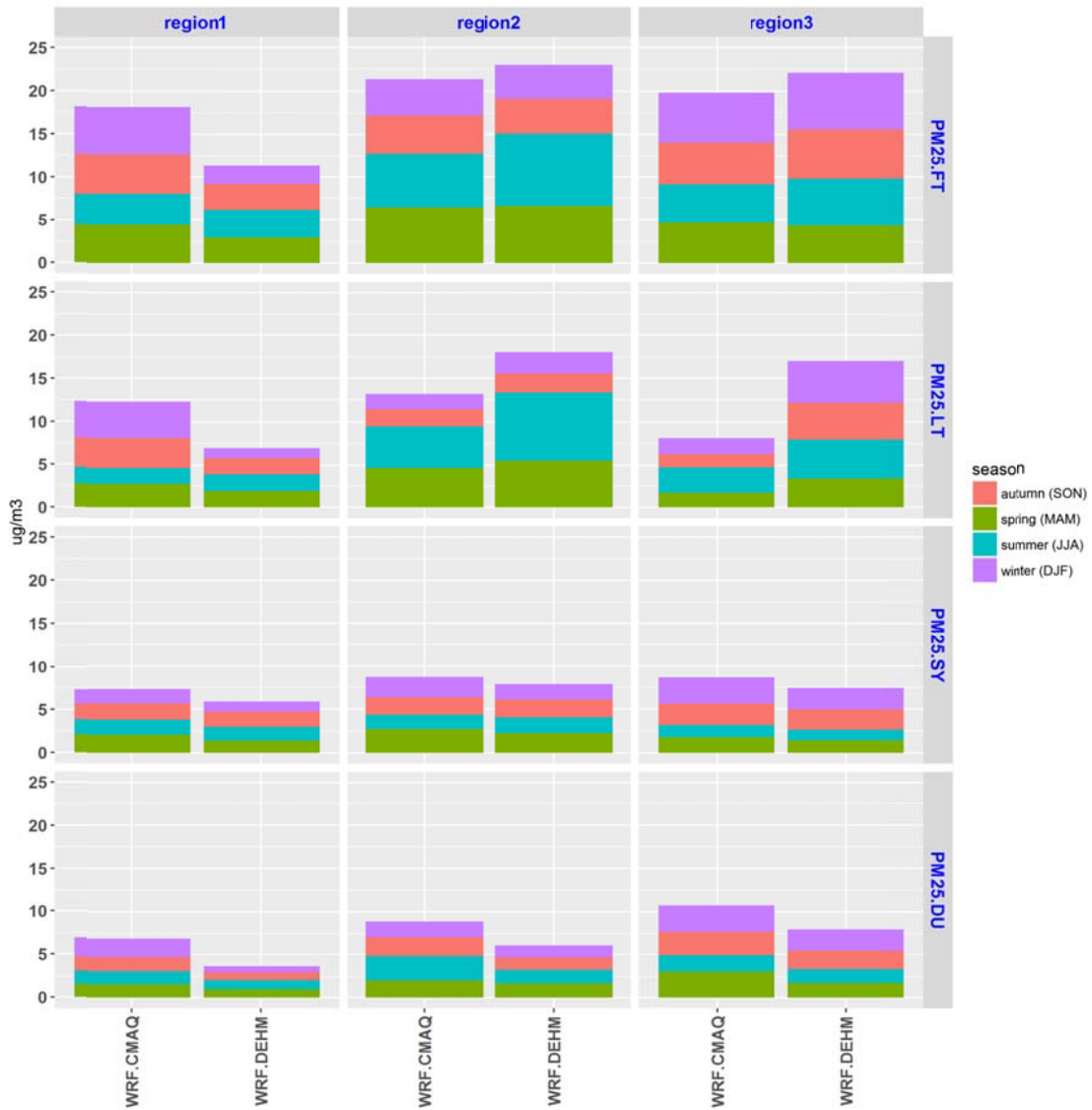
1941   a)

1942

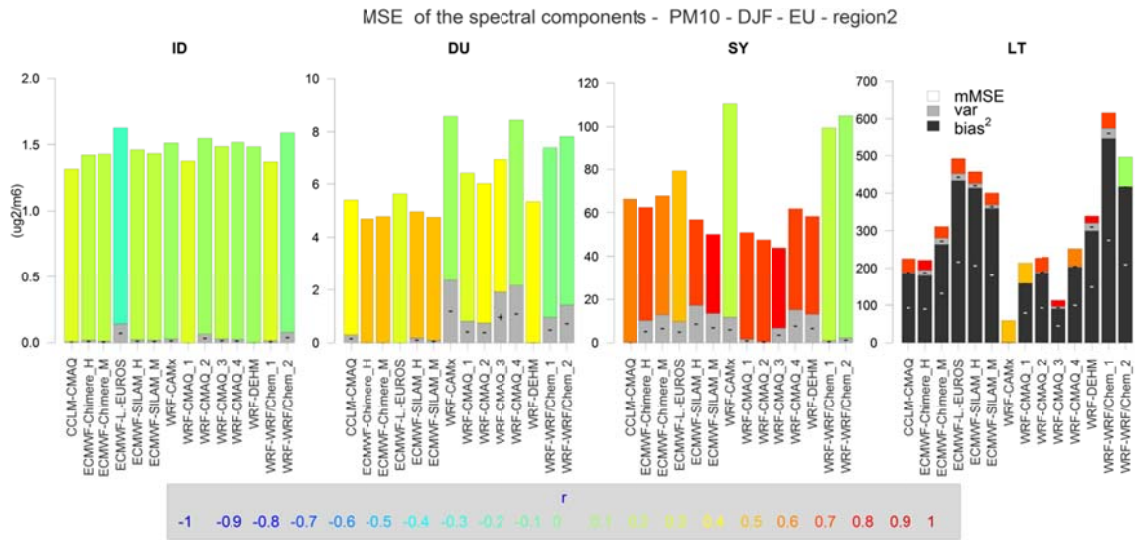
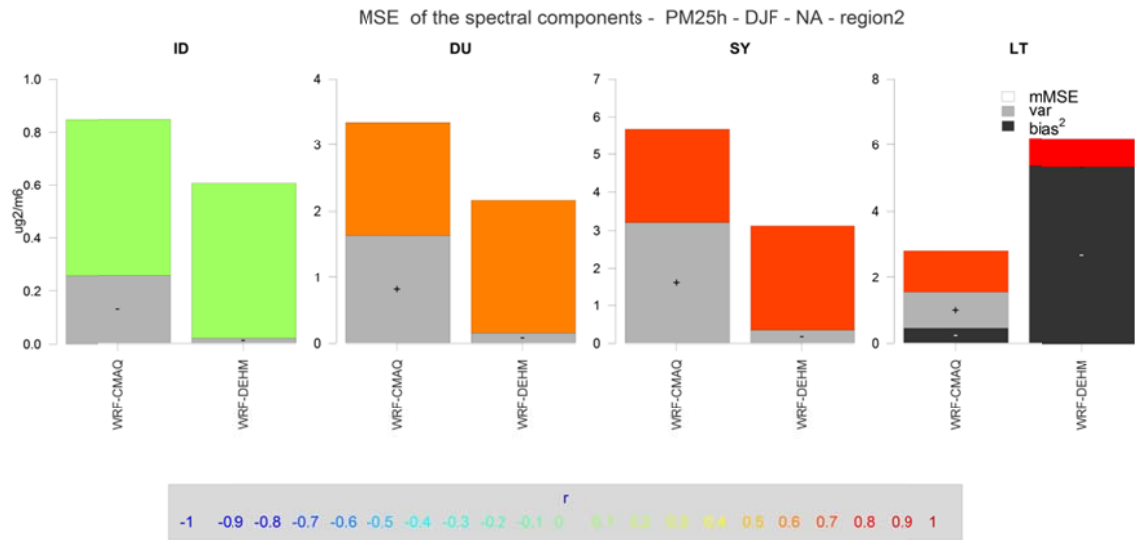
1943   b)

1944   **FIGURE 22. AS IN FIGURE 9 FOR SO₂**

1945

1946   a)

1947

1948   b)

1949   **FIGURE 23. AS IN FIGURE 10 FOR SO₂**

1951 **FIGURE 24. AS IN FIGURE 9 FOR PM$_{10}$ IN EUROPE (ERROR UNITS IN μg/m^3)**

1952

1953   FIGURE 25. AS IN FIGURE 9 FOR PM2.5 IN NORTH AMERICA (ERROR UNITS IN μg/m³)

 FIGURE 26. AS IN FIGURE 10 FOR PM$_{10}$ IN EUROPE (ERROR UNITS in μg/m^3)

 FIGURE 27. **AS IN FIGURE 10 FOR HOURLY PM$_{2.5}$ IN NORTH AMERICA (ERROR UNITS in μg/m^3)**

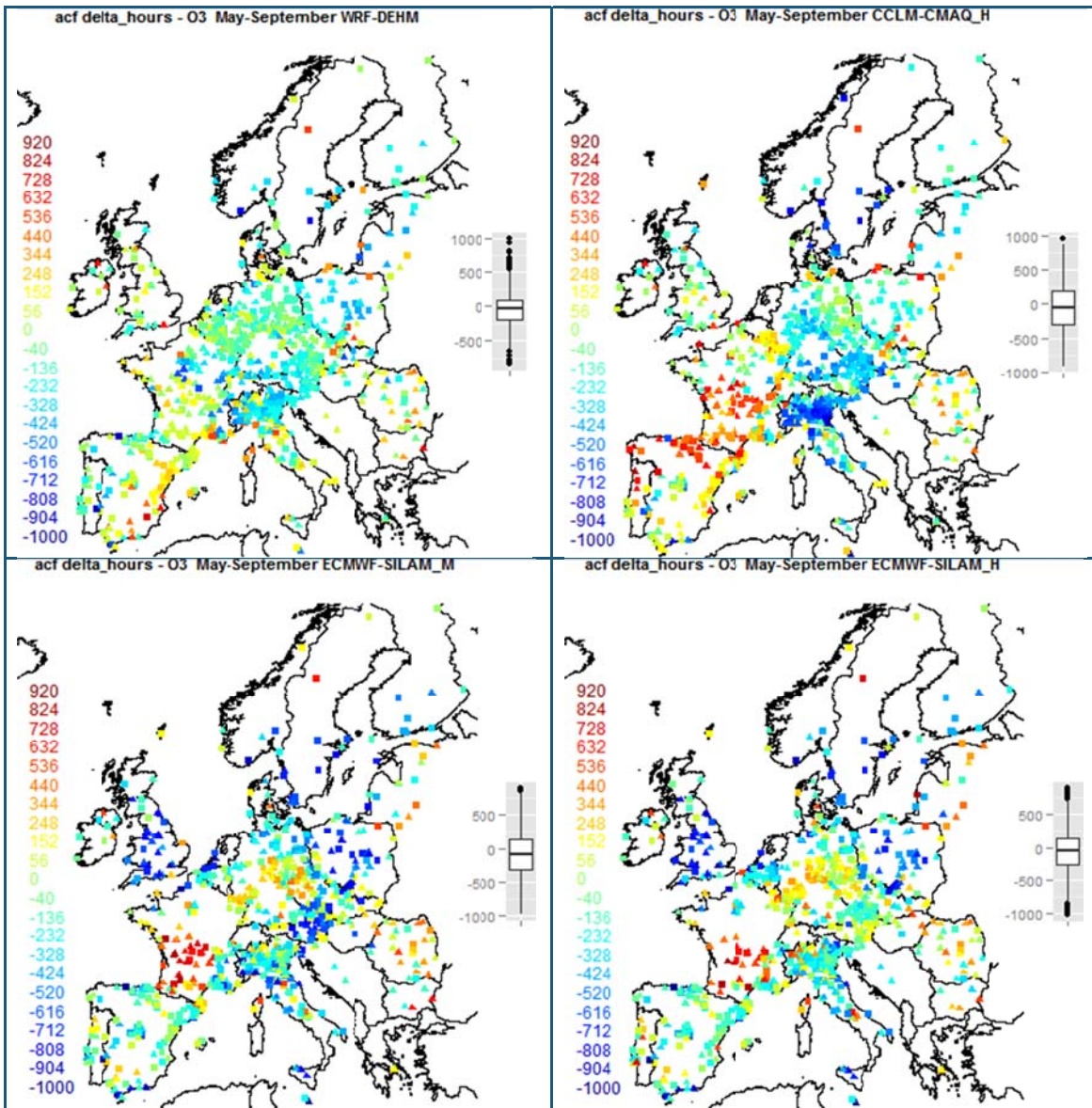
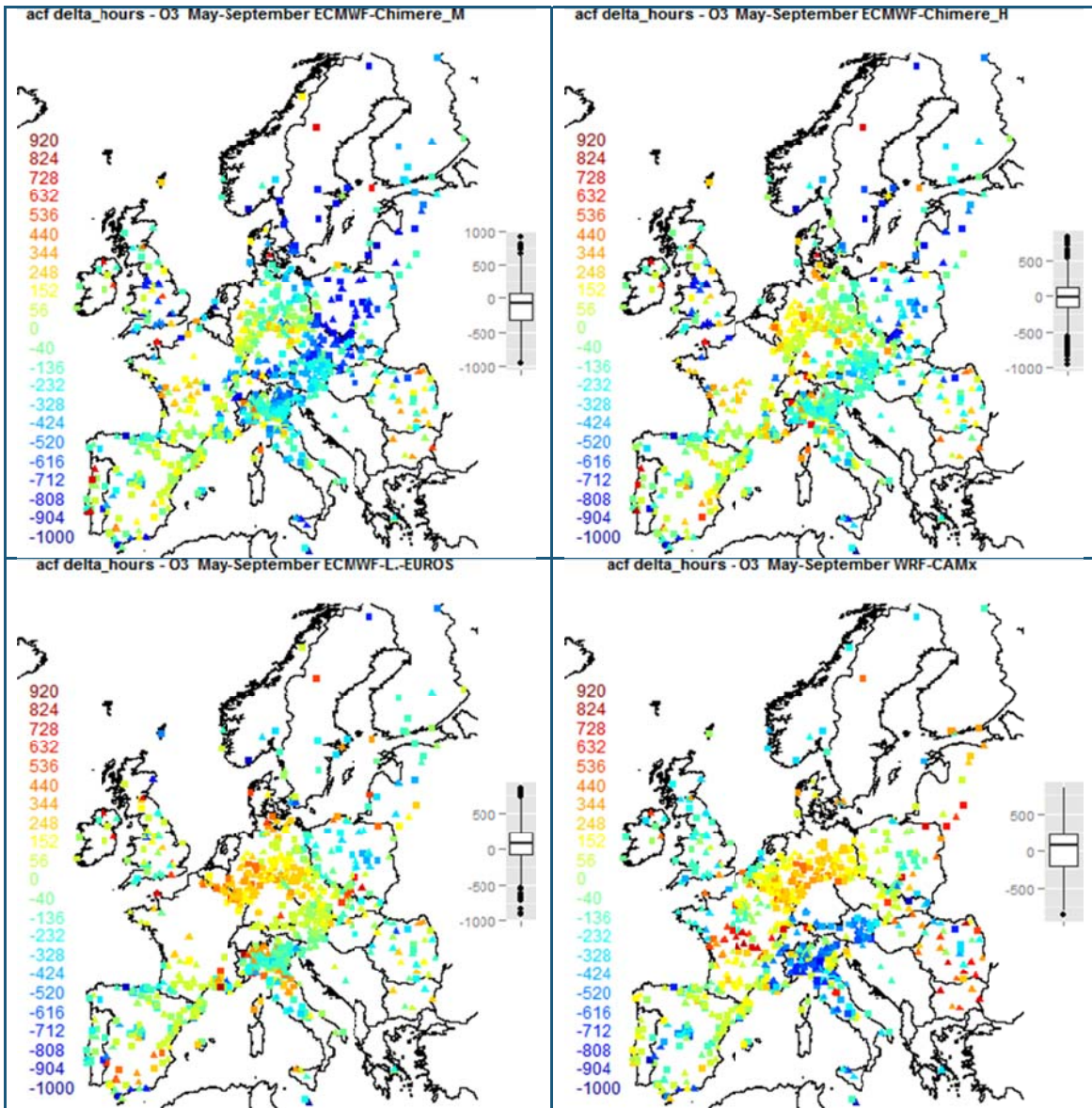
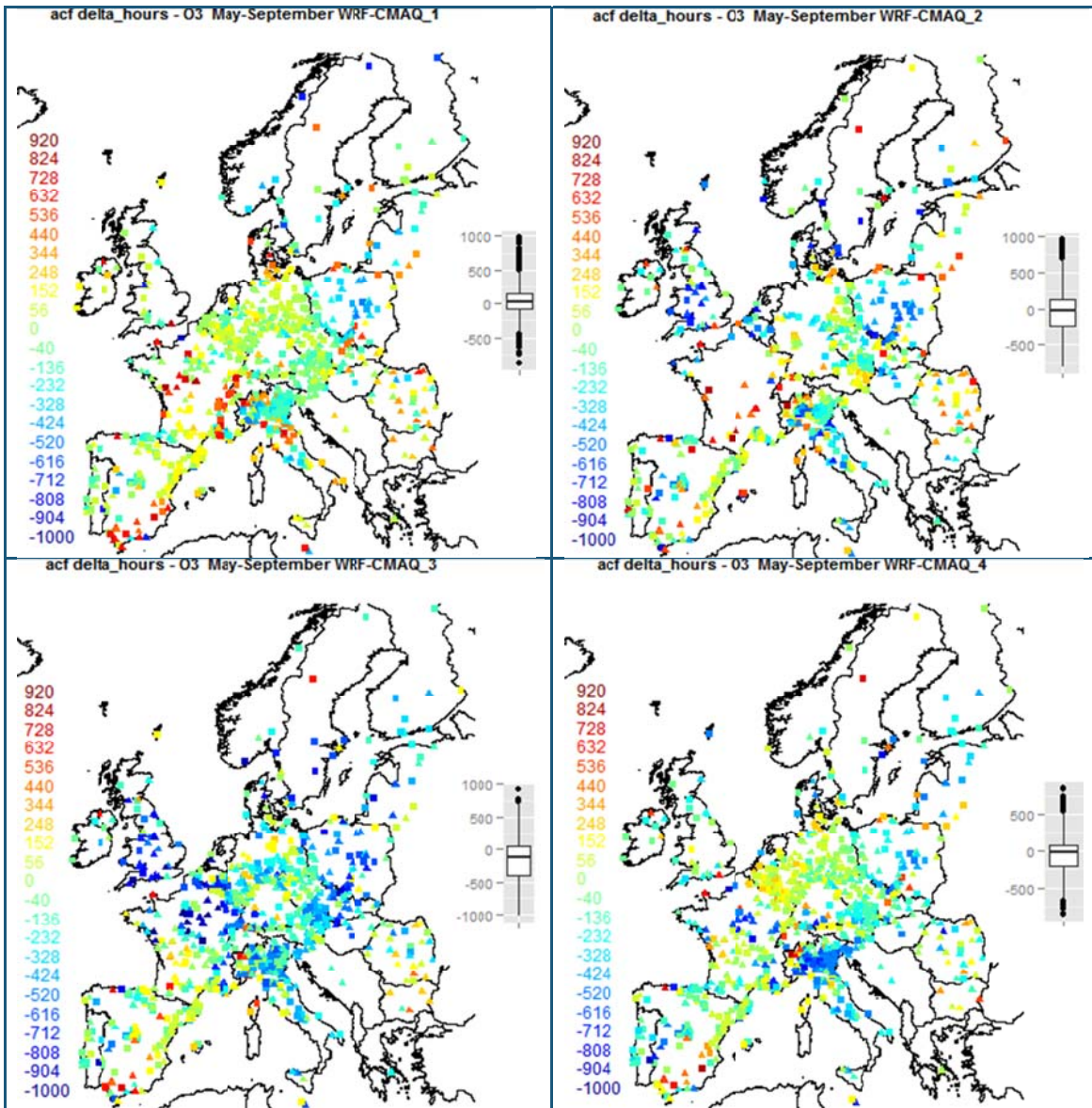
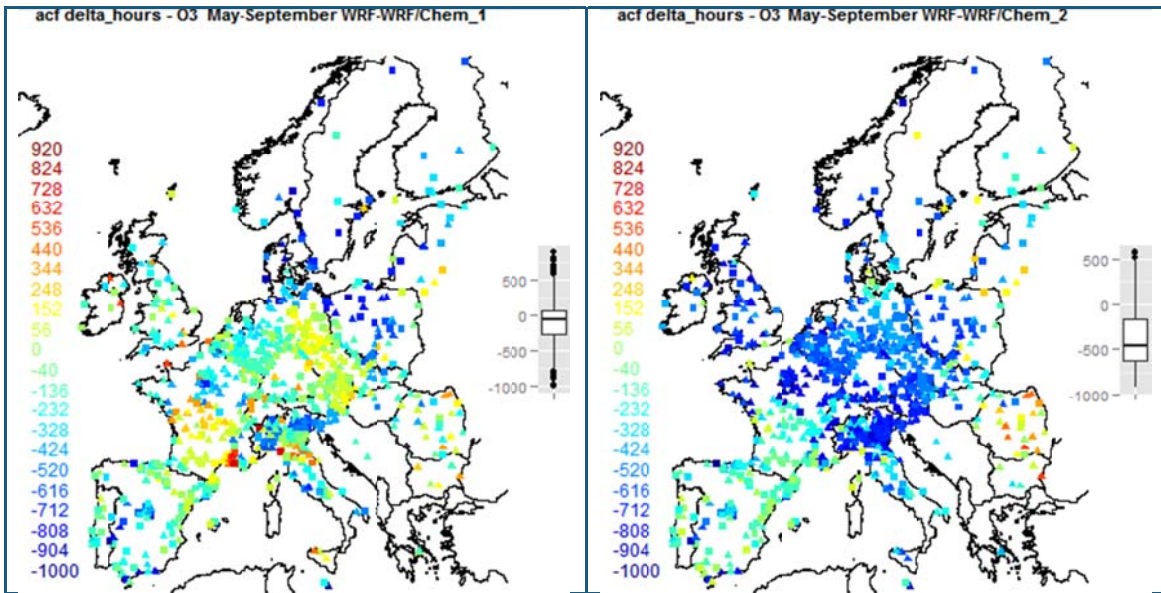



FIGURE 28. SPATIAL MAP OF THE OZONE MONITORING STATIONS COLORED BASED ON THE 'DELTA HOUR' VALUES, I.E. THE DIFFERENCE IN HOURS BETWEEN THE ZERO OF THE AUTOCORRELATION FUNCTION (ACF) FOR THE MODELLED OZONE MINUS THE ZERO OF THE ACF OF THE OBSERVED ONE. THE ACF IS CALCULATED ON THE LONG TERM COMPONENT 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