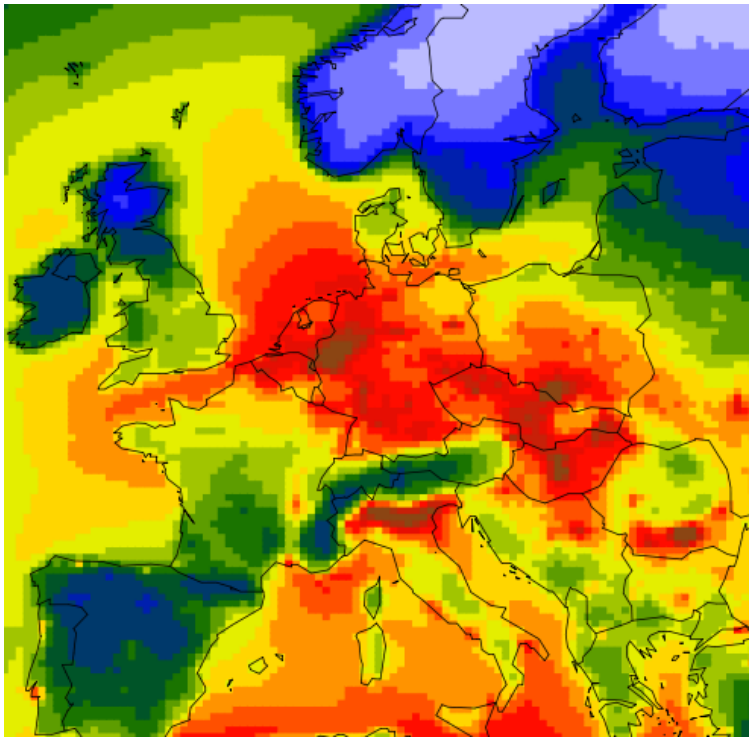
Long term air quality trends in Europe

Contribution of meteorological variability, natural factors and emissions



ETC/ACM Technical Paper 2016/7
January 2017

Augustin Colette, Sverre Solberg,
Maxime Beauchamp, Bertrand Bessagnet, Laure Malherbe, Cristina Guerreiro

The Eurodelta-Trends Modelling Team: A. Andersson, A. Colette, C. Cuvelier, A. Manders, K.A. Mar, M. Mircea, M.T. Pay, V. Raffort, S. Tsyro, M. Adani, R. Bergström, B. Bessagnet, G. Briganti, A. Cappelletti, F. Couvidat, M. D'Isidoro, H. Fagerli, N. Ojha, N. Otero, P. Wind

European Topic Centre
on Air Pollution and
Climate Change Mitigation

The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM)
is a consortium of European institutes under contract of the European Environment Agency
RIVM Aether CHMI CSIC EMISIA INERIS NILU ÖKO-Institut ÖKO-Recherche PBL UAB UBA-V VITO 4Sfera

Front page picture:
PM10 concentrations in the Eurodelta-Trends Ensemble in 1990, see Figure 34 of the present report

Author affiliation:
Augustin Colette, Maxime Beauchamp, Bertrand Bessagnet, Laure Malherbe: INERIS
Sverre Solberg, Cristina Guerreiro: NILU
The Eurodelta-Trends Modelling Team : A. Colette, F. Couvidat, B. Bessagnet (INERIS), M.T. Pay (BSC), S. Tsyro, H Fagerli, P. Wind (Met Norway), C. Cuvelier (JRC), A. Manders (TNO), A. Andersson, R. Bergström (SMHI), M. Mircea, G. Briganti, A. Cappelletti, M. Adani, M. D'Isidoro (ENEA), V. Raffort (CEREA), K.A. Mar, N. Otero, N. Ojha (IASS)

DISCLAIMER

ETC/ACM Technical Paper 2016/6
European Topic Centre on Air Pollution and Climate Change Mitigation
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone +31 30 2748562
Fax +31 30 2744433
Email etcacm@rivm.nl
Website http://acm.eionet.europa.eu/

Contents

1 Executive summary

This report follows up on the work on long-term trends in particulate matter and surface ozone achieved within previous EEA and ETC/ACM studies (EEA, 2009; Colette et al., 2015). The main focus is on long-term trends in observational data and how they could be explained and attributed to separate drivers by use of the model results and analyses available from the Eurodelta-Trends project (Colette et al., 2016b). The model analyses are based on all the models having delivered results at the time the report was initiated: CHIMERE, CMAQ, EMEP, MATCH, MINNI, LOTOS-EUROS, POLAIR, and WRF-CHEM. The report also presents available results from other recent and ongoing projects (in particular the EMEP TFMM Trend Assessment Report (Colette et al., 2016a) and the IGAC Tropospheric Ozone Assessment Report).

1.1 Observations and indicators

The time period addressed here is 1990-2010. This gives certain limitations to the monitoring network and regional coverage as also discussed in the EMEP TFMM Trend Assessment Report even if additional observations from the EEA e-reporting database are included for the period post 2000. For ozone there is a strong bias in the monitoring network when requiring sufficient data capture during the trend period. Some regions/countries are well covered (the United Kingdom, Ireland, Germany, the Benelux, Austria, Switzerland and Scandinavia), whereas in other regions, particularly in the south and the east part of Europe, there are no stations with sufficient number of years with data. For Particulate Matter (PM_{10} and $PM_{2.5}$), a satisfactory coverage of the monitoring network can only be found after 2000, even if some measurements of the inorganic fraction are available at EMEP sites in the 1990s.

For ozone, the selection of the indicator is essential since the long-term trends differ significantly for the various percentiles of the ozone distribution. It is furthermore important to note that some of the indicators the ozone directive and other guidelines (e.g. WHO) are based on, like AOT40 and SOMO35, are not particularly well suited for assessing the effectiveness of abatement policies. Accumulated exposure indicators like SOMO35 and AOT40 are known to be difficult in this respect since they are very sensitive to the hemispheric background ozone level which is just around the threshold values of 35-40 ppb. For most of our analyses we instead used the annual 4th highest max daily running 8h mean (4hMDA8). We consider this one is a better indicator of the annual photochemical episodicity and thereby it reflects the influence of the European anthropogenic emissions better than the exposure indicators mentioned above. We do, however, also show results for the regulatory and exposure indicators in this report.

1.2 Model performances

The comparison of model and observations for ozone for the complete 1990-2010 time period is based on the three models that delivered continuous results at the time the present report was initiated: CHIMERE, EMEP and LOTOS-EUROS (the other models only participated to time slice simulations of the Eurodelta-Trend exercise, or delivered trend results later in the process, as MATCH and MINNI, that were taken into account in the ensemble maps, see 1.3). As a main result, we find that based on all data from the long-term running monitoring stations and the Eurodelta-Trends model results there is a general downward trend in both the

observations and the model results for several ozone indicators during the 1990-2010 period. However, as reported in previous studies, the magnitude of the observational trends is in general smaller than what the models predict and the reason for this discrepancy is not clear. In general, the modelled peaks are higher than the observational peaks in the first part of the 1990s and the other way around in the last part of the 2000s.

Firstly, all three models show an increase in performance when compared to the ozone measurement data during the period 1990-2010. The model bias is reduced and the accuracy is improved with time, and these trends in model performance are statistically significant. There are several possible reasons for such a trend: more realistic emission data or more reliable measurements in the 2000s, chemical (e.g. NOx titration) or physical processes (e.g. stratospheric influence) not well captured by the models, etc. This improvement in model performances is particularly strong in the Mid/East Europe and North Italy regions, whereas we found no significant trends in model performance in the regions to the north (Inflow, England and Scandinavia). The trend in model performance is seen to be largely driven by poor performances in the beginning of the 1990s. An in-depth analyses of model performance is needed to clarify this, but that was beyond the scope of the present report.

Further, regional characteristics are evident. England is the region with the strongest decline in the ozone peaks and exposure indicators and also the region with best agreement between model predictions and observations and, additionally, also a region with the steadiest model performance. Thus, for England, the modelled and observed data strongly supports that the European emission abatement during 1990-2010 has led to a substantial reduction of high surface ozone levels and that the magnitude of these reductions are well explained by state-of-the-art models. However, for some other regions, there are marked differences between the observed and modelled values, particularly for the first part of the 1990s. This applies particularly to the Mid-Europe region and central Europe where the observational data show smaller trends than modelled for all indicators and all models. The reason for these regional differences is not clear and would require more detailed investigations of the representativity of monitoring data, model processes etc. Unfortunately, the monitoring network in Southern part of Europe exposed to high summertime ozone episodes does not offer a good enough temporal coverage over the 20 year period to be compared with the models.

Thirdly, the three models studied (EMEP, CHIMERE and LOTOS-EUROS) show systematic differences when compared to each other and when compared to the observations.

The models exhibit the usual underestimation of PM_{10} and to a lesser extent for $PM_{2.5}$, especially in winter. The multi-model ensemble performs best compared to any individual model. The ensemble spread is largest in areas where PM loads are small, and spread remains smaller than reported in other recent multi-model intercomparison projects. There is no substantial variation in terms of model variability for total PM load, but long terms trends can be found for specific areas and species, such as nitrate over the Iberian Peninsula where model agreement increased significantly between 1990 and 2010. There are no trends in model performances either for PM_{10} and the low bias of the models is quite constant in time. Although all models underestimate PM_{10} concentration, this lack of drift in performances is an important feature to build support for their use in a trend assessment.

The comparison between model results and observations show that the sign of the trend is well captured when looking only at sites where the observed trend is significant, with as much as 100% rate of good detection of the sign of the trend for all models. The range of

variability of the trend is however underestimated in the models compared to observations but only a detailed investigation of individual sites can explain this feature.

The maps of trends in the multi-model ensemble shows that the area where the modelled trend is significant is larger than in the observations. But on average, the models perform very well in capturing the relative decrease of PM_{10} concentrations. Even if the models display an underestimation, this bias is quite constant in time, so that the average relative change in PM_{10} over the observations included in the Air quality e-reporting database is well captured by the Eurodelta-Trend ensemble.

1.3 Ensemble maps of air pollution in 1990, 2000, and 2010

Based on the Eurodelta-Trends ensemble of eight models having produced simulations for 1990, 2000 and 2010, we present composite maps for ozone peaks and annual mean PM_{10} by computing the multi-model median at each grid point. Those composites can be put in perspective with the maps of differences 2000-1990 and 2010-2000 and the maps of 11 year trends over both periods (only for the five models that produced the continuous 21 year simulations). They confirm the steady decrease of ozone peaks in the 1990s and the 2000s and the important reduction of annual mean PM_{10} over continental Europe (especially during the first decade).

1.4 Attribution of the drivers of air pollution changes

Beyond model evaluation, the EuroDelta-Trends modelling experiment was specifically designed to perform an attribution of the relative contribution of (i) European Anthropogenic Emission, (ii) intercontinental inflow, and (iii) meteorological variability to the net air pollution trends over the 1990-2010 period. Emissions and intercontinental inflow can be addressed with different combinations of 1-yr simulations performed by all modelling teams (8 CTMs), while the impact of meteorological variability is derived from two 21-yr simulations performed by 5 modelling groups. An adequate combination of all available information, accompanied by a thorough discussion of the possible caveats of the approach (including an assessment of model uncertainty, quantification of possible interactions, and an alternative methodology based on ozone climatologies), allow to complete the attribution study.

Based on these calculations, we could confirm that the impact of the reduction of European anthropogenic emissions dominate the modelled net reduction in ambient air pollution whereas the background (inflow) conditions have a much smaller influence. The annual mean ozone level over England in the 1990s is an exception to this, though, because of the titration effect with reduced NOx. We confirm the penalty brought about by the intercontinental inflow in the 1990s for annual mean ozone where boundary conditions were found to contribute to an increase of the ozone background. On the contrary, the inflow was found to contribute to decrease annual mean ozone over the 2000s. More surprising is the magnitude of the role of meteorological variability which exceeded the penalty of intercontinental inflow in the 1990s for annual mean ozone.

The summertime ozone peak episodes are caused by the anthropogenic emissions (of NOx and VOC) and at the same time strongly linked to the weather conditions and favoured by episodes of warm, stagnant high-pressure. Since the occurrence of such episodes varies considerably from year to year, the long-term trend in ozone could be driven by a

combination of trends in emissions and in the occurrence of such heat waves. Our analyses indicate that the downward trend in ozone episodes during 1990-2010 was mainly driven by the emission changes, whereas trends in meteorology lead to an additional ozone reduction in the 1990s while for the 2000s the influence of the meteorological variability was smaller.

For particulate matter (annual mean PM_{10}), the impact of emission changes clearly dominates the response. Intercontinental inflow was negligible in the 1990s but some impact is found in the 2000s. Meteorological variability is found to have an impact on PM_{10} that can match the order of magnitude it has on ozone for some regions, eventually leading to an increase of PM_{10} concentration in the 2000s.

The modelled attribution for each aerosol compounds contributing to the PM_{10} mix shows that European anthropogenic emissions dominate for secondary inorganic and organic aerosols. The changes in desert dust and sea-salt are primarily attributed to meteorological variability, their impacts are strongest for the Iberian Peninsula and the Mediterranean (desert dust) and Inflow, England and Scandinavia (sea salt), although it remains below 20% of the total PM_{10} change.

2 Scope and structure

With the important development of air quality monitoring networks in Europe since the 1990s, it is timely to engage into assessments of the long-term evolution of air pollution, in particular to relate this evolution with the efforts devoted to mitigate emission of air pollutants and conclude on the effectiveness of air quality policies.

The present report takes stock of recent and still ongoing trend assessment initiatives, in particular the ETC/ACM Technical Report on trends detected in European routine monitoring network (Colette et al., 2015), as well as the Trend Report of the Task Force on Measurement and Modelling of the LRTAP Convention (Colette et al., 2016a) and the ongoing Tropospheric Ozone Assessment Report.

In order to relate the reported observed trends in air pollution to emission reduction strategies it is essential to take into account the fate of air pollutants in the atmosphere through their atmospheric transport and transformation. Chemistry-Transport models are required to represent these processes and assess ambient air pollutant trends in relation to the evolution of anthropogenic emissions.

This is because the long-term evolution of air quality is driven by a variety of factors; changes in the emissions on a national, regional and intercontinental basis, changes in the atmospheric chemistry, natural variations in meteorology and the interplay between these factors. The meteorological year-to-year variations could, however, influence the atmospheric concentrations of key pollutants like PM and ozone in several ways and thereby also the atmospheric trends. In winter, the occurrence of episodes with cold, stagnant weather situations would typically lead to higher emissions of PM from household heating and traffic and additionally, the surface PM levels would be increased by reduced atmospheric mixing and a shallow boundary layer. Likewise, the occurrence of heat waves in the summer season would promote atmospheric photochemistry leading to enhanced levels of ozone and secondary PM, and could furthermore lead to stronger biogenic emissions adding to the increased levels of ozone. Thus, there is a close link between the day-to-day weather situations and the observed levels of pollutants. Long-term trends in the concentration of PM and ozone would thereby not only reflect the trends in anthropogenic emissions but could also reflect the variability in meteorology during the period studied.

Although empirical relationships between certain meteorological parameters and observations of pollutant concentration levels could be found (Barmpadimos et al., 2011; Rasmussen et al., 2013), such relationships are typically not generic and thus not easy to extrapolate to larger regions or longer time periods. In practice, such relationships should be regarded as a "poor man's model". On the other hand, more and more sophisticated CTMs (chemical transport models) have been developed during the last 30 years giving the quantitative relationships between emissions and concentrations in a deterministic way. CTMs are essential for assessing the relative contributions from different processes to the levels of key pollutants in the atmosphere.

This report is focused on the analyses of modelled and observed long-term trends in PM and surface ozone during the period 1990-2010. It is built upon recent work done within EMEP TFMM assessment and the previous years' ETC/ACM studies for the EEA on trends. It has

been an aim to extract and present main results from ongoing or recently finalized research activities, e.g. with EMEP, Eurodelta, the IGAC TOAR project etc. as well as doing own research based on existing modelling and observational data. The report is structured in the following way:

In Section 3, a brief summary of the main findings of the recent and ongoing air pollution Trend Assessment Reports relevant to the European region (ETC/ACM, TFMM and TOAR) is provided before engaging in the comparison between observed and modelled trends. Section 4 is devoted to the various methods used in the report: the Eurodelta-Trends modelling experiment (4.1), the monitoring data which this work is based on, including station network, data selection criteria, etc. (4.2), the choice of air pollution indicators (4.3) as well as statistical approaches (4.4, 4.5, 4.6). The results concerning the general model performance for different models and indicators and the comparison between the calculated long-term trends as seen by the models and the measurements are presented in Section 5. Composite maps of the modelled air pollution in 1990, 2000 and 2010 as well as its evolution are presented in Section 6. The attribution study focusing on the quantification of the respective role of European anthropogenic emission changes, meteorological variability and intercontinental pollution is discussed in Section 7.

3 Main findings of recent trend assessment reports

3.1 ETC/ACM and TFMM/EMEP Reports: key findings

A thorough assessment of trends in air quality measurements in different types of stations all over Europe (urban, suburban and rural background; traffic; and industrial sites) was provided by the ETC/ACM in 2015 (Colette et al., 2015). The report was focused on the 1990-2012 time period for NO_2 and O_3 trends, and the post-2000 time period for particulate matter, as most of the PM stations in the EEA´s Air Quality e-reporting database[1] were limited to that time period.

In 2016, a Trend Assessment was published by CLRTAP EMEP´s TFMM focusing on EMEP background sites (with complementary information for other station types extracted from the aforementioned ETC/ACM Report). This study included more details on the evolution of ozone for a variety of indicators, as well as some information on the evolution in the period 1990-2012 of secondary inorganic aerosols and heavy metals and persistent organic pollutants.

For ozone trends, the TFMM report highlighted for EMEP background sites a relatively flat trend for annual ozone mean (1990-2012) but a steady 10% reduction in the amplitude of peaks over the period (Figure 1). A 20% reduction in the number of 50 ppb exceedances was reported over the 1990-2012 period and reductions as high as 30% and 37% for SOMO35 and AOT40 were found over the 2002-2012 period. The fraction of EMEP sites where the trend is statistically significant/insignificant increasing/decreasing for the 1990-2001 and 2002-2012 periods is presented in Figure 2. Trends are considered statistically significant when the p-value of their Mann-Kendall statistic is lower than 0.05 and it should be emphasized that insignificant trends should really not be considered as either increasing or decreasing (despite the matching colour in the barplot). The absence of significance in the statistical test shows that adding or removing a single year in the time series might give a very different result. The trends in a number of indicators are provided on the figure: annual mean, SOMO35, AOT40, number of days with the maximum daily 8-hours mean concentration (MDA8) larger than 50 and 60 ppb, respectively, as well as the 4th highest and max MDA8.

Figure 2 shows that most of trends in the various indicators where not statistically significant. In the first period, 1990-2001, the results indicate that significant increases are observed for annual mean O_3 concentration, SOMO35, AOT40 and the number of days with a MDA8 exceeding 50 ppb, whereas extreme indicators like the maximum or 4th highest MDA8 was showing a higher fraction of downward trends. This presumably is related to the fact that SOMO35 and AOT40 reflect, to a larger extent, changes in the hemispheric baseline level whereas the extreme indicators reflect more the photochemical ozone formation and high pollution episodes.

[1] http://www.eea.europa.eu/data-and-maps/data/aqereporting-1, the former AirBase.

Up to 20% of the EMEP sites experienced a statistically significant increase of annual mean ozone in the 1990s but for the other three ozone indicators with statistically significant increase, such increases were observed at less than 10% of the sites. In the 2000's, a statistically significant increase was only observed at one site and only for the annual mean indicator. These features were however undermined by a relatively large fraction of sites where the trend was still statistically non-significant because of the relatively short span of the time period, and the high interannual variability of ozone. Similar features were emphasized in the ETC/ACM study for rural sites, but that report pointed out that the trends at background urban sites appeared to be more frequently positive.

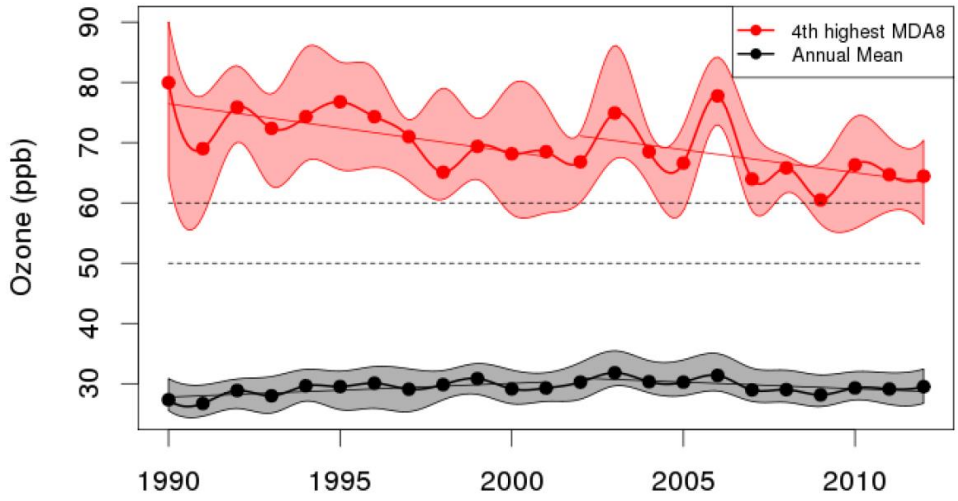


Figure 1: Composite of annual mean ozone (black) and 4th highest MDA8 (red) ozone recorded at 55 EMEP rural monitoring sites between 1990 and 2012. The thick line is the network-wide annual median and lower/higher bounds of the shaded areas are for the 25th and 75th percentiles. Thin straight lines show the linear trend over the 1990-2001 and 2002-2012 periods and dashed lines indicate the WHO air quality guideline (50ppb) and the EU long term objective (60ppb) reference concentrations. Source TFMM Report 2016, (Colette et al., 2016a)

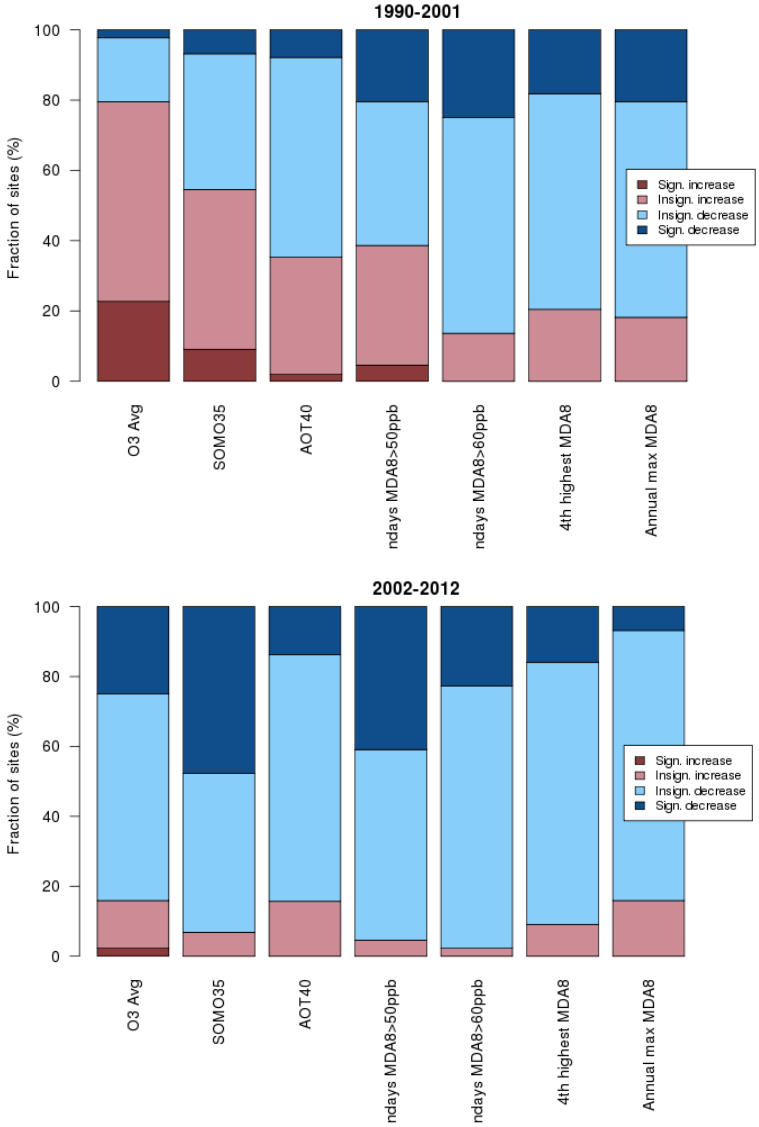


Figure 2: Percentage of EMEP monitoring sites where statistically significant upward trends (dark red), insignificant upward trends (light red), insignificant downward trends (light blue) and significant downward trends (dark blue) were observed in the 1990s (top) and in the 2000s (bottom) for each ozone indicator (O3 Avg: ozone annual mean, SOMO35 and AOT40, ndays MDA8>50 and 60ppb: number of days per year where the O_3 MDA8 is above the 50ppb (WHO) or 60ppb (European) threshold, fourth highest MDA8 day and annual maximum of MDA8). Source TFMM Report 2016, (Colette et al., 2016a)

For particulate matter (PM_{10}), both ETC/ACM and TFMM reports concluded that, in the period 2001-2012, the average rate of decrease in relative terms was of the order of 30% for background sites, with a limited spatial variability. As shown in Figure 3, the trend was more pronounced close to the sources (traffic, industrial and urban background sites) than in rural areas. Finally, the decreasing trend is bigger in summer compared to winter for all types of stations except industrial ones, where trends do not vary throughout the year.

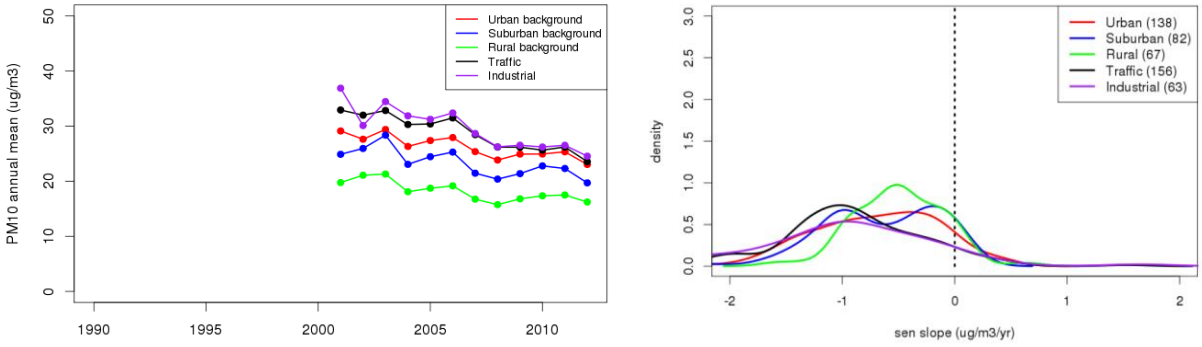


Figure 3: Average PM_{10} time series (left) and distribution of the trends (right) over the monitoring network for the 2001-2012 time period for each station type. Source ETC/ACM 2015 Technical Report (Colette et al., 2015)

For secondary inorganic aerosols and deposition, high reduction rates of 60 to 90% for sulphur compounds were reported in the TFMM Assessment, while the reduction for nitrogen compounds was of the order of 30 to 40% over the 1990 to 2012 period. In agreement with the ETC/ACM Technical Report, 30% reduction rates were found for PM_{10} and $PM_{2.5}$ over the 2000-2012 time period.

3.2 Recent literature studies

A detailed review of scientific papers related to ozone trends and detrending with respect to meteorology was given in (Solberg. et al., 2015). There is, in addition, apparently, a strong effort on ozone trend studies in the scientific community at the moment, and a number of publications have become available since the scoping report. A number of these are revised in the following.

(Watson et al., 2015) looked at the meteorological forcings on ozone and NO_2 within various CTMs, including EMEP and CHIMERE, as part of the IMPACT2C project. They found marked differences in the model performance for the various models and linked that to differences in temperature, PBL height and surface deposition.

(Lefohn et al., 2014) looked at background surface ozone concentrations in the United States based on model-derived source apportionment methods. They applied coupled global and regional CTMs with backward trajectories and surface observations. They found that the so-called emissions-influenced background ozone contributed a major portion for most sites in their analysis, especially during the spring. Their findings were consistent with other studies showing underestimations of the contribution of stratospheric ozone to surface monitoring sites.

(Katragkou et al., 2015) made an evaluation of ozone data based on the MACC reanalyses (Flemming et al., 2015). They highlighted particularly a shortfall of the models to capture the spring peak in ozone at background sites.

(Henneman et al., 2015) published a study on the meteorological detrending of ozone, PM and primary species (NOx, CO and SO_2) for the period of 2000-2012 on data from Atlanta. They used empirical methods to detrend the pollutant signals into long-term, seasonal,

weekly and short-term components. They found a strong increase in the meteorological detrended winter ozone levels whereas for summer the methods revealed only a slight decrease.

(Santurtún et al., 2015) looked at the trends in surface ozone levels in Spain and the relationship with meteorology for the period 2001-2010. They firstly screened the ozone data on local influence, ending up with 9 sites, and then looked at the relationship with daily circulation weather types. Only stations which were free from the direct influence of local point sources and thus representative of background boundary layer air was included. They found indications for an increase in ozone throughout all seasons to a varying degree for the various weather regimes.

In the framework of the CityZen FP7 project, Colette et al. (2011) investigated ozone urban, suburban, and rural observed trends over 705 Air Quality e-reporting database background stations passing quality screening over the 1998-2007 decade. They apply a similar deseasonalised Mann-Kendall and Sen-Theil slope method with the more stringent criteria for detecting significant trends of a p-value smaller than 0.05. The estimated average daily mean ozone trend at urban background, suburban background and rural background sites was 0.37, 0.27 and 0.05 µg m^{-3} yr^{-1}, respectively. The proportion of sites where the ozone trend was found to be significantly positive was 30.8% when considering daily means but this number dropped to 18.5% when considering O_3 daily peaks, reflecting the larger efficiency of mitigation strategies for ozone peaks than for baseline ozone.

3.3 The TOAR project

The ongoing project TOAR (Tropospheric ozone assessment report) is a science effort initiated by the International Global Atmospheric Chemistry (IGAC) project with the aim to provide an up-to-date scientific assessment of the global tropospheric ozone distribution including the surface and the free troposphere. The project is coordinated by Owen Cooper from NOAA, with the support of Jülich Forschungs Zentrum in compiling a global database of ozone observation and related indicators. The work within TOAR is focused on 7 comprehensive scientific chapters and one summary chapter being prepared on a voluntarily basis by researchers in the scientific community:

- Chapter 1: Critical Review of the present day and near future tropospheric ozone budget, Archibald, et al.
- Chapter 2: Tropospheric ozone observations, Tarasick et al.
- Chapter 3: The description of global ozone metrics for climate change, human health, and crop/ecosystem research, Lefohn, et al.
- Chapter 4: Present day ozone distribution and trends relevant to human health, Fleming et al.
- Chapter 5: Present day tropospheric ozone distribution and trends relevant to vegetation, Mills et al.
- Chapter 6: Present day distribution and trends of tropospheric ozone relevant to climate and global model evaluation, Gaudel et al.
- Chapter 7: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Young et al.
- Chapter 8: The Tropospheric Ozone Assessment Report (TOAR): Key findings and recommendations for future research, Cooper et al.

At the moment of drafting this report, the TOAR work was only presented as non-citeable draft reports although several chapters existed in fairly mature versions.

The main part of TOAR is concentrated on observational based studies on a global basis. A separate discussion/review chapter is devoted to ozone modelling and model performance but the TOAR project itself does not include actual model exercises.

4 Methods: measurements and models

4.1 Eurodelta-Trends: a Multi-Model Air Pollution Trend Ensemble

Chemistry-transport models (CTM) are an essential tool to explore the relative impact of key driving factors such as emission or meteorological changes on recent air pollution trends. Individual modelling teams have engaged into such endeavour, for instance (Wilson et al., 2012) in the Geomon project or (Banzhaf et al., 2015). But considering the uncertainties in chemistry-transport modelling, ensemble of models are considered more robust to assess such trends, as done for instance in the European Projects CityZen (Colette et al., 2011).

An up-to-date multi-model trend ensemble was recently designed by the EMEP Task Force on Measurement and Modelling (TFMM). Through the various phases of the Eurodelta initiatives, the TFMM has proposed different aspects of CTM model benchmarking to the EMEP/MSC-W model supporting the works of the LRTAP convention. The ongoing Eurodelta-Trends (EDT) experiment offers a new way to explore the skill of CTMs in capturing air pollution trends, also paving the way for attribution studies allowing to quantify the respective role of the key driving factors: emission changes, intercontinental transport of air pollution, meteorological variability. A complete description of the experiment is available in (Colette et al., 2016b).

The time period covered by the experiment is 1990-2010. The beginning of the period is 1990 because that year serves as reference for the Gothenburg protocol. The end of the period is 2010 because of the availability of underlying forcing data (emission, boundary conditions and meteorology). The modelling domains is the larger European region and the model resolution is approximately 25 km.

In contrast with other model intercomparison initiatives, the EDT framework is quite constrained, with tight standards on the consistency of the setup of participating models in order to allow pointing out the geophysical processes bearing upon the modelled response. Identical trends in anthropogenic emissions are used by all models, they were provided by IIASA and computed for the GAINS model as part of the ECLIPSE project. More details on these emission can be found in (Amann et al., 2012) and (Klimont et al., 2016b; Klimont et al., 2016a). A single source of chemical boundary condition is also used. It consists of a simplified version of those used in the standard EMEP/MSC-W model (Simpson et al., 2012), and is based on in situ observation trends, in particular at the Mace-Head observatory, in Ireland. A common source of meteorological forcing is provided based on a dynamical downscaling of ECMWF reanalyses with the WRF model (Stegehuis et al., 2015).

The EDT model experiment is divided into a number of tiers, targeting various science and policy questions. Tier 1 regards model evaluation for three key years (1990, 2000, 2010) as well as sensitivity to emission changes. Tier 2 is devoted to sensitivity to boundary conditions for those 3 years. And tier 3 consists in the full 21-year trend for the 1990 to 2010 time period.

Eight European teams delivered results to the EDT database for at least one tier of experiment using the following models: Chimere (Menut et al., 2013), CMAQ (Byun and Schere, 2006), EMEP/MSC-W (Simpson et al., 2012), LOTOS-EUROS (Sauter et al., 2012;

Schaap et al., 2008), MATCH (Robertson et al., 1999), MINNI (Mircea et al., 2016), POLAIR (Mallet et al., 2007), and WRF-Chem (Mar et al., 2016; Grell et al., 2005).

In all the ozone comparisons between models and observations we filtered the model data based on the periods of missing observational data to ensure that these parallel time series covered the exact same hours. Furthermore, all model data, originally provided in $\mu g/m^3$ were converted to ppb by use of hourly time series of T and P extracted from the EMEP model data. The ozone observational data in the EMEP database are given in "proxy $\mu g/m^3$" defined as ppb multiplied with a fixed factor using a factor = 2.0.

4.2 Long term air pollution monitoring data

The main basis for calculating trends at monitoring sites in the present report is the EMEP observational data. In particular, we used a data set very similar to the one used within the EMEP TFMM study (Colette et al., 2016). In that data set the EMEP data were screened for data capture, requiring at least a 75 % data capture based on hourly concentrations in at least 75 % of the years in the period 1990-2010. Furthermore, in the model comparison for ozone we avoided mountain sites, defined as stations located above 1000 m asl, due to the problems of matching surface observational data with coarse scale model results in complex terrain. In addition, we did not use data from the Portuguese site PT0004, Monte Velho, due to considerations of data quality and representativity. These selection criteria were modified to fit the EuroDelta period 1990-2010 as opposed to the 1990-2012 period in the TFMM study. For ozone, Air Quality e-reporting database data have not been used in our study presented here. This is justified by the availability of monitoring data in the different databases. Regular reporting of ozone data to the EEA AQ database was initiated in 1997 implying too short timeseries of these data compared to our selection critera of 75 % data capture within 1990-2010. Older ozone data do exist in the EEA AQ database, but as shown in the TFMM study, these data correspond to the data already in the EMEP database. For Particulate Matter, we focus on the period post 2000, because measurements in the 1990s are very scarce anyhow (except for secondary inorganic aerosols). For the period post 2000, the EEA AQ e-reporting database includes much more monitors than the EMEP network, and also urban sites. Therefore, we will include those measurement in the analysis, following the selection criteria presented in (Colette et al., 2015) and keeping only background sites.

4.2.1 Definition of European subregions

Monitoring stations were grouped into nine subregions of Europe that are also used in the model analysis of the present report. Building upon the regional climatic zones originally defined in the PRUDENCE project (Christensen and Christensen, 2007) we propose 9 areas which are relevant to surface air quality trends:

- EN: England: 5W, 2E, 50N, 55N
- IN: Inflow: -10W, -5E, 50N, 60N, and 5W, 2E, 55N, 60N
- IP: Iberian Peninsula: 10W, 3E, 36, 44N
- FR: France: 5W, 5E, 44N, 50N
- ME: Mid-Europe: 2E, 16E, 48N, 55N
- SC: Scandinavia: 5E, 30E, 55N, 70N
- NI: North Italy: 5E, 16E, 44N, 48N
- MD: Mediterranean: 3E, 25E, 36N, 44N

- EA: Eastern Europe: 16E, 30E, 44N, 55N

These regions are presented in Figure 4. Compared to the original Prudence regions, the main differences are: (i) renaming "Alps" by "North Italy" since that region is really dominated by the Po Valley for an air pollution standpoint; (ii) split the "British Isles" between "England" which is a NOx saturated emission areas, and the rather clean oceanic "Inflow" regions of Ireland and Scotland. Note also an extension of the North Italy region 1 degree eastward to ensure we are not missing out any station.

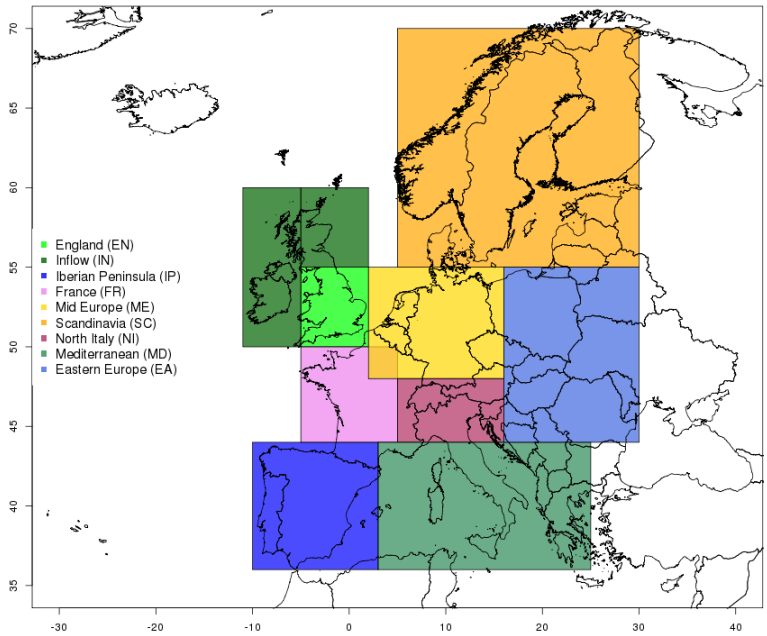


Figure 4: Map of the aggregation European subregions

4.2.2 Ozone

The data capture criteria for the 1990-2010 period imposes a strong geographical bias to the network as shown in Figure 5. This shows the number of years with observational data for the EMEP station network. As seen from this figure, a criterion of at least 75 % capture, i.e. 15 years or more of data, limits the station network to the sites marked with red in Figure 5 covering Northern and Central parts of Europe only with no sites in the southern/Mediterranean area (as PT0004 was not considered).

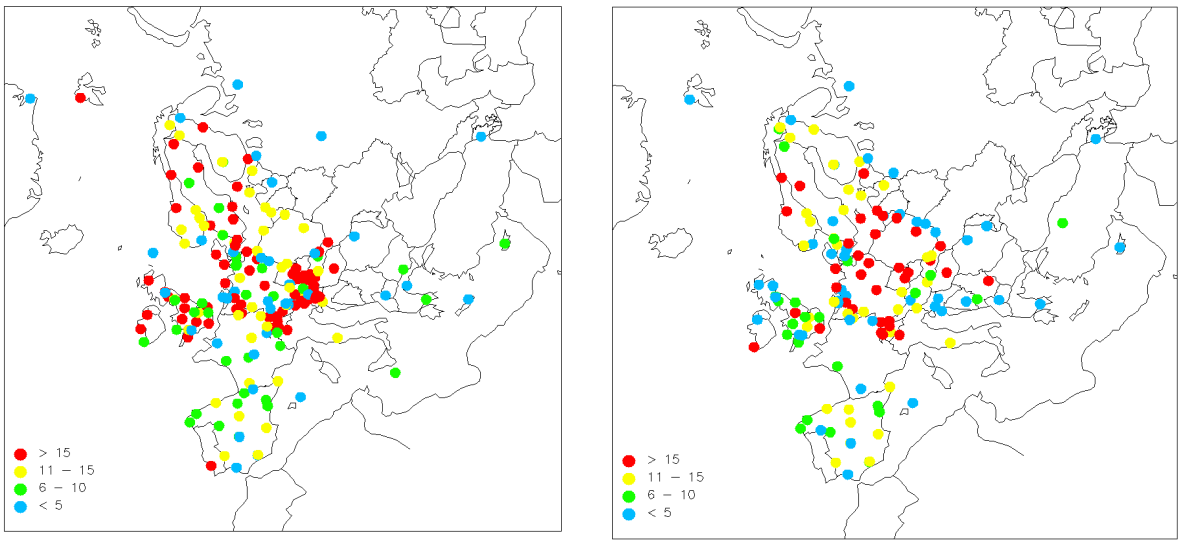


Figure 5: Number of years with monitoring data at EMEP stations during 1990–2010 for ozone (left) and NO_2 (right) (adapted from Tørseth et al., 2012).

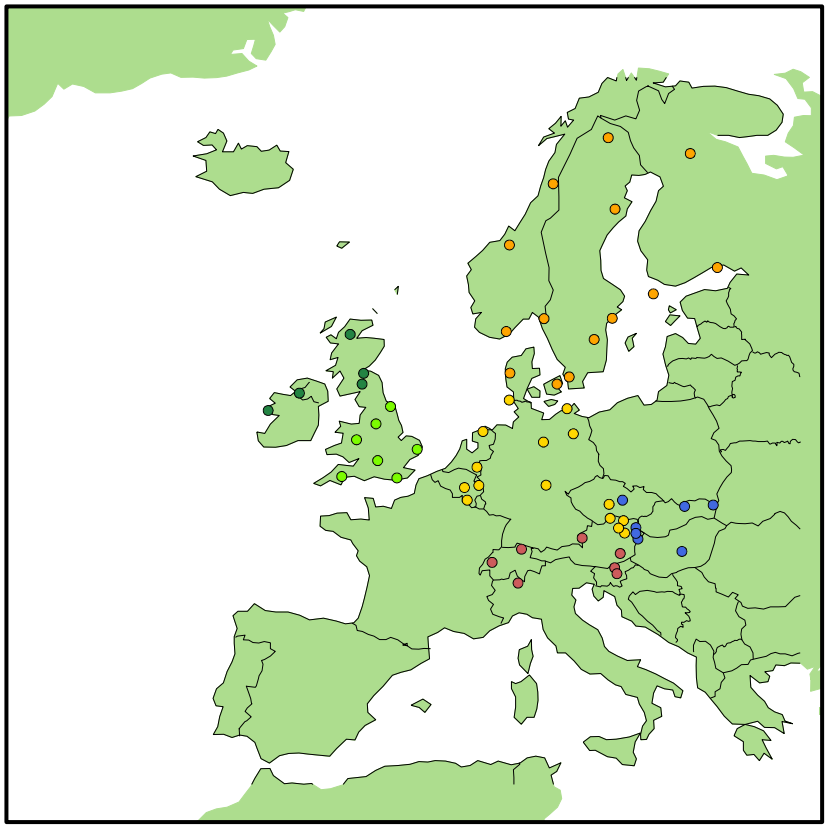


Figure 6. EMEP O_3 sites with at least 75% data capture in at least 75% of the years in the period 1990-2010. Sites located > 1000 m asl are not used, neither is PT0004 (Monte Velho). The coloured sites refer to the European subregions as defined in Figure 4.

4.2.3 Particulate matter

For total particulate matter mass, 29 and 18 sites were selected in the TFMM assessment for PM_{10} and $PM_{2.5}$, respectively. In both cases, only data from 2000 on could be used because of the low coverage before then. More measurements are available in the 1990s in the EMEP network for secondary inorganic aerosol and precipitation chemistry because of their importance in deposition of air pollutants onto ecosystems. For the 1990-2012 time period, 21 stations offered a satisfactory coverage for sulphate trends, while for nitrate and ammonia, 13 and 16 stations were selected, respectively. The EMEP stations available for the comparison between model and measurement presented in Section 5.2 are displayed in the maps of Figure 7.

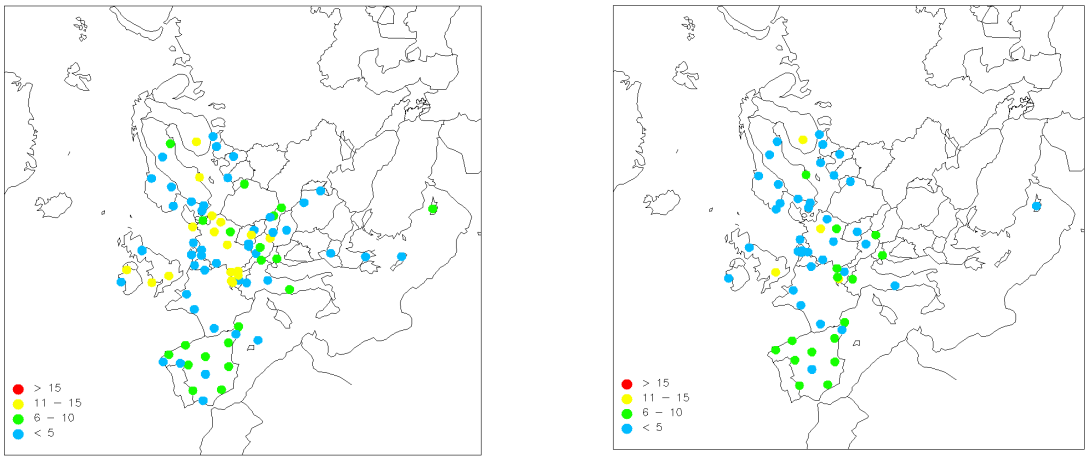


Figure 7: Number of years with monitoring data at the EMEP stations during 1990–2010 for PM_{10} (left) and $PM_{2.5}$ (right) (adapted from Tørseth et al., 2012).

In addition to those EMEP background sites, we also included PM_{10} measurement from the Air Quality e-reporting database in the model evaluation. This database benefits from a more comprehensive spatial coverage especially since 2000, also by including urban and suburban sites. We replicated the selection procedure presented in (Colette et al., 2015) and ultimately selected 170 background rural, suburban and urban sites.

4.3 Air pollution indicators

In this report, we use the word *indicator* for a certain aggregation of the observational time series, like annual mean, SOMO35, AOT40, etc., whereas *model score* refers to a certain statistical quantity used for evaluating model performance, like correlation coefficient, bias, etc.

4.3.1 Ozone indicators

The basic ozone data consist typically of hourly mean values from surface monitoring sites or from 2-D or 3-D geospatial fields from CTMs. These hourly values have to be aggregated into certain *indicators* to be able to proceed with further assessments of the data. This could be simple annual means, monthly means, high and low percentiles, annual mean of daily max etc. or more elaborate indicators such as SOMO35, AOT40, etc. (Lefohn et al., 2017) have studied a large number of ozone observational indicators and demonstrated the sensitivity of each indicator to certain parts of the ozone distribution.

Ozone indicators could be sorted into two main categories:

- Effect based indicators (SOMO35, AOT40, number of exceedances, etc.)
- Process based indicators (e.g. daily max, annual high and low percentiles, etc.)

The effect based indicators have typically been developed within the effect community and are used as proxies for assessing adverse effects from ozone exposure on humans or vegetation. On the other hand, process based indicators are used to assess individual physio-chemical processes controlling the ozone concentration level in certain ways or for certain atmospheric regimes. Ozone high peak indicators (e.g. annual max, 4th highest maximum daily 8-hour mean, etc.) reflect particularly the extent of photochemical episodes in summer; the mean diurnal ozone amplitude is more indicative for the occurrence of night-time inversions or local NOx emissions; and the daytime seasonal mean concentration could be used as an indicator of the influence of the hemispheric baseline level with regional effects on top.

When looking at long-term trends in ozone the selection of proper indicators is particularly important since each individual indicator could have a very distinct trend.

Furthermore, the ozone indicators used for impacts on human health and vegetation and of most interest for abatement policies are not necessarily the indicators best suited for evaluating model performance. Common impact indicators such as AOT40 and SOMO35 are known to be sensitive to the ozone baseline level which typically lies in the range 35-40 ppb in Europe. Thus, small errors in the model assumed baseline level could lead to large errors in the modelled AOT40 and SOMO35 (Avnery et al., 2011; Tong et al., 2009) . This has important consequences for the assessment of present abatement policies. Thus, an assessment of ozone trends should include i) observed trends in impact indicators and ii) modelled trends in process indicators and then, finally, an evaluation of how the model performance as seen by the latter point gives trust to the overall understanding of ozone in the atmosphere. These points are crucial to keep in mind when doing assessments of long-term trends in ozone.

The following indicators have so far been selected as relevant for model-measurement comparisons:

- Monthly mean of the daily average
- Maximum daily 8-hours mean (MDA8) concentrations along with standard deviations percentiles
- Monthly mean diurnal cycle
- Monthly mean of daily minimum and maximum of hourly average

In the following we have used the MDA8 as the main indicator for our evaluation of model performance. Furthermore, for the evaluation of trends we focus the assessment on the 4th highest annual MDA8, denoted 4MDA8 for simplicity. This corresponds to the 99 percentiles of the MDA8 (or the 98 percentiles of the 6-months summer MDA8). This choice of indicator is based on several points:

- Avoiding low night-time values typically determined by local effects
- Focusing on peak values most indicative of the model's chemistry and transport and less on prescribed boundary conditions

- Using a compromise between the most extreme peak (annual max) and lower percentiles more “blurred” by the prescribed boundary

Our conceptual idea behind this choice is that a number of marked photochemical ozone episodes occurs every year in Europe due to the European emissions. Although not exactly defined, this number is typically in the range of 3-6, and choosing the 4MDA8 should then reflect the episodes of that year – not too blurred by the imported background ozone and not too statistically uncertain as when using the maximum. Figure 8 shows the time series of MDA8 in 1990 for a few sites with the 97-, 98- and 99-percentiles on top, supporting the idea of a moderately varying time series with a small number of elevated peaks on top.

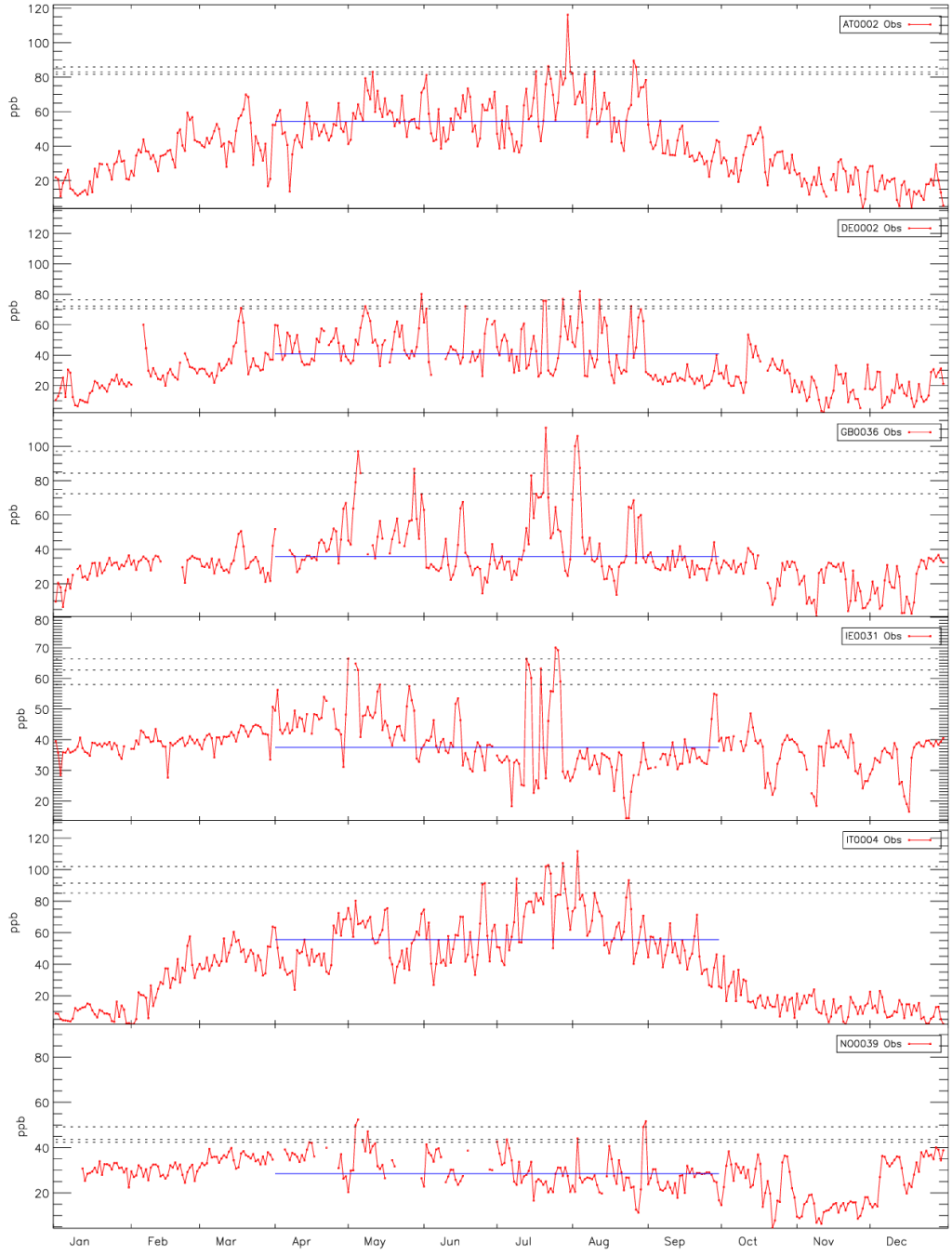


Figure 8. Observed max daily 8-hours mean (MDA8) for few sites in 1990. The dashed lines show the 99-, 98- and 97-percentiles of the annual MDA8. The blue line marks the 50-percentile of the summertime (Apr-Sept) MDA8.

As opposed to the MDA8 defined in the EU directive and the MDA8 defined in US legislation, we allocated the MDA8 to the mid hour of the 8 hours period. It was required that at least 75 % of the 24 hourly values that day were present as non-missing data (i.e. at least

18 values) for the day to be counted and that at least 75% of the 8 values (i.e. at least 6 values) in each 8h period were present as non-missing data for that 8h period to be counted.

Prior to these calculations of the MDA8 values we "filtered" the data such that missing hourly values in the observational data were set as missing also in the model data. This was done to avoid any bias due to differing times in the model vs. the observations.

4.3.2 Particulate Matter indicators

For particulate matter, the only indicator considered in the present report is the annual mean, which is the key indicator for health impact and also the references for one of the limit values used in the Air Quality Directive. In follow-up studies the number of exceedances could also be addressed although it is notoriously more challenging to capture by models.

4.4 Model scores

An evaluation of the performance of an atmospheric model typically involves extracting a time series from a monitoring station and the time series from the corresponding grid point in the model, then computing the relevant indicator (aggregated quantity) as defined in the previous section and then computing some kind of statistical measure of the agreement between the modelled and observed indicators. The statistical measures are called model scores in this report and a number of such model scores are defined in the following.

As pointed out by (Willmott et al., 2012), model scores like RMS (root mean square) based on squaring the errors could lead to an over-weighting of the large errors. (Willmott et al., 2012) instead recommend the use of MAE = mean absolute error and MAD = mean absolute deviation and also propose a new IOAr = refined index of agreement to be used for evaluation of model performance.

(Eskes et al., 2015) applied various model scores when studying model performance within the MACC project (http://www.copernicus-atmosphere.eu), like the modified normalized mean bias (MNMB) and the fractional gross error (FGE). They argue that MNMB and FGE, being normalised quantities, have advantages compared to the more common model scores mean bias and root mean square error. The MNMB applied by (Eskes et al., 2015) is identical to the "mean fractional bias" applied by (Lefohn et al., 2014) for assessing the performance of the coupled GEOS-Chem/CAMx models in North America.

In the following we have used the model scores defined in Table 1.

Table 1. Model scores for evaluation of model performance used in this study. N = number of data values, m_i = modelled data value at time i, o_i = observed data value at time i, om = mean observed value (mean of o_i), mm = mean modelled value (mean of m_i), σ_m = st.dev of modelled data and σ_o = st.dev of observed data.

Mean absolute errror	MAE	$1/N \sum (m_i-o_i)$
Modified normalized mean bias	MNMB	$2/N \sum [(m_i-o_i)/(m_i+o_i)]$

Fractional gross error	FGE	$2/N \sum \lvert(m_i-o_i)/(m_i+o_i)\rvert$
Refined index of agreement	IOA_r	$1- \sum \lvert(m_i-o_i)\rvert/[2\sum \lvert(o_i-om)\rvert]$, if $\sum \lvert(m_i-o_i)\rvert \leq 2\sum \lvert(o_i-om)\rvert$ $2\sum \lvert(o_i-om)\rvert \sum \lvert(m_i-o_i)\rvert - 1$, if $\sum \lvert(m_i-o_i)\rvert > 2\sum \lvert(o_i-om)\rvert$
Linear correlation coefficient	r	$1/N \sum[(m_i-mm)(o_i-om)]/\sigma_m\sigma_o$

MAE expresses the mean difference between the modelled and observed data in the relevant unit (ppb). MNMB is symmetric around zero bias and ranges between ± 200 %. FGE is a linear measure, and has the advantage compared to the more common root mean square measure that it is not dominated by outliers. IOA_r expresses the mean difference between the modelled and observed values relative to the variation in the observed values and is in the range [-1, 1]. Perfect agreement is indicated by a value of 1 whereas a value of 0.5 implies that the mean difference between modelled and observed data is equal to the spread in the observed data. Negative values indicate either poor model agreement or little observed variability.

4.5 Temporal decomposition: the Kolmogorov–Zurbenko filter

The Kolmogorov–Zurbenko (KZ) filter has been used in various studies to extract the various temporal components (long-term, seasonal, etc.) from an ozone time series (e.g., (Ahmadi and John, 2015)). The KZ filter is scientifically sound and robust and at the same time simple to compute. It is basically a plain running mean value applied several times to the original time series. The mathematics are explained in more detail in Annex 11.2.

In the following we used various types of KZ filters to calculate quantities expressing the long-term mean, the short-term mean, peaks, etc., as defined below. Note that the basis for these KZ calculations, i.e. the original time series, were daily data consisting of the MDA8 values through the whole year for each station. As explained above, we believe that the daily MDA8 values excludes many of the locally determined effects of the time series, such as night-time inversions, short-term dips in O_3 due to titration by NOx from local emissions, etc. Based on these daily MDA8 values, we used the KZ filter to estimate the development in the long-term mean concentration by the same definition as in (Ahmadi and John, 2015), using 3 iterations over a 365 days period:

$O_{3,\,LTM} = KZ_{m,k}[O_3(t)]$ m=365, k=3

Furthermore, a smoothed mean time series was defined and calculated as:

$O_{3,\,MEAN} = KZ_{m,k}[O_3(t)]$ m=30, k=3

This corresponds to a 30 days running mean applied three times.

To estimate the development in lower baseline and high peak values we designed a modified KZ filter using the running percentile instead of the running mean in the first iteration, whereas for the subsequent iterations we used the plain running average as in the standard KZ filter:

$KZ^*_{p,m,k=1} = p[O_3(t+s)]$, s=(m-1)/2, (m+1)/2 where p = the percentile function, p = [0,100]

We then defined:

$O_{3,\,PEAK} = KZ^*_{p,m,k}[O_3(t)]$ p=95, m=30, k=3

$O_{3,\,BASE} = KZ^*_{p,m,k}[O_3(t)]$ p=15, m=30, k=3

This definition of $O_{3,\,PEAK}$ was based on a comparison with the annual 98 percentiles of the MDA8 for the 6-months in the summer season (April-September) as discussed in more detail in the chapters below. We found that the 98 percentiles of the summer MDA8 that is used as the main indicator for the photochemical episodes in this paper, agreed closely with the annual peak value in the $O_{3,\,PEAK}$ variable as defined above. An example of that decomposition is provided in Figure 11.

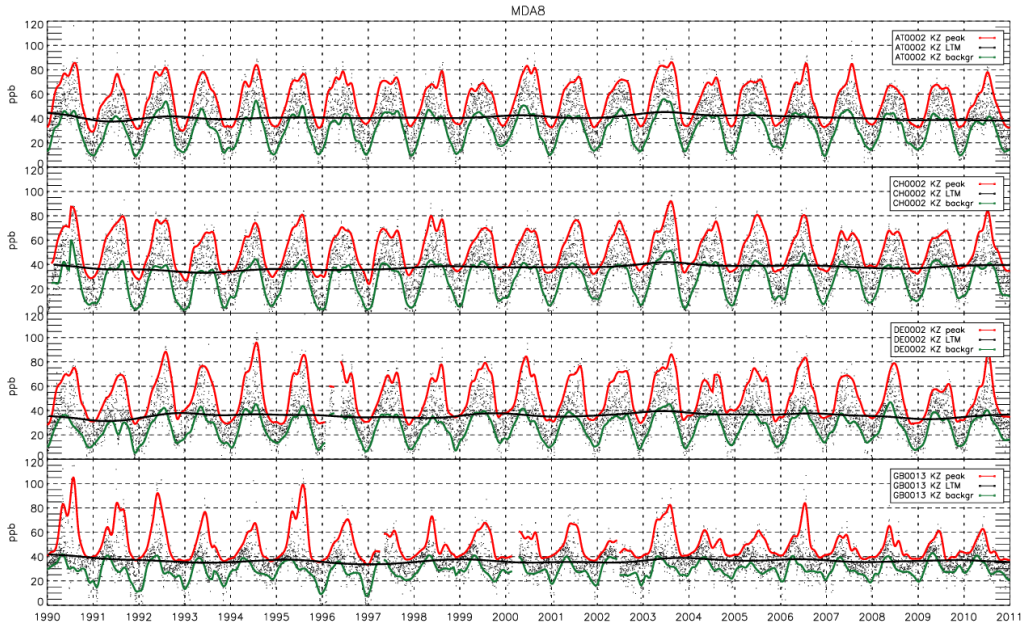


Figure 9: Example of KZ decomposition (LTM stands for long term mean, i.e. 365 days running mean applied 3 times)

4.6 Non-parametric trends tests: Mann-Kendall significance and Sen's slope

The non-parametric Mann-Kendall (MK) method combined with the Sen's slope estimator has been widely used for assessing long-term trends in atmospheric constituents. The advantages of this method is that it does not require the input data to be normally distributed, it is simple to apply and it is not sensitive to outliers in the data (which are typically included in atmospheric data) since it is based on the ranks of the data elements and not on the absolute values. The MK method was the main trend method used within the EMEP TFMM work (Colette et al., 2016a) as well as in previous EEA air quality trend studies (EEA, 2009). The ongoing IGAC initiative TOAR, looking at ozone trends on a global scale, also uses the MK method as the main trend tool in their analyses

A detailed outline of the MK method and the Sen's slope estimator is given in Annex 11.1. The basis of this method is an analysis of the ranks of each element of the time series.

5 Model Evaluation

This section addresses both the long-term evolution of model performances, i.e. the changes in time in their capacity to capture air pollution for a given year, and the performance of the model in capturing the long-term evolution, i.e. the comparison of modelled and observed air pollution trends.

5.1 Ozone

5.1.1 Change of model performances over 1990-2010

Whereas the observational data tells the "true story" of the long-term development in ozone, this is limited to a number of fixed points in space, typically with a bias to certain regions as shown above (Section 4.2) and also limited by data quality issues, station representativity, etc. CTMs are essential for extrapolating the information to regions with poorer data cover and also for making future predictions and for providing an understanding of the quantitative relationships between emissions and surface air pollution levels. To be able to do so, however, the models must be shown to provide reasonable results that agree to a certain extent with what the observations tell. Thus, we need to establish certain criteria for evaluating the models' performance with respect to observed indicators as discussed above.

We chose to focus our analyses on the maximum daily 8-h mean concentrations (MDA8) during the summer half year (April-September). By this selection, we avoid periods of the day (typically night time) with low ozone levels due to surface dry deposition and reduced vertical mixing that are mostly determined by very local conditions (land use, topography, etc.). We also avoid the winter season when the ozone levels are more determined by the hemispheric baseline level and titration by local NOx emission sources. We regard the daily MDA8 as the most representative of the conditions in the planetary boundary layer in a specific day and thus the most comparable to the regional models.

Figure 10 and Figure 11 show the temporal development in the model scores IOAr, MAE, MNMB and FGE, during 1990-2010 for three models, EMEP, CHIMERE (CHIM) and LOTOS (LOTO). All calculations were done on the same set of stations (ST90-10) as described above.
For each year (i.e. each summer season) we calculated the model score X_i (IOAr, MAE, etc.) for every station i in the dataset with sufficient data capture:

$$X_i, \text{ with } i = 1, \ldots, n; \quad X = [\text{IOAr, MAE, MNMB, FGE}]; \text{ and } n = \text{number of stations}$$

The box- and whiskers mark the spread and percentiles in these n values each year. For each model score we looked for trends in the annual medians of these model scores using the Mann Kendall and Sen's slope non-parametric methodology, and the estimated trend lines were superimposed on the plots whenever a statistically significant slope ($p < 0.05$) were found. The Sen's slopes are given (in percentages per year, %/y) in the top titles for each box in Figure 10 and Figure 11. This was done for each model, separately.

These results reveal systematic differences among the models as well as marked changes with time. In general, the results from CHIMERE show the lowest biases (MAE, MNMB and FGE) and the best agreement (IOAr) and LOTOS the poorest, with EMEP in the middle.

Furthermore, a significant improvement in these statistics (increase in agreement and reduction of bias) with time is evident for most of the cases, pointing to a significant trend in the overall model performance for the 1990-2010 period. Marked reductions in the mean absolute error (MAE) are found for all three models. Averaged over the whole period 1990-2010 LOTOs gives a mean MAE of nearly 9 ppb whereas CHIMERE gives a mean MAE of less than 1.

These results show that there is a clear trend in the overall agreement between models and observations during the 1990-2010 period for all the models. Since the observational basis is the same set of stations (although it varies somewhat from year to year depending on the data availability), this trend could not be explained by changes in the monitoring network. The reason for these results is not obvious. Assuming that the atmospheric chemistry and physics (deposition, turbulent mixing, etc.) is fairly homogeneous with time and that the quality and accuracy of the meteorological input data also fairly consistent, such a trend in model/observational agreement must be explained by systematic trends in other controlling factors such as emission data, hemispheric baseline levels or in the quality and representativity of the observational data. Alternatively, the chemistry-driven trend relative to other controlling processes (vertical mixing, biogenic emissions, etc.) could be overestimated by the models.

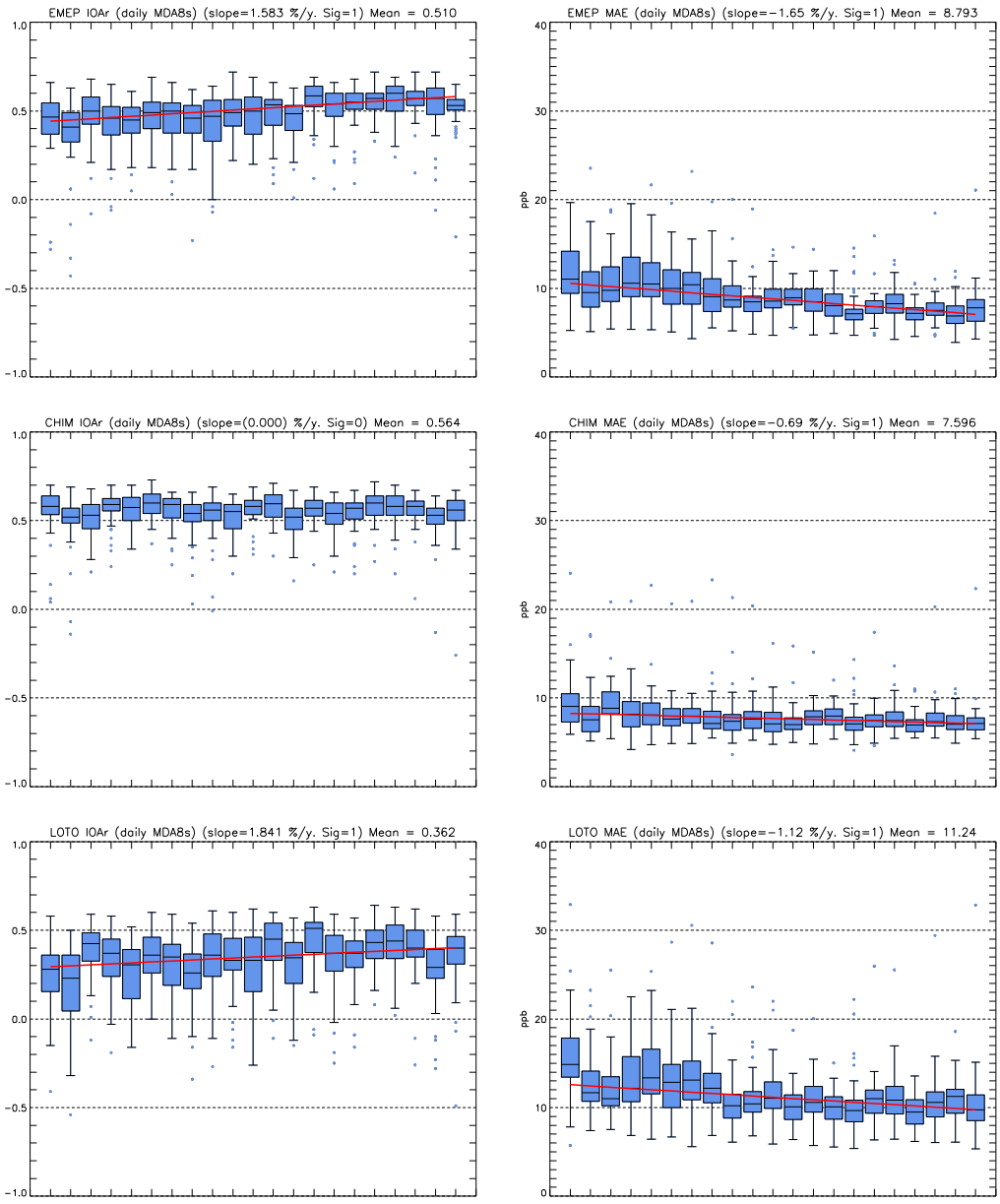


Figure 10. Box- and whiskers plots of model scores (IOAr and MAE, in ppb, see table 1) for model performance based on the same set of stations ($ST_{90\text{-}10}$). Sen's slopes of the median values for each year are indicated in the top title and plotted for significant cases. Sig = 1/0 means significant/non-significant on a p=0.05 level. All these statistics refer to the station-averaged statistic based on the MDA8 for the summer half year (MDA8=maximum daily 8h mean ozone mixing ratios computed for each day in April-September). Results for the EMEP (top), CHIMERE (middle) and LOTOS (bottom) models are shown. The boxes mark the 25- and 75-quantiles with the median given inside. The whiskers extend to the min and max values or to the 1.5 times P25 and P75 if there is data (outliers) beyond this range. Such outliers are identified with small circles.

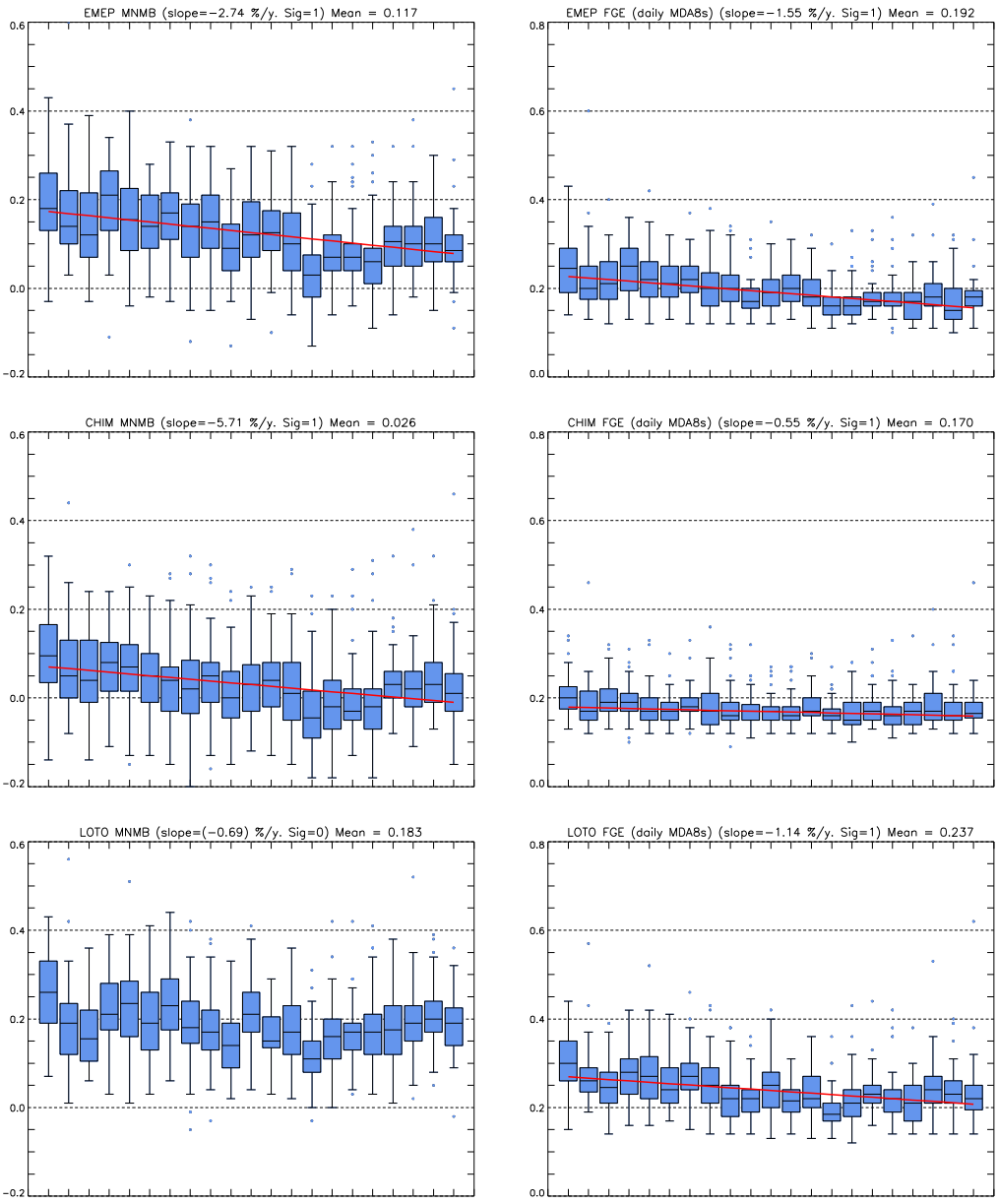


Figure 11. Same as Figure 10 for the modified normalised mean bias (MNMB) and FGE (fractional gross error).

The model performance statistics presented above was furthermore applied for the six regions discussed above and as shown in Figure 6

Figure 12 show the results for MNMB (modified normalized mean bias) for the EMEP, CHIMERE and LOTOS models, respectively, for each of the six regions defined above (see figure 6). The results given in
Figure 12, Figure 13 and Figure 14 indicate clear regional differences in the model biases. All three models give a significant downward trend for the North Italy (NI) and Middle Europe (ME) regions, and both the EMEP and CHIMERE model a significant downward trend for the East Europe (EA) region. For the other regions, there is no significant trend for any of the three models.

These results, using the MDA8 for each day in the summer half year, reflect the overall model bias for the different regions. The downward trends in this model score indicate a general improvement in model performance (as measured by the bias) during the period 1990-2010 when compared to the observations. Furthermore, the marked difference between the regions seen by the three models, points to systematic regional patterns.

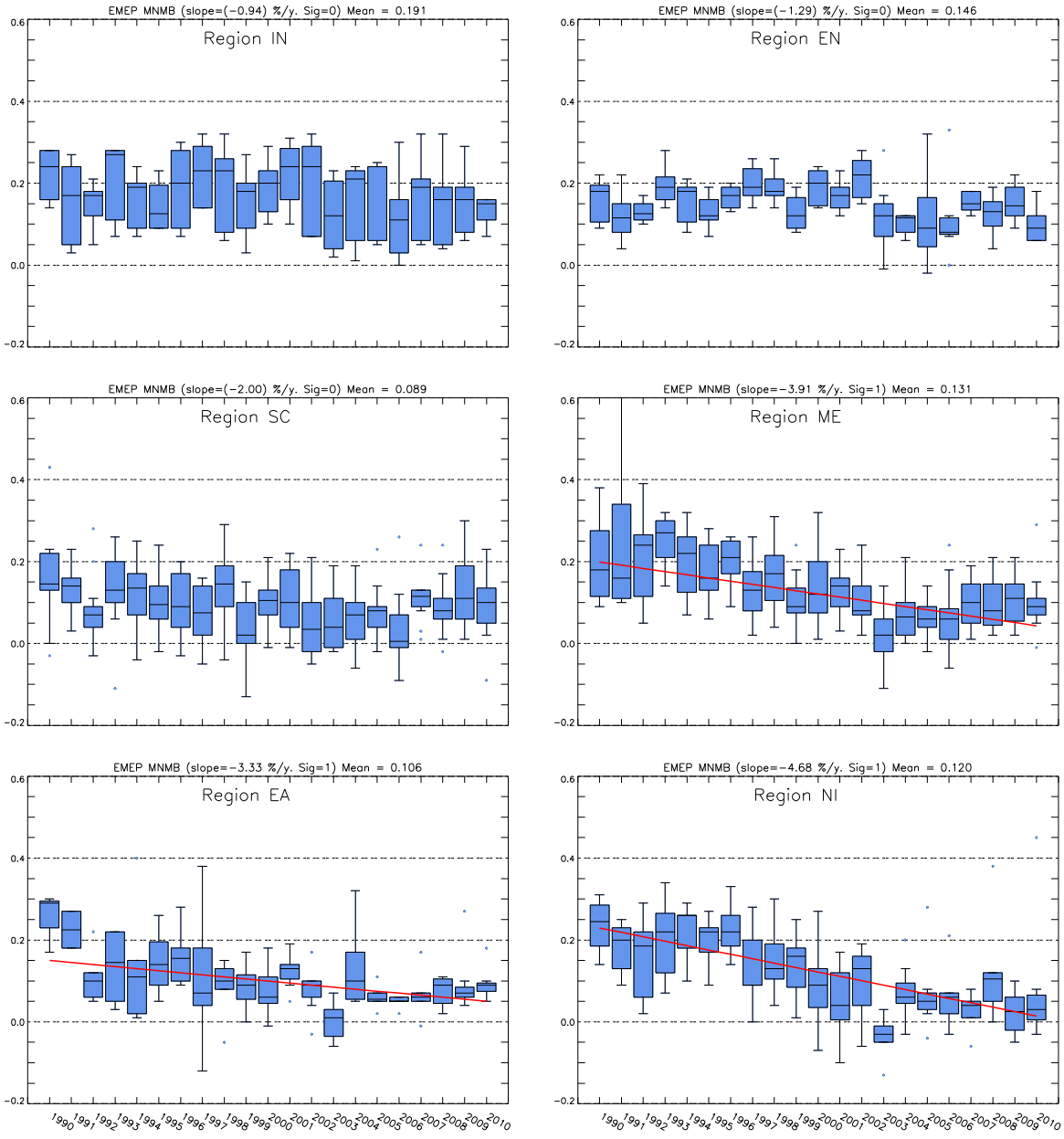


Figure 12. Box- and whisker plots for the MNMB (modified normalized mean bias) based on the MDA8s (max daily 8 h averages from April to September) for the European sub regions where long term ozone measurement are available based on the EMEP model only. The regions refer to the stations given in Figure 4 and defined in 4.2.1

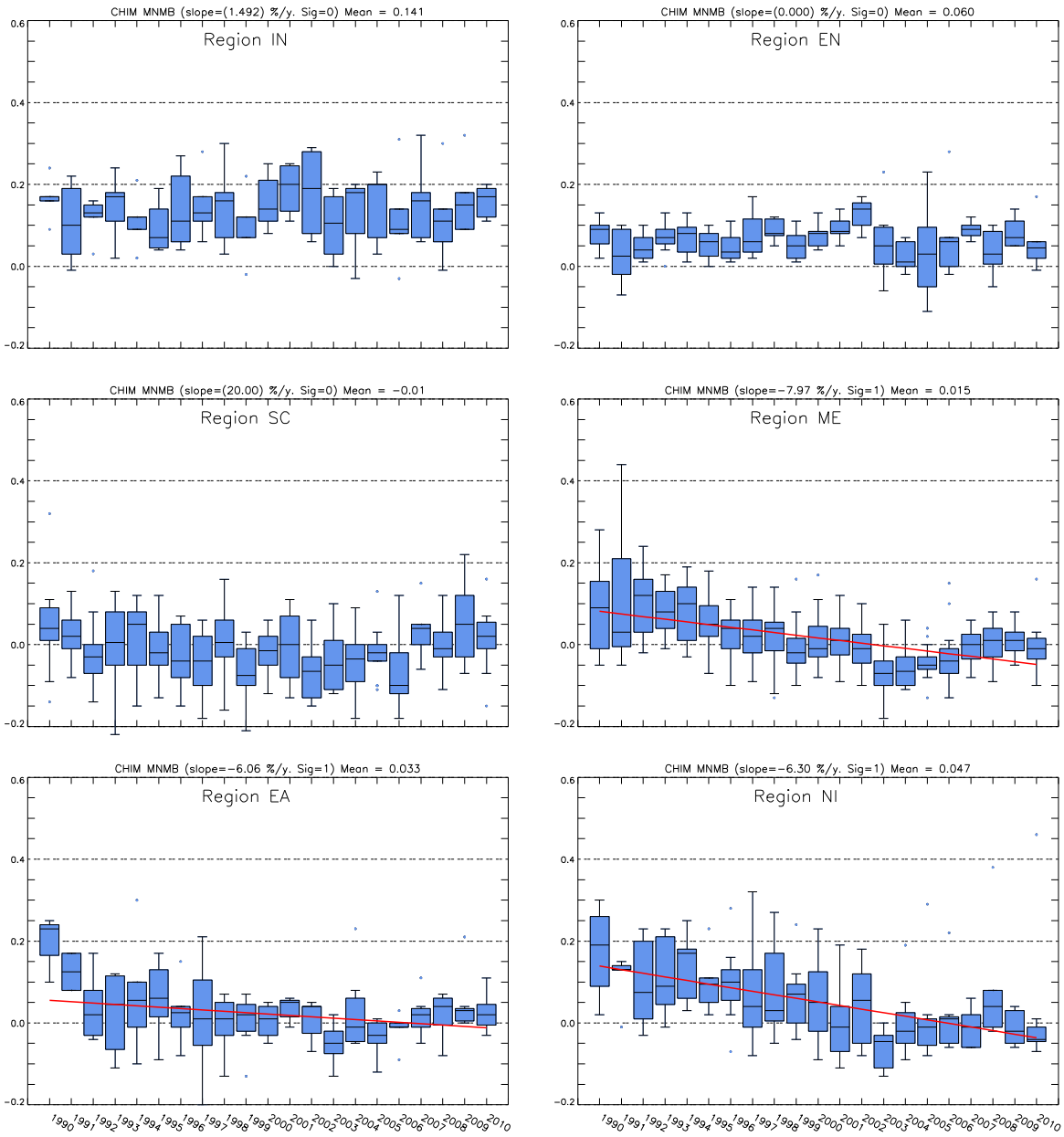


Figure 13. Same as Figure 12 for the CHIMERE model.

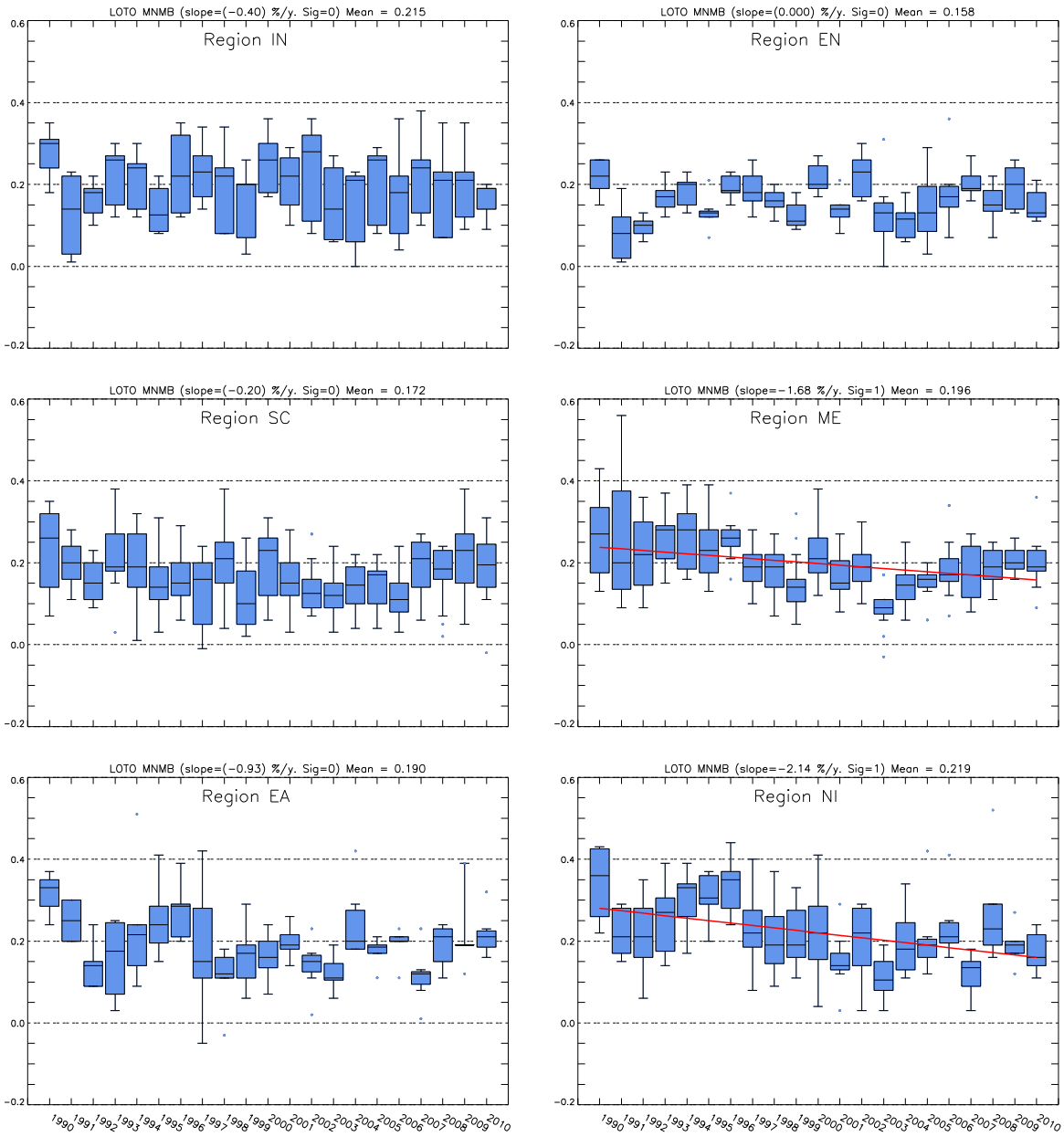


Figure 14: Same as Figure 12 for the LOTOS model

5.1.2 Comparison of observed and modelled ozone trends

An assessment of long-term trends in observed and modelled ozone raises a number of questions:

- How reliable are the historical ozone monitoring data?
- How representative are the ozone monitoring sites?
- Which ozone indicators should we look at?
- How good are the emission data?
- How good are the models?
- How could we measure the model performance?

In the ETC/ACM Working Paper on trend decomposition (Solberg. et al., 2015), the conceptual view of a procedure for an ozone trend assessment based on observations and models combined was illustrated as in Figure 15. An essential part of such an assessment is various kinds of *screening* of data – screening with respect to measurement data quality, ozone indicator, geographical region, weather situation, etc.

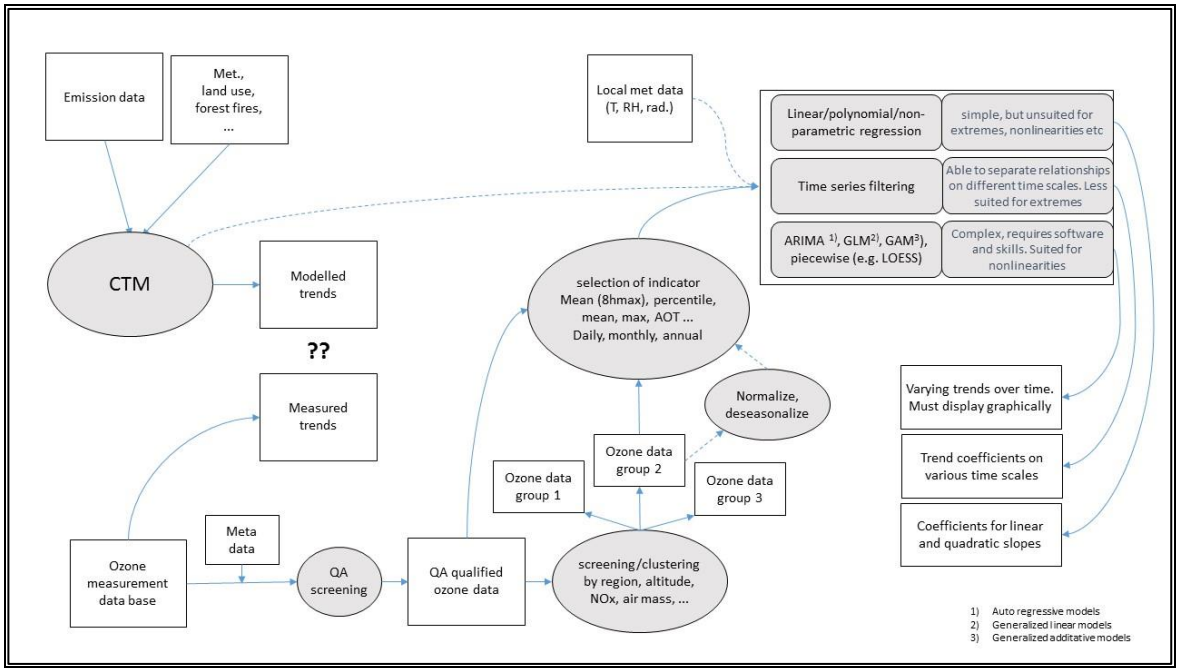


Figure 15 Conceptual view of the complex procedure for an ozone trend assessment based on observations and modelling data in combination (adapted from (Solberg. et al., 2015)).

5.1.2.1 Ozone indicators

As discussed by (Lefohn et al., 2017) the long-term ozone trends could differ substantially for different ozone indicators. They investigated a large number of indicators and showed that the statistical significance and the value of the trend reflects the type of indicator used. This is in line with the results from the EMEP TFMM assessment (Colette et al., 2016) where the trends in various ozone indicators based on European observational data have been documented.

The TFMM assessment was mainly based on observations and modelling was not a central issue in the assessment report. A main aim of the current document is, however, to evaluate to what extent the observed ozone trends corresponds to what to expect as based on trends in emissions and interannual variations in meteorology by use of models. Figure 16 shows the relative trends (% per year) for the Eurodelta period 1990-2010 in some of the indicators also shown in Figure 2 based on EMEP observations for the EMEP model results. The other models were let out just for the purpose of document space. The trend values in Figure 16 express the Sen's slopes relative to the regression intercept at year = 1990:

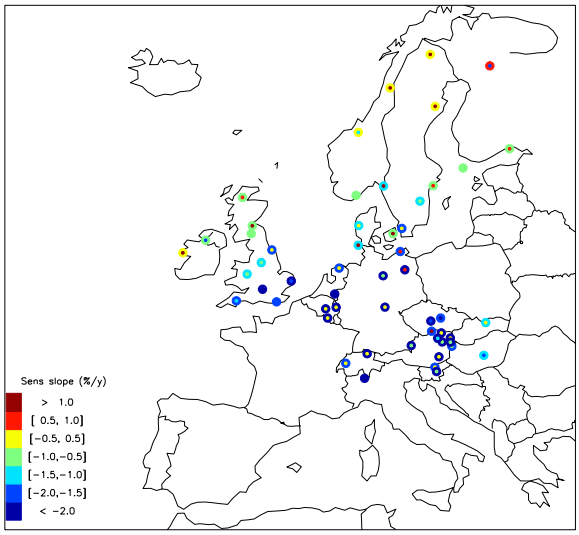
The Sen's slope calculation gives a simple linear equation:

$$X(y) = X(y_0) + c[X(y) - X(y_0)]$$

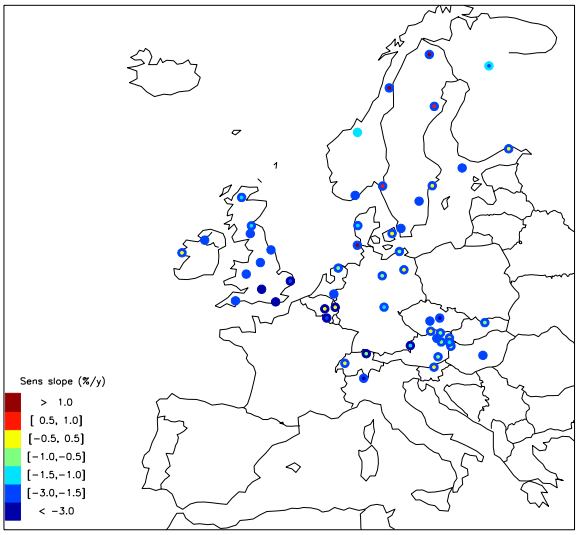
where X(y) = the value of the ozone indicator in year y, y = [1990, …, 2010]
y_0 = the start year, i.e. 1990
c = the Sen's slope

The relative trend is then given by $c_{rel} = c/X(y_0)$

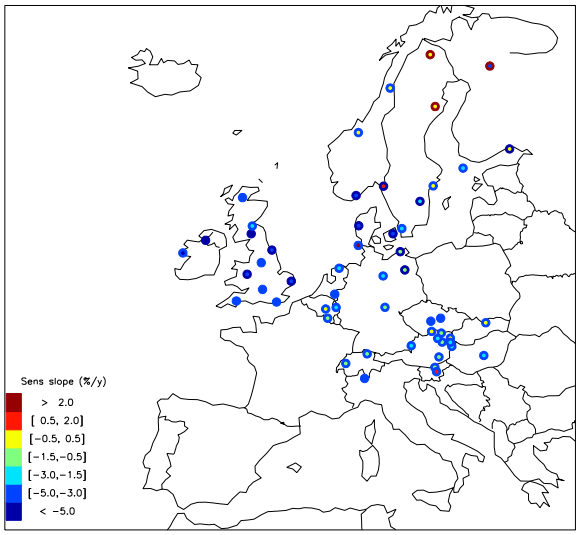
SOMO35



AOT40



N(MDA8)>60ppb



4th high MDA8

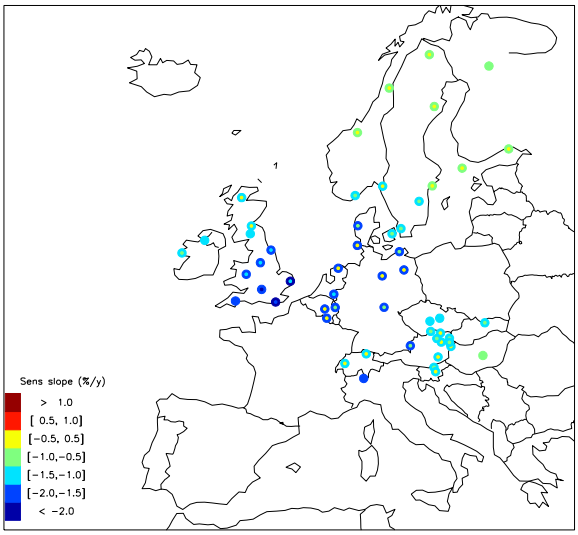


Figure 16. Relative trend (%/year) at EMEP stations ST90-10 for the period 1990-2010 as observed and modelled by the EMEP model in SOMO35 (upper left), 6-months AOT40 (upper right), number of days with a MDA8>60 ppb (lower left) and 4th highest MDA8 (lower right). Outer circle marks the modelled trend and inner circle the observed trend. Note that all trend values are included, independent on the statistical significance. Note that the scales differ.

In Figure 16 green-blue colours mark downward trend, red marks upward trends whereas yellow marks trends around zero (in practice no real trend). Note that all trend values are included independent of the trend being statistically significant or not. Note also that we applied a linear scaling for missing data when calculating SOMO35 and AOT40 meaning that the calculated value was scaled by the inverse of the data capture. The results from Figure 16 is furthermore given in scatter plots in Figure 17 in which the symbols/colours refer to the regions shown in Figure 6.

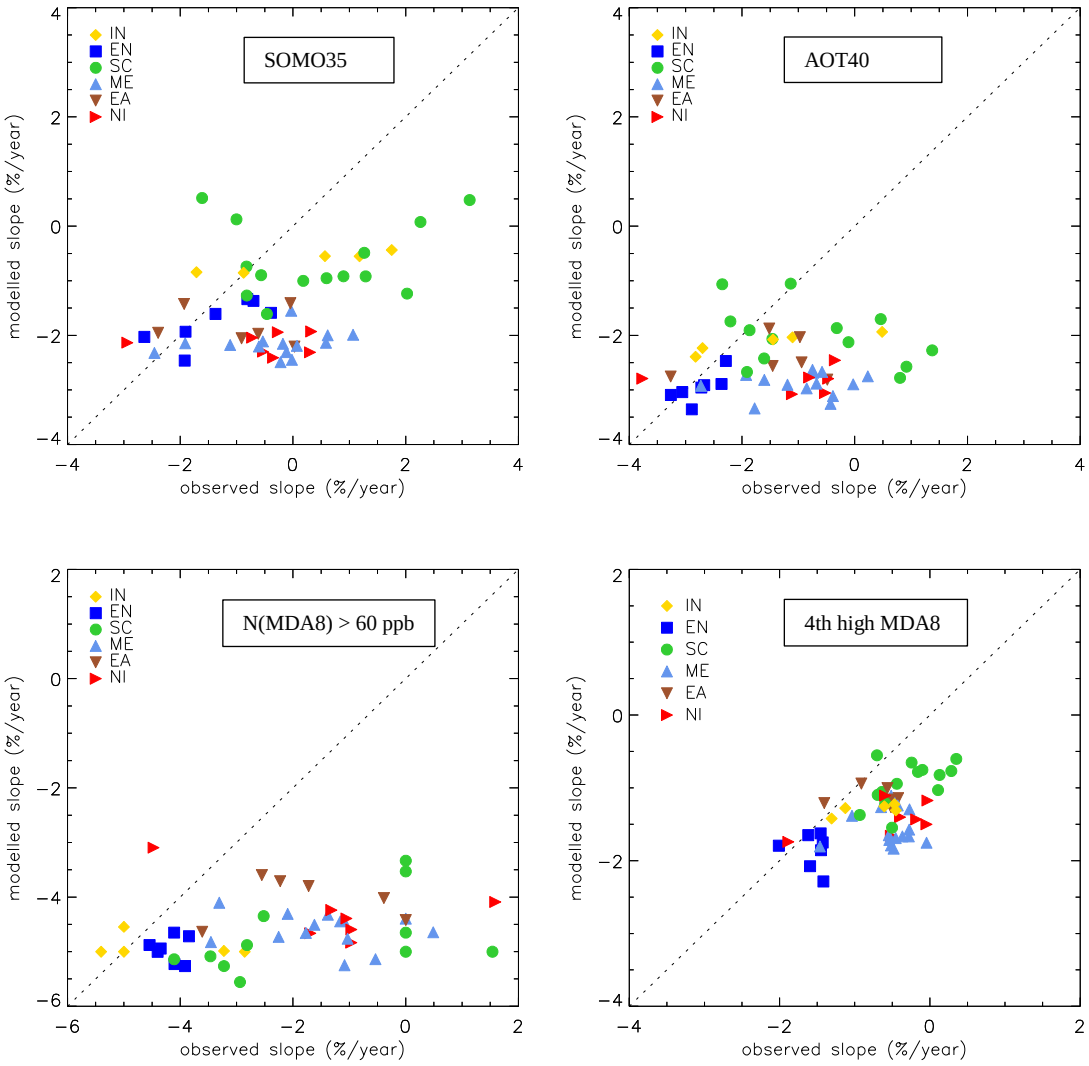


Figure 17 (corresponding to Figure 16). Relative trend (%/year) 1990-2010 in measured values vs. modelled values for the European subregions defined in Figure 4. Only results from the EMEP model are shown for SOMO35 (upper left), 6-months AOT40 (upper right), number of days with a MDA8>60 ppb (lower left) and 4th highest MDA8 (lower right). The different symbols correspond to the station/regions defined in Section 4.2.1. Note that all trend values are included, independent on the statistical significance. Note also that the ranges of the axis differ.

The observational network history implies that only certain parts of Europe is covered when applying the criteria of 75 % data capture for the period 1990-2010. Furthermore, the time periods of Figure 2 and Figure 16 differs which could have a critical influence on the estimated trends. As shown in Figure 2, a higher fraction of significant downward trends was found in the last period, 2002-2012, compared to the first period, 1990-2001 in the TFMM study.

As seen from Figure 16 and Figure 17 the trends vary considerably with region and ozone indicator. In general, the EMEP model gives stronger downward trends than the observations. The results show good agreement between observed and modelled trends in England and at some other sites (Ispra, Birkenes, etc.). In most other regions (ME, EA, NI, SC) clear differences between the observed and modelled trends are seen. Furthermore, the significance and slopes of the trends vary substantially for the different indicators, which is in line with the findings of (Lefohn et al., 2017). For SOMO35, the data show particularly large regional variations with marked increases in some regions (SC) and marked decreases in others (EN, ME).

Although ozone is considered a regional scale pollutant supposedly with a broad scale distribution, the results in Figure 16 show large differences in the estimated trends of a indicator like SOMO35 for nearby sites e.g. when comparing the sites in NE Austria and the Czech Republic. Another example is AOT40 in the area around Denmark where the measured trends go in opposite directions when comparing Westerland (DE01), Vavihill (SE11), Ulborg (DK31) and Lille Valby (DK41).

These results reflect the large uncertainties inherent in trend assessments of ozone indicators as AOT40 and SOMO35 as discussed above, and furthermore points to questions regarding station representativity and data quality.

5.1.2.2 Trend in ozone peak values: 4th highest MDA8

As shown above, the long-term trends differ substantially with the selection of ozone indicator. (Lefohn et al., 2017) showed that an indicator like the 4hMDA8 was most sensitive to high peak episodes and showed clearer and more significant trends than e.g. indicators including more moderate ozone levels. Thus, in the following we study the trends in the 4hMDA8 in more detail.

Whereas the 4hMDA8 is an indicator of the highest photochemical episodes, the median of the MDA8 for the six-months summer season (p_{50}) is a measure of the mean conditions for that period. The observed and modelled 4hMDA8 and p_{50} of the MDA8 are given in Figure 18 for the EMEP, CHIMERE and LOTOS models.

These results reveal systematic differences between the three models and between each model and the observations. For the p_{50}, there is no trend in the observations during 1990-2010. The three models all predict a slight reduction in this indicator, but only significant for the EMEP model (-0.23 % year^{-1}). There are, however, marked differences between the models in this indicator. The EMEP and LOTOS models give consistently higher medians than observed, whereas the CHIMERE medians are close to the observed values.

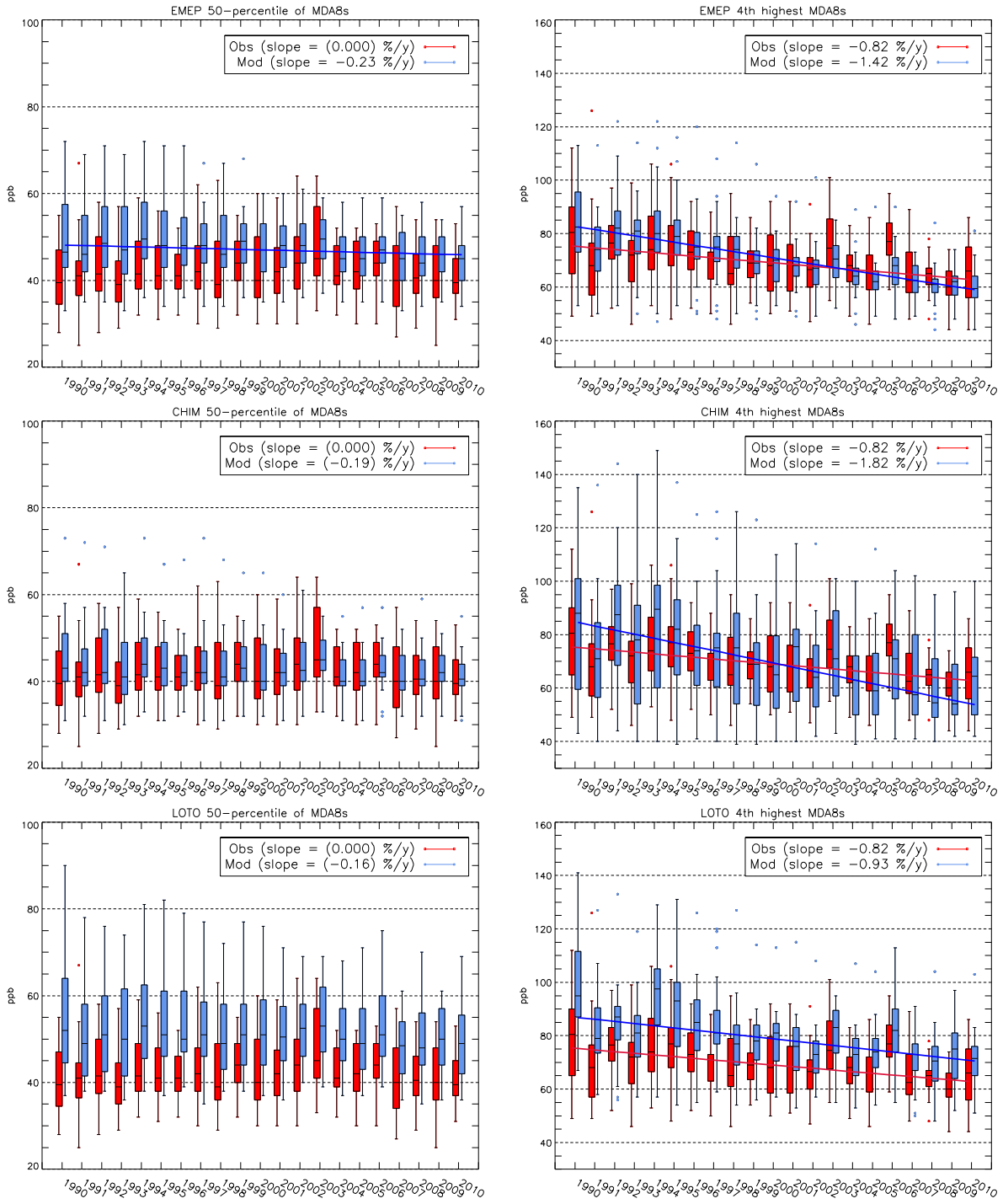


Figure 18. Box- and whiskers plots showing the station-wise spread in the median (left) and 4th highest MDA8 (right) based on the observational (red) and modelled (blue) data for each year separately (MDA8s=maximum daily 8h mean ozone mixing ratios computed for each day in April-September). Sen's slopes based on the median values each year are given in the legend. Non-significant trends (on a p=0.05 level) are given in brackets. Significant trends are also marked by lines with the corresponding colour in the plots. The models shown are EMEP (top), CHIMERE (middle) and LOTOS (bottom).

For the 4hMDA8 there is a significant decline in all the three models' results as well as the observations. However, all three models predict stronger reductions than observed, consistent with the decline in bias seen from Figure 10 and Figure 11. Furthermore, there are clear differences between the three models. The EMEP model gives systematically higher values than observed during the first part (the 1990s), whereas the opposite is seen for the last part (the 2000s). The station-wise spread (as measured by the length of the bars each year) in the EMEP model agree well with the spread in the observations, though. For certain years, there is a stronger bias between the model and the observations, like in 2006 (model underestimation) and 1991 (model overestimation).

The 4hMDA8 values from CHIMERE show a substantially higher station-wise spread than observed, and the overall downward trend (-1.82 % $year^{-1}$) is clearly higher than the observations. In general, all the LOTOS model results show a significant positive bias (as seen also from Figure 10 and Figure 11 above), implying a systematic overestimation by the model. The predicted trend in the 4hMDA8 indicator matches the observed trend nicely, though.

The marked bias in the median 4hMDA8 values seen by the EMEP and LOTOS models are somewhat surprising and indicates a systematic offset in the ozone baseline level. This indicator is more dependent on the assumed background levels than the photochemical episodes. On the other hand, the 4hMDA8 values are, as mentioned, a proxy for the photochemical episodicity.

To investigate further the deviations between the modelled and observed 4hMDA8, we split these results on to the regions as given in Figure 4. The observed and modelled annual 4hMDA8 is shown in Figure 19 (EMEP),
Figure 20 (CHIMERE) and
Figure 21 (LOTOS) for each region and year. These results indicate systematic differences between the regions and between the models.

First of all, the spread in the station-wise values each year (the length of the bars) indicates the homogeneity of the stations included in the various regions as defined above (Figure 4). Short bars, reflecting homogeneous sites within the region, gives trust and meaning to the definition of the regions and the allocation of sites. This is seen in particular for Scandinavia (SC) and to a somewhat less extent the Inflow (IN) region. On the other hand, the Middle Europe (ME) region shows larger spread in the observational data, indicating a less homogeneous region. Also in England (EN) there is a larger spread in data in certain years. For the North Italy region (NI) the lengths of the bars are mostly very small, however we notice one outlier each year and these values mark the results for Ispra, showing constantly higher ozone levels than the rest of this region, presumably reflecting the conditions in the Po Valley being very favourable for photochemical ozone formation.

When comparing the homogeneous regions with the model values, certain characteristics are evident. For the Inflow region, the EMEP model tends to give higher values than observed, particularly in the first part of the 1990s, whereas CHIMERE is closer to real values for most of the years. The LOTOS model, as mentioned above, gives systematically higher values than observed for most regions/years.

For the Scandinavian region, the EMEP model results are closer to the observed data, whereas CHIMERE results are systematically lower and LOTOS are systematically higher.

The measured data show a significant decrease in the 4hMDA8 for all regions except SC (Scandinavia).

In the England region, the observations show a strong and significant downward trend (1.55% y^{-1}). This trend and the annual values are well reflected in all three models.

As mentioned, the spread in the Middle Europe region (ME) is larger than in the other regions, indicating a more inhomogeneous region, particularly in the first part of the 1990s. However, we see that the models reproduce to some extent both the annual spread and the interannual variability, but with a larger downward trend than observed. Both the measurements and the models show a significant downward trend in the 4hMDA8 for the stations in this region during the period 1990-2010 although the modelled trends are substantially stronger than observed. This discrepancy is due to a marked positive bias (model values > observed values) during the first part of the 1990s and a negative bias the last part of the 2000s

Also for the EA and NI regions the modelled trends in 4hMDA8 are larger than observed. This is particularly true for the NI region where the EMEP model predicts a mean downward trend of -1.5 %/y whereas the measurements show a mean downward trend of -0.5 %/y. Furthermore, the results for Ispra (IT04) show up as outliers (small dots) in this figure.

In addition to what is discussed above, Figure 19 to Figure 21

Figure 19 indicate certain deviations for some years. In 1991 particularly low 4hMDA8 values are observed in all regions, and a corresponding dip is reflected in the modelled values although less marked in some regions. It has been speculated (Tang et al., 2013) that the low surface ozone levels in 1991 is linked to the eruption of the Philippine volcano Mt Pinatubo in June that year, the second largest volcanic eruption in the 20th century.

Furthermore, in the years 2003 and 2006 with well documented heat waves associated with elevated ozone episodes in Europe, the EMEP model is seen to strongly underestimate the 4hMDA8 values. An underestimation is seen also for CHIMERE, but to a less extent. For the LOTOS model, the values agree fairly well in these years, however this could be seen as a result of the model giving systematically higher ozone levels than the other models (and the measurements) for the rest of the period. It is a general finding that the distribution of ozone levels become narrower in models than observed, which is explained by systematic weaknesses in the models to capture the extremes.

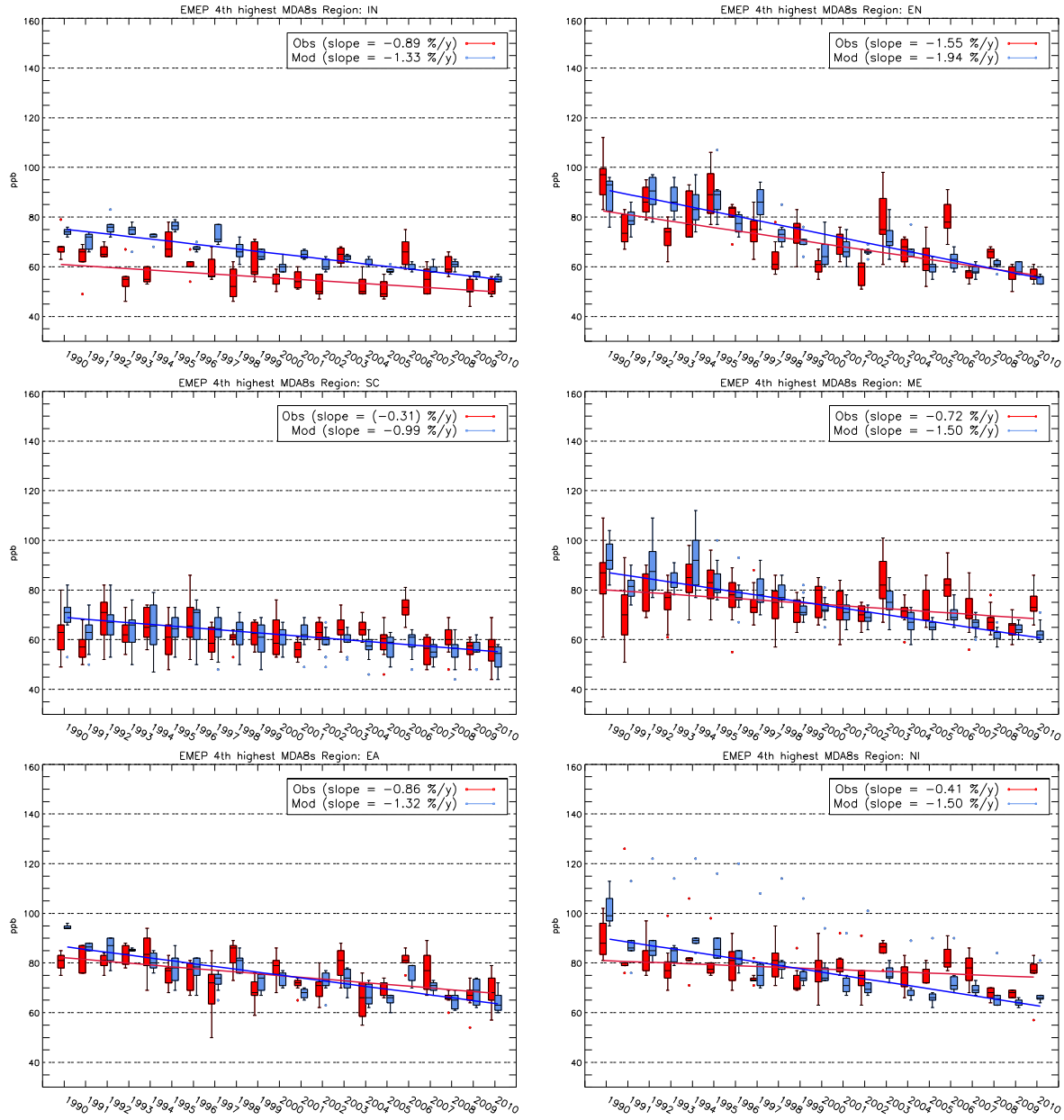


Figure 19. Same type of plot for the 4th highest MDA8 as in Figure 18 for each region, separately, based on the EMEP model results.

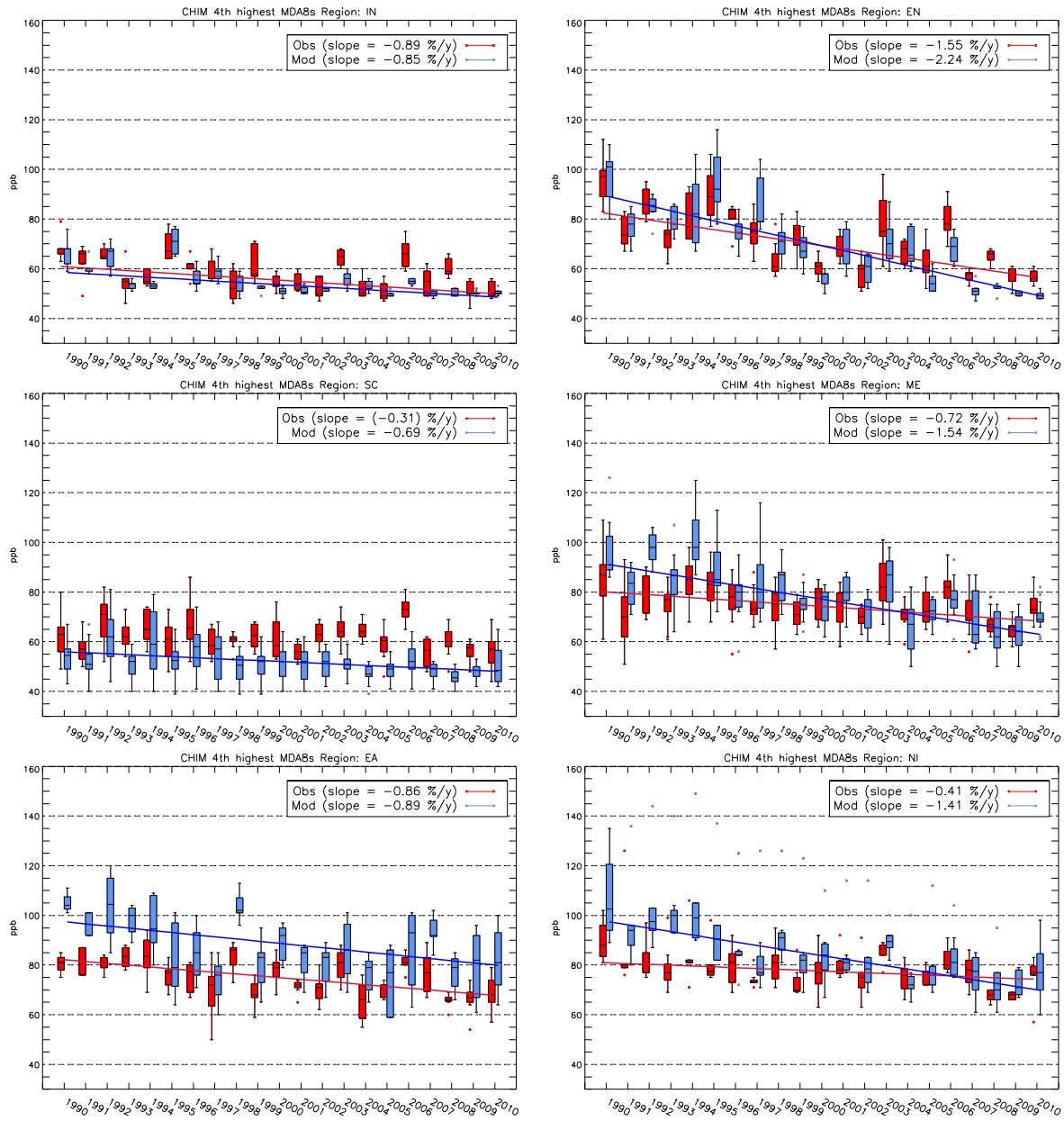


Figure 20. Similar fig as Figure 19 based on the CHIMERE model results.

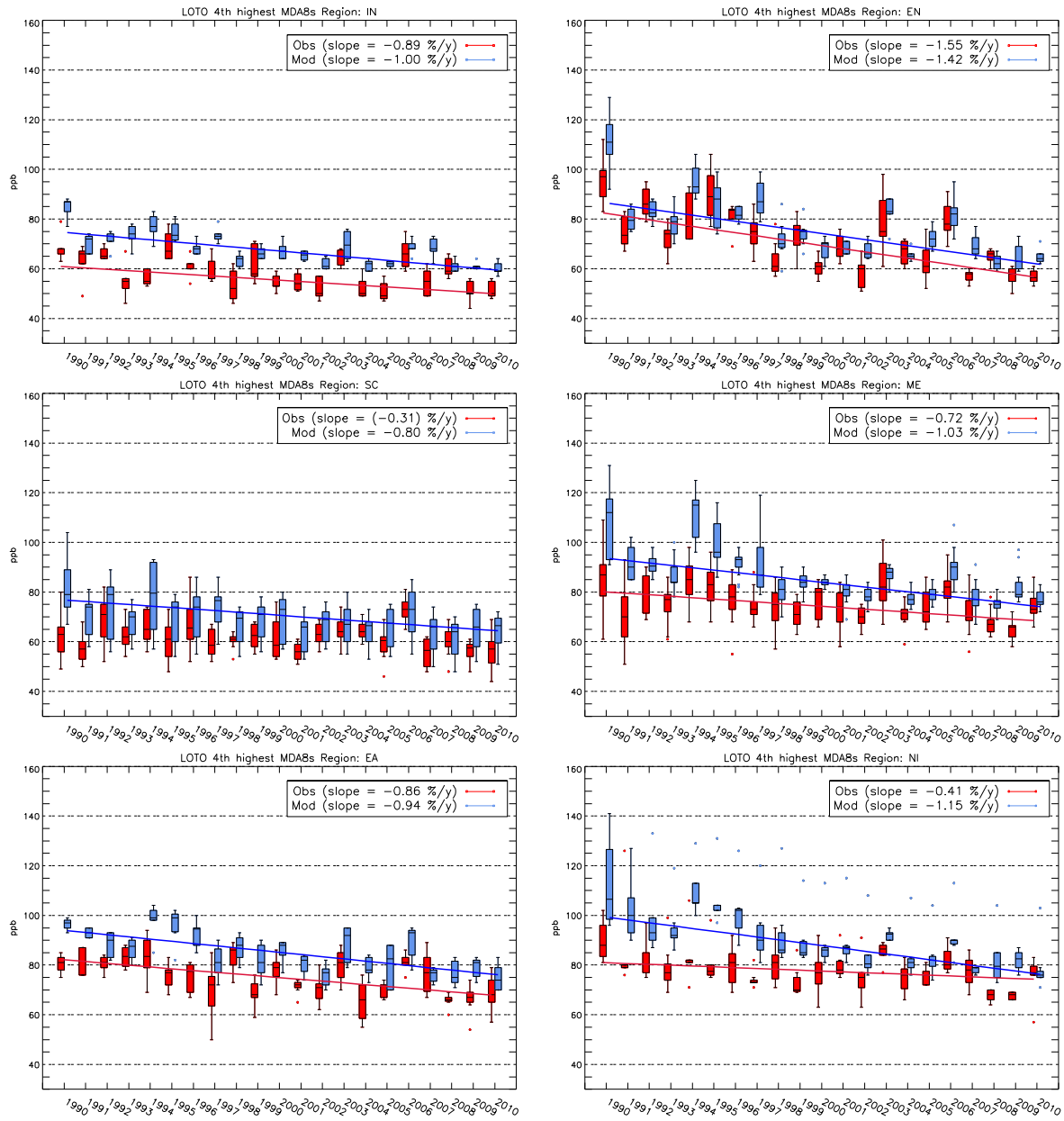


Figure 21. Similar fig as Figure 19 based on the LOTOS model results.

5.1.2.3 Annual mean ozone

Based on the KZ filtering method described above, we calculated the $O_{3,\,LTM}$, i.e. the long-term mean ozone concentration during the period 1990-2010 for all stations. This was done for the observations as well as for each of the three models EMEP, CHIMERE and LOTOS based on the EuroDelta-Trends results. Figure 22 shows the grouped means and spread (min/max) of the $O_{3,\,LTM}$, for the six regions defined in figure 4. The results show marked differences between the regions and models.

The modelled $O_{3,LTM}$ data are in general higher than the observed data for nearly all combinations of model/region. There is furthermore a systematic difference between the

model concentration levels being ordered CHIMERE<EMEP<LOTOS. Thus, in absolute values, the CHIMERE model is closest to the observational data.

The “Inflow” and “England” regions show an almost flat timeline in the observations over the whole period, with indications of just a tiny drop in the long-term mean, and this is well reflected in all the three models, although with a certain bias. We also note that the spread in the observational Inflow region data is larger than we would expect for a homogeneous region representing the background. This may reflect that the Inflow region as defined here is less coherent than believed. This region consists of only five stations from Mace Head at the west coast of Ireland and Lough Navar in Northern Ireland to three sites in Scotland (Strat Vaich, Bush, Eskedalmuir). However, both the long-term trend (or lack of that) and the inter-annual variability are very well captured by the models.

The Scandinavian region (SC) shows an upside-down U-form with a peak around 2000 and 2003 and lower concentrations at the start and end of the period. The models, however, all show a nearly flat timeline. The peaks in 2000 and 2003 and the dip in between is very well reproduced by the models, though. The sites defining this region show a fairly small span in values, indicating coherent and robust data within this area.

The largest discrepancy between the observational and model data is seen for the North Italy (NI) region. The observations in the NI region show very low values in the 1990s compared to the models. Furthermore, the models, particularly EMEP and LOTOS, show a marked drop in this ozone indicator for the NI region, whereas the measurements show a nearly flat development with a small peak around 2003. Differences between the models are seen also in other regions. For the Middle Europe (ME) region, the EMEP model (and to a less extent LOTOS) gives a marked drop in levels whereas the results from CHIMERE show a flatter development, more in line with the measurements.

For the EA region there is a positive bias (model predictions higher than measurements) for all three models in the first part of the 1990s. After 1995 the CHIMERE results agree well with the measurements whereas there is a gradual reduction in bias for the EMEP model results compared to the measurements. LOTOS tends to predict systematically higher values during the whole period.

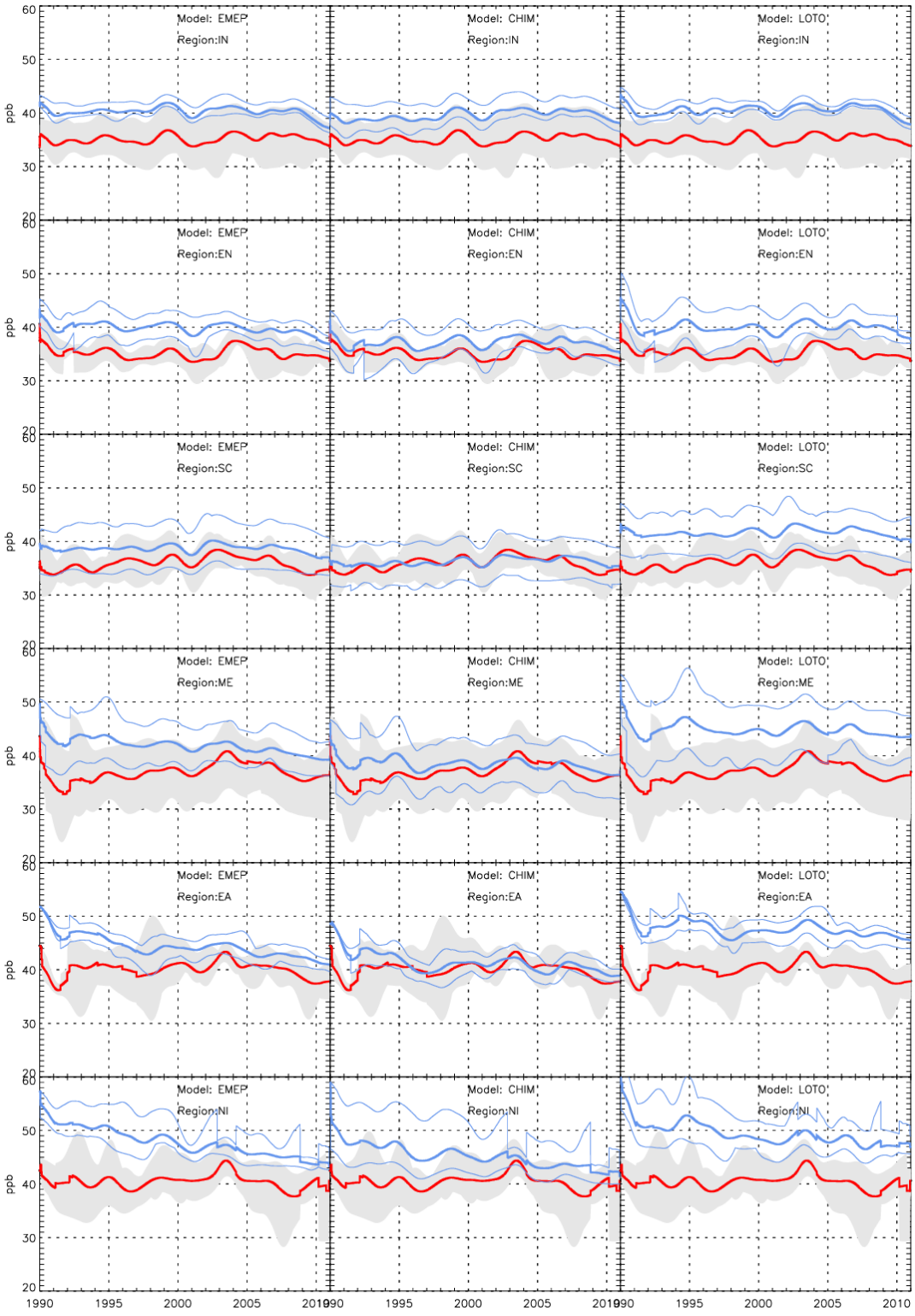


Figure 22. Long-term mean ozone concentrations, $O_{3,\ LTM}$, as calculated by the KZ methodology for the observations and three models for six geographical regions as given in the panels. Red lines mark the mean of the observations and the grey shading the min and max of these data. Thick blue lines mark the mean of the model values while the thin blue lines mark the min and max. The sites included in the various regions correspond to those given in Figure 6 and defined in 4.2.1.

5.2 Particulate Matter Trends

In this section, we assess the evolution of model performances by investigating if there are any long-term trends in their scores when comparing models and observations (Section 5.2.1). We also assess their capacity in capturing the trends of aerosols reported in observations (Section 5.2.2). The main focus is for the 2000-2010 period given the scarcity of the network in the 1990s, especially in terms of total PM_{10} and $PM_{2.5}$, although more years are also included when observations are available.

5.2.1 Change in model performances

A restricted set of stations (rural and sub urban) has been used to calculate usual error statistics (root mean square error, correlation and bias), the results are presented in Figure 23. Since we are interested in the evolution of model performances rather than comparing the trends, we use a subset of the selection presented in Section 4.2.3 to obtain a consistent set of stations for the whole period. The 75% coverage criterion is maintained within any given year for all stations, but the completeness criteria is 100% in terms of number of years covered. WRF-CHEM, CMAQ and POLAIR only delivered the 2000 and 2010 years, so their year-to-year variation cannot be compared to the other models.

A general underestimation is seen for PM_{10} (Figure 23, left) This is due to the usual negative bias observed for this species particularly during the winter period and also to the relatively coarse spatial resolution for the models to be compared to some sub-urban stations. The magnitude of the underestimation is smaller for $PM_{2.5}$ showing that the models have more difficulty in capturing the coarse fraction. Sub-urban stations were included in the analysis to increase the dataset of stations that was too small considering only the rural stations. MINNI and CHIMERE gives the best correlation either for $PM_{2.5}$ and PM_{10} concentrations, the biases are close to -6 µg m^{-3} for the models which delivered the full period. For the common year 2010, it is interesting to note the capability of CMAQ and WRF-CHEM to simulate higher PM_{10} concentrations which seem closer to the observations but WRF-CHEM overshoot for the $PM_{2.5}$ concentrations in winter. POLAIR strongly underestimates winter concentrations in 2010 leading to a poor reproduction of the seasonal cycle (monthly correlation of 0.12 for this year) while CMAQ gives the best shape of the seasonal cycle despite low $PM_{2.5}$ concentrations in summer. The error statistics of the ensemble mean (ENSEMBLE) are the best since they take advantage of each individual model.

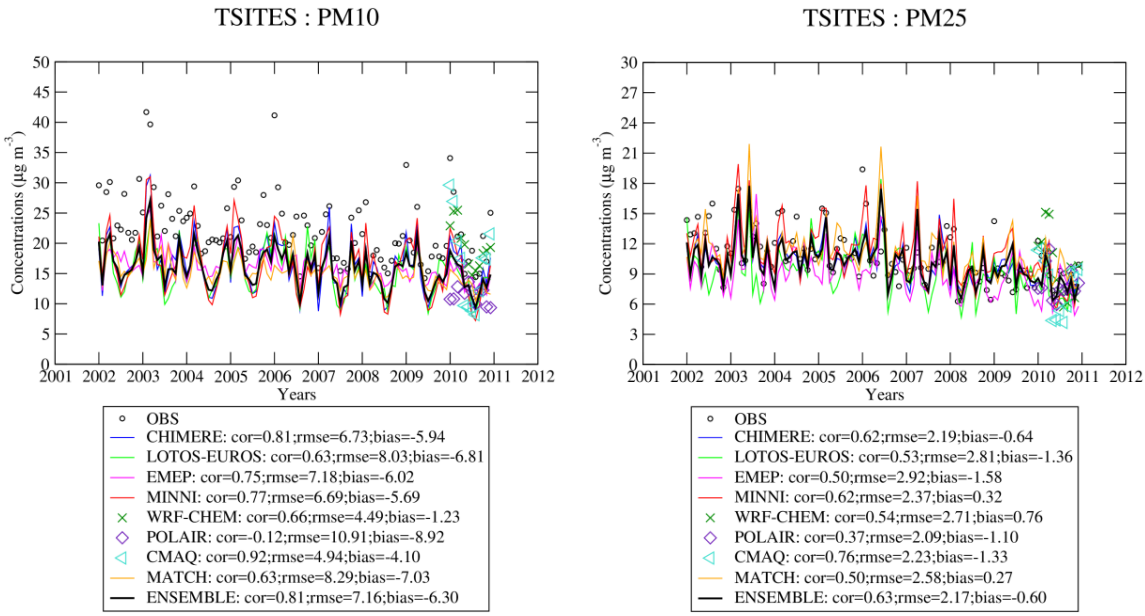


Figure 23: Evolution of monthly PM_{10} (left) and $PM_{2.5}$ (right) concentrations over the 2002-2010 period for rural and suburban stations providing at least 75% of measurements each year.

The model spread is smaller compared to what was reported in (Bessagnet et al., 2016) for short term field campaigns. For the analysis of the "ensemble" of the five models which delivered the full period we use a coefficient of variation *VAR* which quantifies the disagreement between models and is defined as follows:

$$VAR = \frac{1}{C_{ENS}} \sqrt{\frac{1}{M} \sum_{m} (C_m - C_{ENS})^2}$$

where C_m is the concentration of the individual model *m* included in the ensemble (CHIMERE, LOTOS-EUROS, MINNI, MATCH and EMEP), M is the number of models (5), and C_{ENS} is the mean concentration of the *ENSEMBLE*.

The coefficient of variation for the PM_{10} (Figure 24) in each subregion varies from about 12% over EA, FR, NI, ME, EN to about 25% for SC, IN and MD. The latter areas are either exposed to low concentrations or influenced by boundary conditions and therefore affected by natural emissions (sea-salts and desert dusts) that are difficult to estimate. There is no significant trend of this coefficient for PM_{10} except for the Iberian Peninsula (IP) area where a slight increase is observed. For $PM_{2.5}$, the coefficient of variation in each region is usually higher than for PM_{10} and a more significant positive trend is often observed.

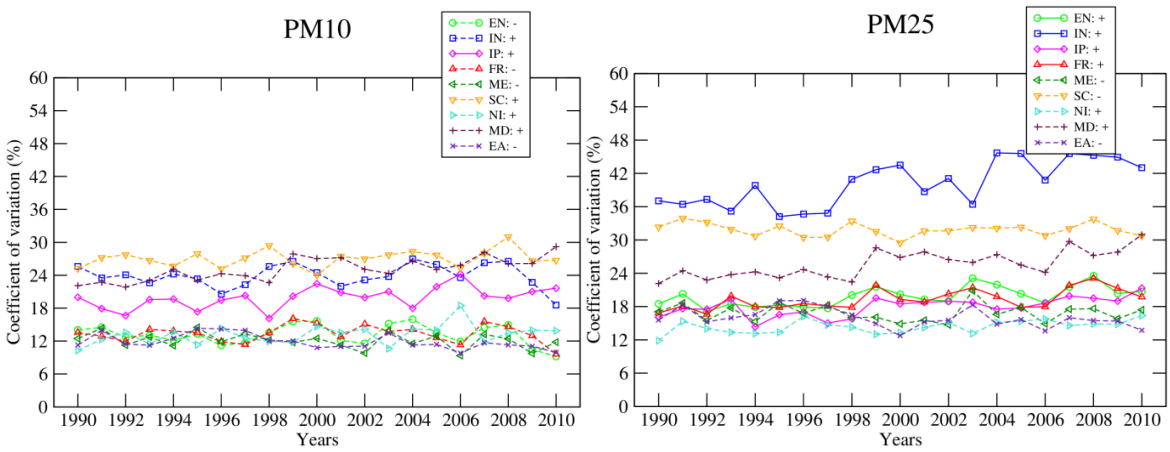


Figure 24: Evolution of the annual mean coefficient of variation of models for PM_{10} and $PM_{2.5}$ for each PRUDENCE zones. Dashed line is used when no significant trend is computed and solid line is used for significant trends.

The coefficient of variation is very different from species to species. Concerning nitrate, sulphate and ammonium concentrations for the IP area, a clear decrease can be observed for nitrates while the ammonium and sulphate show rather flat insignificant trends (Figure 25). For all areas the coefficient of variation is the highest for nitrates (40-60%) suggesting very different way to treat the chemical equilibrium for particles related to the chemical regimes in models (load of total nitrate *versus* total ammonium).

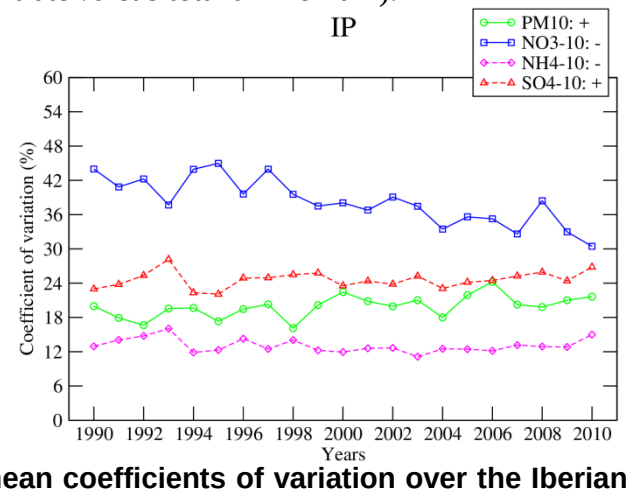


Figure 25: Annual mean coefficients of variation over the Iberian Peninsula zones for PM_{10}, nitrate, ammonium and sulphate concentrations. Dashed line is used when no significant trend is computed and solid line is used for significant trends.

Figure 26, Figure 27 and Figure 28 show the evolution of usual error statistics for PM_{10} (RMSE, Bias and correlation) based on daily values for each subregion zones. Only rural and sub-urban stations are presented. As observed in previous works for PM_{10} concentrations ((Bessagnet et al., 2016) and reference there in), a large negative bias is observed for all models in the range -60 to -40%. EMEP produces the lowest absolute biases while LOTOS-EUROS exhibits the highest negative biases but it is very dependent of the regions, and for instance, CHIMERE gives the highest negative bias over the Iberian Peninsula while MINNI shows the lowest bias over Scandinavia.

On average, there are no significant trends in the bias, the England area shows the lowest biases with an insignificant negative trend, in this area the bias is positive for MINNI and EMEP in 2002. For most models, it is over Scandinavia that the largest negative bias and the worst correlations are found. In this region the concentrations are the lowest and are influenced by boundary conditions and biogenic emissions (biogenic VOC precursors) as shown in (Bessagnet et al., 2016).

In Scandinavia, there is also a clear negative trend on correlations while in most other regions no significant trend is observed. CHIMERE gives on average the best correlations for most regions while LOTOS-EUROS exhibit the lowest ones, but as for the bias this depends on the region, for instance EMEP and MINNI give the lowest correlation over the North of Italy area (NI). The correlations range from 30% to 70% and are close to 10% in Scandinavia in 2010.

The combination of trends in bias and correlation result in the evolution of RMSE presented for information in Figure 28.

Figure 29 gives an alternative view of the trends for the bias and correlations by mapping the trend of those scores over Europe. Over Germany, the absolute bias tends to increase with some significant trends at some stations. For France, CHIMERE, LOTOS-EUROS and MINNI show an increase of correlation while EMEP and MATCH exhibit a slight decrease.

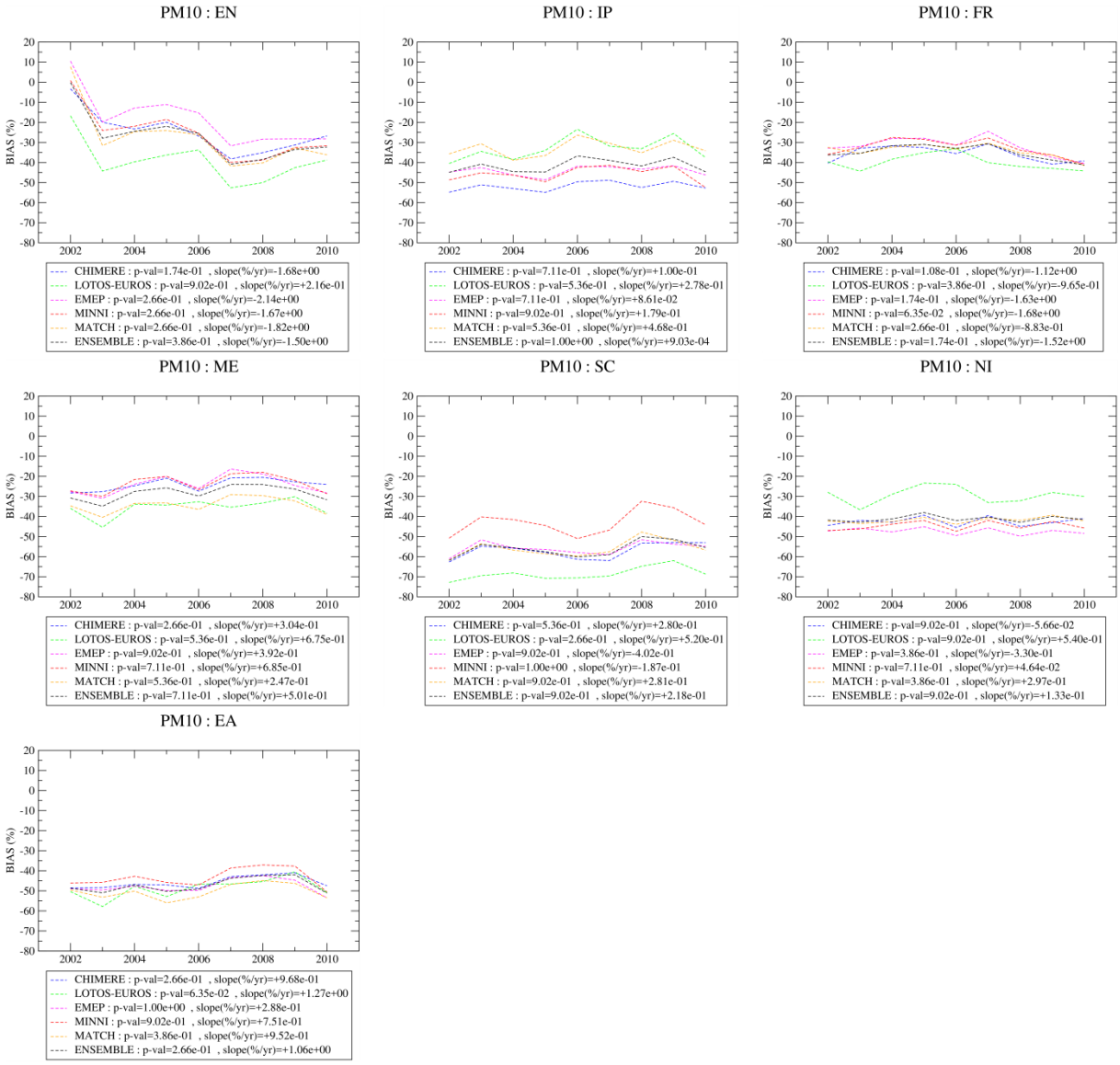


Figure 26: Evolution of the normalized bias for PM_{10} concentrations over the 2002-2010 period for each PRUDENCE zones. Dashed line is used when no significant trend is computed and solid line is used for significant trends (if any). The slopes are given in percentage points per year since error statistics are already expressed in %.

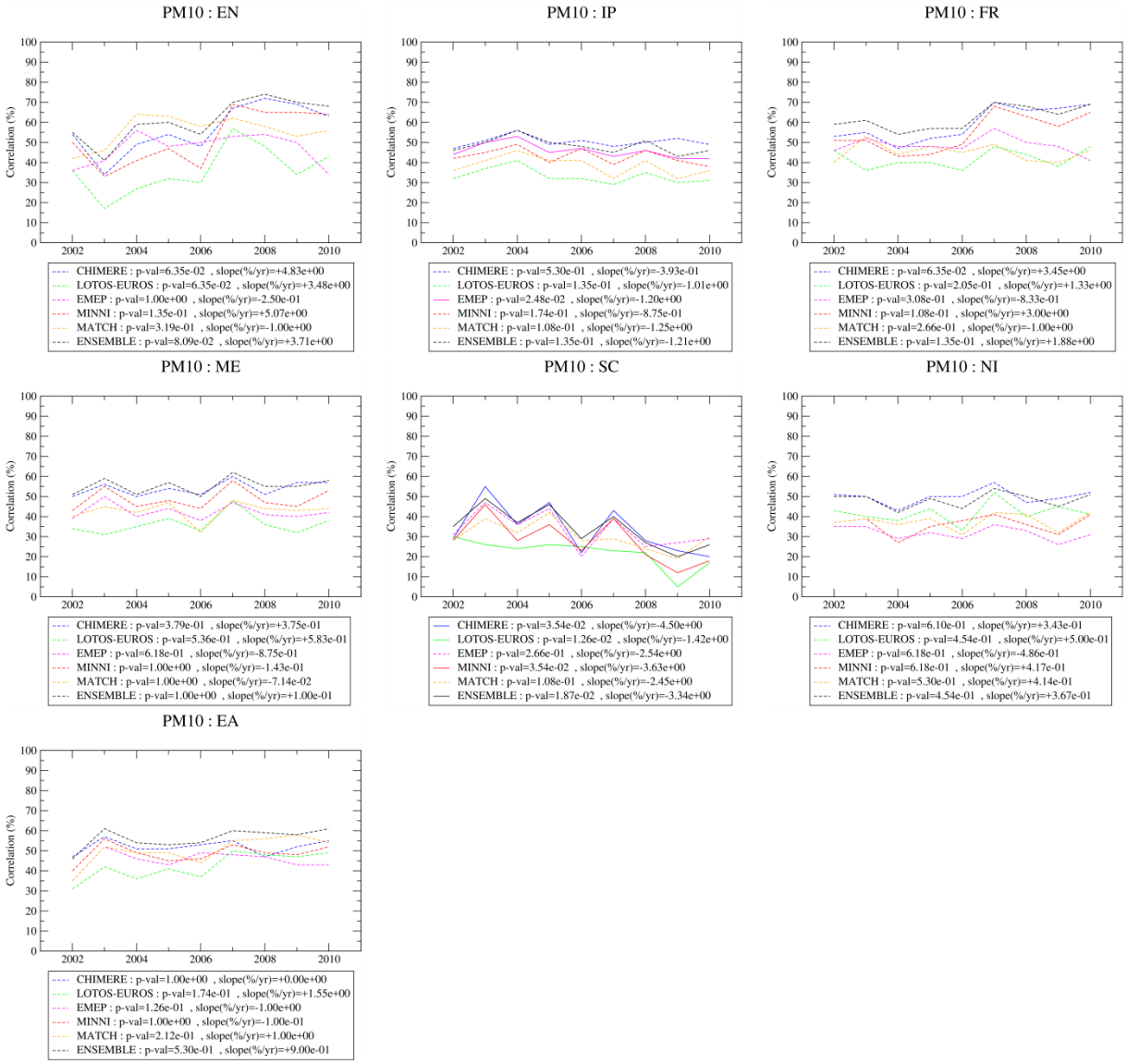


Figure 27: Evolution of the correlation for PM_{10} concentrations over the 2002-2010 period for each PRUDENCE zones. Dashed line is used when no significant trend is computed and solid line is used for significant trends (if any). The slopes are given in percentage points per year since error statistics are already expressed in %.

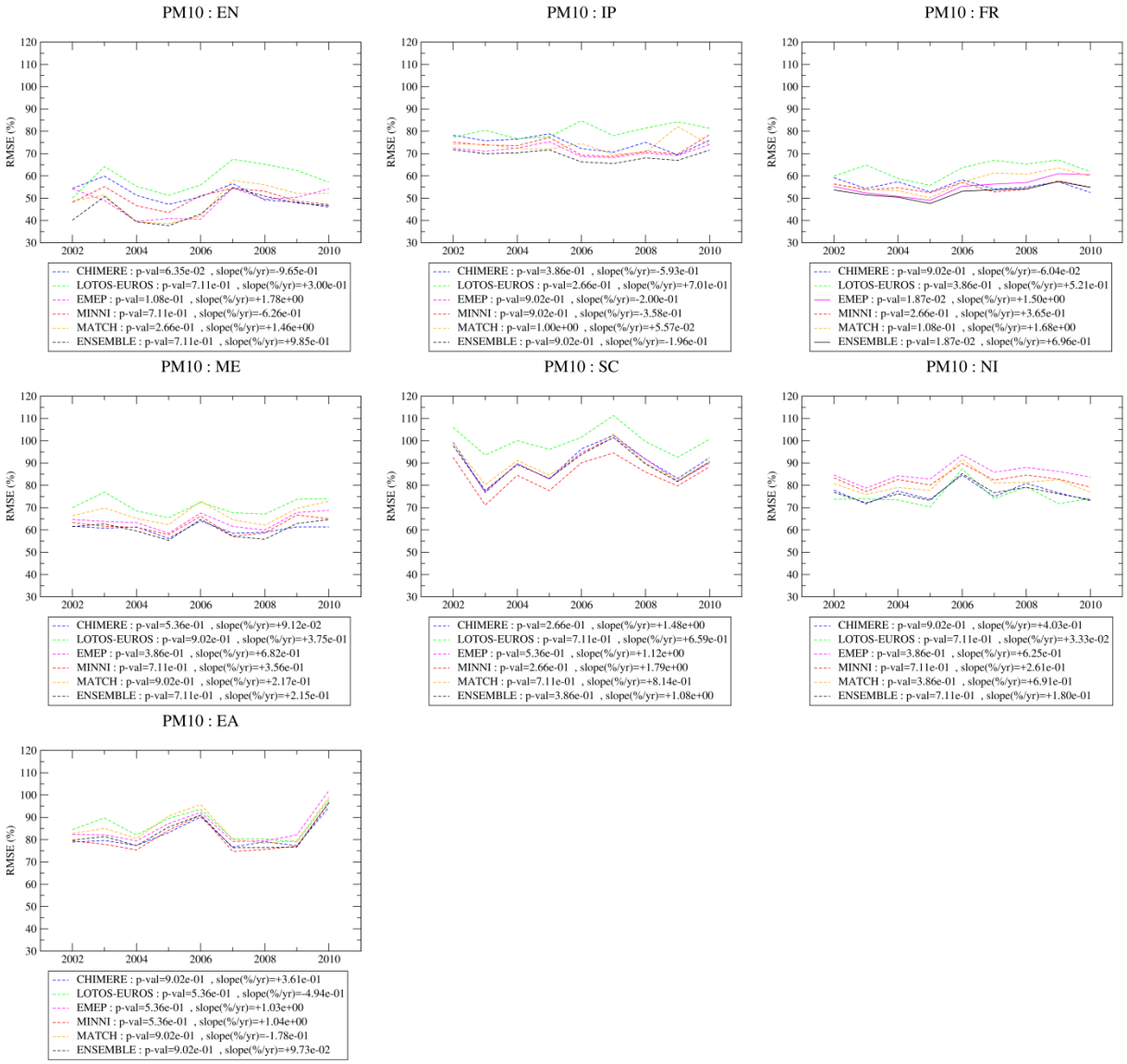


Figure 28: Evolution of the normalized root mean square errors for PM_{10} concentrations over the 2002-2010 period for each PRUDENCE zones. Dashed line is used when no significant trend is computed and solid line is used for significant trends (if any). The slopes are given in percentage points per year since error statistics are already expressed in %.

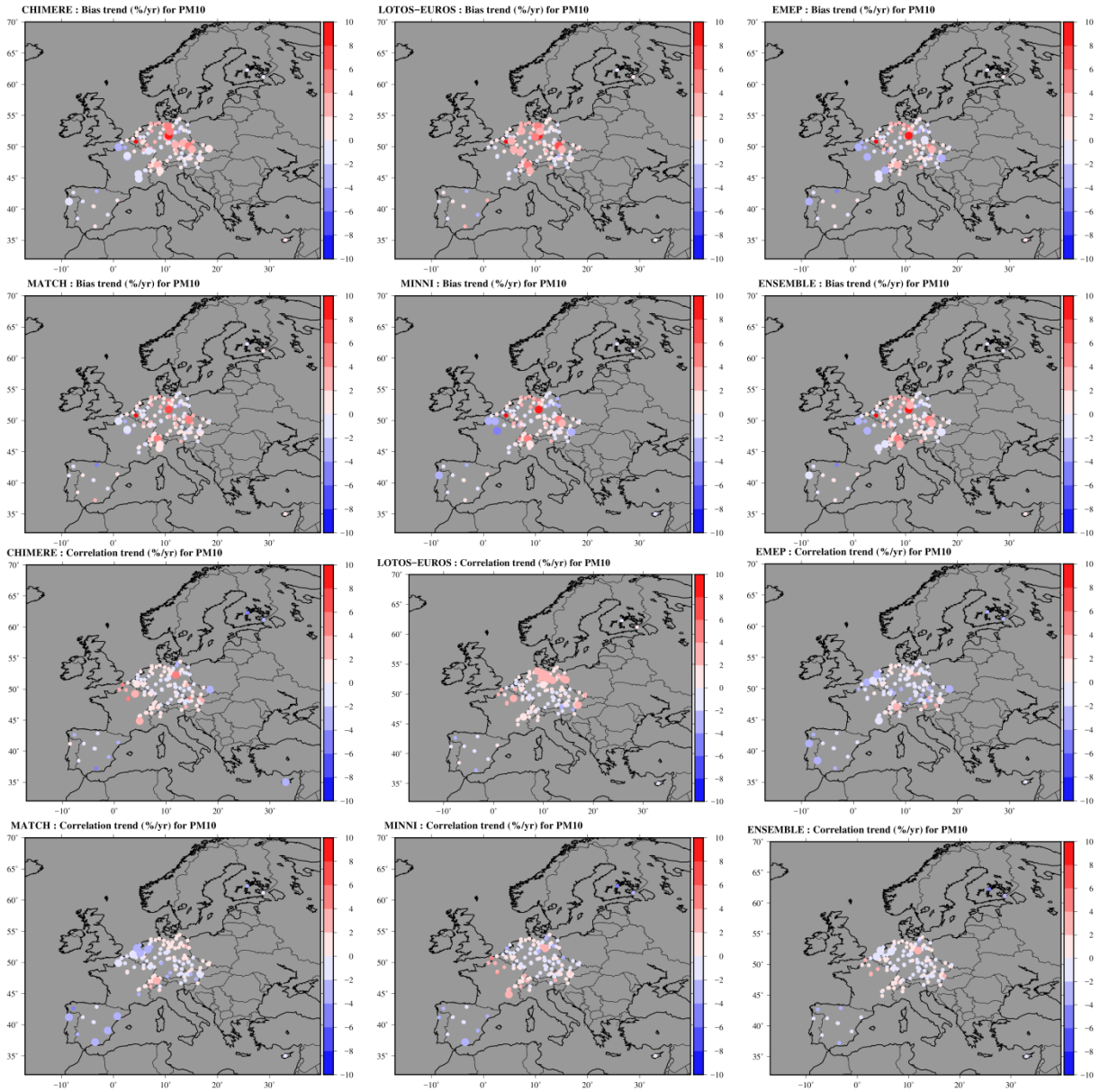


Figure 29: Trends of biases and correlation for the PM_{10} rural and sub-urban stations selected for the computation statistics. Large circles correspond to significant trends while the colour is related to the range and magnitude of trends. The slope corresponds to the %points of the error statistics.

When looking at the performance of models on secondary inorganic compounds (sulphate, nitrate and ammonium), we find a clear underestimation of sulphate concentrations for CHIMERE and LOTOS-EUROS (Figure 30). For CHIMERE, this can explain a large part of the strong positive bias of nitrate concentrations and then on ammonium concentrations. Most of models exhibit good correlations between 0.6 and 0.7 for all inorganic species, only LOTOS-EUROS is clearly outside this range with correlations between 0.4 and 0.6 and particularly low for sulphates. No clear trends on performances are observed, a slight insignificant increase of correlation accompanied by a decrease in bias an RMSE is observed for nitrates. A slight decrease of the bias for sulphate is observed.

Figure 30: Evolution of the biases, normalized root mean square errors and correlations for sulphate, nitrate and ammonium over the 2002-2010 period for all EMEP stations. Dashed line is used when no significant trend is computed and solid line is used for significant trends. The slope corresponds to the %points of the error statistics.

5.2.2 Comparison of observed and modelled PM trends

The models involved in the Eurodelta-Trends exercise demonstrate a good capability in capturing average particulate matter trends over Europe. The time period assessed in this section is 2000-2010 because of the scarcity of the PM network in the 1990s (cf. Section 4.2.3). Since we focus here on the long-term trend, only the models having delivered results for the whole period are included here (CHIMERE, EMEP, LOTOS-EUROS, MATCH and MINNI). At the time of publication of the report, CMAQ, WRF-CHEM and POLAIR have only contributed to the sensitivity experiments for 1990, 2000 and 2010, and they are therefore not taken into account for the analysis.

A comparison with observed Sen-Theil Slopes is given in the scatterplot of Figure 31. It shows that the models capture very well the sign of the trend even if the range of variability

is systematically underestimated (see the much wider range of slopes observed than modelled). The percentage of good detection of the trend is 100% for all models when focussing on sites where the observed trend is significant. Including also the sites where the observed trend is not significant, the rate of good detection of the sign of the trend remains 84% for all models. The 16% of sites where the models fail to capture the sign of the trend do not belong to specific countries or station typology (for example, urban versus rural). And all the models fail systematically for the exact same list of stations. It is therefore difficult to point out a flaw in emission inventories, or representation of a given process in the models whereas uncertainties in the observation records remain a possibility. It should however be emphasised that those missed trends are not significant, thereby minimizing the importance of capturing them in the models.

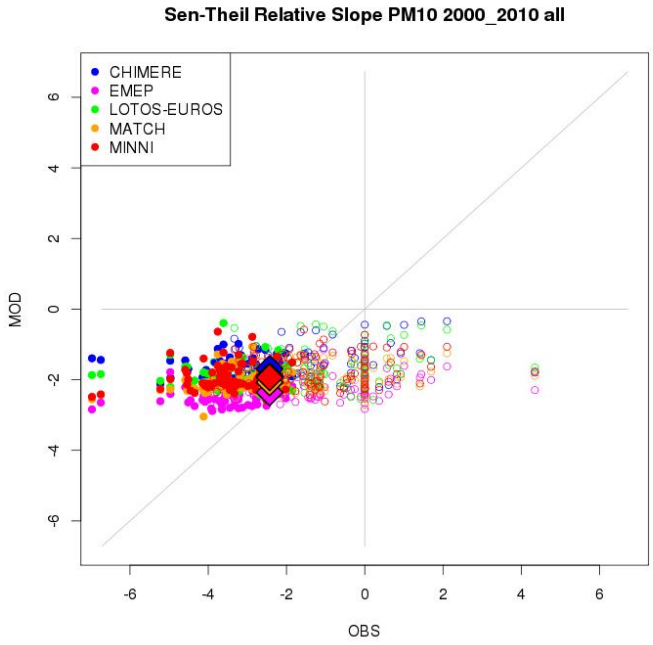


Figure 31: Scatterplot of observed (x-axis) and modelled (y-axis, for CHIMERE, EMEP, LOTOS-EUROS, MATCH and MINNI) relative sen-theil slope (%/yr) for PM_{10} at urban, suburban and rural background sites in the Air Quality e-reporting database. Empty dots are for the sites where the observed trend is not significant. Large diamonds correspond to the average over the whole network.

The map of PM_{10} trends in the composite of CHIMERE, EMEP, LOTOS-EUROS, MATCH and MINNI is given in Figure 32. It shows that the area where trends are significant in the models (coloured areas) is larger than in the EMEP observations, with a number of locations where the observational record indicates a non-significant trend (smaller black dots), whereas the models have a significant decrease.

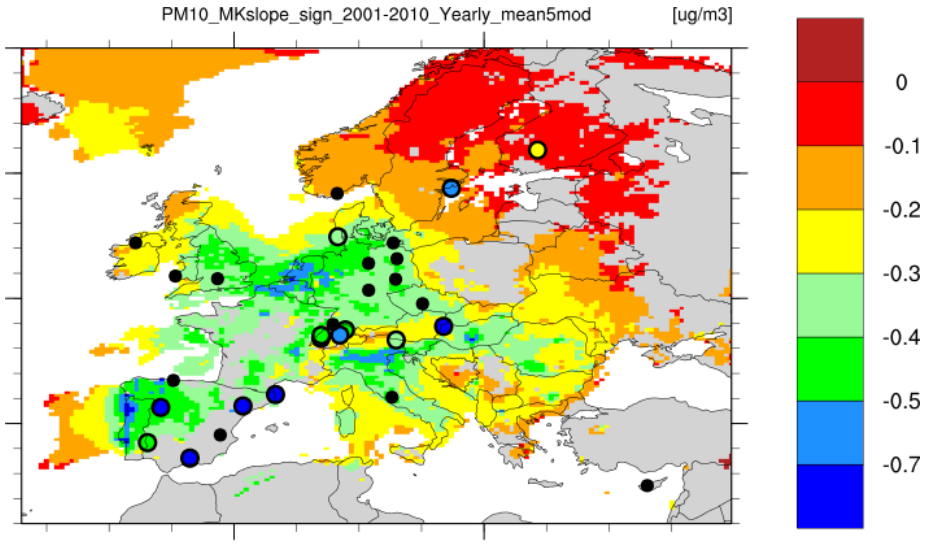


Figure 32: PM_{10} Trends (in µg/m^3) in the models (when significant, the areas are coloured) and observed trends at EMEP monitors (dots, black when insignificant), courtesy Tsyro et al, in prep, 2017

The composite time series of European-wide PM_{10} evolution is given in Figure 33, either in absolute values, or normalised to 2000 levels and presented in %. All the models are interpolated at the location of the urban, suburban and rural background stations, and the composite is the average over all stations. The ensemble (as the average of 5 models) is also displayed. The models display the usual low bias in capturing PM_{10} concentrations, but the positive feature for this trend exercise is that the bias appears constant in time. The relative change presented in the lower panel demonstrates clearly that the models are very good at capturing relative PM_{10} changes since the early 2000. The various models differ in their capacity to capture the interannual variability but the ensemble average performs well. The observed relative change in PM_{10} over the urban, suburban, and rural background sites of the Air Quality e-reporting database is -1.87%/yr and the Eurodelta-Trends ensemble mean gives -1.94%/yr. The relative trend is slightly overestimated for urban sites (-2.04%/yr modelled versus -1.87%/yr observed), but on the contrary it is underestimated at suburban and rural sites (-2.00%/yr modelled versus -2.25%/yr observed).

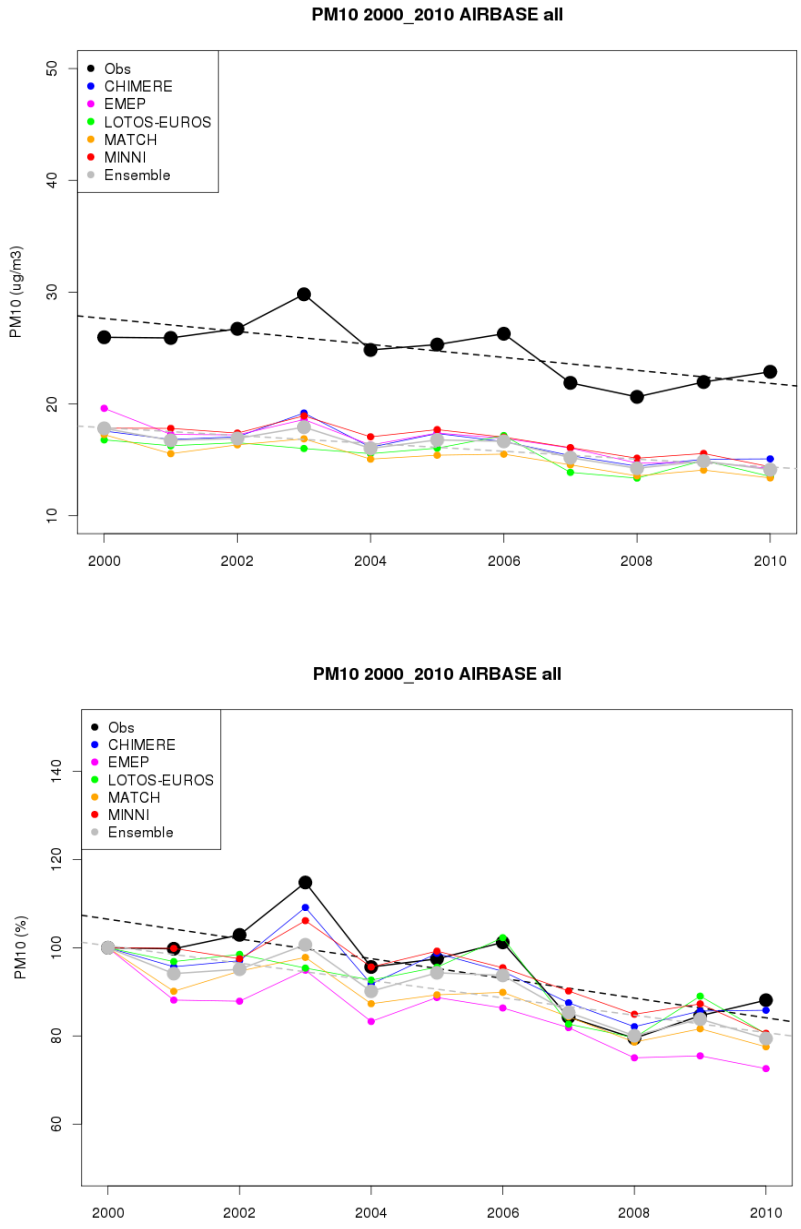


Figure 33: Average time series of annual PM_{10} (absolute values, top; normalized to 2000 levels, bottom) at all background stations in the Air Quality e-reporting database for the observations (black) and individual Eurodelta-Trends models (coloured dots) and the average ensemble (grey).

6 Mapping the evolution of air pollution over 1990-2010

One of the main added values of air quality models compared to in-situ observations is their extensive spatial coverage. The results of the Eurodelta-trend exercise can be used to map the evolution of air quality over Europe over the past period 1990-2010. One should bear in mind the limitations of the models pointed out in Section 5 when comparing them to in-situ observations, but that section also showed the average good performances, especially for the multi-model ensemble that minimises the uncertainty that can be attributed to imperfect knowledge and representation of chemistry-transport processes in state of the art tools.

Spatial composite maps for ozone peaks (as the fourth highest annual value of MDA8: the daily maximum ozone computed on the basis of the 8-hr running mean) and PM_{10} (annual mean) for 1990, 2000 and 2010 are provided in Figure 34 and the corresponding changes are provided in Figure 35 and Figure 36. In all cases, at each grid point the mean over the ensemble is given (note that the median is often used to compute an ensemble but the mean is preferred here given the relatively small size of the ensemble).

In Figure 34, PM_{10} average and O_3 peaks are displayed on common colour scales for 1990, 2000 and 2010 for all eight models having contributed to the reference simulations for 1990, 2000, 2010 (CHIMERE, CMAQ, EMEP, LOTOS-EUROS, MATCH, MINNI, POLAIR, WRF-CHEM).

For the same models, the relative difference between 2000 and 1990, and between 2010 and 2000 are given in the top row of Figure 35 and Figure 36. It is expressed in $\mu g/m^3/yr$ (i.e. the absolute difference divided by 11: the number of years separating either 1990 and 2000 or 2000 and 2010) in order to be compared to the actual trends. The trend itself for the 1990s and the 2000s is displayed in the lower row of Figure 35 and Figure 36, it can only be computed for the 5 models that produced the continuous 1990-2010 simulation (CHIMERE, EMEP, LOTOS-EUROS, MATCH and MINNI – note that because of a delay in their delivery the latter two could not be included in the ozone model evaluation presented above). Although it is less robust in terms of model uncertainty (5 models instead of 8 for the 1990/2000/2010 simulations), those trends account for the interannual variability.

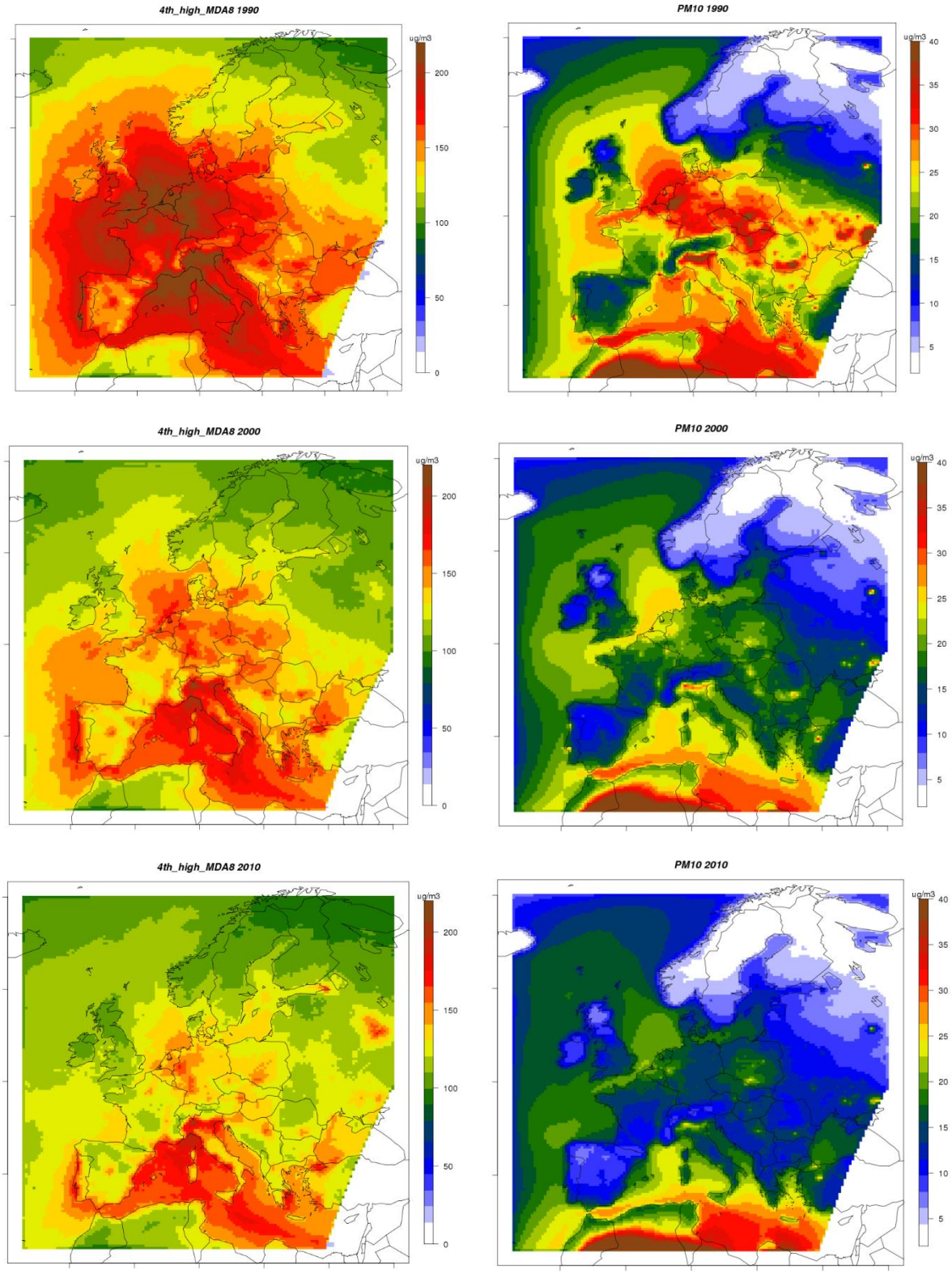


Figure 34: Average of daily ozone maxima (4th highest MDA8) over the April-Sept period (left, µg/m³) and annual mean PM_{10} (right: µg/m³) in 1990, 2000 and 2010 in the Eurodelta-Trend ensemble (8 model mean).

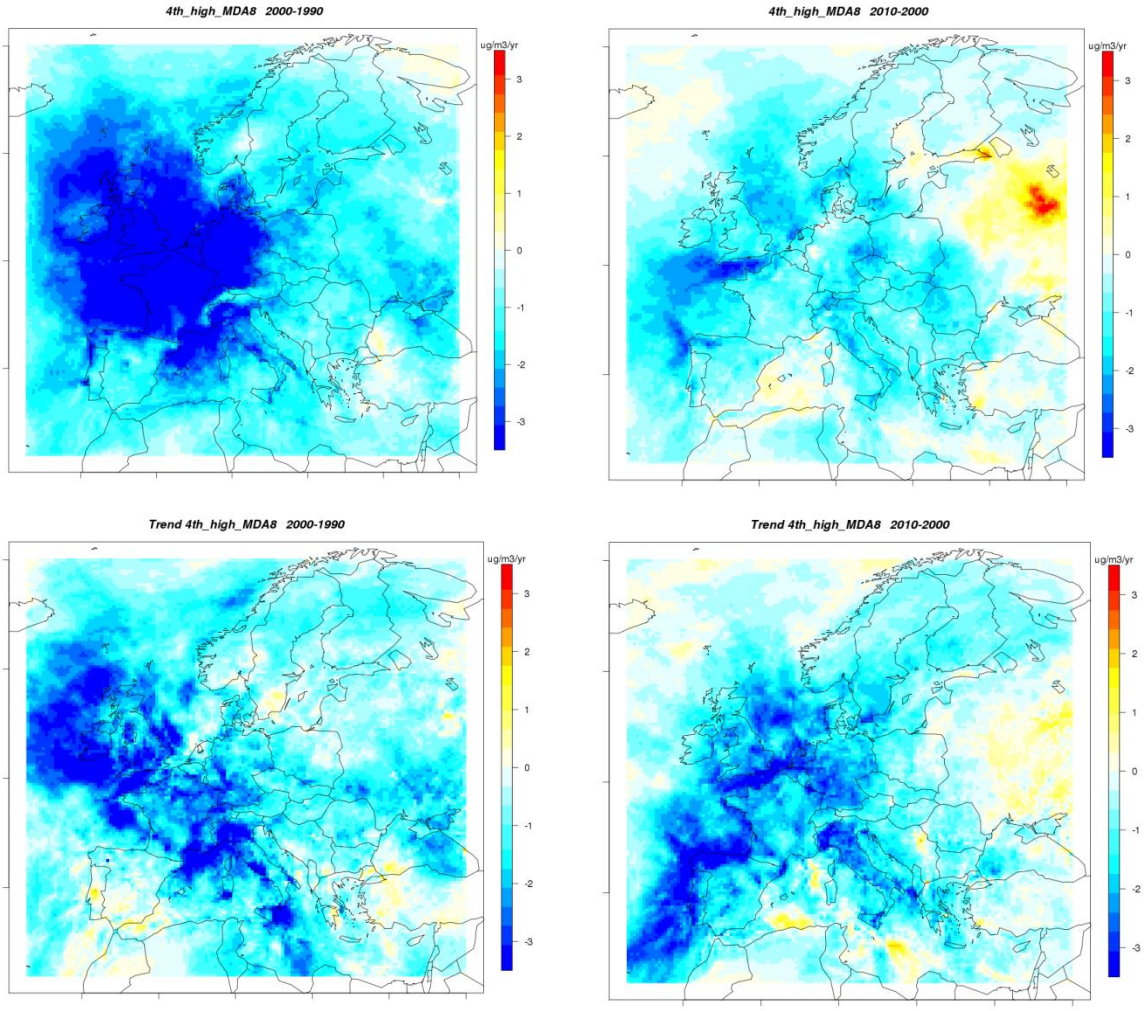


Figure 35: For summertime ozone peaks (4th highest MDA8): top row: the difference 2000-1990 and 2010-2000, respectively, in the same 8 model ensemble. Bottom row: 10-years trend in the 1990s and 2000s for the 5 model having contributed continuous simulations (µg/m^3/yr in both cases).

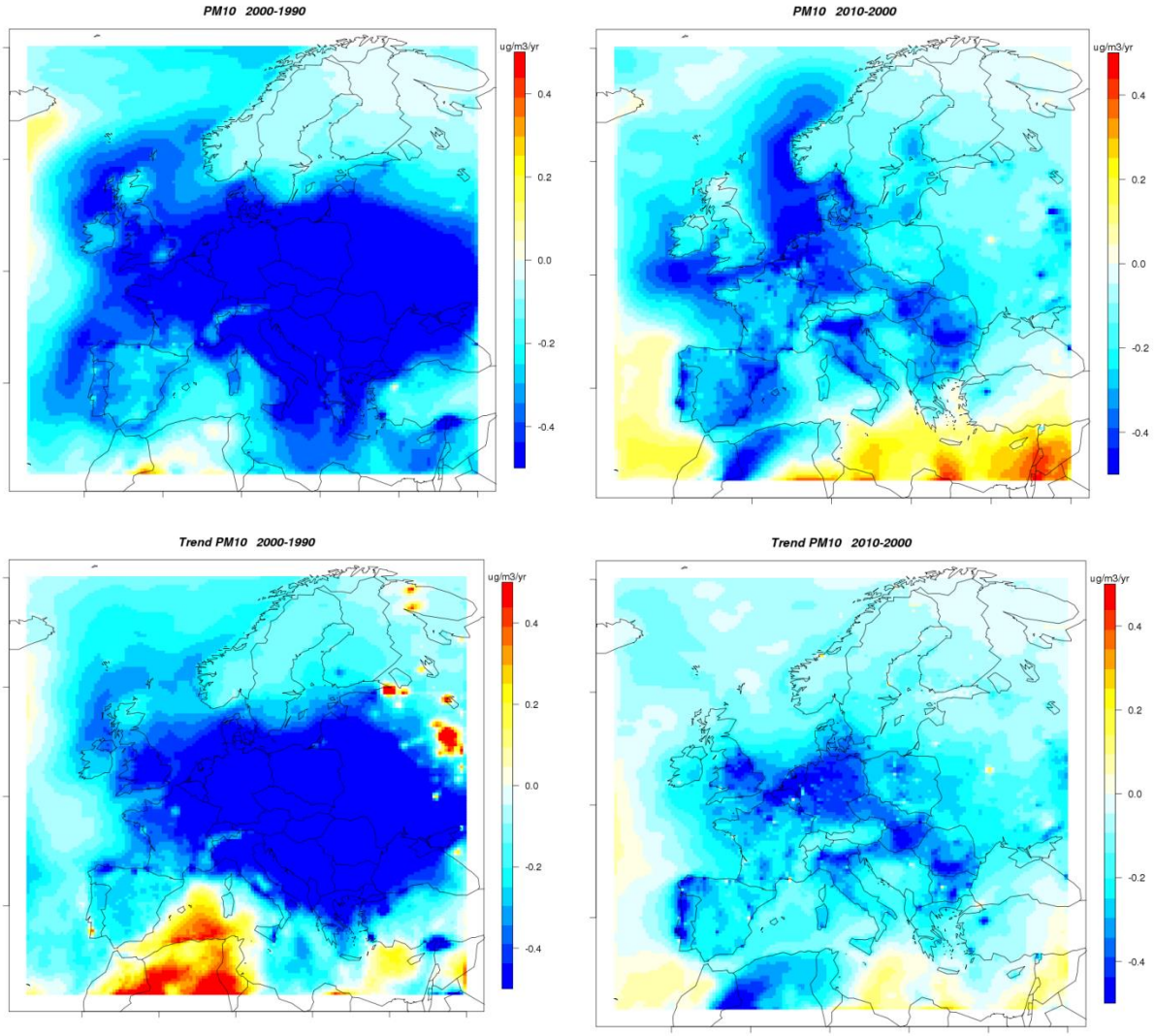


Figure 36: For annual mean PM_{10}: top row: difference between 2000-1990 and between 2010-2000 in the same 8 model ensemble. Bottom row: 10-years trend in the 1990s and 2000s for the 5 model having contributed continuous simulations (µg/m^3/yr in both cases).

The maps showing ozone and PM_{10} on common colour scale for 1990, 2000 and 2010 in Figure 34 are very illustrative of the improvement of surface air quality over this 21 years period. A closer view into the evolution in the 1990s and the 2000s is provided in Figure 35 and Figure 36 by looking at the differences between: 2000-1990 and 2010-2000 (top rows). One should note however that these differences are sensitive to the year selected: because of interannual variability the difference between 2001 and 1991 could be very different from the difference between 2000 and 1990. That is why we present on the same figures in bottom rows the actual annual trends during either 1990-2000 or 2000-2010, for the five models having delivered the full continuous simulation: CHIMERE, EMEP, LOTOS-EUROS, MATCH and MINNI. These figures must therefore be analysed by comparing the top and bottom row to investigate the importance of using a full trend or individual years to assess the evolution of air quality over the 1990s and 2000s.

The evolution of summertime ozone peaks (as the 4th highest annual MDA8) is quite different if assessed from the difference between 2000 and 1990, or 2010 and 2000 than from the actual trend for either of these two 11-years periods. According to the differences, the

decrease would be much larger in the 1990s than in the 2000s. On the contrary, the trends give a very similar picture for both periods. This discrepancy highlights the need to account for the interannual variability when assessing long term changes. The steadier evolution assessed from the trends is also more consistent with the observations, as reported for instance in the TFMM Trend Assessment Report (Colette et al., 2016a), where the decrease of the 4th highest MDA8 in the EMEP network was of the order of 10% for both the 1990s and 2000s decades. Note that there are still some differences: for instance, a slightly larger decrease in France in the 1990s, or an increase in Russia in the 2000s, but those cannot be compared to observations since no long term EMEP stations were selected in those areas in the TFMM Report.

For particulate matter, the overall pattern is very similar in the trends and annual changes with larger decreases over the 1990s than over the 2000s. Larger discrepancies are found for the 2000s between the absolute difference and the trend, although they remain limited to oceanic areas and north Africa. Such differences are therefore illustrative of the fact that the choice of the meteorological year for the sensitivity analysis (2000 and 2010) has a strong impact on the natural contribution to PM_{10} through sea salt and desert dust. Note that this difference is not just due to the use of different subset of models as can be inferred from the difference between 2010 and 2000 in the five models used in the trend analysis (not shown).

In the 1990s, decreasing trends are widespread in continental Europe, the only areas where they are smaller are Spain, Scotland, Ireland and Scandinavia, with some increase modelled in large Russian conurbations. Although it is not possible to compare such modelled PM_{10} evolution with in-situ measurements, it should be emphasised that such larger trends in the 1990s are at least compatible with the observed evolution of the inorganic fraction of PM (e.g. -0.05 µg/m^3/yr in the 1990s versus -0.025 µg/m^3/yr in the 2000s for sulphate ions in the EMEP network (Colette et al., 2016a)).

7 Attribution Analysis

7.1 Introduction

One of the key motivations for investigating air pollution trends is to assess the efficiency of air pollution mitigation measures. The mitigation strategies implemented over the past couple of decades led to the reduction of anthropogenic emissions of air pollutants. In order to compare such trends in emissions and the observed evolution in ambient air, one must also take into account the impact of meteorological variability (e.g. the frequency of air pollution episodes) and the long-range transport of air pollutants. A very legitimate question is then the quantification of each factor, for instance to assess the relative importance of local mitigation measures and changes in the background air pollution.

Taking the example of ozone, average levels and exceedances vary from year to year due to the combined effect of weather pattern, inter-continental transport and European emission of ozone precursors. The sole influence of meteorology on ozone is thus not easy to separate since these processes are so tightly linked together. Actually, it's not conceptually obvious what is meant by a "meteorological adjustment" of the ozone levels or of the long-term ozone trends. Meteorology affects the ozone levels in many ways; high temperature and strong solar radiation promotes ozone formation, strong winds increase the atmospheric mixing and reduce the ozone peaks (but may lead to more effective ozone formation overall), biogenic emissions of organic precursors of ozone (isoprene) is typically increased with high temperature, etc.

Various methods have been used to try to separate and quantify the sole influence of meteorological variation on air pollution trends. This was discussed in detail in (Solberg. et al., 2015) with a specific focus on ozone. Such "meteorological adjustment" of surface ozone are generally based on some kind of regression between ozone and local meteorological parameters. However, the ozone levels at a point in space and time is determined by the history of the air mass during several days whereas the in-situ meteorological conditions at the monitoring site is not necessarily a relevant parameter for the link between ozone and meteorology. In addition, the non-linear couplings between air pollution, meteorology, emissions and chemistry challenges the design of simple "subtraction" of the meteorological influence of ozone.

Instead we propose here to rely on chemistry-transport model experiments performed in the framework of Eurodelta-Trends, which was specifically designed to perform such attribution analyses. As presented in Section 4.1, eight modelling groups contributed to long term simulations covering Europe and the 1990-2010 period. The experiment was divided in several tiers for the reference simulation of 1990/2000/2010, sensitivity analysis investigating the role of emissions and boundary conditions, and 21 years continuous trends.

The methodology for the decomposition is presented in Section 7.2 and first results in relation with the decomposition of the changes (the difference, not the trends) in air pollution between the years 1990/2000/2010 are given in Section 7.3. Specific focus on the possible sources of uncertainties inherent to the Eurodelta-Trend design is shown in Section 7.5, in particular with an alternative approach based on the use of an ozone climatology in Section 7.4.3. The results devoted to the attribution of the actual trends are shown in Section 7.5 for

ozone, and total PM_{10}, including a focus on the natural contribution to the particulate matter mix.

7.2 Methodology

The Eurodelta-Trends multi-model experiment was designed specifically to allow for an attribution of the respective effect of European emission changes, meteorology and boundary conditions through the three tiers of experiments introduced in Section 4.1

Taking as example the evolution between 1990 and 2010, we can proceed to the decomposition illustrated in Figure 37 (note that with the EDT experiment design, such a decomposition can also be performed for the 1990-2000 and 2000-2010 time periods). The simulations are labelled for instance as M10B10E10 when the reference year 2010 is used for the meteorology (M), boundary conditions (B) and emissions (E). The last two digit of the relevant year change for each combination.

The complete list of simulations performed in the Eurodelta-Trends experimental plan are summarised in Table 2.

Table 2: List of model simulations available in the Eurodelta-Trend experiment for each Tier. The labels are for the Meteorological (M), Boundary (B) and Emission (E) year indicated with its two later digits.

Tier	Label
1A	M10B10E10
	M00B00E00
	M90B90E90
1B	M10B10E00
	M10B10E90
2A	M10B00E00
	M10B90E90
2B	M10C10E10
	M00C00E00
	M90C90E90
2C	M00B90E90
	M00B00E90
3A	MyyByyEyy, yy=1990 to 2010
3B	MyyByyE10 yy=1990 to 2010

Tier 2B is devoted to the sensitivity to the input data used as boundary conditions and is not used in the present report. Whereas in the remainder of the Eurodelta-trend experiment the boundary conditions are constrained with observations, in tier 2B, a global model is used – CamChem, hence the "C" in the label. Using a global model at the boundaries of the regional model allows being more exhaustive in terms of chemical forcing, but it also possibly introduces biases due to the global model itself.

Considering in a first stage the change between 1990 and 2010 (and not the continuous trend over the whole period), in principle 2^3 combinations are required to disentangle each contribution. All the possible combinations are indicated in the schematic of Figure 37.

In order to limit the number of required simulation, only 4 simulations where selected in the modelling experiment (M90B90E90, M10B90E90, M10B10E90, M10B10E10, displayed in black in Figure 37).

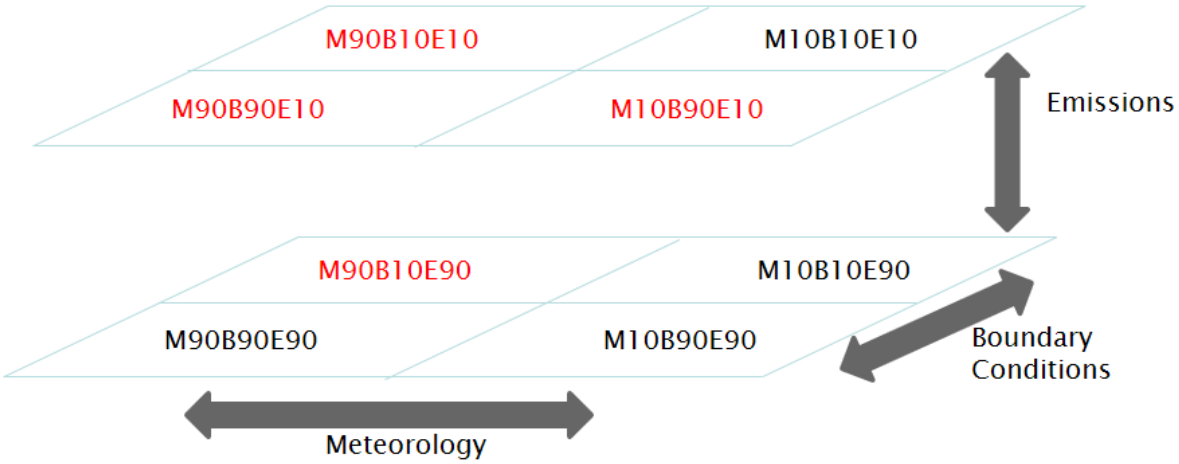


Figure 37: Attribution of air pollution changes between 1990 and 2010 using the sensitivity experiments in the Eurodelta-Trends modelling plan. Meteorology (M), Boundary Conditions (B) and Emission (E) changes are represented in the x, y, and z directions, respectively. The simulations are displayed in black/red when included/excluded from the experiment.

On the basis of these simulations, we can propose the following decomposition, where Δ stand for the absolute difference:

$$\text{Total Change (2010 – 1990)} = \Delta(\text{meteorology}) + \Delta(\text{boundaries}) + \Delta(\text{emissions}) + \text{residual} \quad (7.1)$$

Where

Total Change (2010 – 1990) = (M10B10E10 – M90B90E90): Tier 1A
Δ(meteorology) = (M10B90E90 – M90B90E90): Tier 2A
Δ(boundaries) = (M10B10E90 – M10B90E90): Tier 2A
Δ(emissions) = (M10B10E10 – M10B10E90): Tier 1B

One caveat of this decomposition lies in its sensitivity to the selected years (1990 and 2010). It is especially a concern for the change attributed to meteorology (which exhibits a higher interannual variability than the other two factors) where the 21 years change could be very different if addressed with the 2010-1990 and 2011-1991 changes. This point is addressed below when moving from differences to actual trends.

Note that the choice of using 2010 meteorology and boundary conditions to explore the sensitivity to emission changes is arbitrary (similarly using 1990 boundary conditions and emissions for the meteorology, likewise for the boundary condition sensitivity). Other combinations of those factors could have been used to explore the parameter space. The impact of this choice can be significant, in particular because the magnitude of the meteorological factor will depend on available anthropogenic precursors. The sensitivity to the meteorological factor, will be larger if computed with 1990 than 2010 emissions, so that the present decomposition is probably on the higher end. This hypothesis is further tested in Section 7.4.2 where we present the result of additional simulations from one of the participating models (CHIMERE) volunteered to produce the whole set of possible combinations. We also test this hypothesis by proposing in Section 7.4.3 an alternative method to assess the impact of meteorological variability based on 21 years simulations using either 1990 or 2010 emissions from another volunteer model (EMEP).

Performing a full decomposition for the whole period would require 21^3 annual simulations, which is too demanding in terms of computing resources. It is however possible to propose an alternative decomposition of the actual trend (τ) by involving the Tier 3 simulations that consists of Tier 3A: a full simulation of the 1990-2010 time period with changing emissions and Tier 3B: a full simulation of the 1990-2010 time period with constant 2010 emissions. The decomposition of the trend would therefore read:

$$\tau(1990\text{-}2010) = \tau(\text{meteorology}) + \tau(\text{boundaries}) + \tau(\text{emissions}) + \text{Residual} \qquad (7.2)$$

The total trend over the period (τ (1990 - 2010)) is computed with the tier 3A simulations (MyyByyEyy with yy=1990 to 2010).

Given their limited interannual variability, Trend_Emissions and Trend_Boundaries can be approximated using the difference used for 1990/2010 change: respectively (M10B10E10 – M10B10E90) and (M10B10E90 – M10B90E90) divided by the number of years. (Note that we use here again the four sensitivity experiments chosen for the sensitivity of changes, but other combinations could have been used, the impact of this choice is discussed in Section 7.4.2.).

In order to quantify the trend attributed to meteorology, we note that the Tier 3B simulation includes the effect of both meteorology and boundary conditions (with constant anthropogenic emissions). One can thus subtract the effect of boundary conditions obtained above (thanks to the comparison between tier 2A sensitivity experiment and tier 1B) to isolate the contribution of meteorology.

To summarise, we have the following decomposition that relies on an optimal decomposition of the continuous 21 years trend numerical experiments and differences in sensitivity scenarios:

- Total trend between 1990 and 2010

    - Computed from the 21-yr reference simulation in tier 3A using time varying meteorology, boundary conditions and emissions (Myy,Byy,Eyy)
- Trend due to the boundary conditions
    - Difference in the tier 2A and tier 1B sensitivity experiment with changing only boundary conditions for 1990 and 2010 (divided by the number of years to be comparable with the trends: either 11 or 21 where decadal or 2-decade trends are considered)
- Trend due to emissions
    - Difference in the tier 1B sensitivity experiment (also including M10B10E10 as reference) with changing only emissions for 1990 and 2010 (divided by the number of years to be comparable with the trends)
- Trend due to the meteorology: τ (meteorology)
    - trend in the sensitivity 21 years simulation in tier 3B using time varying meteorology, boundary conditions but constant emissions (Myy,Byy,E10) minus the difference attributed to boundary condition changes assessed in the previous bullet point.
- The residual is computed from the difference between the total change and individual factors.

7.3 Decomposition of the 1990/2000/2010 change

Important insights on the robustness of the decomposition proposed here can be gained by analysing the decomposition of the change between the individual 1990/2000/2010 years. More definitive conclusions on the respective role of each driver to the air pollution trend will be assessed in Section 7.5.2 on the basis of the results of the five models that delivered the full suite of requested simulation (including the computationally demanding sensitivity trends simulations).

This section includes a discussion on the robustness of the decomposition technique as well as a discussion on the variability of the decomposition according to the eight CTM of the Eurodelta-Trends ensemble.

7.3.1 Ozone

The decomposition of the main driving factors contributing to the change in summertime ozone peaks between 1990 and 2000 and between 2000 and 2010 according to the methodology presented in Section 7.2 are given in Figure 38 and Figure 39. For both time periods, maps are provided for the contribution attributed to:

- all factors (Total Change (2010 – 2000 and 2000-1990): Total)
- emissions (Δ(emissions): Emis.)
- boundary conditions (Δ(boundaries): Bound. Cond.)
- meteorology (Δ(meteorology): Meteo)
- interaction, or the residual computed with the difference between the simulation with all factors changed and individual tests (Resid.).

For each of these factors, the change is given in $\mu g/m^3/yr$. Even if we are actually computing the difference between two given years, that difference is divided by the number of

separating years in order to provide an information that can ultimately be compared with the annual trends.

As noted before, this decomposition is sensitive to the year selected for the sensitivity experiment. Because of year to year variability, the difference between 2001 and 1991 could be very different from the difference between 2000 and 1990. This is particularly true for the impact of meteorology which is highly variable from year to year, whereas emissions and boundary conditions evolve more gradually. The maps of total change (Total) and change attributed to meteorology (Meteo) should thus be interpreted with care at this stage (more robust conclusions will be drawn on the basis of the trend decomposition in Section 7.5 and 7.5.2).

The main feature to be noted in Figure 38 and Figure 39 regards the more uniform decrease attributed to emission changes throughout Europe. The only regions where the impact of emission changes was limited where the Iberian Peninsula in the 1990s and England in both the 1990s and 2000s. The residuals are very small, which gives support to the robustness of the decomposition analysis. The only point of caution regards boundary conditions whose contribution is of the same order of magnitude as residuals, yielding a larger uncertainty for this factor.

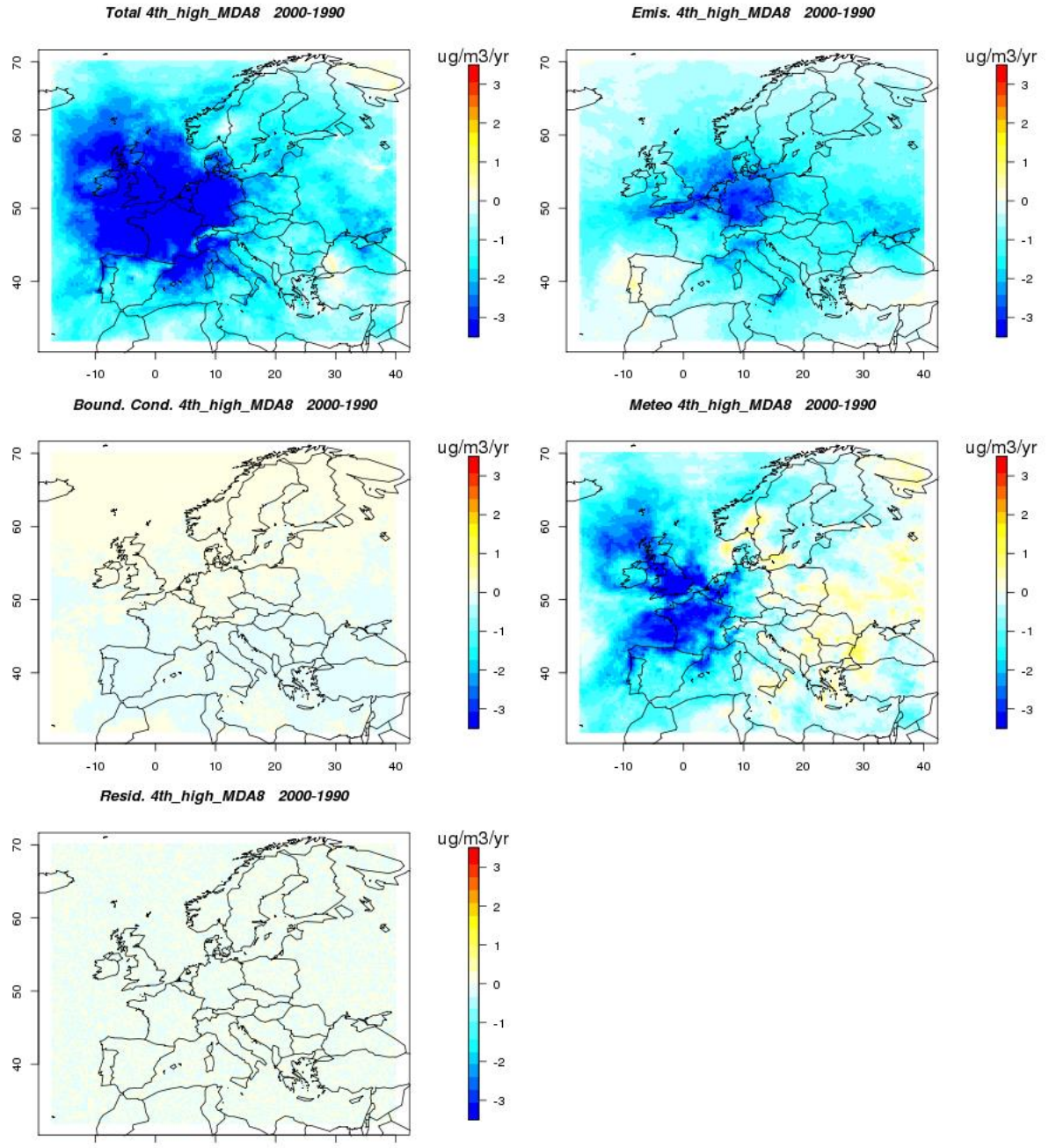


Figure 38: Maps of the contribution of the main driving factors (Total: all factors together, Emis.: anthropogenic emissions, Bound. Cond.: boundary conditions, Meteo: meteorology, Resid.: interaction terms) to summertime O_3 peaks (4th highest MDA8) change (µg/m^3/yr) between 1990 and 2000 (as 2000 minus 1990) in the eight-model Eurodelta-Trend ensemble.

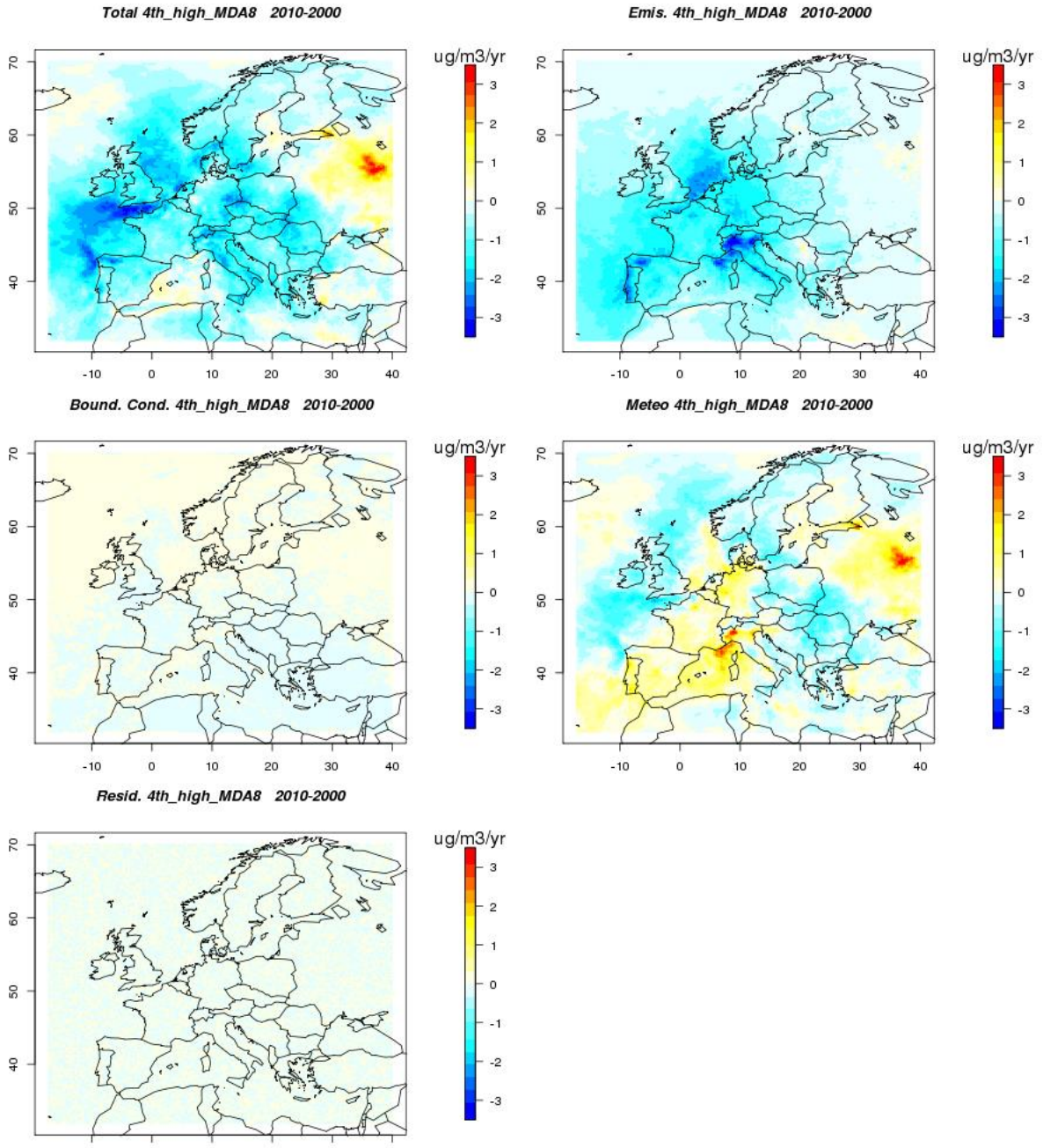


Figure 39: Maps of the contribution of the main driving factors (Total: all factors together, Emis.: anthropogenic emissions, Bound. Cond.: boundary conditions, Meteo: meteorology, Resid.: interaction terms) to summertime O_3 peaks (4th highest MDA8) change (µg/m^3/yr) between 2000 and 2010 (as 2010 minus 2000) in the eight-model Eurodelta-Trend ensemble.

7.3.2 Particulate Matter

The analyses of the decomposition for particulate matter (PM_{10}) in Figure 40 and Figure 41 (for 2000-1990 and 2010-2000, respectively) show that emission changes contributed to reduce PM_{10} concentrations throughout Europe in the 1990s, the only areas where that change was more limited were the Iberian Peninsula and South-Western France. For the

2000-2010 period, emission changes led to more limited improvement, especially over Eastern Europe. Substantial reductions are still modelled over Portugal, France, Italy, England, Benelux, Germany, and the Czech Republic. The impact of boundary conditions is very limited over most of the domain, except in the southernmost boundary in the 2000s because of the impact of desert dust. The choice of the meteorological year is important and it will be discussed in 7.5 given the sensitivity of the result presented here to the choice of the year used for the sensitivity analysis. Last, a small impact of residuals is found, supporting the relevance of the approach.

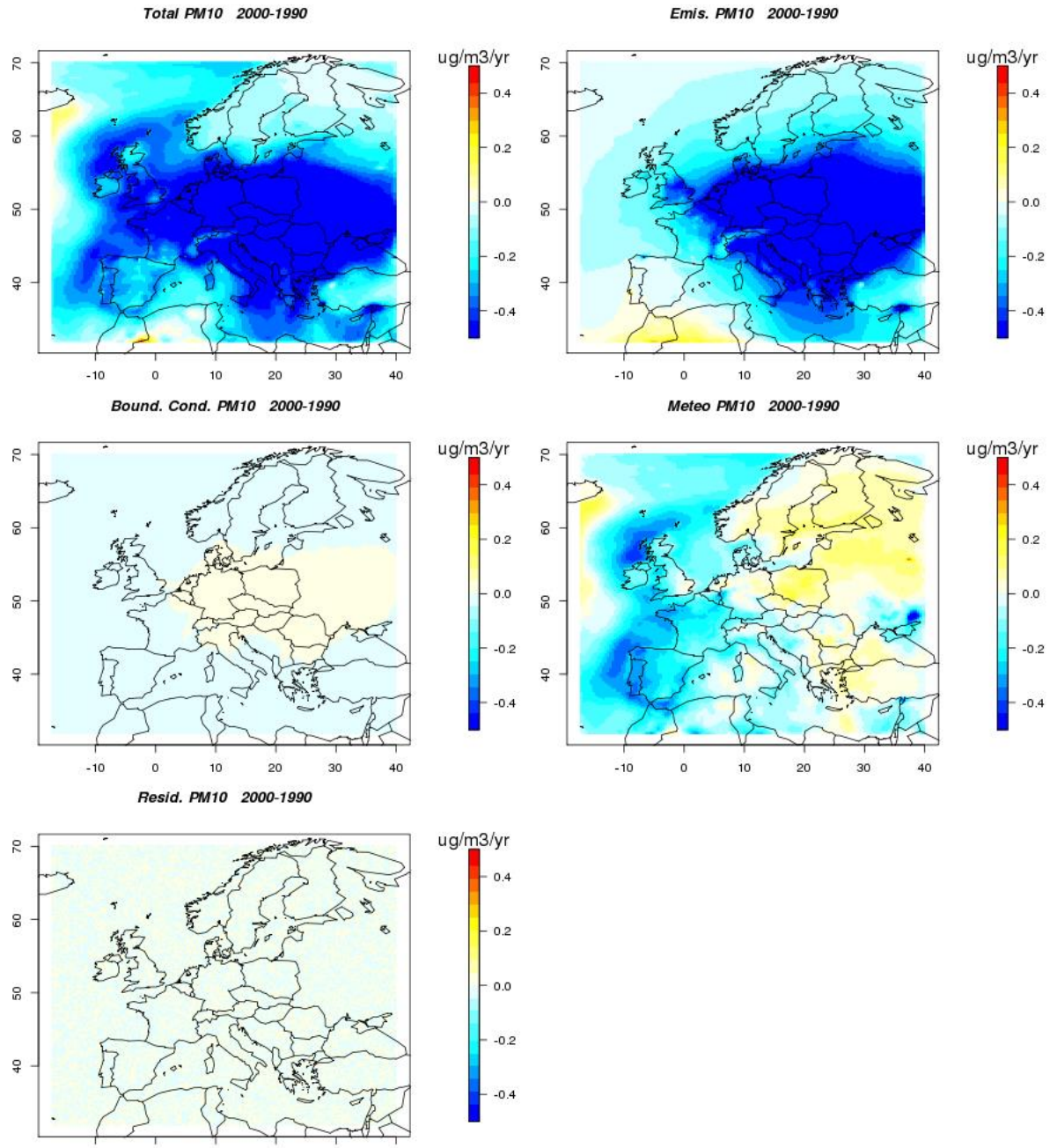


Figure 40: Maps of the contribution of the main driving factors (Total: all factors together, Emis.: anthropogenic emissions, Bound. Cond.: boundary conditions, Meteo: meteorology, Resid.: interaction terms) to annual mean PM_{10} change (µg/m^3/yr) between 1990 and 2000 (as 2000 minus 1990) in the eight-model Eurodelta-Trend ensemble.

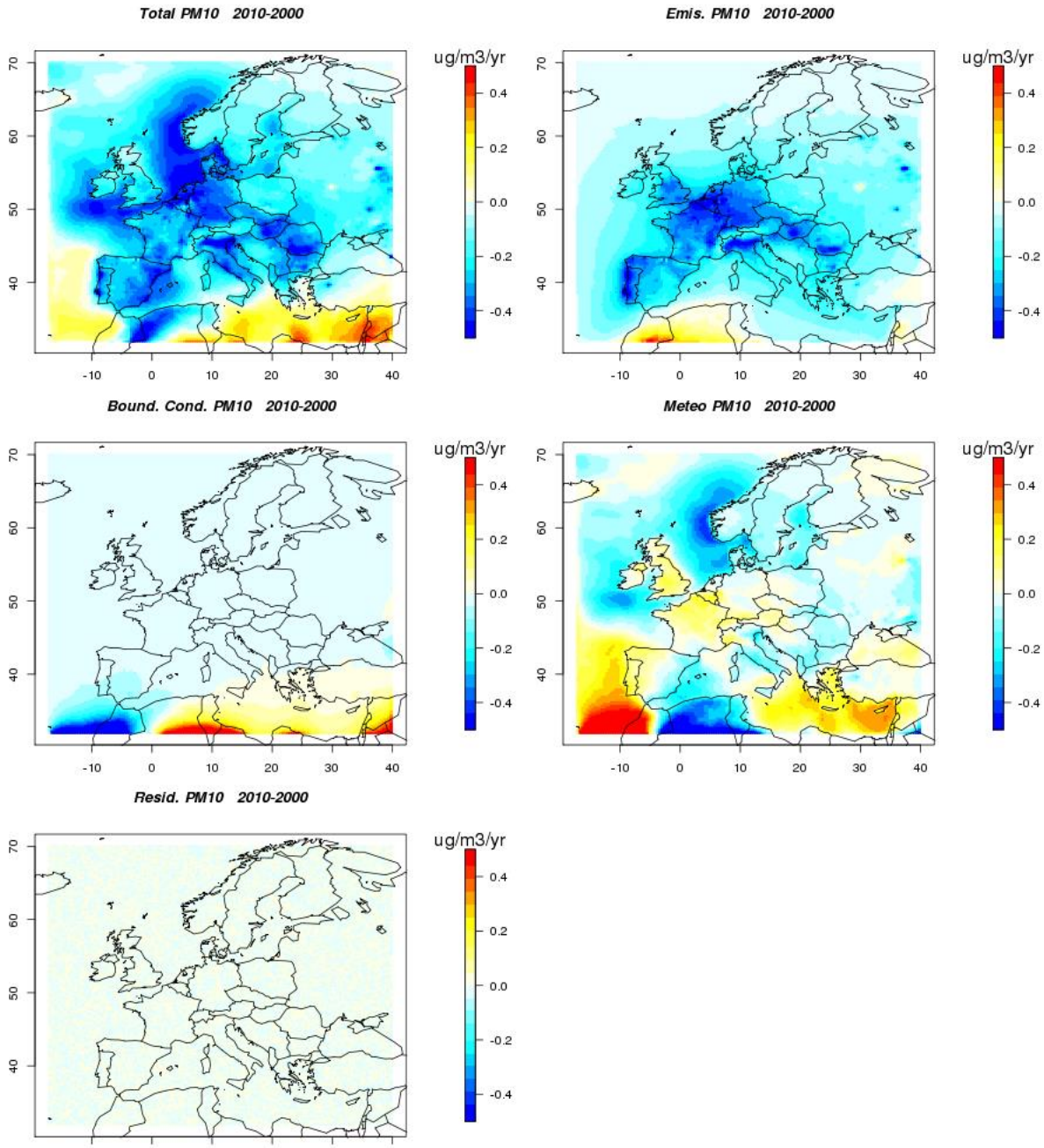


Figure 41: Maps of the contribution of the main driving factors (Total: all factors together, Emis.: anthropogenic emissions, Bound. Cond.: boundary conditions, Meteo: meteorology, Resid.: interaction terms) to annual mean PM_{10} change ($\mu g/m^3/yr$) between 2000 and 2010 (as 2010 minus 2000) in the eight-model Eurodelta-Trend ensemble.

7.4 Uncertainty of the decomposition analysis

7.4.1 Model uncertainty in the decomposition analysis

For each factor, the maps discussed in Section 7.3.1 and 7.3.2 are the 8-model mean across the Eurodelta-Trend Ensemble. It is however essential to account for possible differences in the model response to the changes attributed to emissions, boundary conditions or meteorology, and it is also one of the key strengths of the Eurodelta-Trend experiment to rely on such an ensemble.

To assess the inter-model spread in the above-mentioned attribution, we averaged each factor over the Mid-Europe region (approximately North-Eastern France, Benelux, German and Western Czech Republic). The inter-model variability, for each driving factor is given in the boxplot of Figure 42. Similarly to the discussions of Section 7.3.1 and 7.3.2, those results are very sensitive to the year selected. Because of the role of interanual variability, the "total" impact as well as that of the "meteorological" factor should be interpreted with care. The ambition of those figures is only to assess the model uncertainty in quantifying these trends.

It appears that the signal attributed to European emission changes is robust across the ensemble since the whole range of responses does not overlap the zero line. The interquartile distance is also robust for the impact of meteorology on ozone peaks, but the range of variability is actually larger for the impact of the meteorology in annual mean PM_{10}. The Boundary Conditions and Residuals are more uncertain, but also very small in magnitude. It would be important for the modelling community to understand the reason for such a variability but that investigation falls outside of the scope of this report.

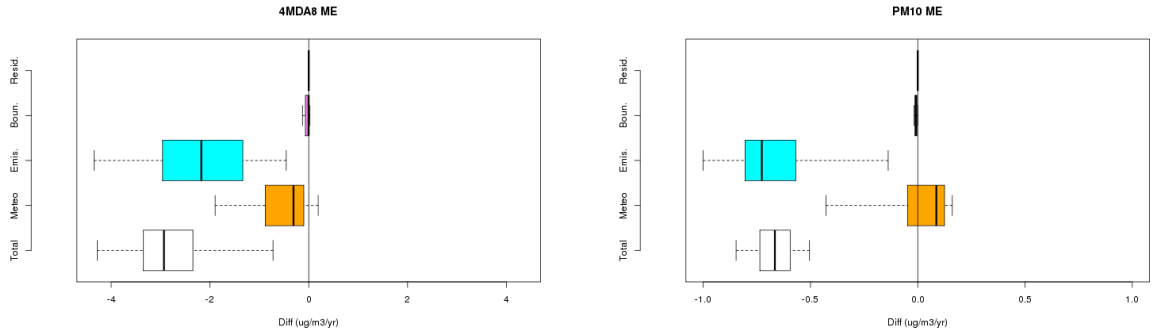


Figure 42: Model uncertainty in the Eurodelta-Trend ensemble of eight models in the attribution of the contribution of Meteorology (Meteo), Emissions (Emis), Boundary Conditions (Boun), and residuals (Resid.) to the total (Total) change between 1990 and 2010 (as 2010 minus 1990 in µg/m^3/yr) in the Mid Europe (ME) region for ozone summertime peaks (4th highest MDA8) (left) and PM_{10} annual mean (right)

7.4.2 Interactions terms in the decomposition analysis

As explained in Section 7.2, in order to minimise the computing demand of the Eurodelta-Trend exercise, the number of sensitivity experiment required to decompose the three main factors was reduced from eight (2^3) to four (2^2). For instance, in equation (7.1), we choose to assess the impact of using either the 1990 and 2010 meteorological year using fixed 1990

boundary conditions and anthropogenic emissions, whereas that impact could be different if 2010 emissions had been used.

The impact of using only 4 out of the 8 required sensitivity simulations is assessed by comparing the results of the decomposition for the CHIMERE model, which delivered the required additional simulations.

The maps of the importance of each driving factor according to the CHIMERE model using all 2^3 combinations for the 1990 to 2010 change for both ozone and PM_{10} annual means are given in Figure 43 and Figure 44, respectively. For each factor, those figures also provide the standard deviation across the various sensitivity simulations required to assess the role of a given factor. For instance, the impact of emission changes can be assessed with the four following differences:

- M10B10E10 – M10B10E90
- M10B90E10 – M10B90E90
- M90B90E10 – M90B90E90
- M90B10E10 – M90B10E90

Only the first of these differences was computed by all the models of the ensemble, and therefore included in the multi-model decomposition introduced in Section 7.3.1 and 7.3.2. But having the CHIMERE simulations for all four combinations allow mapping the average change across all four combinations of those sensitivity simulations as well as the standard deviation. So that the importance of interaction terms can be assessed.

From these results, it is clear that the sensitivity to the selected year remains limited compared to the signal we are trying to quantify. The difference in the estimated contribution of boundary conditions is negligible when different years are selected. The sensitivity to the selected year is slightly larger for emissions and meteorology, but always at least a factor 10 smaller than the average impact of those factors, highlighting the robustness of the decomposition analysis, even though only part of the parameter space was explored.

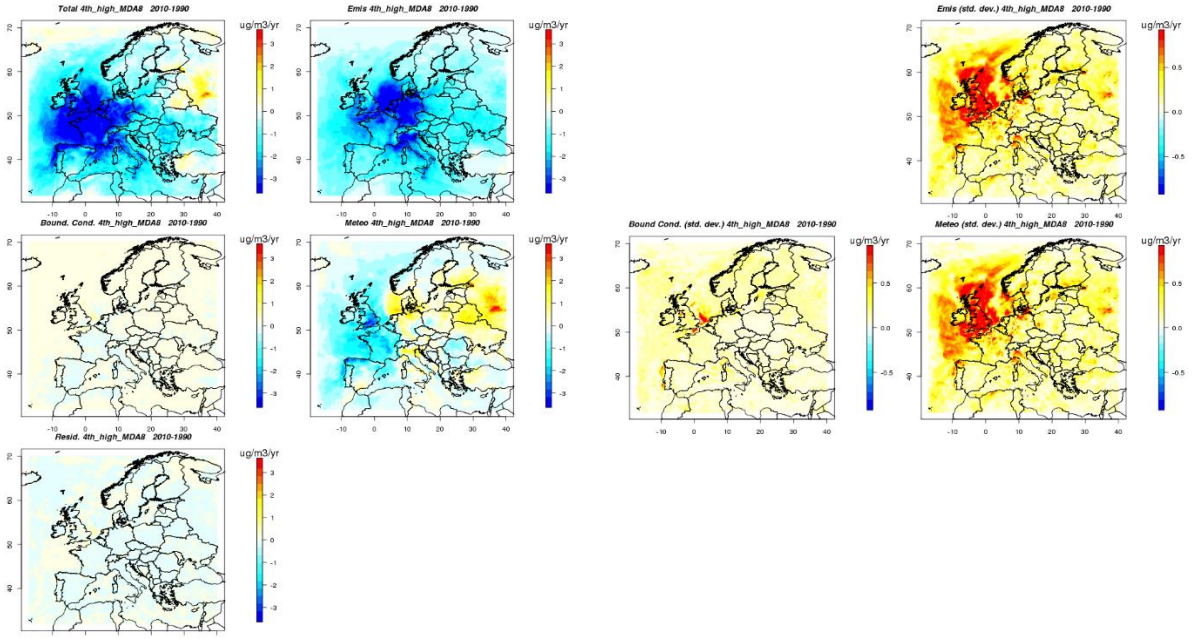


Figure 43: For summertime ozone peaks: Left: average change between 1990 and 2010 attributed to all factors (Total), Emissions (Emis), Boundary conditions (Bound. Cond.), meteorology (Meteo) and residual (Resid) for the CHIMERE model and all possible combinations of the aforementioned factors. Right: standard deviation in the four possible combination of sensitivity experiments used to assess the impact of emissions, boundary conditions and meteorology (µg/m^3/yr).

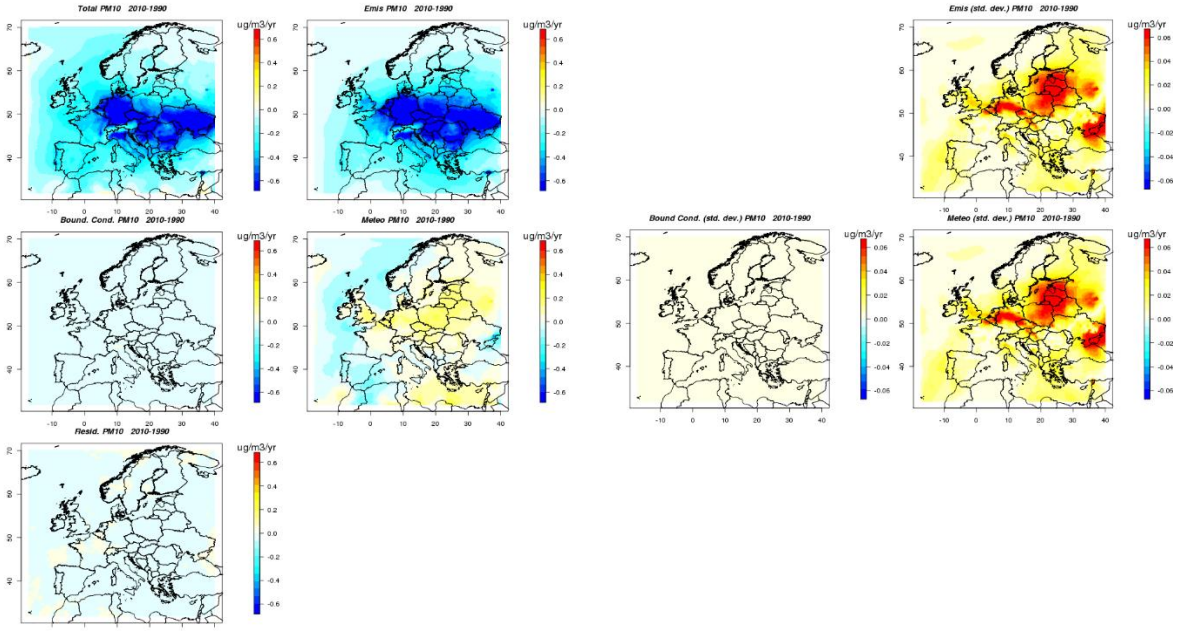


Figure 44: Same as Figure 43 for annual PM_{10}.

7.4.3 Focus on the meteorological influence on ozone trends

Complementary to the analyses presented above, we have estimated the "meteorological influence" on ozone each year by the extent to which the ozone levels that year deviates from a climatological mean, similarly to the common way of relating one year's weather to the long-term climate normals. The idea is that we estimate each year's deviation from an ozone long-term mean climatology based on model scenarios with fixed emissions. As the typical weather climate periods are 30 years long, we would in principle need a full 30 years CTM climatology with fixed emissions for each year. Thus, for assessing the meteorological influence in year X we would need to run a CTM for a 30 years period with actual meteorology each year and fixed emissions as of year X. For a trend study like e.g. the Eurodelta looking at the development during 1990-2010, we would then need to run the 30 years climatology for each year in the period, i.e. 21 x 30 = 630 annual CTM runs. This is not feasible in terms of time and resources, and thus a simplified method is explained below.

In addition to the Eurodelta Tier 3B scenario with fixed 2010 emissions, the EMEP model was run in a similar scenario with fixed 1990 emissions. This extra scenario could be called Tier 3C and is complementary to the scenarios listed in Table 2:

Tier 3C: MyyByyE90, yy = 1990 to 2010

A basic assumption is then that the ozone climatology for the years between 1990 and 2010 could be estimated by linear interpolation between the Tier 3B and Tier 3C scenarios. We used the following equation:

$$X_{ih,jy,ky} = [(2010\text{-}ky)\, X_{ih,jy,1990} + (ky\text{-}1990) X_{ih,jy,2010}]/(2010\text{-}1990), \quad (7.3)$$

where

$X_{ih,jy,ky}$ = modelled ozone at hour *ih* in year *jy* assuming emissions for year *ky*
ih = 1, …, number of hours in year *jy*
jy = 1990, …, 2010
ky = 1990, …, 2010

In this way, we estimated the ozone level every hour in each actual year and for every emission year 1990, …, 2010, giving us a complete ozone climatology of 21 years. In other words, we have an estimate of how the modelled ozone concentration would have been each hour using emission data relevant for any of the years 1990 to 2010. By the linear assumption explained above, we reduced the number of required model scenarios from 21 x 21 = 441 (as explained above) to 21 x 2 = 42.

The assumption of linearity, i.e. that each model scenario $X_{ih,jy,ky}$ could be estimated by a linear combination of the E90 and E10 scenarios, was tested by comparing the $X_{ih,jy,ky}$ based on Eq 7.3 with the standard model scenario for the situation of *jy* = *ky*. The results for one site/year is shown in Figure 45 as example.

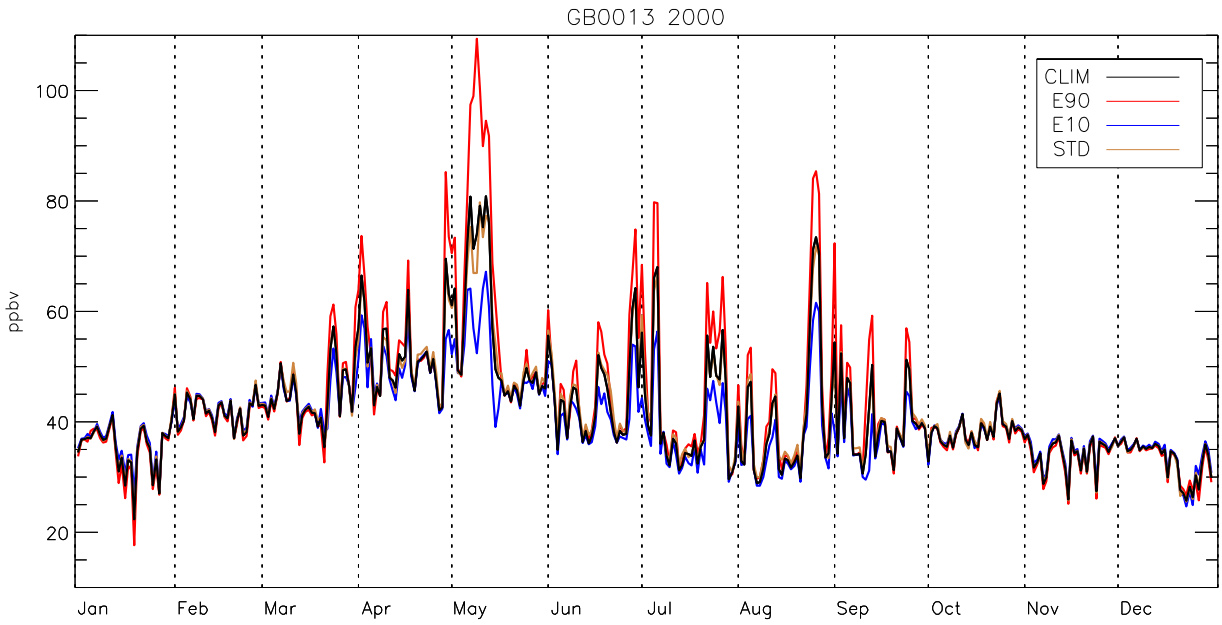


Figure 45. Daily modelled MDA8 values at GB0013 (Yarner Wood, UK) in 2000 as example. The "CLIM" scenario shows the values based on linear interpolation between the E90 and E10 scenarios (based on hourly data) as explained in the text whereas "STD" shows values from the standard model run.

The time series in Figure 45 shows the basic concept of the idea, i.e. that the time series consists of a baseline with a number of small and large perturbations on. These perturbations are controlled by the precursor emissions and other factors (boundary conditions, etc.). Since there has been a gradual decline in emissions over the 1990-2010 period, the E90 and E10 scenarios given by the blue and red lines in Figure 45 mark the span of these perturbations, i.e. the max and min, whereas the standard model scenario is somewhere between these two extremes. The results from the linear interpolation ("CLIM") based on Eq 7.3 is seen to be very close to the standard model scenario. A further inspection of the relationship between the daily data obtained from the linear relationship (Eq. 7.3) and the standard model showed a correlation coefficient typically of the order of 0.98 and a root mean square error of around 1.0 giving strong confidence to our approach. A comparison showing the CLIM values (stemming from a linear interpolation between E90 and E10) and the other scenarios are shown in Figure 46 for two sites and three years as examples. This shows a very good agreement between the interpolated values (CLIM) and those from the STD scenario, and the RMSE values for the STD scenarios are much smaller than for E90 and E10. This supports our idea that the modelled daily ozone values could as a good approximation be seen as a baseline with superimposed perturbations on and that these perturbations could be estimated by a linear combination of the end points, i.e. the E90 and E10 scenarios.

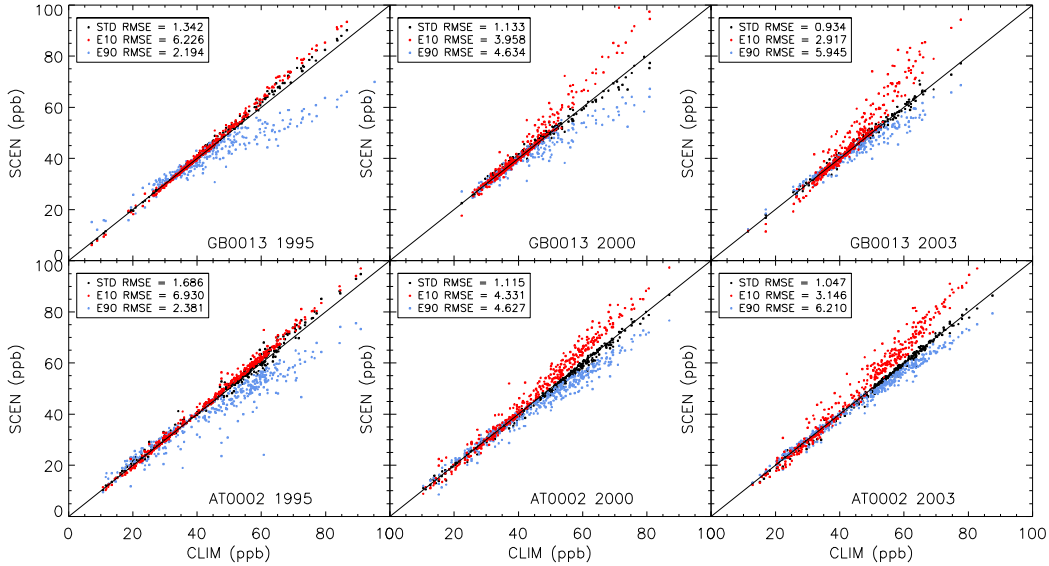


Figure 46: Scatter plots between the "CLIM" scenario resulting from a linear interpolation between the E90 and E10 scenarios as explained in the text and the main model scenarios for two selected sites and three years as examples. The symbols mark the daily MDA8 values through the whole year and the full line marks the 1:1 line. The root mean square error (RMSE) between the CLIM and the other scenarios are also given. All data are based on the EMEP model results.

This implies that the interpolated values, i.e. the $X_{ih,jy,ky}$ data obtained from Eq. 7.3 using the Tier 3B and Tier 3C scenarios could be used as a proxy for the full ozone model climatology for the period 1990-2010 as defined above. With the ozone climatology defined in this way, the "meteorological influence" on ozone in a specific year could be defined as the deviation of that year's ozone levels from the long-term climatology, using the emissions and meteorology for that specific year.

Thus, for each year *my* =1990, …, 2010, we define the climatological mean value as the mean of the 21 annual values:

$$Y_{my} = \text{mean } [F(X_{ih,my,ky})],\ ky = 1990, \ldots, 2010$$

where $F(X_{ih,my,ky})$ is an annual statistic like e.g. AOT40, SOMO35 or the 4hMDA8 based on emissions for the specific year *my* and meteorology for each year *ky* = 1990, …, 2010.

The "meteorological influence" defined as the deviation of the ozone indicator in the year *my* from the 21-year climatological mean is then simply the difference:

$$\Delta Y_{my} = [F(X_{ih,my,ky}) - Y_{my}],\ \text{for } ky = my$$

With this approach we could calculate the meteorological influence (or deviation from the mean) for each year in the period 1990 to 2010 and look for possible trends in this influence. There are a few points worth mentioning though:

- Ideally, the climatological means should be based on a longer period, i.e. 30 years, and the period should be outside the actual year of interest. In comparison, in weather climatology, the standard periods 1961-1990 and 1981-2010 are used as climate normals.

- All the values discussed here refer to model results without taking the actual observed data into account. Thus, the approach is purely model based, and systematic model biases and deficiencies with respect to the measured data are not considered.

Figure 47 and Figure 48 shows the results for the 4hMDA8 indicator for either the 1990-2000 or 2000-2010 periods. Figure 49 shows the results for the whole period, 1990-2010. In this Figure the annual modelled and observed data are compared to the median values interpolated at the station location of the climatological means ("Clim") each year using the actual emission year. It should be said that the results in Figure 47 and Figure 48 don't separate between the influence of the boundary conditions and the meteorology, or in other words that these contributors are taken together. The E90 scenario is based on the boundary conditions of 1990 and the E10 scenario on 2010. However, as seen above, the influence of the boundary conditions on ozone peaks is seen to be fairly minor in most regions thereby justifying our approach here.

The results in Figure 47 and Figure 48 could be interpreted in the following way:

- Firstly, one would expect the climatological data (annual black dots in the Figures) to follow a straight line only reflecting the trend in the overall (total) precursor emission since these data are based on 21 years of averaged meteorology. The reason we see deviations from a straight line is that the monitoring sites with valid data vary somewhat from year to year. If certain low- or high-ozone sites are lacking one year, the mean value would reflect this.
- Secondly, differences between the standard model runs (blue bars with median) and the climatology (black dots) for a certain year indicate to what extent the meteorology reduced (std model > climatology) or increased air pollution (std model < climatology) that specific year. Examples of this are 1994 and 2003 for the ME region (ozone reduction) and 1991 for the EN, SC and ME regions (ozone increase).
- Thirdly, comparisons between the variations in the observed data (red bars) with the differences mentioned above (deviation from the mean climatology) then indicates to what extent the observed variation could be explained by a special meteorology in certain years or changes to the monitoring network or other factors.

Furthermore, the results in Figure 47 and Figure 48 show a somewhat smaller downward trend in the climatological data compared to the standard model results for all regions except Scandinavia (SC). Since the trend in the climatology data in Figure 47 and Figure 48 only reflects the changes due to the emission changes (each year of data is based on 21 meteorological years), the difference between the standard model data and the climatology data indicates the influence of the meteorology year by year and overall (the trend).

Thus, since the model data from the STD run show a stronger downward trend than the climatology (which only reflects the emissions) these results indicate that for all regions except SC the meteorological variability during the 1990-2000 and 2000-2010 periods lead to an even stronger reduction in this ozone indicator than the emission reduction alone. The emission induced part was, however, the major influence whereas the meteorologically driven trend only contributed with a small fraction of the total trend. Thus, over the 21-year period studied, the year-to-year variations in meteorology did contribute to the decline in

ozone; however the reductions in ozone precursor emissions was the clearly dominant cause for the ozone decline.

As far as differences between the 1990s and 2000s are concerned, the largest differences between the STD model and the proxy climatology is in the 1990s where the CLIM trend is generally smaller than the STD. This implies that the actual inter-annual meteorology during this decade combined with the downward trend in the emissions contributed to a larger decline in ozone than the emissions alone. For the 2000 decade, there is less differences between CLIM and STD and for most regions the downward CLIM trend is somewhat larger in absolute value than the STD implying that the additional trend imposed by the meteorology was small and – if present – worked in the opposite direction as the emission trend and gave slightly less reductions in ozone than the emission trend alone would have given.

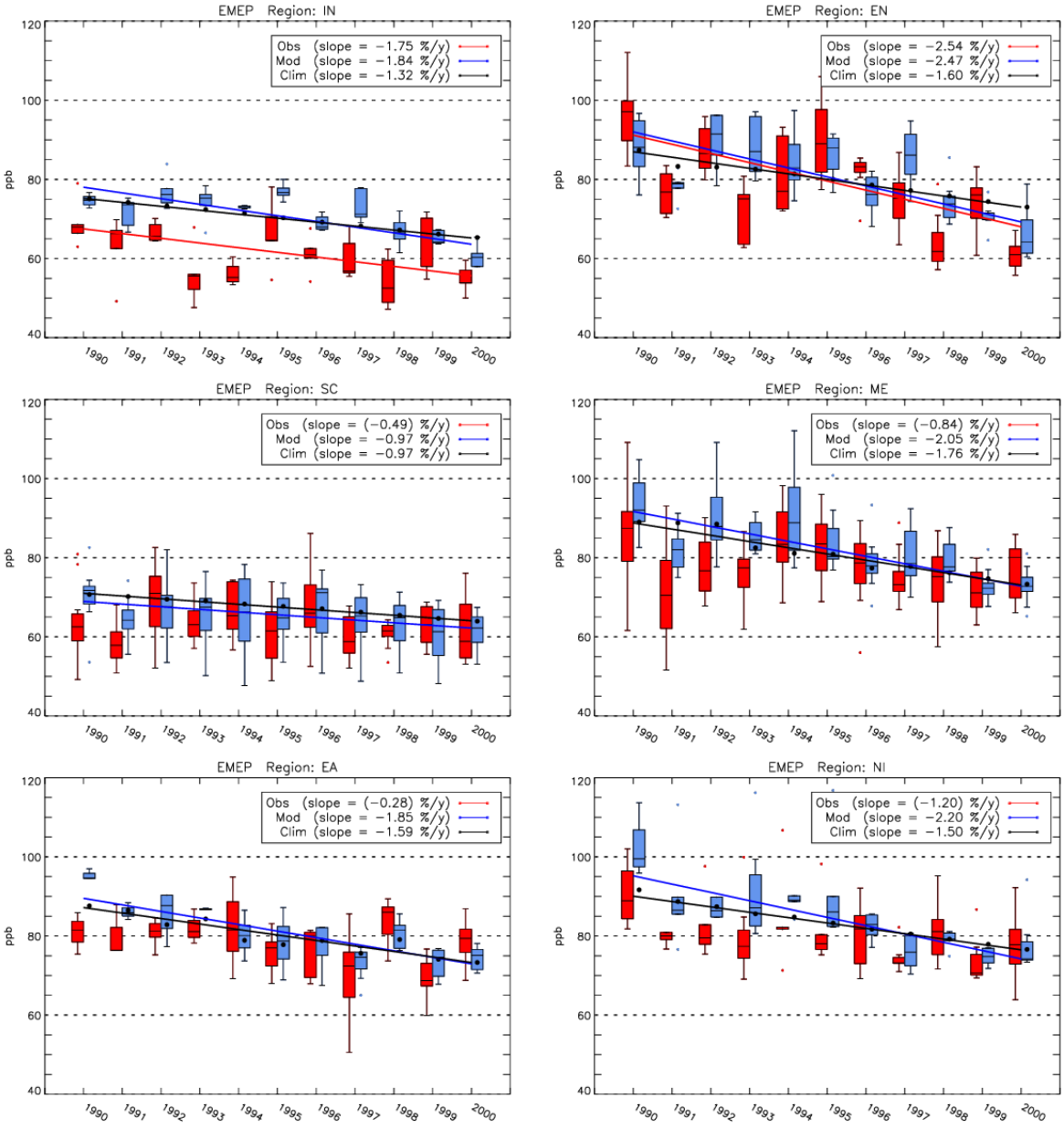


Figure 47. For the 1990 – 2000 period: Box-whisker plots of the 4th highest MDA8 for the regions based on the observations (red), the EMEP model (blue) and

the 21-years climatology each year (black dots mark the median of the station values) as explained in the text. Regression lines (black) for the median values are plotted when a significant slope is found.

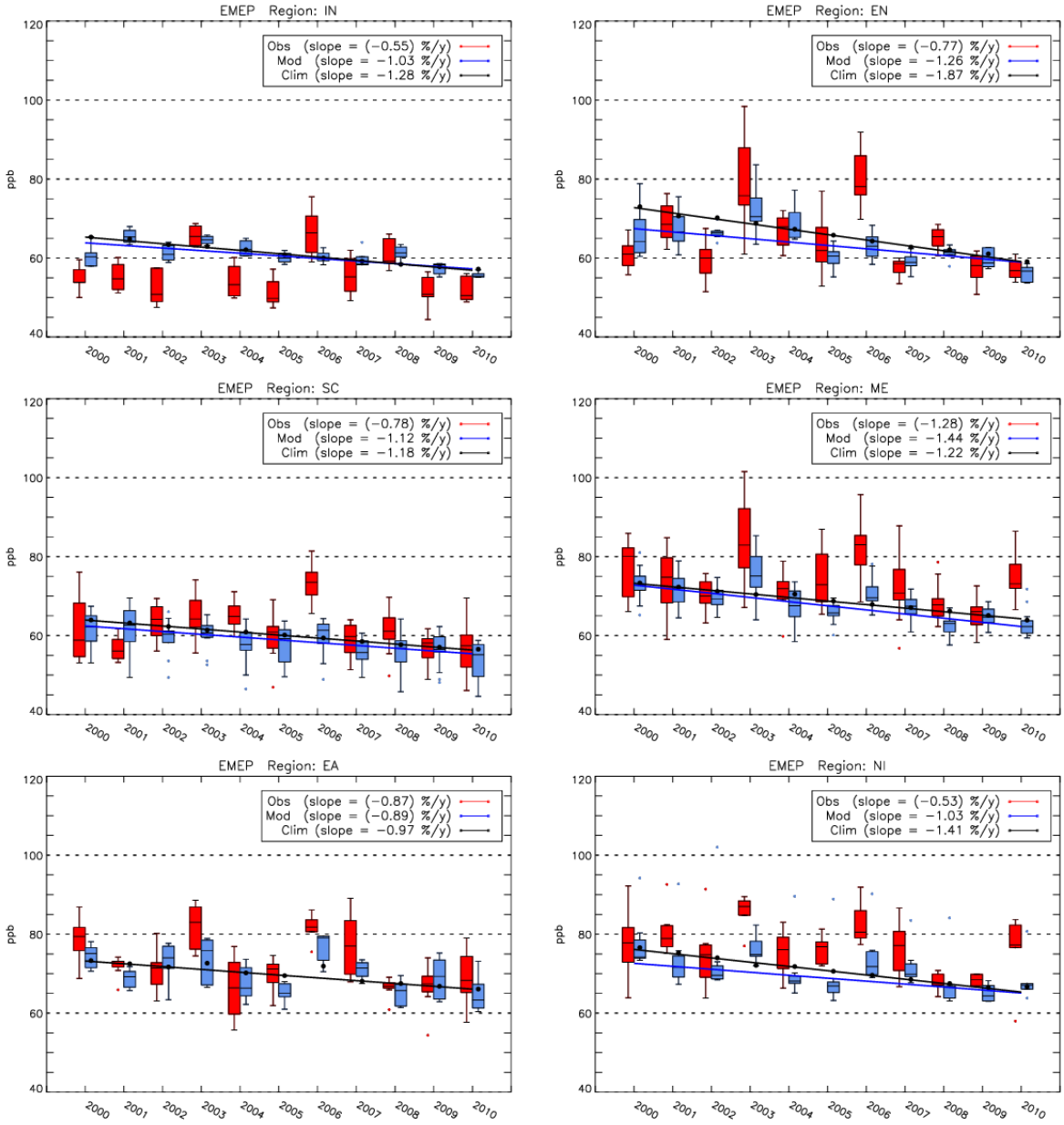


Figure 48: Same as Figure 47 for the 2000-2010 period.

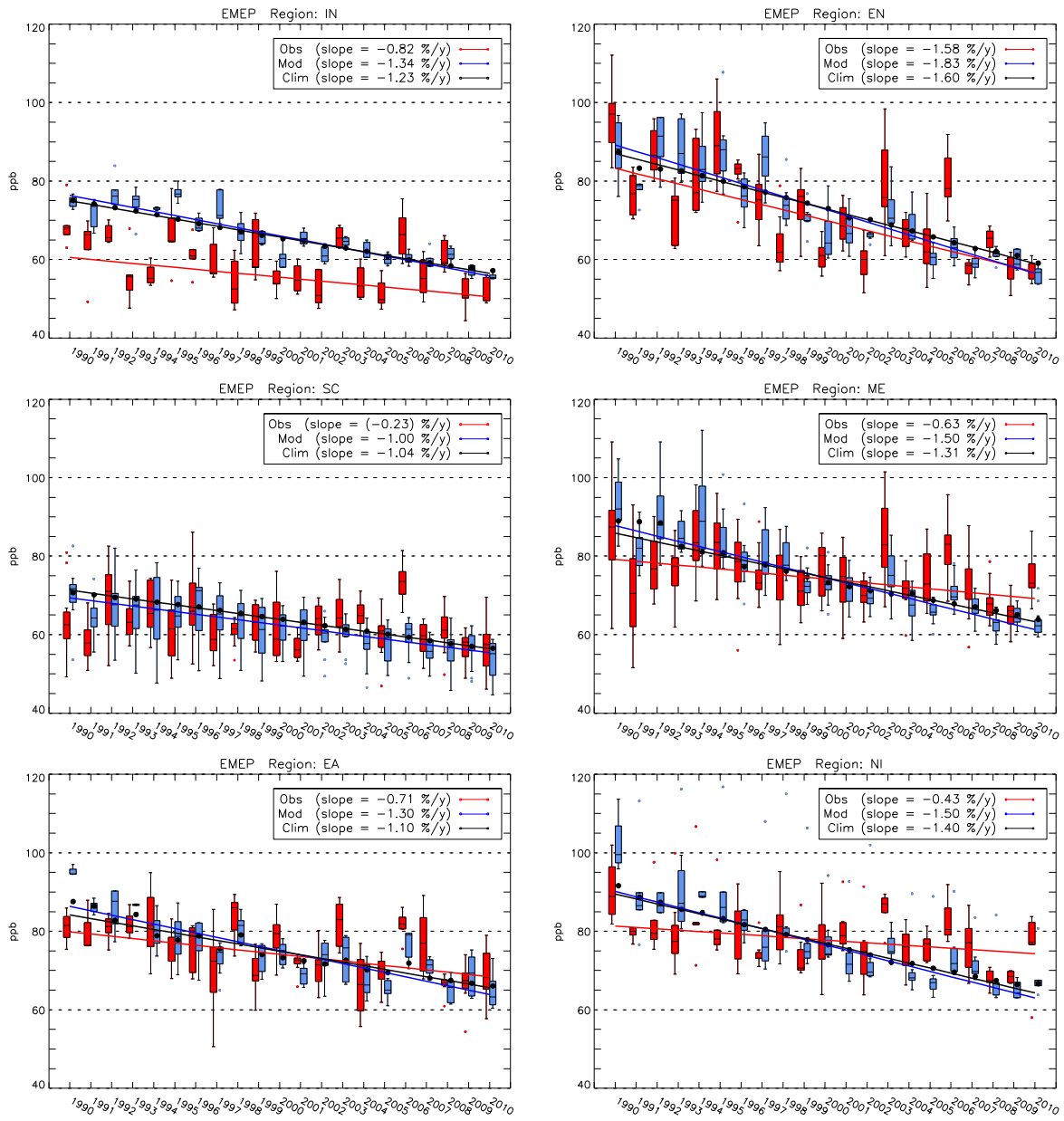


Figure 49: Same as Figure 47 for the 1990-2010 period.

7.5 Decomposition of the 1990-2010 Trend

In order to assess the relative contribution of emission, meteorology and boundary conditions on the long-term evolution of European Air quality, we involved the 21-year trend simulations in addition to the annual sensitivity experiment using the methodology introduced in Section 7.2, in particular equation (7.2). For concision purposes, only results aggregated over a set of nine European regions are discussed (see definition in Figure 4): England (EN), Inflow (IN, Ireland and Scotland), Iberian Peninsula (IP), France (FR), Mid-Europe (ME), Scandinavia (SC), North Italy (NI), Mediterranean (MD), Eastern Europe (EA).

7.5.1 Ozone

The decomposition of the key driving factors influencing annual mean and summertime ozone for each region is displayed in Figure 50. The overall evolution can be assessed on these plots with the net trends displayed with a diamond sign: the magnitude of the decrease differs depending on the region but no increase of summertime MDA8 mean is found. For ozone annual mean, increases are found over the following regions: England, Inflow, Mid-Europe, Scandinavia and Eastern Europe in the 1990s, but no-where in the 2000s.

The main feature to be pointed out in these graphs regards the importance of European anthropogenic emission changes on ozone peaks, which appears to dominate for most areas and time periods. The smallest impact is found for the Iberian Peninsula in the 1990s but everywhere else emissions are the dominating factor. As we will see below, emissions can compete with other factors, but this analysis demonstrate that they did have a substantial impact. The picture is slightly different for ozone annual mean, where emission reduction led to an increase in the England region for both the 1990s and the 2000s (and also, to a lesser extent in the Inflow region in the 1990s). This annual mean ozone increase is due to lower titration effect of the reducing of NOx emission in such a saturated area.

The impact of meteorological variability and boundary conditions is smaller, but still substantial. Depending on the region and indicators, these two factor can reach similar order of magnitude, but meteorology dominates overall, which can be considered as a surprise given the importance of background ozone on long term trends often pointed out (e.g., (Maas and Grennfelt 2016)).

In the 1990s, meteorological variability contributed to increase annual mean ozone levels in all regions, however because of the mix of processes accounting for annual mean ozone, it is difficult to explain this meteorological impact. The impact of meteorology is more straightforward for ozone peaks, where stagnation and heat waves favour surface ozone production. In almost all region and time periods, meteorological variability contributed to decrease ozone peaks. That can for instance be explained by the fact that for the 2000s, the main heat wave occurred in the beginning of the period (2003), therefore contributing to a decreasing trend. This feature must also be highlighted given that the limited decrease of ozone peaks compared to precursor trends is sometimes speculatively attributed to unfavourable meteorological conditions. Our analysis demonstrate that it is not the case.

Boundary conditions constitute an important factor for annual mean ozone, it contributed to increase the trend in the 1990s throughout Europe but on the contrary, it decreased ozone levels in the 2000s. Tropospheric ozone trends indicate an increase in the 1990s and a flattening of the trend in the 2000s, but such a contribution to an actual decrease of annual mean ozone would deserve further investigation. Boundary conditions have an opposite impact on summertime peaks (decrease in the 1990s, increase in the 2000s), though of smaller value (actually, close to zero in the 1990s), which demonstrate the complexity of understanding the impact of background ozone changes on baseline and extreme of surface ozone. This impact of the background on annual mean ozone levels was excepted for instance in the previous HTAP experiments (Wild et al., 2012). Although of smaller importance, the contribution of boundary conditions to the trend in summertime peaks was not expected, especially since it is of an opposite sign in the 2000s.

As mentioned in the decomposition analysis of Section 7.3.1, the residuals are limited, although they can sometime compete with some factors in given regions and time periods. The main surprise in these results regards the contribution of the meteorological factor, which exceeds that of boundary conditions and even sometimes compete with emission changes. In particular, the evolution of meteorological variability in the 1990s was at least as important in contributing to increase annual mean ozone trends as the intercontinental import through boundary conditions. And over both the 1990s and 2000s, meteorological trends contributed to decrease summertime peaks.

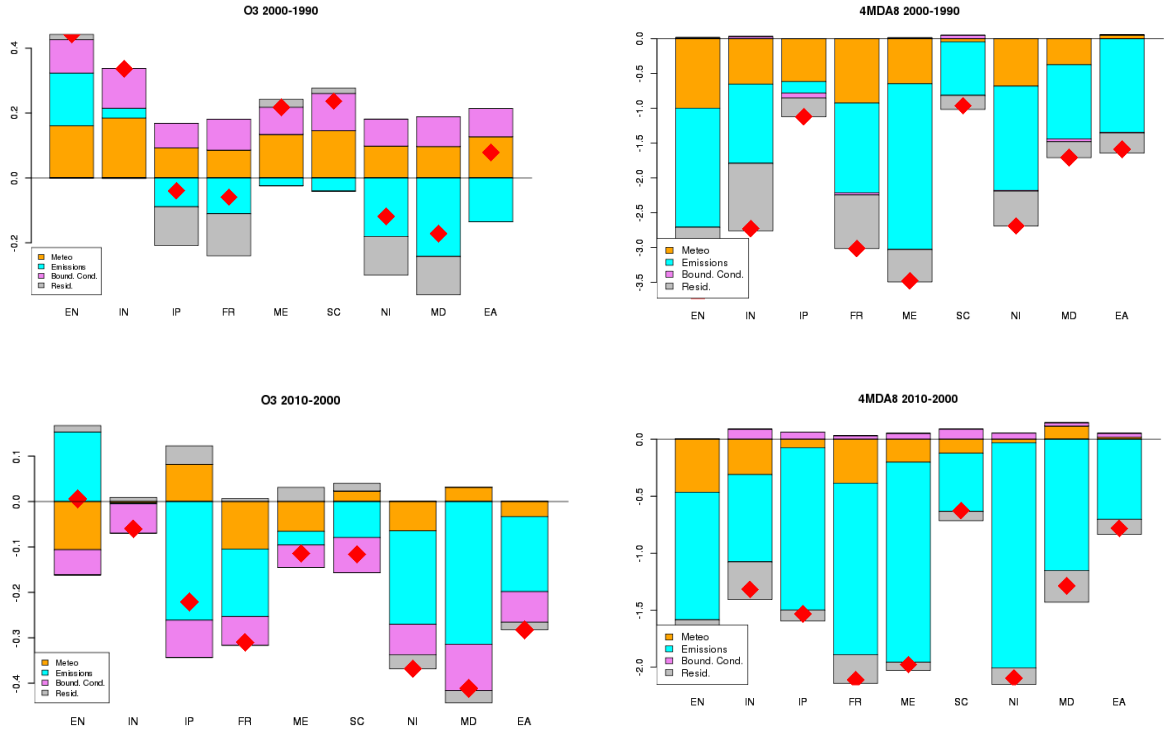


Figure 50: Trend attribution of the key driving factors (Meteo: meteorology, Emissions: European anthropogenic emissions, Bound. Cond.: boundary conditions, Resid.: residual) to the modelled trend of ozone annual mean (left) and summertime ozone peaks (April-Sept mean of MDA8, right) in the 1990s (top) and in the 2000s (bottom). The red diamond indicates the net trend (µg/m^3/yr), all factors considered.

7.5.2 Particulate Matter

The same decomposition for annual mean PM_{10} is given in Figure 51. Here the contribution of emission changes clearly dominates, the contribution of boundary conditions was negligible in the 1990s and remains marginal in the 2000s but acts to reduce slightly PM levels throughout Europe. Meteorology has an impact but it remains much smaller than the impact of European emission changes, except for the Iberian Peninsula, and to some extent France, in the 1990s where both effects compete in magnitude. Meteorology contributes to decrease PM_{10} average levels in the 1990s (except in the Mediterranean) but on the contrary, it led to an increase in the 2000s, except for the Iberian Peninsula, Scandinavia and Eastern Europe. Although this feature appears on annual mean PM_{10}, it should be assessed in

conjunction with the long-term evolution of PM_{10} episodes during that period. The residuals of the decomposition are larger than for ozone and generally exceed the factors attributed to boundary conditions and meteorology.

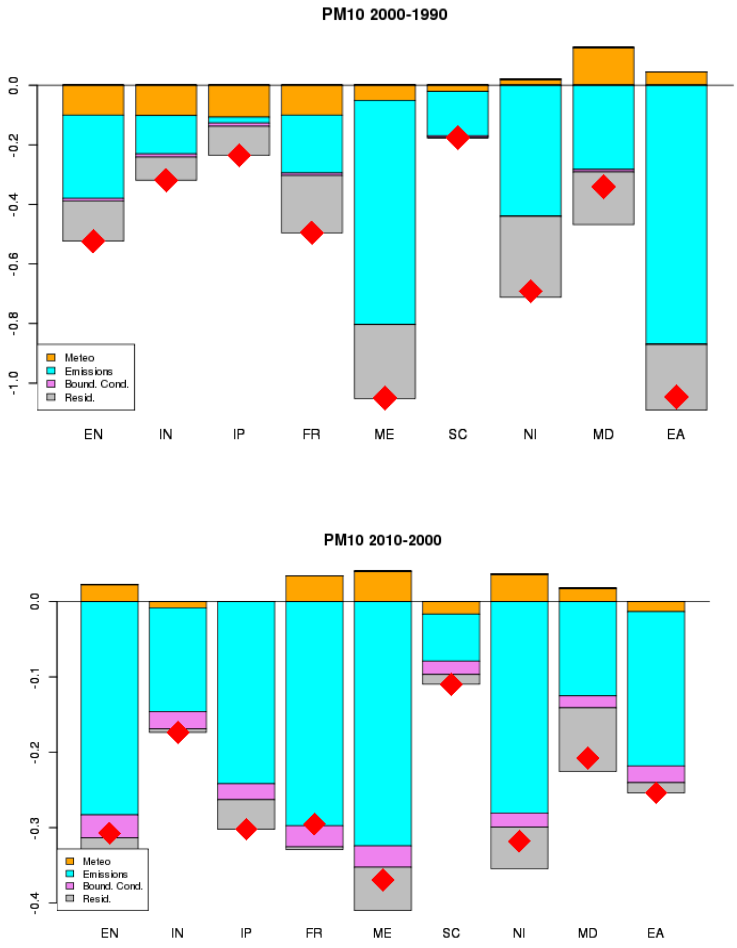


Figure 51: Same as Figure 50, only for annual mean PM_{10}.

7.5.3 Natural factors influencing particulate matter

The assessment of the respective role of European anthropogenic emission changes, meteorology and boundary conditions to the net change of the various aerosols compounds contributing to the PM_{10} mix is presented in Figure 52.

As mentioned in Section 7.5.2, anthropogenic emission changes dominate the evolution of PM_{10}. This change is particularly seen for secondary inorganic aerosols (sulphate, nitrate and ammonium) as well as secondary organic aerosols (SOA). For these compounds, the contribution of boundary conditions and meteorology is similar and much smaller than that of anthropogenic emissions, especially for anthropogenic SOA. Albeit of smaller magnitude, the contribution of anthropogenic emissions to biogenic SOA is the biggest one, because they share some precursors with anthropogenic SOAs.

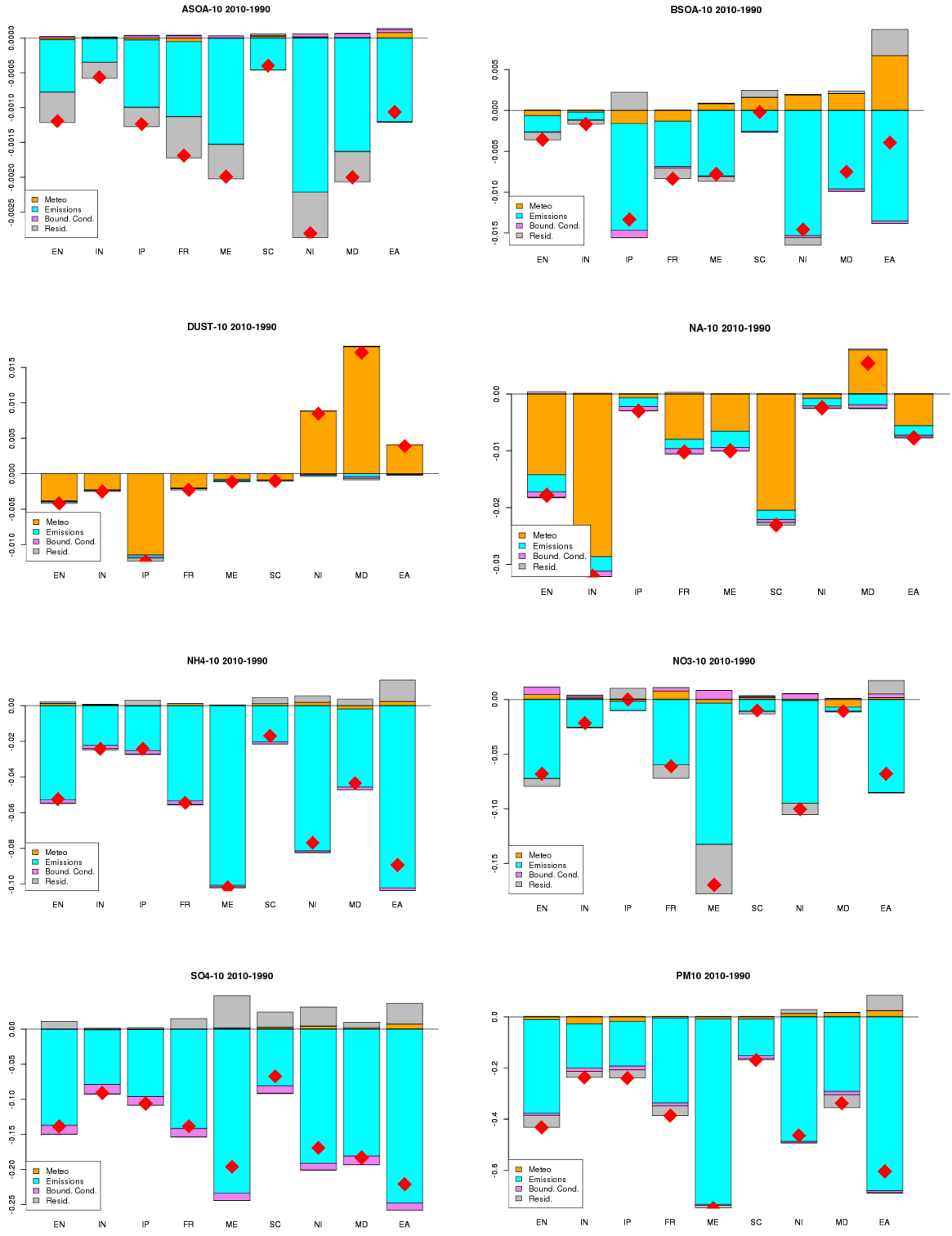


Figure 52: Trend attribution of the key driving factors (Meteo: meteorology, Emissions: European anthropogenic emissions, Bound. Cond.: boundary conditions, Resid.: residual) to the modelled trend between 1990 and 2010 for Anthropogenic (ASOA-10) and Biogenic (BSOA-10) Secondary Organic Aerosols, Desert Dust (DUST-10), Sea-Salt (NA-10), Ammonium (NH4-10), Nitrate (NO3-10), Sulfate (SO4-10) and PM_{10}. The red diamond indicates the net trend ($\mu g/m^3/yr$), all factors considered.

Desert dust and sea-salt changes are primarily attributed to the meteorology factor, partly because of the Eurodelta-Trend setup, where boundary conditions were kept constant for those compounds due to the lack of precise information to constrain their trends, and their inflow into various European region depends on meteorological conditions. Their contribution to the total PM_{10} change is, for dust, highest for the Iberian Peninsula (decreasing) and the Mediterranean (increasing) and for sea salt, highest in the "Inflow" region, England and Scandinavia (both decreasing), but it remains below 20% of the total change.

8 Conclusion

Thanks to the development of air pollution monitoring networks and modelling capacities, it is now possible to engage into the analysis of the long-term evolution of atmospheric composition. The spatial coverage of monitoring networks has developed since the 1990s, so that we can assess the evolution of air pollution over the past 20 years. At least that is the case for ozone and inorganic aerosols, whereas for PM_{10} and $PM_{2.5}$, the development of monitoring capacities really emerged after 2000.

Observation-based trend assessments were recently performed by several technical bodies such as the European Topic Centre on Air Pollution and Climate Change Mitigation (Colette et al., 2015), the Working Group on Effects (De Wit et al., 2015) of the Geneva Convention on the Long Range Transboundary Air Pollution, or the EMEP Task Force on Measurement and Modelling of the same convention (Colette et al., 2016a).

In terms of modelling, numerical capacities have also improved, so an ensemble of teams can engage into long terms simulations of air quality over Europe.

In the present report, we assessed the capacity of those air quality models to capture the main features of ozone and particulate matter air pollution trends in Europe since 1990. By comparing modelled and observed trends, and investigating any drift in model performances, we could demonstrate the following key features:

- Models tend to overestimate the decrease of ozone peaks and inorganic aerosol load in the 1990s, the agreement is better in the 2000s.
- The capacity of models in capturing ozone peak improved in time.
- Models capture very well the average decrease in PM_{10} for the 2000-2010 period but they underestimate the range of spatial variability of decreasing trends in PM_{10}.

Last, we presented an attribution analysis designed to quantify the relative importance of (i) European anthropogenic emission changes, (ii) meteorological variability, and (iii) global background chemistry on surface air quality in Europe. This analysis shows the dominating impact of emission changes on observed air quality trends, thereby demonstrating the effectiveness of air quality management strategies. We also show the substantial impact of meteorological variability and background chemistry, therefore showing the need to account for those factors to reconcile the quantitative comparison between trends in emission of precursors and ambient concentrations.

The models can also be used to map the long-term evolution of air quality, and we presented such ensemble maps for 1990, 2000 and 2010, that can now be used for further analysis, for instance in impact studies focusing on the evolution of health and ecosystem impacts of air pollution since 1990.

9 Acknowledgement

Modelling data used in the present analysis were produced in the framework of the EuroDelta-Trends Project initiated by the EMEP Task Force on Measurement and Modelling of the Convention on Long Range Transboundary Air Pollution. EuroDelta-Trends is coordinated by INERIS and involves modelling teams of BSC, CEREA, CIEMAT, ENEA, IASS, JRC, MET Norway, TNO, SMHI. The views expressed in this study are those of the authors and do not necessarily represent the views of Eurodelta-Trends modelling teams.

10 References

Ahmadi, M., and John, K.: Statistical evaluation of the impact of shale gas activities on ozone pollution in north Texas Sci. Total Environ., 536, 457–467, http://dx.doi.org/10.1016/j.scitotenv.2015.06.114, 2015.

Amann, M., Borken-Kleefeld, J., Cofala, J., Heyes, C., Klimont, Z., Rafaj, P., Purohit, P., Schoepp, W., and Winiwarter, W.: Future emissions of air pollutants in Europe - Current legislation baseline and the scope for further reductions., IIASA, Laxenburg, Austria, 2012.

Avnery, S., Mauzerall, D. L., Liu, J., and Horowitz, L. W.: Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage, Atmos. Environ., 45, 2284-2296, 2011.

Banzhaf, S., Schaap, M., Kranenburg, R., Manders, A. M. M., Segers, A. J., Visschedijk, A. J. H., Denier van der Gon, H. A. C., Kuenen, J. J. P., van Meijgaard, E., van Ulft, L. H., Cofala, J., and Builtjes, P. J. H.: Dynamic model evaluation for secondary inorganic aerosol and its precursors over Europe between 1990 and 2009, Geosci. Model Dev., 8, 1047-1070, 10.5194/gmd-8-1047-2015, 2015.

Barmpadimos, I., Hueglin, C., Keller, J., Henne, S., and Prévôt, A. S. H.: Influence of meteorology on PM10 trends and variability in Switzerland from 1991 to 2008, Atmos. Chem. Phys., 11, 1813-1835, 2011.

Bessagnet, B., Pirovano, G., Mircea, M., Cuvelier, C., Aulinger, A., Calori, G., Ciarelli, G., Manders, A., Stern, R., Tsyro, S., García Vivanco, M., Thunis, P., Pay, M. T., Colette, A., Couvidat, F., Meleux, F., Rouïl, L., Ung, A., Aksoyoglu, S., Baldasano, J. M., Bieser, J., Briganti, G., Cappelletti, A., D'Isidoro, M., Finardi, S., Kranenburg, R., Silibello, C., Carnevale, C., Aas, W., Dupont, J. C., Fagerli, H., Gonzalez, L., Menut, L., Prévôt, A. S. H., Roberts, P., and White, L.: Presentation of the EURODELTA III intercomparison exercise – evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology, Atmos. Chem. Phys., 16, 12667-12701, 10.5194/acp-16-12667-2016, 2016.

Byun, D. W., and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, Appl. Mech. Rev., 59, 51-77, 2006.

Christensen, J., and Christensen, O.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7-30, 2007.

Colette, A., Granier, C., Hodnebrog, O., Jakobs, H., Maurizi, A., Nyiri, A., Bessagnet, B., D'Angiola, A., D'Isidoro, M., Gauss, M., Meleux, F., Memmesheimer, M., Mieville, A., Rouïl, L., Russo, F., Solberg, S., Stordal, F., and Tampieri, F.: Air quality trends in Europe over the past decade: a first multi-model assessment, Atmos. Chem. Phys., 11, 11657-11678, 2011.

Colette, A., Beauchamp, M., Malherbe, L., and Solberg, S.: Air Quality Trends in AIRBASE in the context of the LRTAP Convention, ETC/ACM Technical Paper 2015/4, 2015.

Colette, A., Aas, W., Banin, L., Braban, C. F., Ferm, M., González Ortiz, A., Ilyin, I., Mar, K., Pandolfi, M., Putaud, J.-P., Shatalov, V., Solberg, S., Spindler, G., Tarasova, O., Vana, M., Adani, M., Almodovar, P., Berton, E., Bessagnet, B., Bohlin-Nizzetto, P., Boruvkova, J., Breivik, K., Briganti, G., Cappelletti, A., Cuvelier, K., Derwent, R., D'Isidoro, M., Fagerli, H., Funk, C., Garcia Vivanco, M., Haeuber, R., Hueglin, C.,

Jenkins, S., Kerr, J., de Leeuw, F., Lynch, J., Manders, A., Mircea, M., Pay, M. T., Pritula, D., Querol, X., Raffort, V., Reiss, I., Roustan, Y., Sauvage, S., Scavo, K., Simpson, D., Smith, R. I., Tang, Y. S., Theobald, M., Tørseth, K., Tsyro, S., van Pul, A., Vidic, S., Wallasch, M., and Wind, P.: Air pollution trends in the EMEP region between 1990 and 2012, NILU, Oslo, 2016a.

Colette, A., Andersson, A., Manders, A., Mircea, M., Mar, K., Pay, M.-T., Raffort, V., Tsyro, S., Cuvelier, C., Bessagnet, B., Bergström, R., Butler, T., Couvidat, F., Doumbia, T., Fagerli, H., Garnier, C., Heyes, C., Klimont, Z., Ojha, N., Otero, N., Sindelarova, K., Stegehuis, A. I., Roustan, Y., Vautard, R., van Meijgaard, E., Vivanco, M. G., and Wind, P.: EURODELTA-Trends, a multi-model experiment of air quality hindcast in Europe over 1990-2010., Geosci. Model Dev. Discuss., in prep, 2016b.

De Wit, H. A., Hettelingh, J. P., and Harmens, H.: Trends in ecosystem and health responses to long-range transported atmospheric pollutants, NIVA, Oslo, Norway, 92, 2015.

EEA: Assessment of ground-level ozone in EEA member countries, with a focus on long-term trends, European Environment Agency, Copenhagen, 56, 2009.

Eskes, H., Huijnen, V., Arola, A., Benedictow, A., Blechschmidt, A.-M., Botek, E., Boucher, O., Bouarar, I., Chabrillat, S., Cuevas, E., Engelen, R., Flentje, H., Gaudel, A., Griesfeller, J., Jones, L., Kapsomenakis, J., Katragkou, E., Kinne, S., Langerock, B., Razinger, M., Richter, A., Schultz, M., Schulz, M., Sudarchikova, N., Thouret, V., Vrekoussis, M., Wagner, A., and Zerefos, C.: Validation of reactive gases and aerosols in the MACC global analysis and forecast system, Geosci. Model Dev., 8, 3523–3543, doi:10.5194/gmd-8-3523-2015, 2015.

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.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 8, 975-1003, 10.5194/gmd-8-975-2015, 2015.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957-6975, http://dx.doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

Henneman, L. R. F., Holmes, H. A., Mulholland, A., and Russell, A. G.: Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000-2012) data in Atlanta, Atmos. Environ., 119, 201-210, 2015.

Katragkou, E., Zanis, P., Tsikerdekis, A., Kapsomenakis, J., Melas, D., Eskes, H., Flemming, J., Huijnen, V., Inness, A., Schultz, M. G., Stein, O., and Zerefos, C. S.: Evaluation of near-surface ozone over Europe from the MACC reanalysis, Geosci. Model Dev., 8, 2299–2314, doi:10.5194/gmd-8-2299-2015, 2015.

Klimont, Z., Hoeglund-Isaksson, L., Heyes, C., Rafaj, P., Schoepp, W., Cofala, J., Purohit, P., Borken-Kleefeld, J., Kupiainen, K., Kiesewetter, G., Winiwarter, W., Amann, M., Zhao, B., Wang, S., Bertok, I., and Sander, R.: Global scenarios of air pollutants and methane: 1990-2050, in prep., 2016a.

Klimont, Z., Kupiainen, K., Heyes, C., Purohit, P., Cofala, J., Rafaj, P., and Schoepp, W.: Global anthropogenic emissions of particulate matter, in preparation, 2016b.

Lefohn, A., Emert, C., Shadwick, D., Wernli, H., Jung, J., and Oltmans, S.: Estimates of background surface ozone concentrations in the United States based on model-derived source apportionment, Atmos. Environ., 84, 275–288, doi:10.1016/jatmosenv.2013.11.033, 2014.

Lefohn, A. S., Malley, C. S., Simon, H., Wells, B., Xu, X., Zhang, L., and Wang, T.: Responses of human health and vegetation exposure metrics to changes in ozone

concentration distributions in the European Union, United States, and China, Atmospheric Environment, 152, 123-145, http://dx.doi.org/10.1016/j.atmosenv.2016.12.025, 2017.
Maas, R., and Grennfelt , P.: Towards Cleaner Air - Scientific Assessment Report 2016, EMEP-Steering body and Working Group on Effects - Convention on Long-Range Transboundary Air Pollution 2016.
Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H.: Technical Note: The air quality modeling system Polyphemus, Atmos. Chem. Phys., 7, 5479-5487, doi:10.5194/acp-7-5479-2007, 2007.
Mar, K. A., Ojha, N., Pozzer, A., and Butler, T. M.: Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: Model evaluation and chemical mechanism comparison, Geosci. Model Dev. Discuss., 2016, 1-50, 10.5194/gmd-2016-131, 2016.
Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J. L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, Geosci. Model Dev., 6, 981-1028, 2013.
Mircea, M., Grigoras, G., D'Isidoro, M., Righini, G., Adani, M., Briganti, G., Ciancarella, L., Cappelletti, A., Calori, G., Cionni, I., Cremona, G., Finardi, S., Larsen, B. R., Pace, G., Perrino, C., Piersanti, A., Silibello, C., Vitali, L., and Zanini, G.: Impact of grid resolution on aerosol predictions: a case study over Italy, Aerosol and Air Quality Research, 1253–1267, doi: 10.4209/aaqr.2015.02.0058, 2016.
Rasmussen, D. J., Hu, J., Mahmud, A., and Kleeman, M. J.: The Ozone-Climate Penalty: Past, Present, and Future, Environmental Science & Technology, 47, 14258-14266, 2013.
Robertson, L., Langner, J., and Engardt, M.: An Eulerian Limited-Area Atmospheric Transport Model, Journal of Applied Meteorology, 38, 190-210, 1999.
Santurtún, A., González-Hidalgo, J. C., Sanchez-Lorenzo, A., and Zarrabeitia, M. T.: Surface ozone concentration trends and its relationship with weather types in Spain (2001–2010), Atm. Env., 101, 10–22, 2015.
Sauter, F., van der Swaluw, E., Manders-Groot, A., Wichink Kruit, R., Segers, A., and Eskes, H.: LOTOS-EUROS v1.8 Reference Guide, TNO, Utrecht, The Netherlands, 2012.
Schaap, M., Timmermans, R. M. A., Roemer, M., Boersen, G. A. C., Builtjes, P., Sauter, F., Velders, G., and Beck, J.: The LOTOS-EUROS model: description, validation and latest developments, International Journal of Environment and Pollution, 32, 270-290, 2008.
Simpson, D., Benedictow, A., Berge, H., Bergstrom, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyiri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J. P., Valdebenito, A., and Wind, P.: The EMEP MSC-W chemical transport model - technical description, Atmos. Chem. Phys., 12, 7825-7865, 2012.
Solberg., S., Colette, A., and Guerreiro, C.: Discounting the impact of meteorology to the ozone concentration trends ETC/ACM, Bilthoven, 2015.
Stegehuis, A. I., Vautard, R., Ciais, P., Teuling, A. J., Miralles, D. G., and Wild, M.: An observation-constrained multi-physics WRF ensemble for simulating European mega heat waves, Geosci. Model Dev., 8, 2285-2298, 10.5194/gmd-8-2285-2015, 2015.
Tang, Q., Hess, P. G., Brown-Steiner, B., and Kinnison, D. E.: Tropospheric ozone decrease due to the Mount Pinatubo eruption: reduced stratospheric influx, Geophys. Res. Lett., 40, 5553–5558, doi:10.1002/2013GL056563, 2013.

Tong, D. Q., Mathur, R., Kang, D. W., Yu, S. C., Schere, K. L., and Pouliot, G.: Vegetation exposure to ozone over the continental United States: Assessment of exposure indices by the Eta-CMAQ air quality forecast model, Atm. Env., 43, 724–733, 2009.

Watson, L., Lacressonnière, G., Gauss, M., Engardt, M., Andersson, C., Josse, B., Marécal, V., Nyiri, A., Sobolowski, S., Siour, G., and Vautard, R.: The impact of meteorological forcings on gas phase air pollutants over Europe, Atm. Env., 119, 240-257, 2015.

Wild, O., Fiore, A. M., Shindell, D. T., Doherty, R. M., Collins, W. J., Dentener, F. J., Schultz, M. G., Gong, S., MacKenzie, I. A., Zeng, G., Hess, P., Duncan, B. N., Bergmann, D. J., Szopa, S., Jonson, J. E., Keating, T. J., and Zuber, A.: Modelling future changes in surface ozone: a parameterized approach, Atmos. Chem. Phys., 12, 2037-2054, 2012.

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, Int. J. Climatol., 32, 2088 –2094, 2012.

Wilson, R. C., Fleming, Z. L., Monks, P. S., Clain, G., Henne, S., Konovalov, I. B., Szopa, S., and Menut, L.: Have primary emission reduction measures reduced ozone across Europe? An analysis of European rural background ozone trends 1996-2005, Atmos. Chem. Phys., 12, 437-454, 2012.

11 Annex: Statistical analyses

11.1 The Mann-Kendall test

For analysing a possible trend in observed time series the non-parametric Mann-Kendall test (Gilbert, 1987) has been used. This test is particularly useful since missing values are allowed and the data need not to conform to any particular distribution. Moreover, as only the relative magnitudes of the data rather than their actual measured values are used, this test is less sensitive towards incomplete data capture and/or special meteorological conditions leading to extreme values.
In the trend analyses a *consistent set* of stations is used. Requirements for a consistent set are:

- for each year within the time period a minimum data coverage of 75% is required;
- annual data is available for at least 75% of the years within the time period.

The Mann-Kendall statistic *S* is defined as:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}\left(x_j - x_k\right) \qquad [1]$$

where

$$\begin{aligned} \operatorname{sgn}(x_j - x_k) &= 1 \quad \text{if } (x_j - x_k) > 0 \\ &= 0 \quad \text{if } (x_j - x_k) = 0, \\ &= -1 \quad \text{if } (x_j - x_k) < 0 \end{aligned}$$

x_j is the observable (concentration, number of exceedance days, exposure) in year *j*; *n* is the available number of years with a valid measurement. In other words, *S* is the number of positive differences minus the number of negative differences. If *S* is a large positive number measurements taken later in time tend to be larger than those taken earlier in time. Similarly, if *S* is a large negative number, this indicates a downward trend. The Mann-Kendall statistic is only calculated for consistent sets of stations.
If a linear trend is assumed, the time series can be given as:

$$C_t = B + Q \cdot t + \varepsilon_t \qquad [2]$$

where C_t is the concentration in year *t*, *B* is the intercept, *Q* is the slope and the residuals ε_t have a zero mean. The slope is estimated by Sen's non-parametric procedure (Gilbert, 1987). For each time series with *n* valid measurements a set of slope estimates Q_{jk} is computed for each of the $n(n\text{-}1)/2$ data pairs:

$$Q_{jk} = \frac{x_j - x_k}{j - k} \qquad [3]$$

Sen's slope estimate equals the median of the $N=n(n\text{-}1)/2$ slope estimates.
To calculate the 95% confidence interval about the true slope, first

$$D_\alpha = Z_{1-\alpha/2} \sqrt{VAR(S)}$$

is calculated. *VAR(S)* is the variance of the Mann-Kendall statistics *S* and is, when no ties are present, calculated by:

$$VAR(S) = \frac{1}{18}\left[N(N-1)(2N+5)\right]$$

$Z_{1-\alpha/2}$ is obtained from the standard normal distribution ($Z_{0.975} = 1.96$).

Next $M_1 = (N-D_\alpha)/2$ and $M_2 = (N+D_\alpha)/2$ are calculated. The lower, Q_L, and upper limits, Q_U, of the confidence interval are the M_1-th largest and (M_2 + 1)-th largest of the n ordered slope estimates. If M_1 or M_2 is not an integer value, the lower or upper limit is interpolated. The standard uncertainty σ_i is given by:

$$\sigma_i = \frac{1}{4}(Q_U - Q_L) \qquad [4]$$

To obtain an estimate of B (the concentration at $t=0$), equation [2]) the n values of the differences $x_i - Q\,t_i$ are calculated. The median of these values gives an estimate of B.

The uncertainty of region-averaged trends are reported with 2σ errors calculated through error propagation from the uncertainties in trends at the individual stations (σ_i, equation [4]) by means of Eq [5] where N_R is the number of individual stations in region R:

$$\sigma_R = \frac{\sqrt{\sum \sigma_i^2}}{N_R} \qquad [5]$$

To test for homogeneity of trend direction at M multiple stations, compute the homogeneity chi-square statistic, χ^2_{hmg}, where

$$\chi^2_{hmg} = \chi^2_{total} - \chi^2_{trend} = \sum_{j=1}^{M} Z_j^2 - M\bar{Z}^2$$

$$Z_j = \frac{S_j}{\sqrt{VAR(S_j)}}$$

S_j is the Mann-Kendall trend statistic for the j-th station, and

$$\bar{Z} = \frac{1}{M}\sum_{j=1}^{M} Z_j$$

If the trend at each station is in the same direction, then χ^2_{hmg} has a chi-square distribution with M-1 degrees of freedom (df). To test for rend homogeneity between stations at the α significance level, the calculated value of χ^2_{hmg} is referred to the α critical value with M-1 degrees of freedom in a chi-square distribution table. If χ^2_{hmg} exceeds this critical value, the H_o hypothesis of homogeneous station trends is rejected. In this case no regional-wide statements should be made about trend direction. If χ^2_{hmg} does not exceed the α critical value in the chi-square distribution table, then the statistics $\chi^2_{trend} = M\bar{Z}^2$ is referred to the chi-square distribution with 1 df to test the null hypothesis that the (common) trend direction is significantly different from zero.

11.2 The Kolmogorov–Zurbenko filter

The definition of the KZ filtering is given in the box below, adapted from Wikipedia.

The Kolmogorov-Zurbenko (KZ) filter with parameters m and k can be defined as k time iterations of a moving average (MA) filter of m points. It can be obtained through iterations.

First iteration is to apply a MA filter over process $X(t)$

$$KZ_{m,k=1}[X(t)] = \sum_{s=-(m-1)/2}^{(m-1)/2} X(t+s) \times \frac{1}{m}$$

The second iteration is to apply the MA operation to the result of the first iteration,

$$KZ_{m,k=2}[X(t)] = \sum_{s=-(m-1)/2}^{(m-1)/2} KZ_{m,k=1}[X(t+s)] \times \tfrac{1}{m}$$

$$= \sum_{s=-2(m-1)/2}^{2(m-1)/2} X(t+s) \times a_s^{m,k=2}$$

Generally the *kth* iteration is an application of the MA filter to the *(k-1)th* iteration. The iteration process of a simple operation of MA is very convenient computationally.

11.3 Glossary

AOT40:	Accumulated ozone exposure over the threshold of 40 ppb
FGE:	Fractional gross error
IOAr:	Refined index of agreement
MAE:	Mean absolute error
MDA8:	Max daily 8-hour mean concentration
MDA8s:	Max daily 8-hour mean concentration during the summer half year
MNMB:	Modified normalized mean bias
SOMO35:	Yearly sum of max daily 8-hour mean concentration over 35 ppb
4hMDA8:	The annual 4th highest MDA8