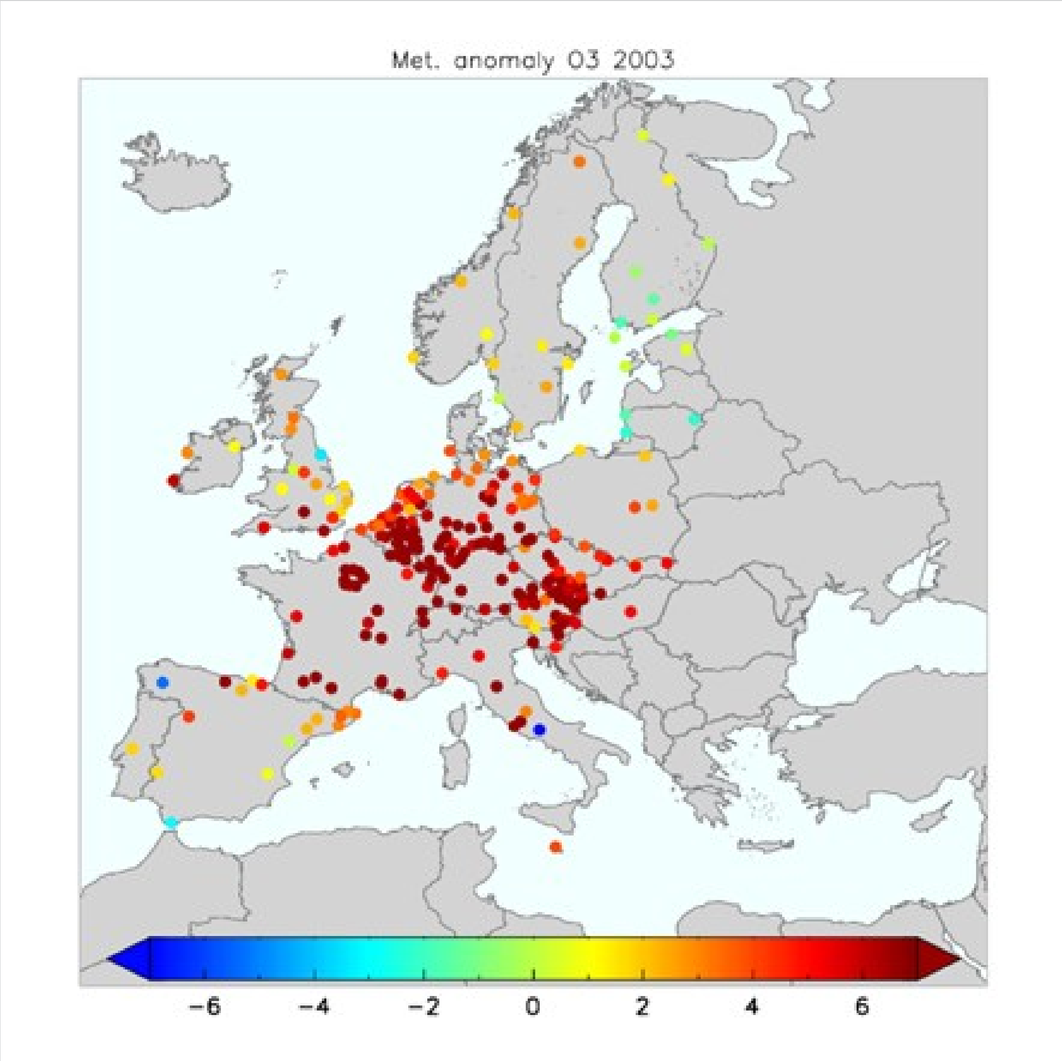
Eionet Report - ETC/ATNI 2019/14

Statistical modelling for long-term trends of pollutants

Use of a GAM model for the assessment of measurements of O_3, NO_2 and PM

December 2019



Authors:
Sverre Solberg (NILU), Sam-Erik Walker (NILU), Cristina Guerreiro (NILU), Augustin Colette (INERIS)

ETC/ATNI consortium partners:
NILU – Norwegian Institute for Air Research, Aether Limited, Czech Hydrometeorological Institute (CHMI), EMISIA SA, Institut National de l'Environnement Industriel et des risques (INERIS), Universitat Autònoma de Barcelona (UAB), Umweltbundesamt GmbH (UBA-V), 4sfera Innova, Transport & Mobility Leuven NV (TML)

European Environment Agency
European Topic Centre on Air pollution, transport, noise and industrial pollution

Cover picture: The meteorological anomaly of the maximum daily 8-h running mean surface ozone concentration in 2003. Source authors.

The withdrawal of the United Kingdom from the European Union did not affect the production of the report.
Data reported by the United Kingdom are included in all analyses and assessments contained herein, unless otherwise indicated.

Author(s)
Sverre Solberg (NILU), Sam-Erik Walker (NILU), Cristina Guerreiro (NILU) and Augustin Colette (INERIS)

ETC/ATNI c/o NILU
ISBN 978-82-93752-12-7

European Topic Centre on Air pollution,
transport, noise and industrial pollution
c/o NILU – Norwegian Institute for Air Research
P.O. Box 100, NO-2027 Kjeller, Norway
Tel.: +47 63 89 80 00
Email: etc.atni@nilu.no
Web : https://www.eionet.europa.eu/etcs/etc-atni

Preface

This report was prepared as part of ETC/ATNI Task 1.1.2.2 Long term air quality measurement trends in Europe in 2019. It constitutes the deliverable to subtask 2 : «ETC Technical Paper on the role of meteorological variability in driving air quality».

Contents

Executive Summary

The current report provides a short overview of previous years' studies on long-term trends in O_3, NO_2 and PM and the role of meteorological variability for the concentration of these pollutants. The report furthermore includes a comprehensive study applying the final version of the statistical GAM model to monitoring data of O_3, NO_2, PM_{10} and $PM_{2.5}$ and model meteorological data covering the entire time period 2000-2017.

The previous studies on the link between trends and meteorology has shown that these links could be estimated by a careful design of model setups using CTMs (chemical transport models). This was a main focus of the EuroDelta Trends exercise where the relative influence of boundary conditions vs that of meteorological variability and vs that of emissions were estimated from the differences between various model scenarios. In this way, one could attribute the individual contributions from each of these three processes to the concentration levels of pollutants such as O_3, NO_2 and PM. This procedure as well as other associated procedures based on the same model scenarios indicates that emission changes have been the main driver for the downward pollutant trends during 1990-2010 while meteorological variability lead to an additional decline in pollutants during the first decade. For the 2000-2010 period, the influence of the meteorological variability on the trends was smaller.

The CTMs are certainly useful tools for explaining pollutant trends but require multi-year scenario calculations designed in very specific ways. This approach could also be sensitive to the years selected for calculating the perturbations in boundary conditions and meteorology.

The statistical GAM model that has been developed during the last years provides an alternative or complementary method for separating the influence of meteorological variability from other processes. This model represents a completely different approach that is based only on links between local meteorological parameters (like temperature, wind etc) and observed pollutant concentration levels. Thus, the model does not contain any representation of the real processes in the atmosphere. The GAM model is therefore an efficient tool that only require access to meteorological data. The main limitation is that this model relies on the assumption of in-situ relationships between meteorology and pollutant concentration which is not really valid for most pollutants. In spite of this essential limitation, it turns out that the model provides good agreement with the observed data in some regions as explained in more detail below. In other regions, mainly those which are located further away from the main emission regions, the performance of the statistical model is poorer. Furthermore, the statistical model seems less capable of predicting the high peak levels which often is of more interest than the mean levels.

Based on the experience with the statistical modelling we see three main applications of the method and the report is focussed on these applications:

1. Separate the long-term trend in observed concentration levels from the variations induced by meteorological variability and additionally look for any trends induced by meteorology alone.
2. Evaluate to what extent the pollutant levels in one specific year deviates from the expected level due to meteorological anomalies that year.
3. Identify possible flaws in the measurement data. Since the statistical model is based on systematic patterns between meteorology and concentration levels, a particularly poor model performance could indicate errors in the observational data.

Monitoring data were extracted from the data stored in EEA's Airbase and e-Reporting data base as well as from EMEP's database EBAS. Meteorological data were extracted from ECMWF as daily or 6-h data with a spatial resolution of 0.3° x 0.3°.

We found differences in model performance both with respect to geographical area and atmospheric species. In general, the best performance was found for O_3 with gradually lower performance for NO_2, PM_{10} and $PM_{2.5}$ in that order. With respect to area, the model generally produced the best predictions for Central Europe (Germany, Netherlands, Belgium, France, Austria, Czech Republic) and the poorest for southern Europe. There were some differences for the performance of the individual species. For summertime ozone, the best agreement with observed data was found in the Mid- and East-Europe while the poorest agreement was found for the Iberian Peninsula and the Mediterranean. Furthermore, for England we found a slightly poorer agreement with ozone than with wintertime NO_2. For the Scandinavian sites, a fair agreement was found for summertime O_3. The timeseries of daily data revealed, however, that this could be explained by the fact that the GAM model reproduced the seasonal cycle very well although the episodes of peak ozone tend to be significantly underestimated.

For wintertime NO_2, particularly poor agreement between the GAM model and the measurements was found for the North Italian region. The model agreement for southern Europe and the Iberian Peninsula was also fairly low, although variable from site to site.

The number of stations with measurements of PM_{10} and $PM_{2.5}$ with sufficient length was substantially lower than for O_3 and NO_2, and thus, a region-by-region comparison of the model performance was not really possible. In general, the PM_{10} data indicated a better agreement between the model and the measurements for summer than for winter. Furthermore, the GAM model seemed to perform better for the background urban than for rural sites. Poorest performance for PM_{10} was found for background rural sites in winter.

Over the 18 years period studied (2000-2017) we found very few cases for which the meteorology alone caused a statistically significant trend in the data. One exception is the O_3 sites in Mid Europe. For the sites in this region taken together, it is estimated that meteorology alone caused a slight increase in the summer mean MDA8 levels. The general lack of meteorology induced trends in the data reflects the length of the time series in this study. A time period of 18 years is presumably a sufficiently long period that interannual variations in meteorology is reduced and still too short for climate change to have a noticeable effect.

Withdrawing the meteorological factor by the GAM model can help in identifying significant trends. The main reason for this is that meteorology introduces a year-to-year variability which could mask the underlying trend. Meteorology could also induce a trend in the concentrations but this is a matter of length of the timeseries. For short time periods (typically less than 10 years) the variations in meteorology could lead to spurious effects reflecting the weather conditions at the start and end year. On a long timescale the effects of climate change (trends in temperature, precipitation etc) will certainly lead to trends in the concentration of pollutants, but this is beyond the scope of this report.

For rural ozone, a decline is calculated in the meteorologically adjusted (hereafter named meteorology adjusted) trends in all regions except for the inflow region at the north-western boundary of Europe that shows only minor variations during the 2000-2017 period. Some regions show a steady decline in ozone while others show a curve peaking in the early 2000s. Many of the regions indicate a flattening of the ozone trend in the last part of the period. The meteorology adjusted trends for NO_2 show a similar pattern as for O_3 with decreasing levels in all regions. As for O_3 the NO_2 trends are seen as a steady decline in some regions and a curve peaking in the early 2000s in other regions. Marked downward meteorology adjusted trends are found for PM_{10} as well as substantial variability from year to year caused by meteorology.

Acknowledgements

Sabine Eckhardt at NILU is acknowledged for extracting and providing the ECMWF meteorological data. Christoffer Stoll at NILU is acknowledged for the preparation of air quality monitoring data from AIRBASE and e-Reporting.

EEA project manager was Evrim Oztürk.

Abbreviations and definitions

CAMS: Copernicus Atmosphere Monitoring Service

CTM: Chemical transport model

ECMWF: European Centre of Medium-Range Weather-Forecasts

EEA: European Environment Agency

EMEP: European Monitoring and Evaluation Programme

ETC/ACM: European Topic Centre for Air Pollution and Climate Change Mitigation

ETC/ATNI: European Topic Centre on Air pollution, noise, transport and industrial pollution

EDT: EuroDelta Trends

GAM: Generalized Additive Model

MDA8 : The maximum daily 8-h running mean concentration defined as in the AQ directive. This metric was only used for ozone.

NMGE : Normalised mean gross error. NMGE= $\Sigma|M_i - O_i| / \Sigma O_i$ (where M_i = model value at day *i* and O_i = observed value at day *i*).

NO_2: Nitrogen dioxide

O_3: Ozone

PBL: Planetary boundary layer (the height of the layer near the ground that could be regarded well mixed vertically during the day, typically in order of a few hundred meters up to 2 km height).

PM: Particulate matter

$PM_{2.5}$: Particulate matter with a diameter of 2.5 µm or less

PM_{10}: Particulate matter with a diameter of 10 µm or less

ppb: Parts per billion (unit of concentration)

r: Linear correlation coefficient

R^2: Coefficient of determination

US-EPA: United States Environmental Protection Agency

UTC: Coordinated Universal Time

µg/m^3: Microgram(s) per cubic metre

1 Introduction

During the last few years, several studies on trends in the air pollutants O_3, NO_2 and PM have been carried out within the framework of EEA (Colette et al., 2015; Colette et al., 2016; Solberg et al., 2018a, Solberg et al., 2018b). In addition to the issue of the magnitude of the trends themselves, a central question has been the influence of meteorological variability vs that of anthropogenic emissions for the observed trends.

Normally, such questions are analysed by the use of chemical transport models (CTMs) which aims at predicting the atmospheric levels of pollutants through parameterizations of advection, turbulent diffusion, chemical reactions, surface interactions etc. and which then could be used to investigate the sole influence of each process. There is a scientific consensus that state-of-the-art CTMs provide the best approach for predicting and analysing trends and variabilities of atmospheric pollutants. However, using CTMs for simulating very long time periods in a multi-scenario approach could be a costly and time-consuming task. Secondly, the analyses of the model results become non-trivial when there are significant discrepancies between the model predictions and the measured levels of pollutants. One background for the EEA trend studies was a statement in the 2013 Air Quality Report (EEA, 2013): "… t*here is a discrepancy between the past reductions in emissions of O_3 precursor gases in Europe and the change in observed average O_3 concentrations in Europe*". This raised the question whether the discrepancy was due to errors in the emission data, lack of performance by the CTMs or simply the range of uncertainty of the links between precursor emissions and ozone levels.

In parallel with the long-term development of CTMs into present day's highly complex deterministic models, several studies in the scientific literature has been devoted to the application of statistical models. Such models aim to link the level of pollutants to several input explanatory variables on the basis of correlations alone without any attempt of reproducing the actual causal physio-chemical processes. Thus, there is a fundamental difference between these two approaches. Whereas a CTM tries to simulate the dependencies in a cause-effect relationship, the statistical model simply look for patterns in the data, i.e. correlation between two or more physical quantities. Such statistical models are commonly met by the critic that they are in some sense worthless and only offering a "poor man's model" that was popular before complex 3-D CTMs could be applied. A large number of scientific papers have shown, however, that statistical based models are useable when they are designed in a careful way and that they could be used as an addition to the much more advanced CTMs (e.g. Thompson et al., 2001; Ordonez et al., 2005; Camalier et al., 2007; Zheng et al., 2007; Chan 2009; Davies et al., 2011; Fix et al., 2018; Otero et al., 2018; Pernak et al., 2019).

A literature review of such statistical models applied to surface ozone was prepared through an ETC task in 2014 (Solberg et al., 2015). In 2015, a trend study for NO_2, O_3 and PM (Colette et al., 2015) was applied to Airbase monitoring data as a complement to the ongoing trend work within UN-ECE for EMEP stations. One conclusion from this study was that it was difficult to conclude on the relative importance of European emission policies and externalities such as intercontinental transport or meteorological variability, and the use of CTMs was recommended. In 2016, this recommendation was followed up in an ETC task (Colette et al., 2016) devoted to the analyses of trends of O_3 and PM_{10} within the project EuroDelta-Trends (EDT). The ETC task in 2016 and the associated EDT exercise was looking at both measured and modelled data, but the two separate ways of isolating the influence of meteorology (and boundary conditions) from the influence of emissions that were applied were based on modelled data only.

Then, in 2017, the trend work was continued in an ETC task (Solberg et al., 2018a) that was focused on the use of a certain statistical method, a so-called generalized additive model (GAM), to O_3 monitoring data. The purpose of this study was to see if a procedure applied by the US-EPA (Camalier et al., 2007) on a regular annual basis to their O_3 data for subtracting the influence of interannual variations in meteorology was applicable also to the European O_3 data. In 2018, this work was continued by applying a similar GAM method to Airbase data for NO_2 and PM_{10}.

With these various studies as a background, the purpose of the present report is to provide a synthesis and recommendation for the possible future use of statistical models for “removing” the influence of meteorological anomalies on measured pollutants and their associated trends.

Key results of the various studies are given below together with the different methods’ strengths, weaknesses, limitations and requirements.

2 Short overview of previously applied methods within ETC

Chapters 2.1 to 2.3 provide the main features of the trend studies carried out in 2016-2018.

2.1 Studies based on the EuroDelta Trends model exercise (2016)

These results were documented by Colette et al. (2016).

Time period: 1990-2010.
Pollutants: O_3, PM_{10}, $PM_{2.5}$ (after 2000). Two methods applied as described in Ch. 2.1.1. and 2.1.2.

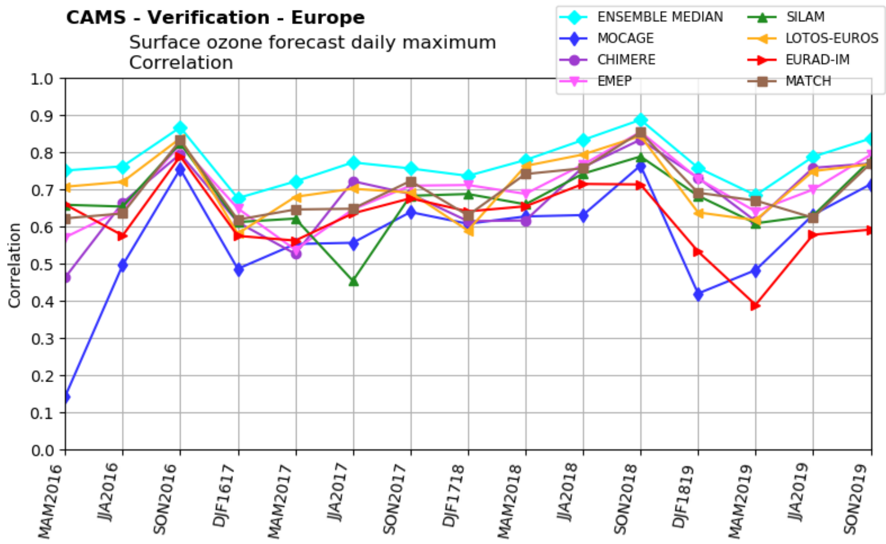
2.1.1 EuroDelta Trends (EDT) attribution procedure based on model scenarios

This was a purely CTM based approach. It was designed specifically to allow for an attribution of the impact of European emission changes vs that of meteorology and boundary conditions, respectively. An ensemble of models was used with the same predefined trends in anthropogenic emissions. One specific set of boundary conditions was used for all models.

The approach was based on the concept of ozone being the result of annual variations in three processes: emissions, boundary conditions and meteorology. By designing model runs corresponding to various linear combinations of these three processes, the trend due to meteorology alone was calculated as the additive combination of these specific model runs.

We could not retrieve ozone correlation scores in the EuroDelta Reports since these data were not included in the reports,, but referring to CAMS, the correlation of daily maximum ozone in state of the art CTMs is of the order or 0.6-0.8, reaching 0.8-0.9 for the CAMS ENSEMBLE model (Figure 1).

Figure 1: Ozone correlation in the CAMS Operational Regional Forecasts, for each quarter since spring 2016, courtesy V. Petiot, Meteo-France.



The metrics studied were the following: Annual mean PM_{10} and 4th highest MDA8 O_3 for the summer half year (April-September).

The results were aggregated and presented for nine European regions.

2.1.2 *Estimation of trends as deviations from a model climatology*

It is possible to define a term called "pollutant climatology" or "ozone climatology" similar to the more familiar weather climatology. The weather climatology is calculated as the 30-years mean of various weather characteristics (temperature, wind, precipitation etc) as a function of time of year at a defined site. A corresponding ozone climatology would refer to the 30-years mean of ozone as a function of time of year, but since this depends on the weather conditions as well as on the ozone precursor emissions (NO_x and VOC), we have to define the ozone climatology as the mean (and spread) in the ozone levels at a certain site and time of year given the 30 years of different meteorologies and given the emissions for that specific year. Thus, the ozone climatology would have to be defined as a model quantity. It could not be determined by observations. The ozone climatology could be written as:

$O_{3,clim}$ ($year_j$) = $1/n\Sigma O_3(i)$ i = 1, …, n (years), assuming emissions for $year_j$

In other words, the ozone climatology for one specific year would have to be calculated by a CTM, running the model for 30 years of meteorological data using the same emission data applicable to that specific year. In practice we will seldom have access to 30 years of meteorological data and thus we will have to limit the climatology to the number of years available. This concept of ozone climatology was applied by Colette et al. (2016).

In Colette et al. (2016), this approach was based on three sets of model runs from the EMEP model, two of them corresponding to the EuroDelta Trend scenarios as mentioned above. Results from these three model runs were used to estimate the climatology of the pollutant levels each year and the deviation from this climatology each year.

It was shown that the actual model scenario for one specific year could be estimated with a good accuracy as the linear combination of these runs. The possible influence of changes in input boundary conditions had to be neglected in this procedure, though.

The metric studied was: The 4^{th} highest MDA8 of O_3.

2.2 Statistical study based on the US-EPA methodology for observed O_3 (2017)

This method was fundamentally different from the approaches outlined above with respect to both the input data and the model concept. The main aim of this study was to apply a statistical method named GAM (Generalized Additive Model) that has been documented and applied on a routine basis by the US-EPA. The GAM is essentially a multiple regression method where one dependent variable (in this case the daily ozone levels) is estimated as a function of several input explanatory variables (in this case meteorology and time). The GAM differs from a standard multiple regression in the way that the relationship between each input variable and the dependent variable is a smooth function and not a constant as in multiple regression as outlined in detailed in Ch. 5 below.

The GAM was applied to rural Airbase and EMEP ozone data (May-August) for two periods separately: 1990-2000 and 2000-2010. Gridded daily meteorological data from EuroDelta Trends were used as input explanatory variables to the statistical method. The performance of the method was evaluated by looking at daily values of MDA8 O_3 station by station.

In the original US-EPA study (Camalier et al., 2007) their method was validated against daily maximum O_3 concentrations based on stations merged into 39 urban agglomerates for which they found that the predictive power as measured by R^2 ranged from 0.56 to 0.80. In our approach, we used no merging of stations and we evaluated the predictions against the daily MDA8 levels for individual rural and suburban sites. It was found that the R^2 (coefficient of determination) statistic for the individual European sites in general was lower than found for the urban agglomerates in the US. The best agreement between the daily observed ozone levels and the values predicted by the GAM was found in Central Europe with R^2 of 0.65-0.70 which seems very satisfactory. On the other hand, significantly poorer agreement was found in Spain, southern France and various areas in the east, particularly Baltic and Slovakian site, with R^2 values of 0.3-0.5. The RMSE (root mean square error) showed particularly high values in Northern Italy, Portugal and the few sites in the southeast, indicating a strong bias in the predicted MDA8 levels in these areas.

2.3 Similar method as for O_3 applied to NO_2 and PM (2018)

In this study, we applied the statistical model developed for O_3 in 2017 to NO_2 and PM_{10} for the periods 1990-2000 and 2000-2010. The $PM_{2.5}$-data were too limited for the GAM to be used for that pollutant. As for ozone, the gridded meteorological data from EuroDelta Trends were used as input. Various adaptions were needed when switching the focus from O_3 to NO_2 and PM_{10}. Firstly, the daily mean values were used as dependent statistic instead of the max daily 8h average used for O_3. Secondly, also data from urban sites were included.

Best performance for NO_2 was found in the areas typically associated with the main European emission area, i.e. Belgium, the Netherlands, North-Western Germany and the United Kingdom. Significantly poorer performance was found for e.g. Austria and areas in southern Europe. Marked differences were also found for PM_{10} although there was a less clear spatial pattern. The amount of PM_{10} data were, however, much less than for NO_2 for this time period.

3 Findings – strengths, weaknesses, uncertainties, feasibilities, requirements

3.1 Conclusions from the studies based on the EuroDelta Trends model exercise in 2016

As documented by Colette et al. (2016), all methods based on the EuroDelta Trends model exercise provided similar results with emission changes being the main driver for the downward trends in peak O_3 levels whereas the meteorological variability also lead to reduced ozone but at a smaller magnitude.

3.1.1 Attribution procedure

Findings: Emission changes are the most important driver of trends in ozone summer peaks. However, in almost all region and time periods, meteorological variability contributed to decreased ozone peaks whereas boundary conditions were less important. For annual mean PM_{10}, emissions were the clearly most important driver in both periods, whereas the impact of meteorology was much smaller. Meteorology lead to slightly reduced PM_{10} levels in the 1990s except in the Mediterranean, and a very slight increase all over in the 2000s.

Requirements: Require gridded model results for at least two meteorological years with different boundary conditions and different emissions

Limitations: Strictly model based. Observational data are not used

Recommendations: The method relies on a very specific set of model scenarios for the entire time period as it was defined within the EuroDelta Trends exercise. The model data (as well as the input meteorological data) from the EuroDelta exercise ends in 2010. One strength of the approach lies in the multi-model assessment, which is not scheduled for an update in the short term, although CAMS is considering performing such a task in a future phase. Furthermore, it is not really possible to evaluate the "performance" of this method compared to the statistical observational methods applied later. Making this method an operational service is in principle possible but will require a substantial modelling effort.

3.1.2 Ozone model climatology

Findings: Emissions are clearly most important, but the results indicate that for all regions except Scandinavia the meteorological variability during the 1990-2000 and 2000-2010 periods lead to an even stronger reduction in ozone (measured by the 4th highest MDA8) than the emission reduction alone.

Requirements: Require gridded model results for the entire period (each year) with three emission scenarios – fixed emissions for the start year, fixed emissions for the end year and true emissions varying from year to year.

Limitations: Strictly model based. Observational data are not used.

Recommendations: Much of the same conclusions as for the previous chapter (Ch. 3.1.1). This method is very similar to the EuroDelta Trends set-up, except the multi-model aspect. An updated EMEP/MSC-W Trend simulation was performed late 2019 and will be investigated by ETC/ATNI in 2020.

3.2 Conclusions from a statistical GAM model based on observational ozone data (2017)

Findings: When the predicted daily O_3 levels were compared to the measured O_3 data, a good to very good agreement was found in central Europe and in Germany in particular. In the Nordic countries, the seasonal cycle and the timing of ozone episodes were well reproduced, but the high ozone peaks were substantially underestimated. In southern Europe poorer agreement with observed values were found.

Requirements: The GAM method relies on the availability of a suite of meteorological data provided on a daily or finer temporal resolution for each station, either from a gridded data set or from local measurements. The necessary parameters typically include temperature, humidity, wind speed, and the height of the mixed layer.

Limitations: A main limitation of this method as for all locally based statistical models is the assumption that the ozone concentration at a station could be explained by the local meteorological conditions at the site. This is normally not a valid assumption since the ozone level depends on the history of the air mass over a period of several days including physio-chemical processes such as surface deposition, photochemical formation, vertical mixing etc. Another limitation is that the GAM model is designed for predicting mean quantities. In the 2017 study, the method was used for predicting the seasonal mean of the MDA8 during May-August. The results from the study indicated that the high peaks in ozone were generally underestimated and thus the GAM method should not be used for predicting exceedances of limit values or other extreme values. The fact that the performance of the method is less suited for peak ozone levels and that certain regions of Europe are not so well modelled is a concern.

Recommendations: When already established and adapted to the data set, the GAM is a straight-forward and efficient tool which could be used on a routine basis if the required input meteorological data are available. It is important to be aware of the strengths and limitations of the method though. The method could be applied to predict daily ozone levels and the influence of meteorological variability for parts of Europe, mainly central Europe, to a fairly accurate degree. On the other hand, it will probably not be possible to use this method for the prediction of exceedances of extreme values.

3.3 Conclusions from a GAM model based on observational NO_2 and PM data (2018)

Findings: The performance of the GAM model showed similar behaviour for NO_2 as for O_3 with highest scores near the emission source areas and poorer scores elsewhere. For PM fewer clear patterns were found, but this could reflect more sparse station network in the period 1990-2010 and that the applied season (4-months winter period) was too limited.

Requirements: To apply the GAM method for NO_2 and PM requires similar input data as for O_3. For PM, representation of natural sources of dust would have been an advantage, but is difficult to think of in practice without involving a CTM.

Limitations: As for ozone, the assumption that the concentration at a station could be explained by the local meteorological conditions at the site is in principle not a valid assumption although the method is seen to work well for certain regions. On the other hand, one could argue that the assumption of local relationships would be more valid for a primary pollutant as NO_2 than a secondary pollutant as O_3, at least in the emission areas. Another limitation is, as for O_3, that the GAM is not really suited to predict exceedances of extreme values since the method tends to underestimate the highest levels.

Recommendations: The recommendations for applying the GAM for NO_2 and PM is similar as for O_3; it is a straight-forward method which could be run on a routine basis (annually) without strong demands on computer resources etc. Access to input meteorological data is required though.

4 GAM model applied to O_3, NO_2 and PM for the period 2000-2017

4.1 Method description

In this chapter, we show results with the GAM method using data extracted from ECMWF for the period 2000-2017 together with data from the consolidated monitoring data from EBAS and Airbase/e-reporting for O_3, NO_2, PM_{10} and $PM_{2.5}$.

The methodology is mainly based on the same approach that was developed in the previous projects on trends for these pollutants as described in Solberg et al. (2018a, 2018b), and the reader is referred to those reports for more details.

The basic method is a so-called GAM, i.e. generalized additive model, which is a statistical regression model linking expected values μ_i of the given response variable Y_i to a number of explanatory variables x_{ij} through the following set of relations:

$$g(\mu_i) = \beta_0 + \sum_{j=1}^{p} \beta_j (x_{ij}); \quad \mu_i = E(Y_i) \tag{1}$$

where β_0 is a constant (the intercept), and $\beta_j(\cdot)$, for $j = 1,...,p$, represent smooth functions of the covariates x_{ij}, where p is the number of such covariates. The regression model (1) is introduced for each compound (O_3, NO_2, PM_{10} and $PM_{2.5}$) and for each monitoring site separately.

The response variable Y_i in (1) represents measured daily mean value of concentration at day number i while x_{ij} represents explanatory variables at the same site and day, where the latter consist of meteorological variables (temperature, wind speed etc.) and time variables (day of week, day of season etc.), for $i = 1,...,n$, where n is the number of data, i.e. days, in the period 2000-2017.

In Eq. (1) $g(\cdot)$ is a link function linking the statistical expected value of the response variable Y_i, i.e. μ_i, to the explanatory variables x_{ij}. In the GAM, the response variable Y_i is also assumed to have a certain probability distribution, with mean μ_i and variance V_i , known as the response distribution. The GAM is an extension of a multiple linear regression (MLR) where each β_j is a smooth function of x_{ij} and not a constant to be multiplied with x_{ij} as in an MLR, and where the mean value μ_i is more generally related to the covariates through a given link function $g(\mu_i)$.

Like in the previous trend study (Solberg et al., 2018b) we apply a unit link function $g(\mu) = \mu$ and a Gaussian distribution as a response distribution for O_3, and a log link function $g(\mu) = \log \mu$ and a Gamma distribution as a response distribution for NO_2, PM_{10}, and $PM_{2.5}$. The reason for this choice is that O_3 has a relatively small range of concentration variations where the variance of Y_i, i.e. V_i, does not change very much with the mean μ_i and thus the response distribution is well represented with a symmetric distribution such as the Gaussian. On the other hand, NO_2, PM_{10}, and $PM_{2.5}$ have a much larger range of concentration variations of several orders of magnitude, and for these compounds, the variance of Y_i, i.e. V_i, is more proportional to μ_i^2 . Thus, for these species it is better, and also common practice in GAM modelling, to choose a distribution that is skewed to the right such as a Gamma distribution, as a response distribution for Y_i.

The other choice of settings and parameters in the present study (Table 1) is closely related to the settings and parameters in the previous trend study (Solberg et al., 2018b) with some exceptions:

1. Absolute humidity instead of relative humidity was introduced as a specific covariate for O_3 since it is found that the O_3 levels were better explained by this variable.
2. Global radiation was found to be superfluous based on a concurvity analysis of all covariates since it was found to be well represented non-linearly by the other meteorological covariates. This covariate has thus been removed from the GAM model.

For the other explanatory variables, the same set as in the previous study (Solberg et al., 2018b) was used, i.e. daily mean values of temperature, wind speed and direction, planetary boundary layer height, relative humidity (for NO_2, PM_{10} and $PM_{2.5}$), day of week, day of season and time in fraction of years, the latter of which represents the trend term in the GAM model. New in the current study, is that this trend term is represented as a smooth function of time rather than as a straight line as was the case in the previous studies (Solberg et al., 2018a, 2018b). The main reason for this choice is that the period 2000-2017 is a relatively long period and thus it is less relevant to represent the whole trend during this long period as a simple straight line. However, in order not to introduce too much unwanted variability due to noise in the residuals into the trend term, we chose to model this with a smaller number of degrees of freedom (4) for this term than the default which is 10. The same smaller number of degrees of freedom (4) is also applied for the same reason to the other two time-covariates, i.e. weekday and day of season. However, for all meteorological covariates, we still apply the default of 10 degrees of freedom for the smooth functions.

Table 1: List of explanatory variables used in the GAM (Eq. (1)) for O_3, NO_2, PM_{10} and $PM_{2.5}$ in this study. The short names refer to the legends used in the map plots shown below.

	Associated explanatory variable	**Short name in the plot legends**	**Not used by**
x_1	Daily temperature at 18 UT	temp	
x_2	Daily mean 10 m wind speed	ws	
x_3	Daily mean 10 m wind direction	wd	
x_4	Daily mean PBL height	pblh	
x_5	Daily relative humidity at 18 UT[1)]	rh	O_3
x_5	Daily absolute humidity at 18 UT	h2o	NO_2, PM_{10}, $PM_{2.5}$
x_6	Daily total precipitation	prec	O_3, NO_2
x_7	Weekday number	dayofweek	
x_8	Day number in season	dayofseason	
x_9	Continuous time in fraction of years (0.0 = 1 Jan at start of period). This is the trend term.	years	

1) Relative humidity was not given in the ECMWF data but was calculated based on the absolute humidity, temperature and pressure (Vaisala, 2013).

Through the GAM optimisation, we calculate the smooth β functions and their significance levels for each station/period as well as various measures of the GAM model evaluation performance such as RMSE, R^2, etc. For this latter part, we use the `openair` library (Carslaw and Ropkins, 2012; Carslaw, 2015).

Examples of the response curves for O_3 and NO_2 are given in Figure 2 and Figure 3, respectively. These response curves show the partial dependency of O_3 and NO_2 with respect to each input parameter when all the other parameters were kept at their mean values. Figure 2 indicates a marked positive relationship between temperature and ozone as expected as well as a negative relationship between O_3 and wind speed and absolute humidity. The trend term shows reduced concentrations with time for this site except for the very last part of the period. For NO_2 (Figure 3) the results for this site show a marked negative relationship with temperature, relative humidity, wind speed and PBL height, as is to be expected. A clear downward meteorologically adjusted trend is also seen for this site.

Figure 2: Example of response curves for one station, AT0002 (Illmitz), for O_3 based on daily data from the summer half year (April-September) during 2000-2017. The daily data are based on the MDA8 (max daily 8h running mean concentration). Each panel shows the response function between the input parameter and ozone when all the other input parameters are kept at their mean level. The short names of the input parameters are explained in Table 1 above. Illmitz is a well-established background site in central Europe, having a very long monitoring history, thus a god candidate to be used as an example site.

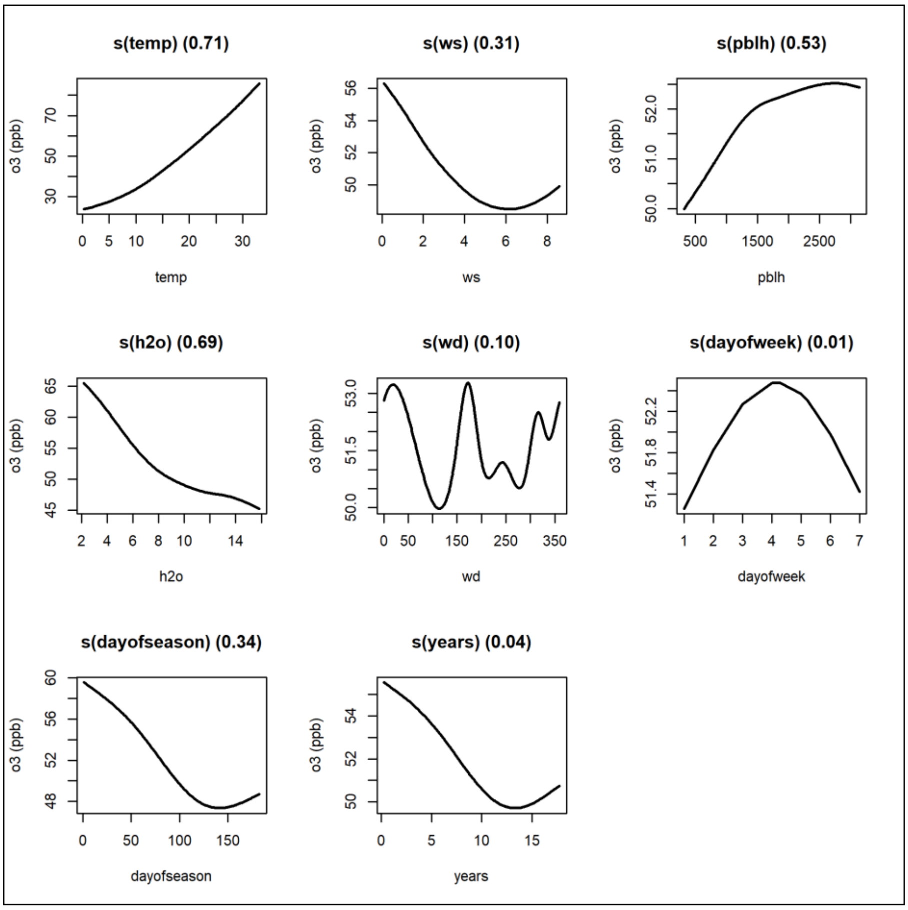
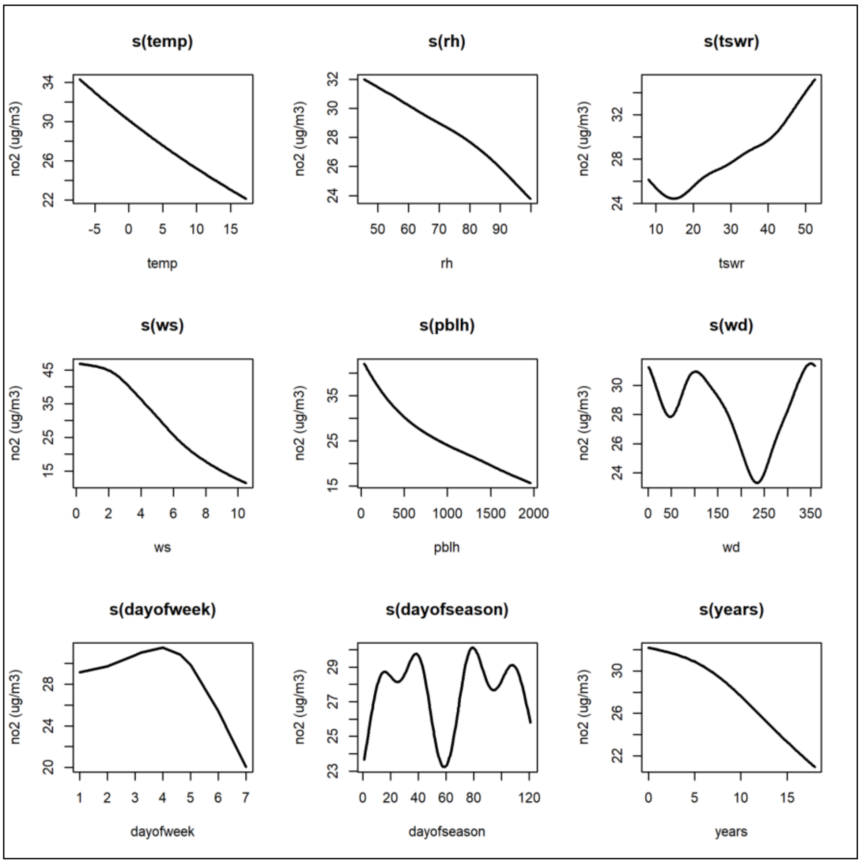


Figure 3: Same type of plot as in Figure 2. Example of response curves for BETBE011 for NO_2 for the winter period, defined as November-February. This is an example of a site with a marked reduction in NO_2 and also a good agreement between modelled and observed levels.



The GAM model was applied to O_3 only for the summer period (1 April - 30 September) and to NO_2 only for the winter period (1 November - 28 February). For PM_{10} and $PM_{2.5}$ the model was applied for both the summer and winter periods.

The reason for looking at separate season was based on the criteria of homogeneity. The use of a GAM relies on an overall assumption of homogeneity, i.e. that the links between the explanatory variables and the output parameter are homogeneous within the time period considered. For ozone, it is well known that that the links to various meteorological parameters are reversed going from summer to winter. In summer, anticyclonic conditions associated with weak winds and a shallow PBL typically lead to enhanced ozone levels, whereas such conditions lead to reduced ozone in winter due to titration with NO_2. Similar shifts from summer to winter are seen for NO_2 and PM, and thus applying the GAM model to individual seasons make sense and will produce more robust results.

As in the previous studies, the GAM performance is evaluated each year by excluding the data for that year, i.e. the “target year” and fitting the model to the rest of the data (the left-out years). This means that e.g. when using the GAM model to predict the daily concentration levels in 2013, we skip the measurement data for 2013, but use the data for all the other years in the 2000-2017 period to calculate the smooth response curves. Then, these response curves are combined with the meteorological data for 2013 to predict the daily concentration levels in 2013. Thus, the model values are true predictions and not in any way influenced by the observed concentration values in the target year.

A more detailed overview of how the GAM model is solved numerically is found in Appendix A.

4.2 Input data

The meteorological data used in this report were retrieved from the European Centre of Medium-Range Weather-Forecasts (ECMWF). The Era-Interim (Berrisford et al., 2011) reanalysis dataset was used with a 6-hourly analysis (00:00, 06:00, 12:00, 18:00 UTC), retrieved on a 0.3° x 0.3° resolution. For precipitation and PBL height daily data were extracted (daily total precipitation and daily mean boundary layer height). These data were consistent through the whole period 2000-2017, i.e. there were no shifts in methodology or resolution which is crucial for the statistical modelling that assumes of systematic relationships between input meteorological data and output pollutant levels.

The measurement data were extracted from the air quality monitoring databases hosted by the European Environment Agency (EEA) and by EMEP. Up to 2012, the datasets from EEA were gathered in the AIRBASE database, for which we used the v8 release. After 2013, the EEA database moved to the Air Quality e-reporting system. A technical difficulty lied in matching these two databases because many stations changed names and codes over time. Instead of station names, the matching is performed using the Sampling Point information, which is the most reliable information about the consistency of a given record.

The data from EEA were combined with measurement data from EBAS hosted by EMEP. Whereas the EEA data contain stations at rural, suburban and urban locations, the EBAS data are from rural stations only. The merging of the EEA data with the EMEP data was not a trivial task and due to limitations on time and resources we ended up with combining EEA and EMEP data only for ozone whereas we used only EEA data for NO_2 and PM. As mentioned above, the merging of the AIRBASE database with the e-reporting system was a considerable task on its own. A further combination of these data with data from EBAS turned out to be more demanding than anticipated due to several factors:

- Different station names and station codes used by EMEP and EEA
- Differences in the geographical coordinates for the same stations
- Differences in the measurement data themselves such as
    - 1 or 2 h shifts in data due to uncertainties linked to UTC or local time
    - Different procedures for flagging data
    - Periods of invalidated data in one database while being valid in the other
    - Errors due to wrong scaling etc in one database and not the other

This reflects the problems when the same measurement data are reported to several separate databases. Due to differences in quality assurance (QA) procedures, file conventions, metadata requirements etc it is almost impossible to avoid differing data. Merging data from two or more databases containing partially the same stations, could thus become time consuming and challenging to carry out.

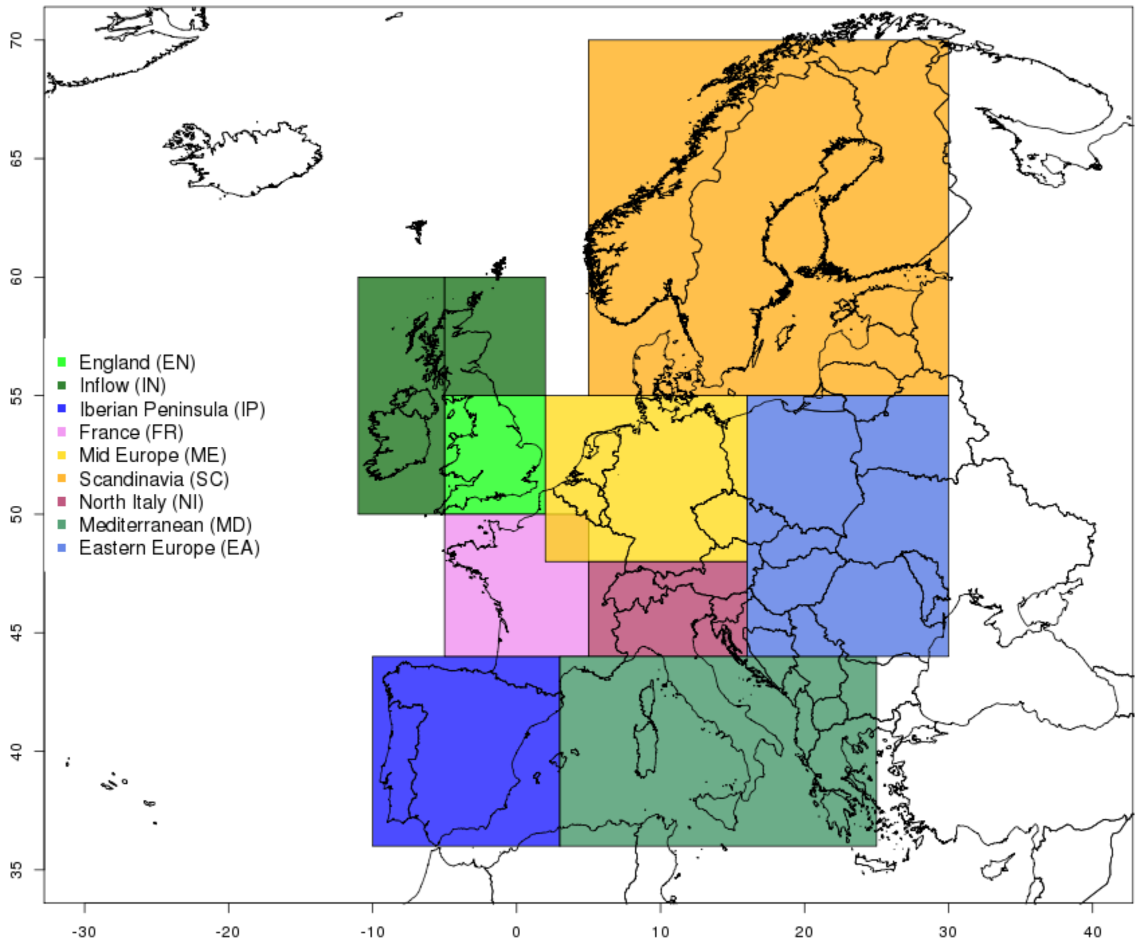
We chose to merge data for O_3 since the EEA data includes significantly more rural sites than EBAS and since urban and suburban sites where not included in the statistical modelling for O_3. Ideally, for the future, the EEA and EMEP data should be combined also for NO_2 and PM.

All the statistical modelling was based on daily values, and for O_3 we used the daily MDA8 values whereas for NO_2 and PM we used the daily means. For NO_2 and PM data there is an additional complication that these measurements could be given as either daily or hourly data and furthermore that the monitoring could shift from one to the other during the time period and even that in some years the same station could have parallel measurements with both hourly and daily data. The merging of the daily and hourly data NO_2 and PM from the EEA databases turned out to be non-trivial and at the end we had to drop this merging and therefore used only the hourly based measurements which was aggregated into daily mean values. For the future it is strongly recommended that a list of preferred hourly and daily data sets to be used are provided in beforehand.

4.3 Screening of stations

In this and subsequent chapters parts of the results were aggregated into the nine geographical regions shown in Figure 4. Figure 4 is copied from the report by Colette et al. (2016) and shows the regions used in the EuroDelta Trends exercise. This definition of regions represented a slight adjustment to the regional climate zones as defined in the original PRUDENCE project (Christensen and Christensen, 2007). We adjusted these regions a bit further to include stations further east and south, and thus the Eastern Europe (EA) region was extended further east and the Mediterranean (MD) region further east and south compared to the original definitions.

Figure 4: The adjusted Prudence regions, copied from Colette et al. (2016) used in the EuroDelta exercise. In the present work, the EA region was extended to the east and the MD region was extended to the east and south.



The GAM method described above was applied for each monitoring station individually. Based on the results of the GAM and inspection of the time series of the observed monitoring data, it was clear that a certain fraction of the time series included various kinds of flaws. As discussed in more detail below, it turned out that the GAM could be used to identify time series with dubious data. Ideally, all input data extracted from the databases should be cleaned for errors, but it became obvious that this was not the case. This is outlined in more detail in Ch. 5.7 below.

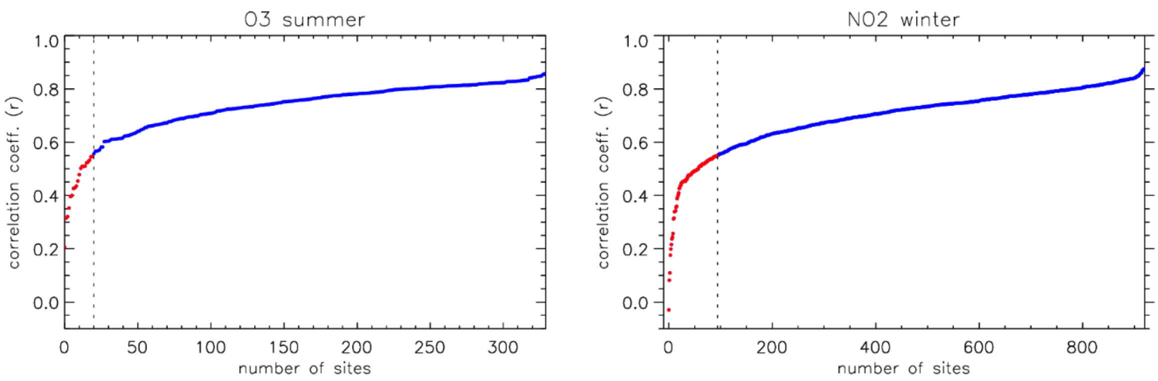
In the further use and evaluation of the observational data for O_3 and NO_2 we included only the time series fulfilling two criteria:

- The height of the monitoring site should be below 1000 m altitude
- The linear correlation coefficient, r, should be 0.55 or higher

These selection criteria were based on a close inspection of the time series. It turned out that the GAM in general did not perform well for high mountain sites which presumably reflects that the pollutant levels are not linked to the local meteorological conditions at these sites and furthermore, that the rather coarse gridded meteorological data from ECWMF are not representative for the mountain sites.

The distribution of the linear correlation coefficients for the individual stations are shown in Figure 5 for O_3 and NO_2. This indicates that a certain fraction of the sites showed particularly low correlation between the GAM model and the measured values and the cut-off value at 0.55 is marked in the plots.

Figure 5: The distribution of linear correlation coefficients, r, for the GAM model vs the observations for O_3 during the summer half year (Apr-Sept) and NO_2 during winter (Nov-Feb). The cut-off at r = 0.55 for the filtering of the stations is marked.



The criteria for the linear correlation coefficient (r > 0.55) are a subjective choice based on inspection of individual time series. Ideally, the observational data should be "cleaned" prior to our use by methods independent on the GAM, but the time did not allow us to do this and besides, this was not part of the task of the present project. In the following, we show aggregated results from the GAM based on the criteria given above.

The inspection of the time series revealed that the issue of time series with possible flaws was most pronounced for NO_2 and to a somewhat less extent for O_3. For the PM data, we did not see a similar link between the linear correlation coefficient (r) and dubious time series as for NO_2 and thus the PM data was only screened for the station altitude. Table 2 gives the number of sites within each of the nine adjusted Prudence regions.

Table 2: The number of monitoring sites within each of the adjusted Prudence regions (see Figure 4) for the different pollutants and seasons. The number in brackets refer to the number of so-called filtered sites (r > 0.55 and altitude < 1000 m asl). The filtering was only applied to O_3 and NO_2.

	O_3 (Apr-Sep)	NO_2 (Nov-Feb)	PM_{10} (Nov-Feb)	PM_{10} (Apr-Sep)	$PM_{2.5}$ (Nov-Feb)	$PM_{2.5}$ (Apr-Sep)
EN	14 (14)	27 (27)	-	-	16	18
IN	8 (7)	5 (5)	-	-	-	-
IP	42 (26)	140 (107)	31	41	25	26
FR	11 (11)	66 (65)	-	-	-	6
ME	138 (133)	390 (384)	153	201	69	64
SC	28 (28)	32 (31)	9	9	12	11
NI	48 (29)	148 (95)	-	5	-	-
MD	14 (10)	48 (43)	-	-	-	-
EA	26 (25)	64 (63)	39	48	12	20

In Figure 6 and Figure 7 the spread in correlation coefficients for the unfiltered and filtered NO_2 and O_3 data are shown for the adjusted PRUDENCE regions defined above. The results (median, 25-percentile and 75-percentile) are also given in Table 3 and Table 4. This shows that the filtering of sites is mainly applied to sites in the North Italy (NI) and Iberian Peninsula (IP) regions. For NO_2, the best GAM model performance as measured by the linear correlation coefficient, r, is seen for the Mid-Europe, France, England and Inflow regions (Figure 6). For O_3, the best performance is seen for the Mid-Europe and East Europe regions.

Figure 6: The spread in linear correlation coefficients, r, for the GAM model prediction vs the observed daily mean values of NO_2 (Nov-Feb) for the adjusted PRUDENCE regions. The number of sites is given on top. Blue refers to the unfiltered data set and red to the filtered data set.

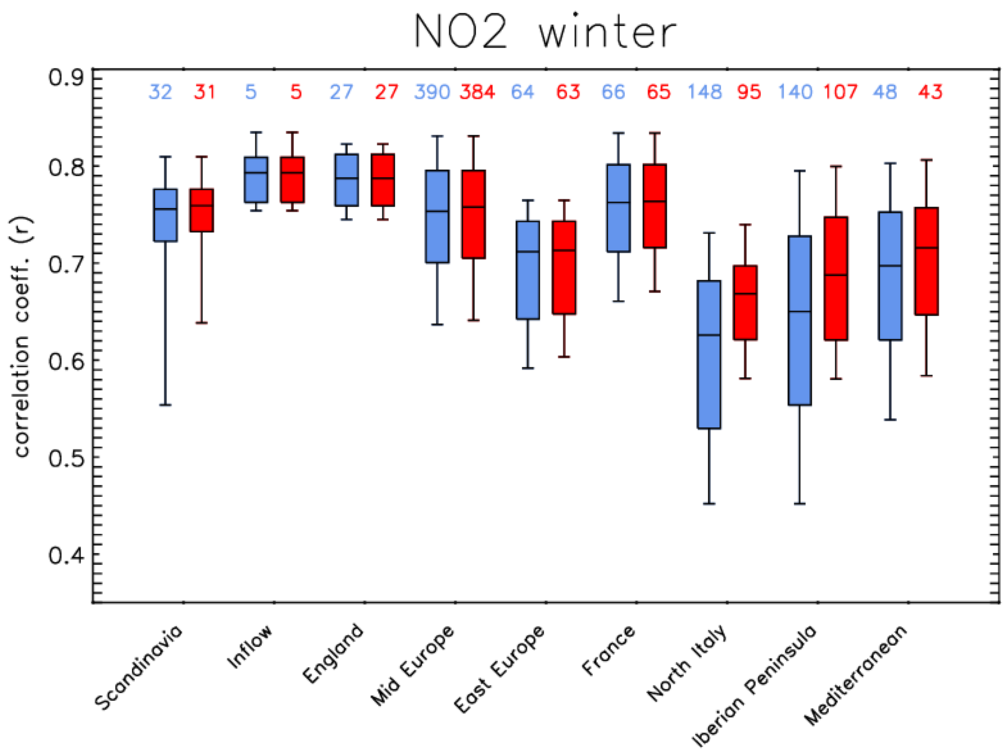


Figure 7: Same as Figure 6 for O_3 during summer (April-Sept.).

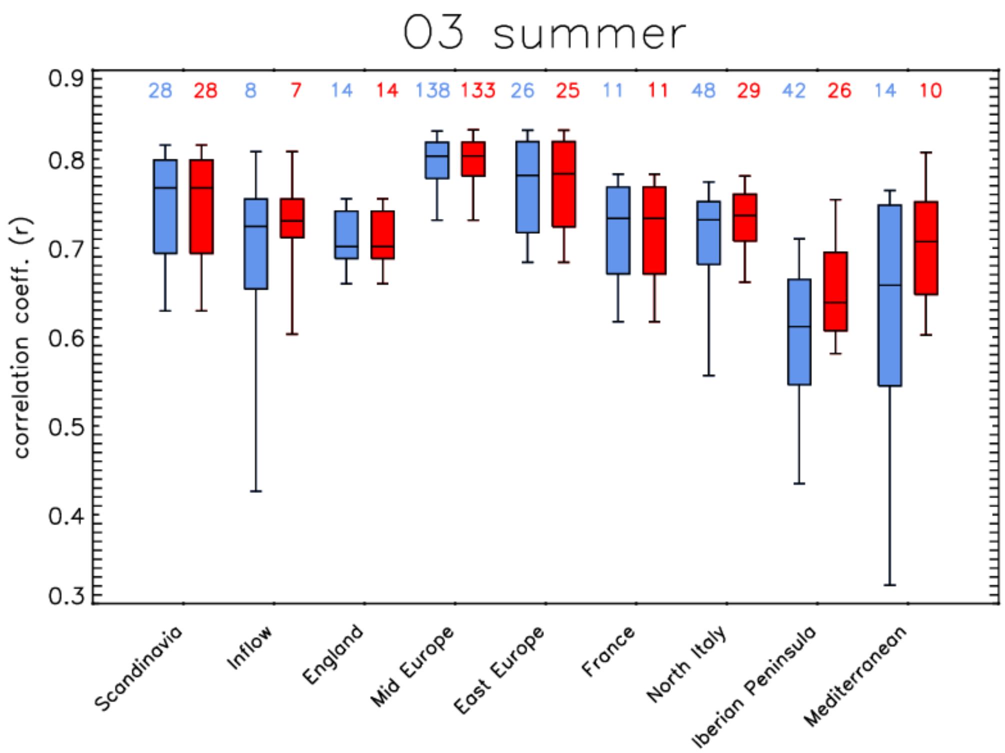


Table 3. The data from Figure 6 in tabulated form showing the median (p50), 25-percentile (p25) and 75-percentile (p75) of the linear correlation coefficients, r, for the GAM model prediction vs the observed daily mean values of NO_2 (Nov-Feb) for the adjusted PRUDENCE regions for the filtered and unfiltered data. The number of sites is given by 'n'.

Region	**Filtered data**				**Unfiltered data**			
	p50	**p25**	**p75**	**n**	**p50**	**p25**	**p75**	**n**
Scandinavia	0.76	0.73	0.78	31	0.76	0.72	0.78	32
Inflow	0.79	0.76	0.81	5	0.79	0.76	0.81	5
England	0.79	0.76	0.81	27	0.79	0.76	0.81	27
Mid Europe	0.76	0.71	0.80	384	0.75	0.70	0.80	390
East Europe	0.71	0.65	0.74	63	0.71	0.64	0.74	64
France	0.76	0.72	0.80	65	0.76	0.71	0.80	66
North Italy	0.67	0.62	0.70	95	0.63	0.53	0.68	148
Iberian Peninsula	0.69	0.62	0.75	107	0.65	0.55	0.73	140
Mediterranean	0.72	0.65	0.76	43	0.70	0.62	0.75	48

Table 4. The data from Figure 7 in tabulated form showing the median (p50), 25-percentile (p25) and 75-percentile (p75) of the linear correlation coefficients, r, for the GAM model prediction vs the observed MDA8 values of O_3 (Apr-Sep) for the adjusted PRUDENCE regions for the filtered and unfiltered data. The number of sites is given by 'n'.

Region	Filtered data				Unfiltered data			
	p50	**p25**	**p75**	**n**	**p50**	**p25**	**p75**	**n**
Scandinavia	0.77	0.69	0.80	28	0.77	0.69	0.80	28
Inflow	0.73	0.71	0.75	7	0.72	0.65	0.75	8
England	0.70	0.69	0.74	14	0.70	0.69	0.74	14
Mid Europe	0.80	0.78	0.82	133	0.80	0.78	0.82	138
East Europe	0.78	0.72	0.82	25	0.78	0.72	0.82	26
France	0.73	0.67	0.77	11	0.73	0.67	0.77	11
North Italy	0.74	0.71	0.76	29	0.73	0.68	0.75	48
Iberian Peninsula	0.64	0.61	0.69	26	0.61	0.55	0.66	42
Mediterranean	0.71	0.65	0.75	10	0.66	0.55	0.75	14

4.4 GAM performance

The maps below show the R^2 and the normalised mean gross error, NMGE, aggregated into grid squares of 1° x 1°. For grid squares with more than one monitoring station, the aggregated data shows the mean of the statistics in the case of two stations and the median in the case of more than two sites.

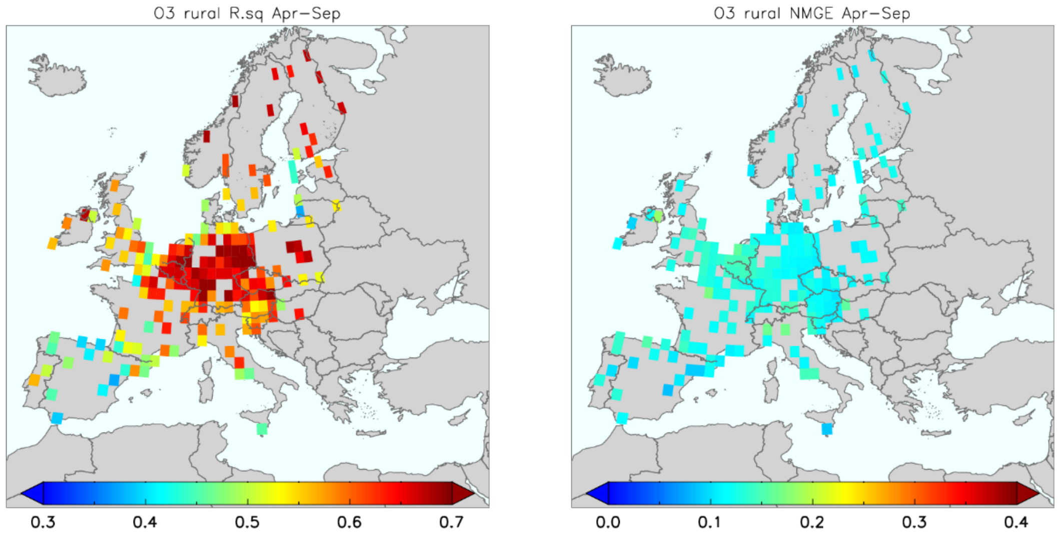
R^2, also called R squared, gives the part of the variance explained by the GAM model while NMGE is a measure of the bias, or mean absolute error, relative to the mean value.

4.4.1 O_3 April - September

Figure 8 shows the mapped values of R^2 and NMGE for rural background ozone sites below 1000 m fulfilling the minimum criteria of $r \geq 0.55$. This shows that the GAM performance varies considerably over the continent with best scores for northern and central Europe (Germany, Belgium, Netherlands, Poland Czech Republic, Austria and to a less extent Scandinavia). For southern Europe (Spain, southern France and southern Italy) the performance as measured by R^2 is clearly poorer. The area in central Europe with best performance matches the area of main emissions of ozone precursors and reflects the fact that the GAM model is based on local relationships (a link between observed O_3 and local meteorological conditions). For stations further away from the main precursor source regions, and thereby more dependent on long-range transport it is to be expected that the performance of a statistical method as the GAM that relies on local relationships will perform poorer. A typical feature at mid- and northern Scandinavian sites is that most of the ozone variability is due to the marked seasonal cycle which is reproduced well by the GAM model, while the peaks in ozone is substantially underestimated by the GAM. This may lead to a fairly high correlation between the GAM model and the observations while the time series of daily data reveal that this is mainly due to the average seasonal cycle and not the episodes.

The values of NMGE are mostly smaller than 0.15, implying that the mean relative error is less than 15% which reflects that the general range of O_3 levels are small (and much smaller than the range of NO_2 and PM levels).

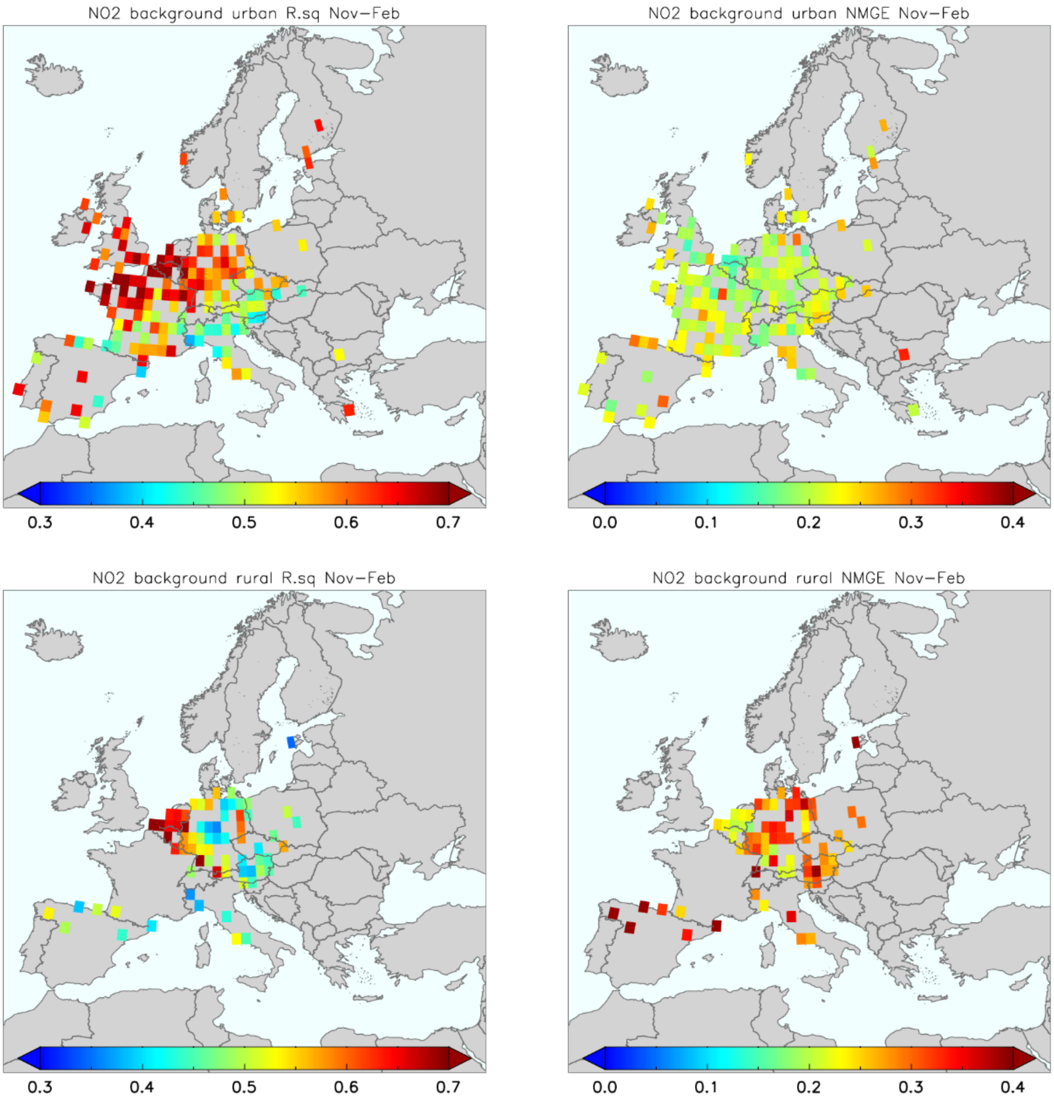
Figure 8: R^2 (left) and NMGE, the normalised mean gross error (right), aggregated into grid squares of 1 °x 1 ° for the performance of the GAM model applied to surface ozone data from the summer half year for rural background stations.



4.4.2 NO_2 November - February

Figure 9 shows the mapped values of R^2 and NMGE for NO_2 for four winter months (November - February) for background urban and background rural stations, respectively. The number of rural sites is however, much lower than the number of urban sites. The R^2 field for the urban NO_2 sites shows a similar pattern as for O_3 with highest score in central and northern Europe, but clearly higher NO_2 scores are seen for the urban Spanish NO_2 sites compared to the rural O_3 sites. The relative errors as given by NMGE are, however, much higher for NO_2 than for O_3. This is to be expected due to the much larger span in concentration levels for urban NO_2 than rural O_3.

Figure 9: Same as Figure 8 for NO_2 at urban background sites (top) and rural background sites (bottom) during four winter months (November-February).



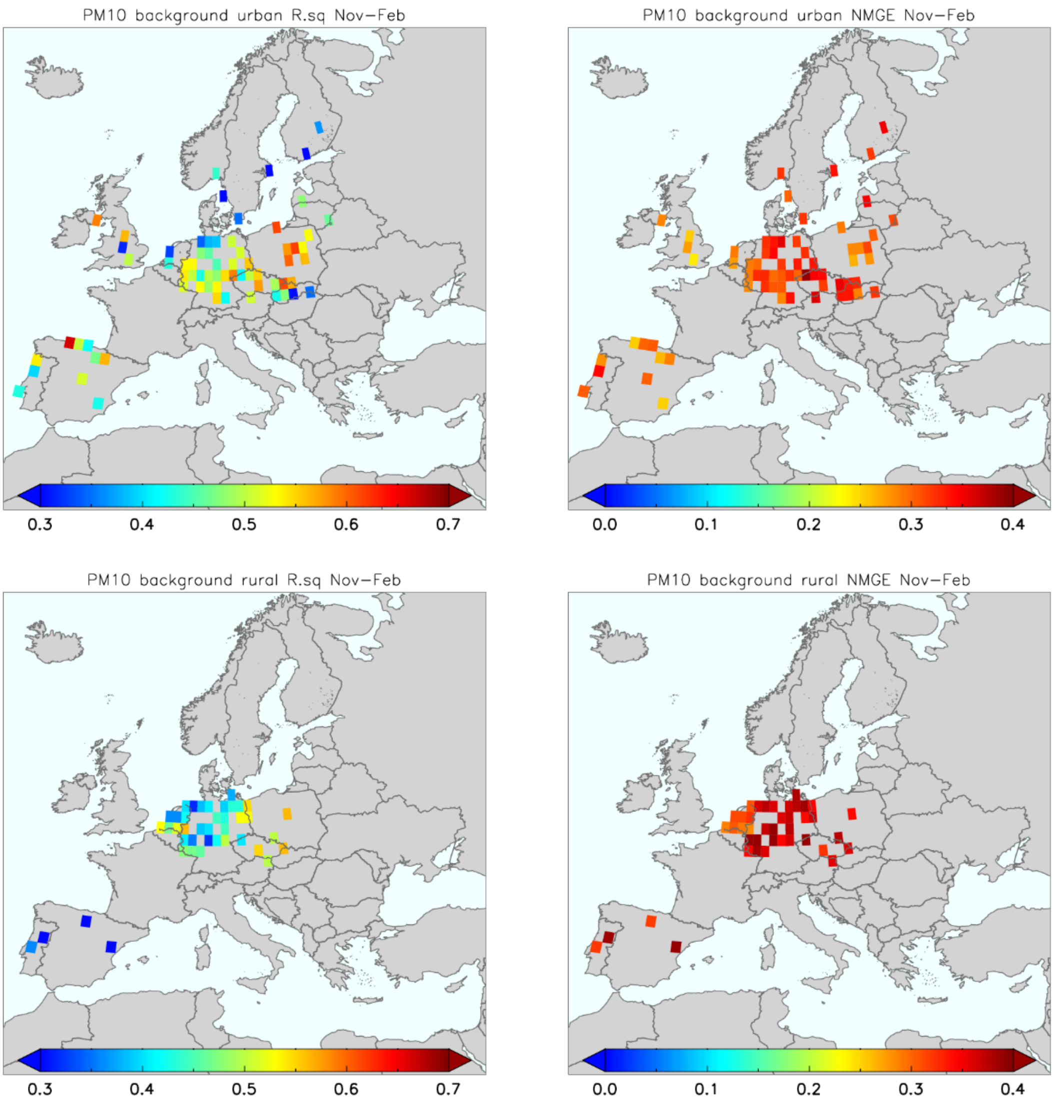
4.4.3 PM_{10} November - February

Figure 10 shows the mapped values of R^2 and NMGE for PM_{10} for four winter months (November - February) for background urban and background rural stations, respectively. The number of stations with long-term PM_{10} measurements is substantially lower than for NO_2 and O_3 as seen in Figure 10, and there are no French sites and very few in Scandinavia. Most of the stations are in central Europe (Germany, Poland, Czech Republic and Austria) and a few on the Iberian Peninsula. The general performance of the GAM model is poorer for PM_{10} compared to NO_2 and O_3. One should keep in mind, though, that these data were not screened for the stations with the lowest r-values (as was NO_2) and thereby more sites with R^2 values in the range of 0.3-04 are seen. Few PM_{10} sites show high R2 scores (> 0.6). Furthermore, these results indicate that the GAM model performs poorer for rural sites compared to urban sites in winter although the number of rural sites is too low to really judge this.

The sites with the highest R^2 values in winter are in Poland, the Czech Republic and one on the north west coast of Spain. More or less all NMGE values are higher than 0.3 implying that the general bias of the GAM model is as 30 % or more relative to the mean.

The fact that the GAM performance for PM_{10} for the winter months are poorer than for NO_2 and O_3 reflects that further physical processes than those included in our GAM model are important for the PM_{10} levels.

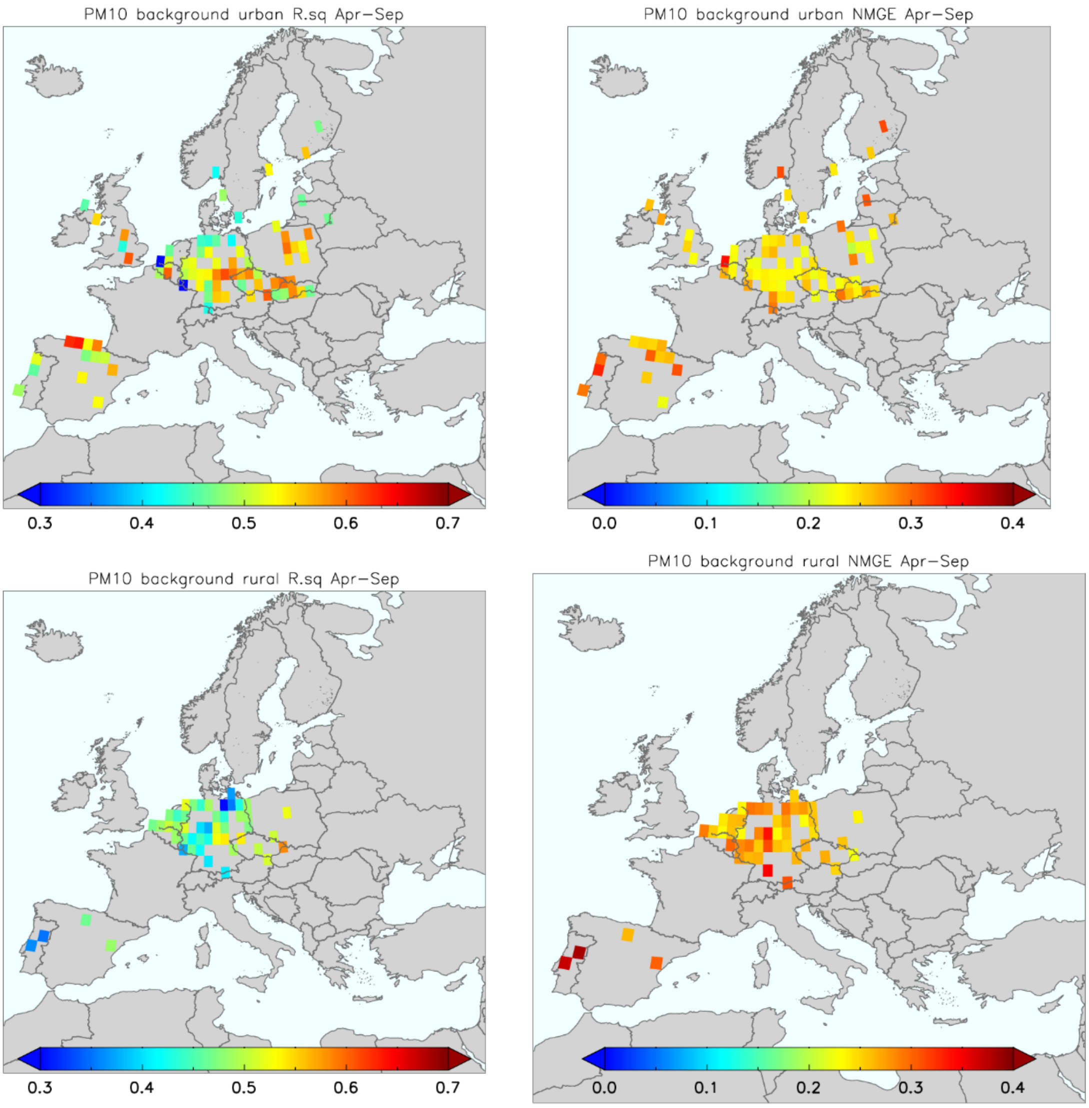
Figure 10: Same as Figure 8 for PM_{10} at urban background sites (top) and rural background sites (bottom) during four winter months (November-February).



4.4.4 PM_{10} April - September

Figure 11 shows the mapped values of R^2 and NMGE for PM_{10} for the summer half year (April - September) for background urban and background rural stations, respectively. These results indicate that the GAM model in general performs somewhat better in summer than in winter for PM_{10} at least for the background urban sites. This probably reflects that very local meteorological conditions such as temperature inversions are much less frequent in summer than in winter and thus that the ECMWF data used as input to the GAM is better suited in summer than in winter. The NMGE, expressing the mean relative error is considerably lower in summer than in winter which is a further sign of a better model performance in summer.

Figure 11: Same as Figure 8 for PM_{10} at urban background sites (top) and rural background sites (bottom) during the summer half year (April-September).



4.4.5 Examples of time series

The aggregated maps in the previous chapters give an indication of the overall performance of the GAM model. By experience, we know however, that simple time series with observed and predicted levels are much more suited to show what these statistical metrics really mean. With 18 years of data, many hundred stations and several pollutants, we could only provide an absolute minimum of these time series as example which are done in the following.

Figure 12: Observed and predicted ozone concentrations (MDA8 = max daily 8h running average) during the summer half year for station with the highest r score, DESN080, 2004-2017.

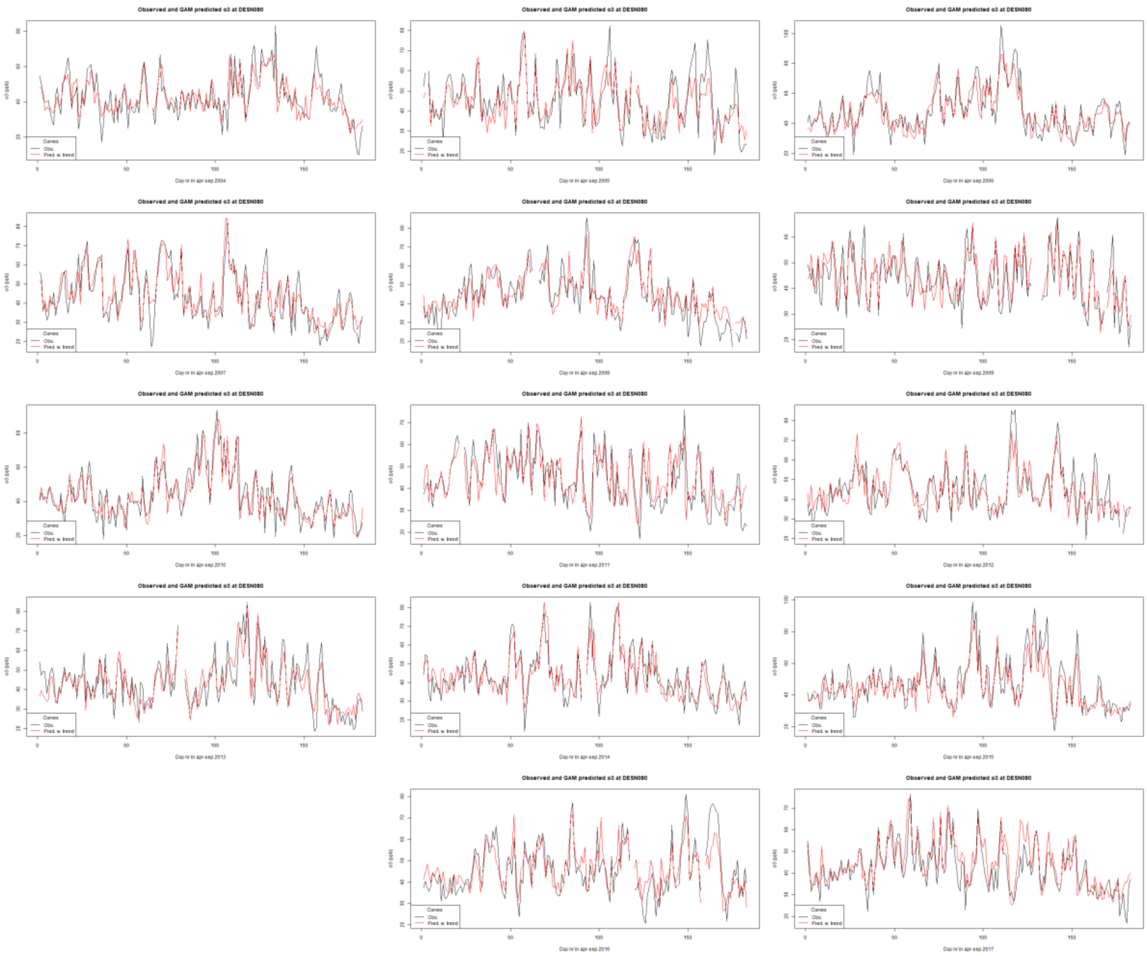
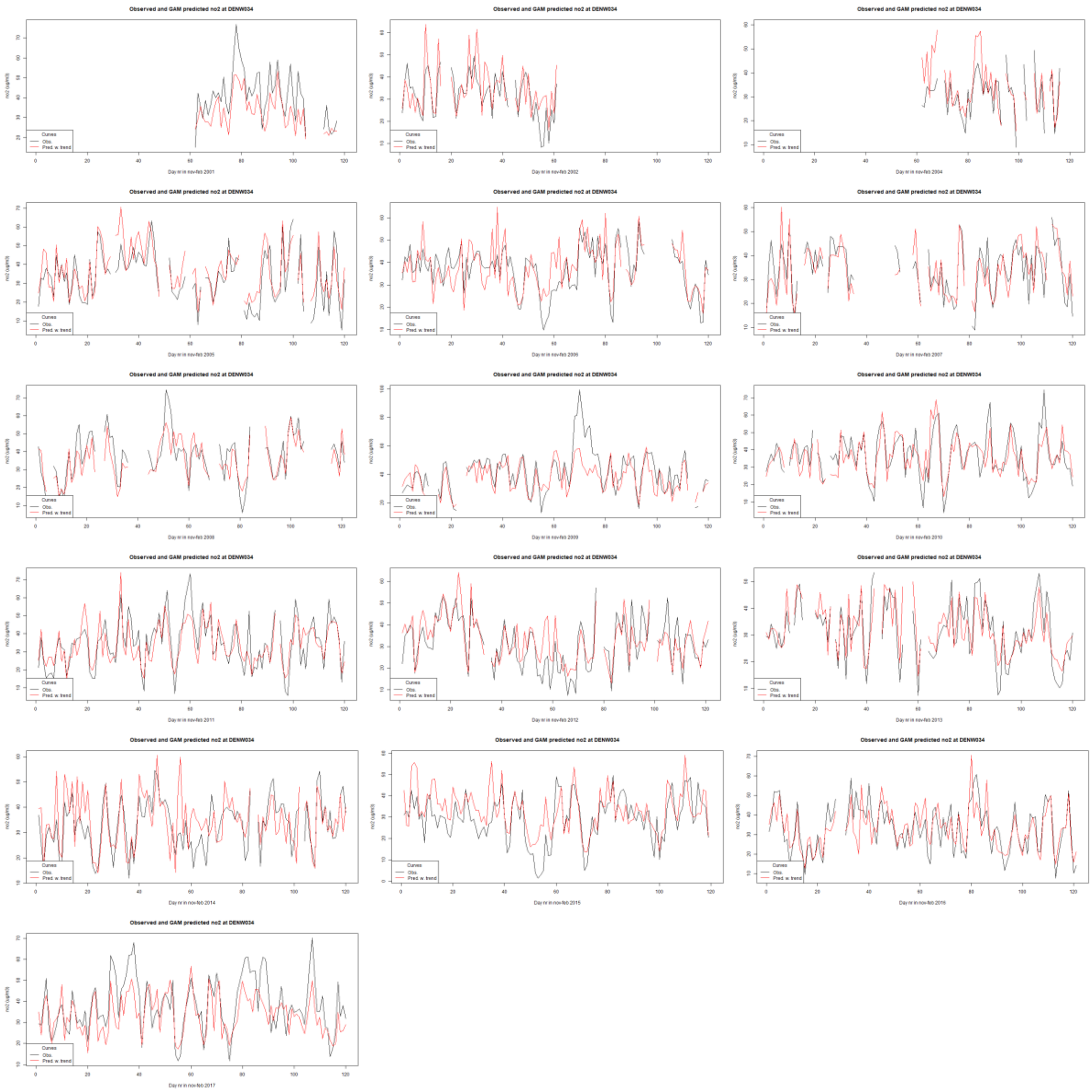


Figure 13: Observed and predicted NO2 daily mean concentrations during the winter half year for the station with the highest r score, station DENW134, 2001-2017.



4.5 Trends

The GAM function (Eq. 1) includes the trend term (X_9 in Eq. 1) and all the other terms as smooth functions which could thus not be tabulated as single numbers as in the case of a linear function. In the following, the results from the GAM were merged for the 9 adjusted PRUDENCE regions (Colette et al., 2016) as shown in Figure 4.

Since the pollutant levels vary considerably between the stations, we focussed on the trends in the relative concentration levels, i.e. relative to the mean concentration for each station, individually, making the individual trends comparable. The GAM model finds a smooth trend function through the whole period and for each station we scaled this function by that station's mean of the same function. Then we calculated the summer and winter seasonal means of these scaled trend values.

A similar scaling of the observed pollutant levels was done. For each site individually, we calculated the seasonal mean each year and then scaled those values to the mean value for the whole period 2000-2017.

This provided n relative GAM trend functions and n number of observed relative seasonal mean values. The smooth GAM trend functions represent the long-term trends adjusted for the meteorological variability and thus we calculated the difference:

$$\Delta = Xobs_{in,iy} - Xgam_{in,iy} \text{ for in} = 1, \ldots, \text{n stations and iy} = 2000, \ldots, 2017$$

where Xobs = the observed relative seasonal mean at one station, one year and Xgam = the relative GAM trend value for the same station and year.

In the following, we show box-whisker plots for the relative GAM trend values, i.e. the meteorologically adjusted trend function, as well as the Δ values, year by year. The latter in principle represents the impact of meteorology alone, although one should consider that this impact could also be estimated directly from the GAM model and that these different methods for evaluating the meteorological impact could lead to different results.

One main outcome from the GAM model is the meteorologically adjusted long-term trends, and another is to what extent the observations in individual years differ from the expected mean due to meteorological anomalies. In principle, this quantifies the split between the effect of emissions, chemistry, boundary conditions on one side and the influence of weather patterns on the other. In addition, a Theil-Sen slope was calculated for the Δ values in those cases where a Mann-Kendall trend test indicated a statistically significant positive or negative linear trend ($p < 0.05$). A significant linear tend in the Δ values would be the case if the inter-annual variations in meteorological effects alone have led to a significant trend in the observed pollutant levels during the whole period. The Theil-Sen slope is added to the box-whisker plots in the cases of significant trends. It turned out, though, that such a significant trend was found in very few cases which may reflect that the period we are looking at, 2000-2017 are sufficiently long that meteorological effects are cancelled out and too short to see any effects of climate change on the pollutant levels.

4.5.1 O_3 April – September

The meteorologically adjusted relative GAM trends and the annual Δ values (the meteorology impact) for summer ozone for each of the nine Prudence regions are summarized in Figure 14. As explained above, these ozone data are based on the daily MDA8 values (max daily 8h running mean concentration) through April-September. Note that the number of stations vary substantially between the regions. A decline in the meteorologically adjusted trends is seen in all regions except the Inflow region that shows only small variations. The shape of the trend curve varies though, from a steady decline in some regions to a curve peaking in the early 2000s in other regions. Many of the regions show, however, an indication of a flattening of the trend the last part of the period.

Another striking difference is the spread in the data among the stations. Whereas many of the regions show a narrow span in the data, like in East Europe, England, Mid Europe, North Italy and Scandinavia, indicating homogeneous data and a robust GAM trend, the stations from the Mediterranean and Iberian region show a much larger spread. For the latter regions, the results reflect the poorer GAM performance, implying that the method is less able to separate the meteorological influence from other processes. Additionally, it could reflect a higher difference between the stations and thus less homogeneous regions.

The Δ values show a marked positive outlier in 2003 and less marked positive offsets in 2006 and 2015. In these years the high levels of summer ozone were the results of certain meteorological conditions, of which the 2003 anomaly is the most famous one. In one of the regions, the Mid Europe region with

129 sites, we calculate a statistically significant increase in ozone due to meteorology alone marked by a red line in the plot, while none of the other regions show a significant trend due to meteorology.

Figure 14: Box-whisker plots for the relative GAM meteorologically adjusted trends of MDA8 ozone levels (April-September) for each of the different regions to the left and the corresponding difference between the relative observed MDA8 levels and the GAM trends to the right.

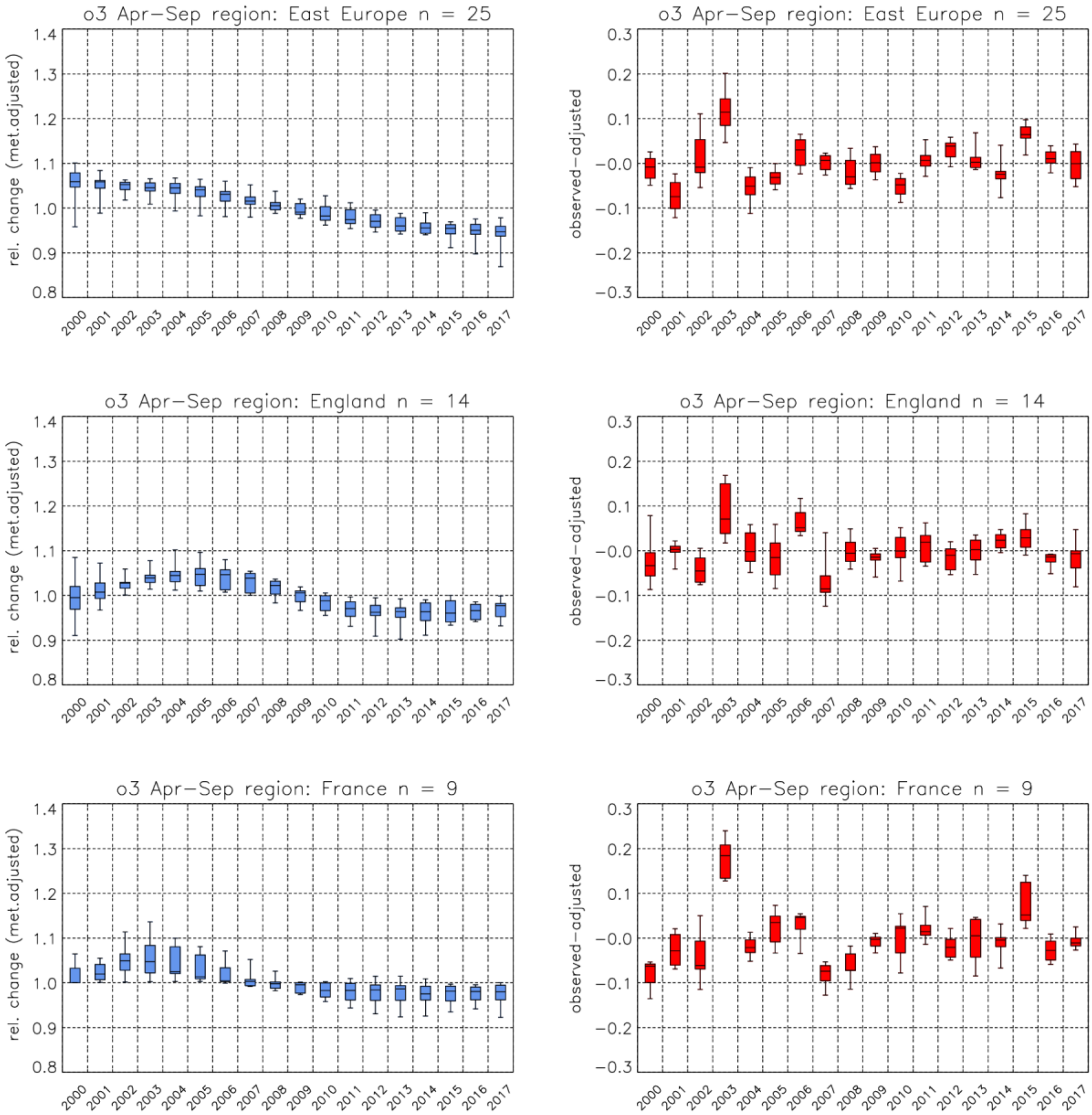


Figure 14 (contd.)

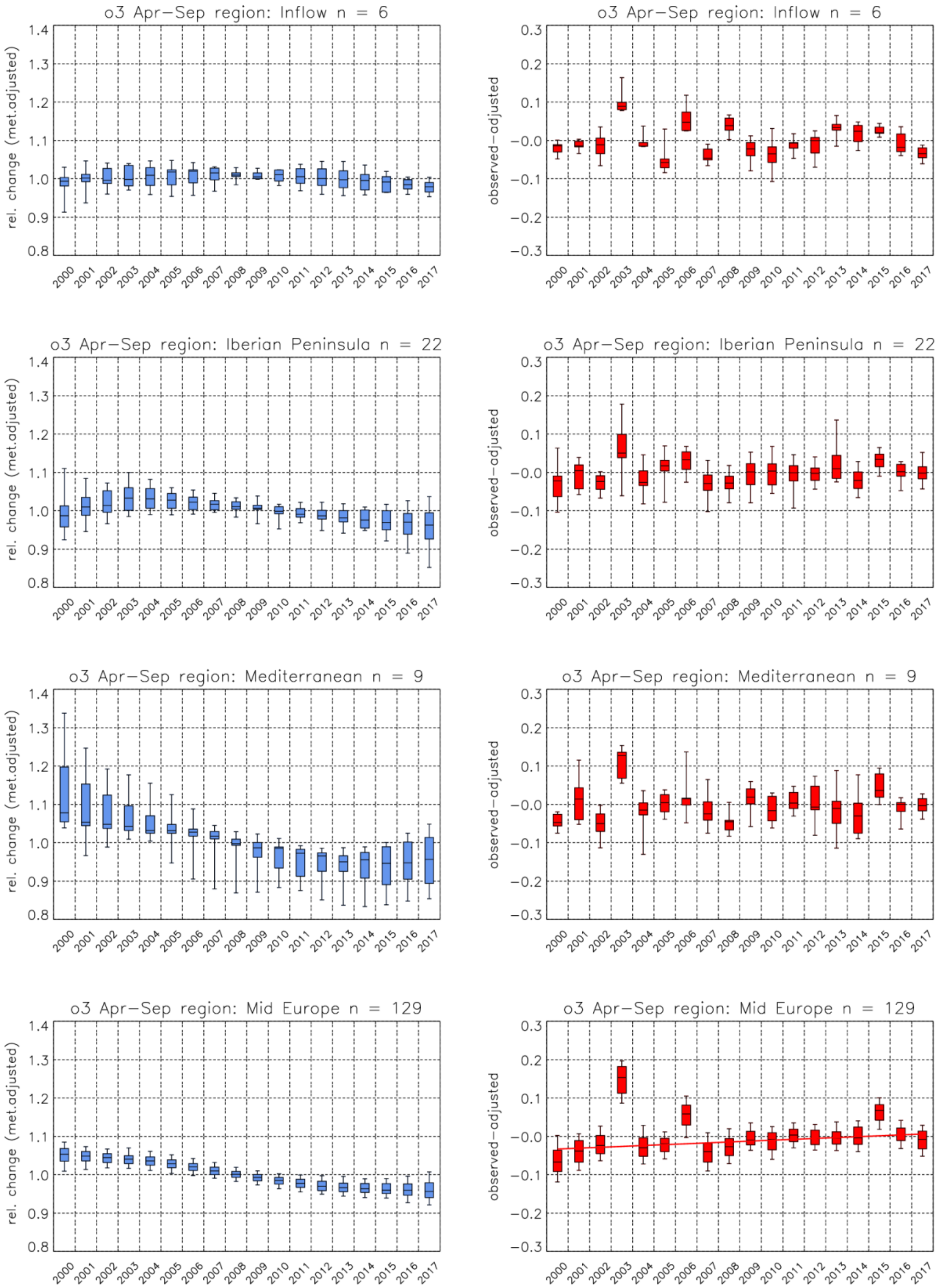
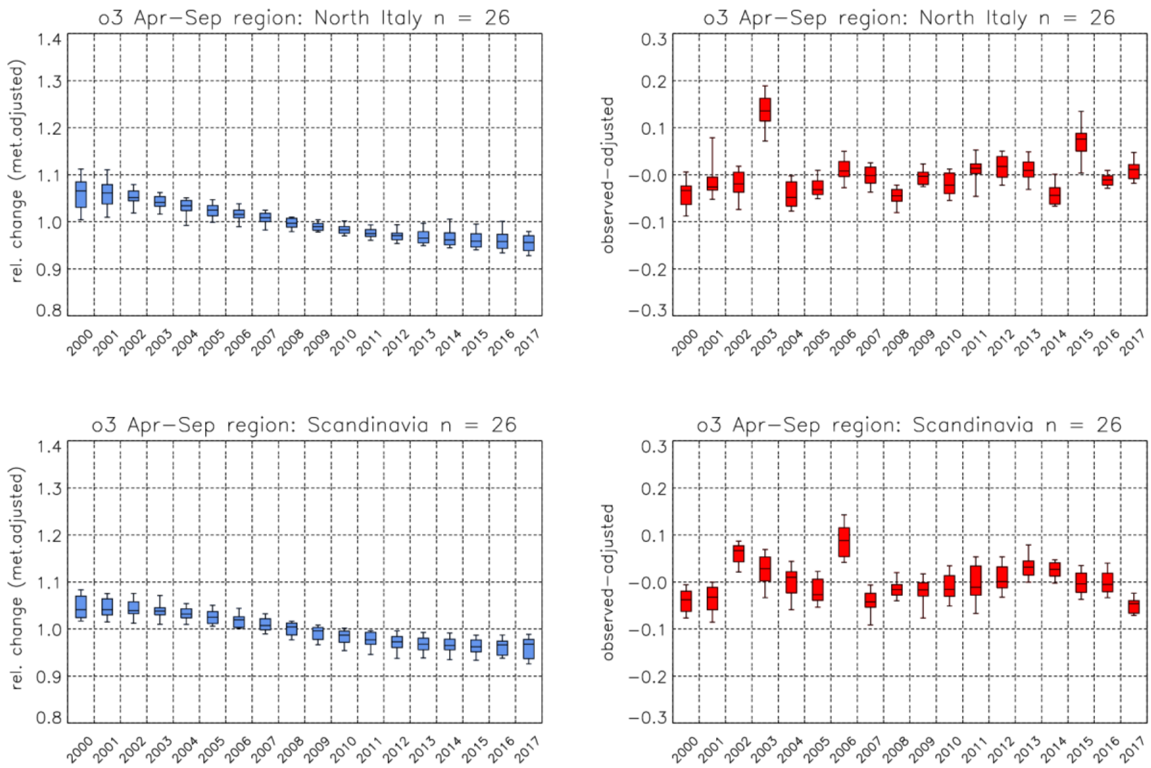


Figure 14 (contd.)



4.5.2 NO_2 November - February

The meteorologically adjusted relative GAM trends and the annual Δ values (the meteorology impact) for the mean wintertime NO_2 concentrations for the Prudence regions are summarized in Figure 15. As for the O_3 data in the previous section, the number of sites in the various regions vary substantially, from 9 in the Scandinavian region to 273 in the Mid European region. The Inflow region had too few sites to apply the box-whisker plots.

In all regions, the meteorologically adjusted trends show decreasing levels after 2007-2008. In the first part of the 2000s some regions show steady values or a slightly increasing trend, while other regions show a decline during the whole period 2000-2017.

The Δ values, expressing the impact of meteorological anomalies varies considerably between the regions with peak values in 2006 and 2017 in some regions. None of the regions show a significant linear trend (as estimated by the Mann-Kendall test) in the meteorological impact but the variability induced by the meteorology on wintertime NO_2 is much larger than for summer-timer ozone.

The rather small span in the meteorologically adjusted GAM trend in many of the regions indicates that the estimated trend is robust and that the GAM method is successful in separating the influence of meteorology vs that of other factors.

Figure 15: Box-whisker plots for the relative GAM meteorologically adjusted trends of mean wintertime NO_2 levels (November-February) for each of the different regions to the left and the corresponding difference between the relative observed levels and the GAM trends to the right.

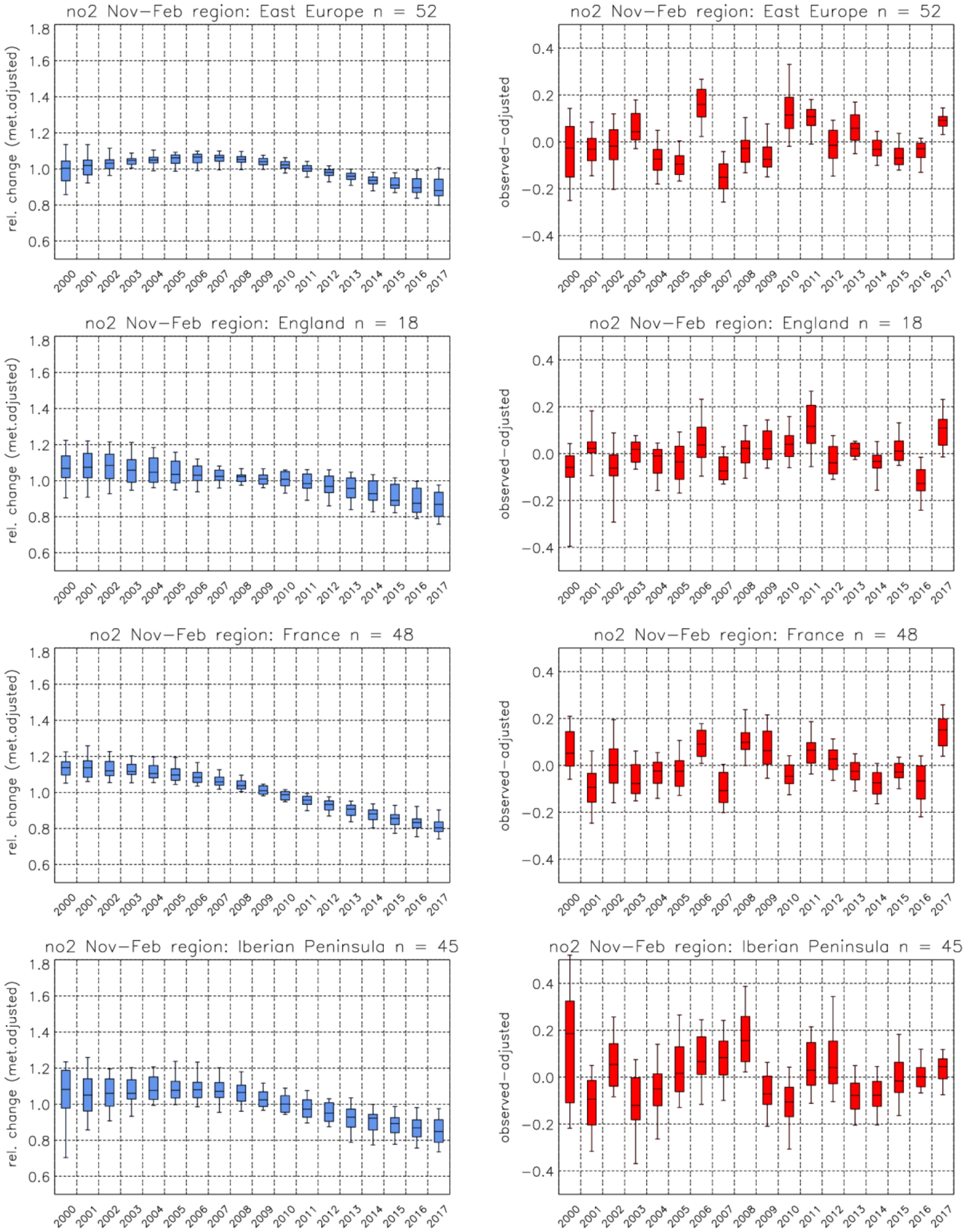
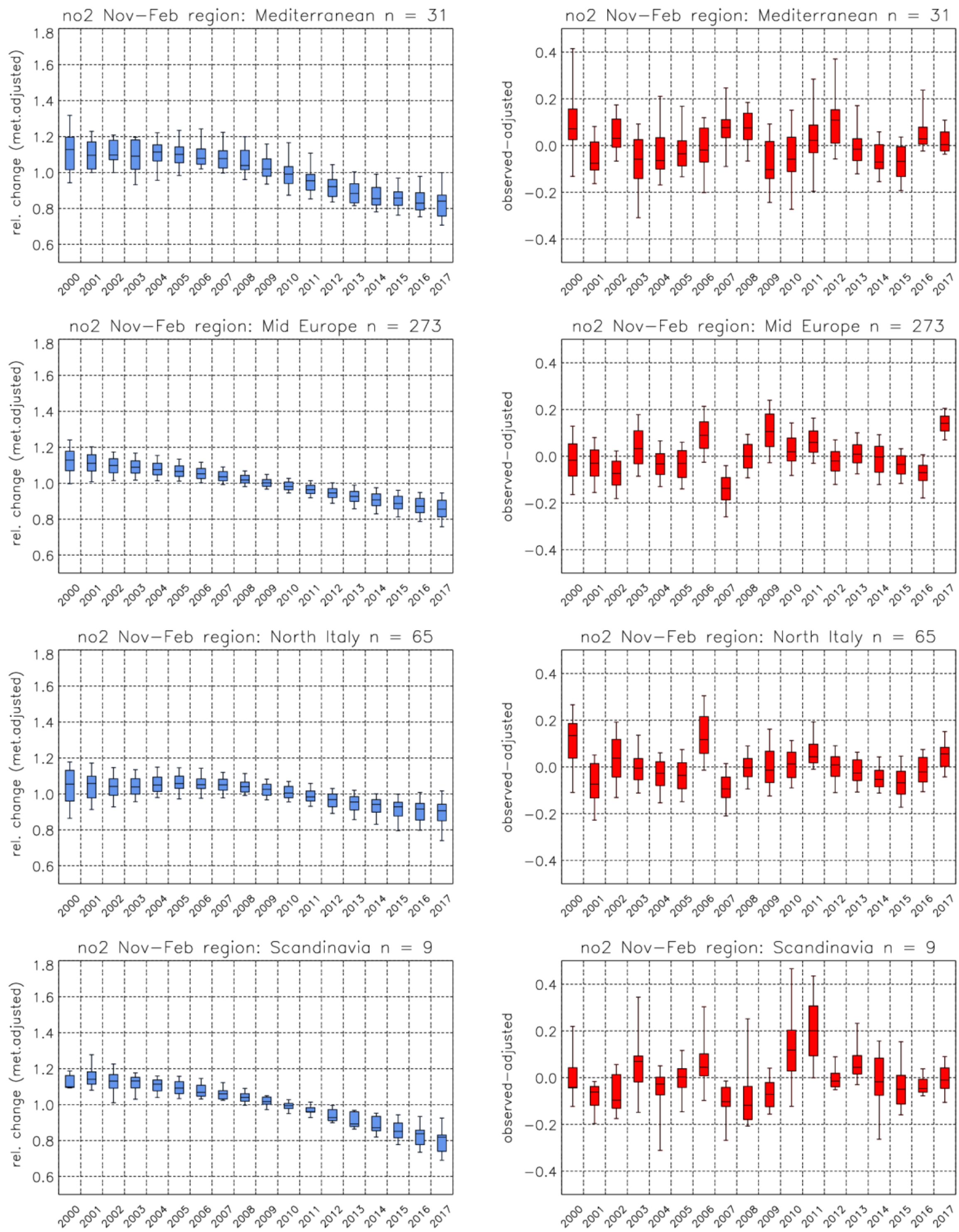


Figure 15 (contd.)



4.5.3 PM_{10} November – February

The meteorologically adjusted relative GAM trends and the annual Δ values (the meteorology impact) for the mean wintertime PM_{10} concentrations for the Prudence regions are summarized in Figure 16. The amount of PM_{10} data is much less than for NO_2 and O_3 and thus only four of the nine regions have enough sites to apply the box-whisker plots. For these four regions (East Europe, Iberian Peninsula, Mid Europe and Scandinavia) the meteorologically adjusted relative GAM trends show a marked decline as well as large spread in the Δ values, reflecting a larger meteorological influence on the wintertime PM_{10} concentrations from year to year than for NO_2. For none of these four regions, there is a significant linear trend in the meteorology induced anomalies (the Δ values) as estimated by the Mann-Kendall trend test.

Some patterns seen in these Δ values are the same as for the wintertime NO_2 data. The Mid European region which has by far the largest number of stations show high positive anomalies in 2006 and 2017 both in NO_2 and PM_{10}. The PM_{10} data also indicate a positive anomaly in 2003 while the similar peak in NO_2 is much smaller. Furthermore, marked negative anomalies are seen in 2001 and 2003 in the Iberian region both for NO_2 and PM_{10}, reflecting the winter weather conditions promoted lower pollutant concentrations these winters.

Figure 16: Box-whisker plots for the relative GAM meteorologically adjusted trends of mean wintertime PM_{10} levels (November-February) for each of the different regions to the left and the corresponding difference between the relative observed levels and the GAM trends to the right.

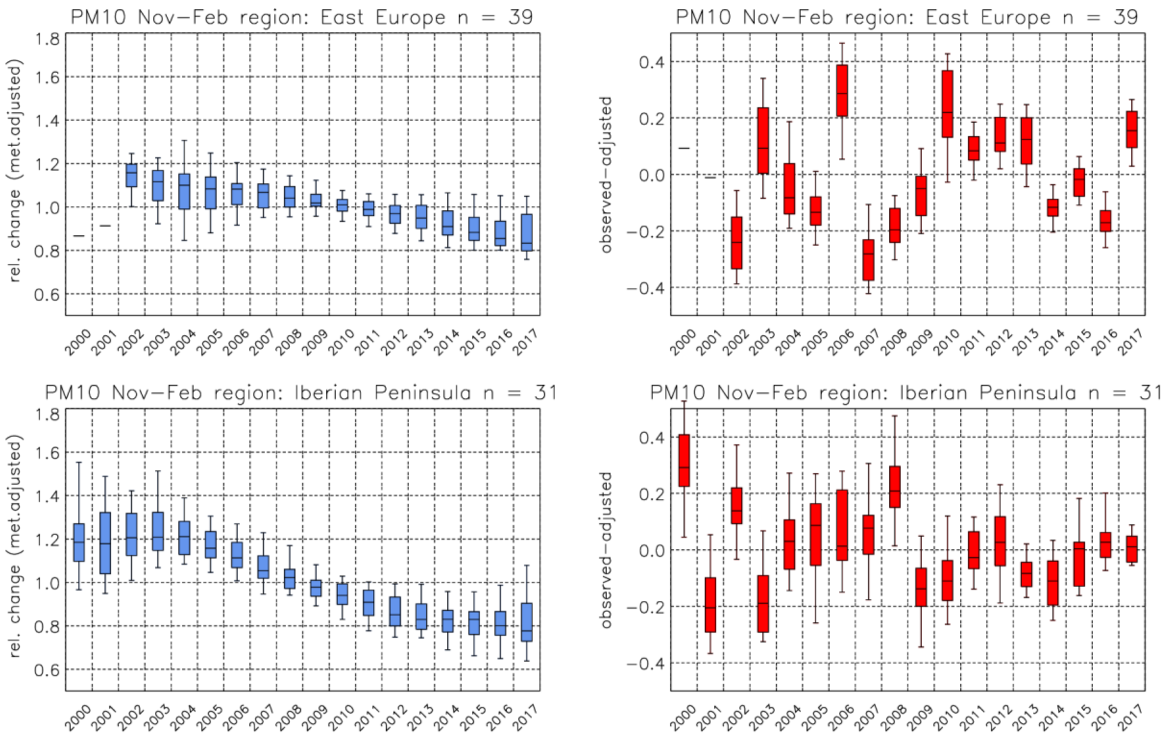
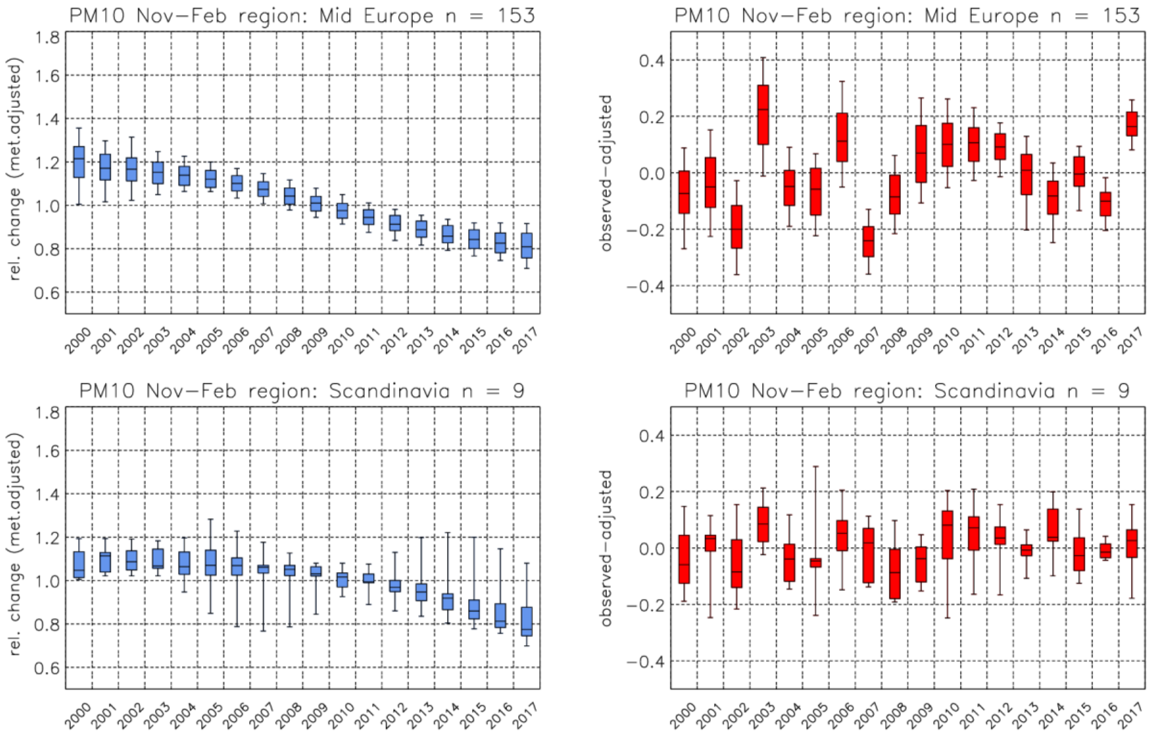


Figure 16 (contd.)



4.5.4 PM_{10} April - September

The meteorologically adjusted relative GAM trends and the annual Δ values (the meteorology impact) for the mean summertime PM_{10} concentrations for the Prudence regions are summarized in Figure 17. Only five of the nine regions have enough sites to apply the box-whisker plots.

As for the wintertime PM_{10} values, the meteorologically adjusted GAM trends indicate a marked decline for all regions during the period. Stable levels are seen however, for the Scandinavian region in the first part of the period. The spread in the Δ values, reflecting the meteorological induced variability is smaller for the summertime PM_{10} data than for wintertime, and as for wintertime, no significant linear trend is found by the Mann-Kendall trend test in any of the regions.

Figure 17: Box-whisker plots for the relative GAM meteorologically adjusted trends of mean summertime PM_{10} levels (April-September) for each of the different regions to the left and the corresponding difference between the relative observed levels and the GAM trends to the right.

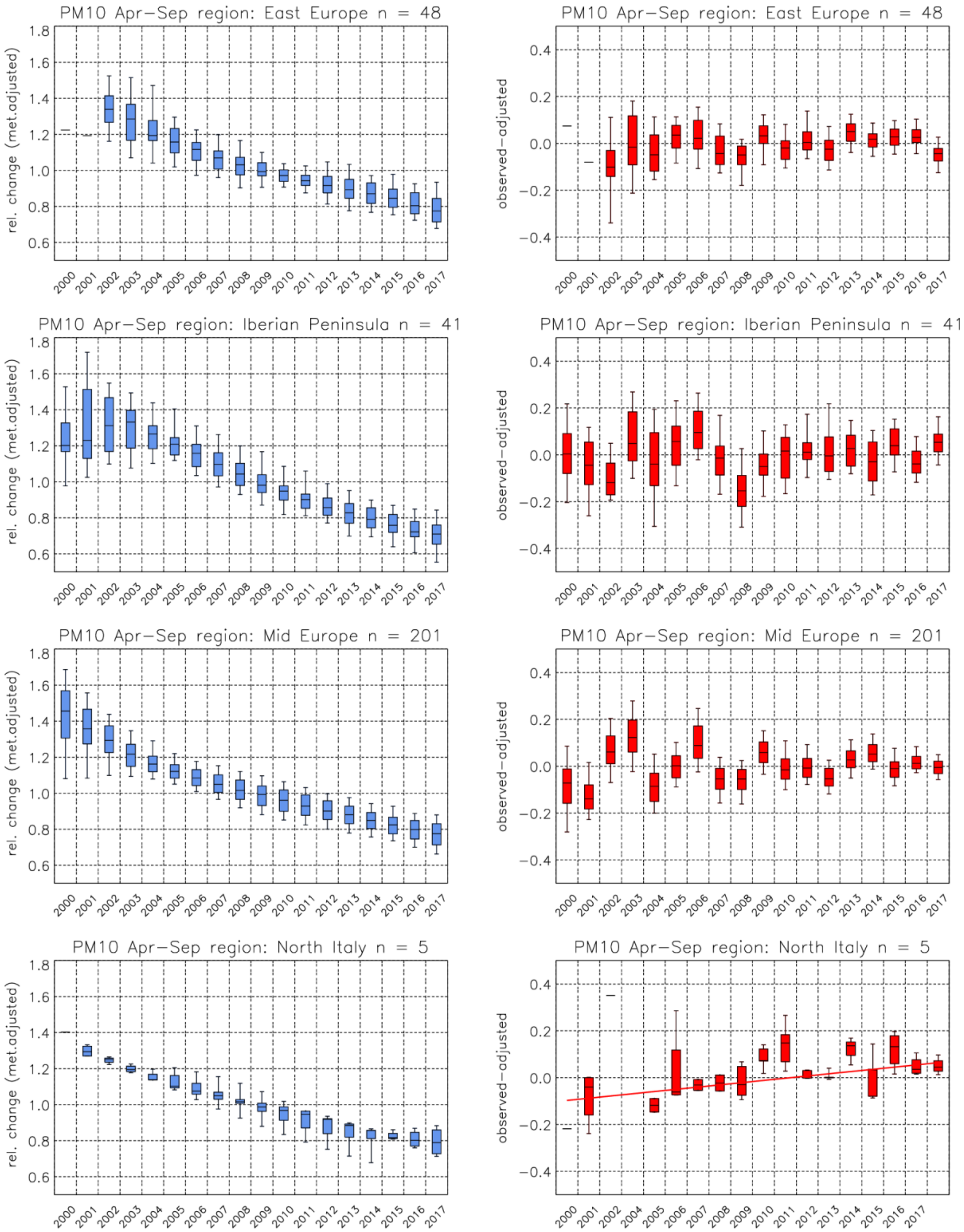
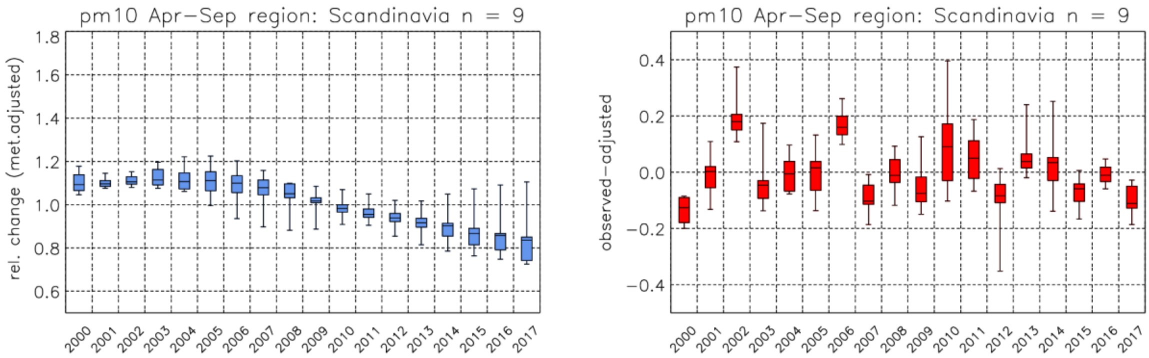


Figure 17 (contd.)



4.5.5 $PM_{2.5}$ November – February

The meteorologically adjusted relative GAM trends and the annual Δ values (the meteorology impact) for the mean wintertime $PM_{2.5}$ concentrations for the Prudence regions are summarized in Figure 18. The amount of $PM_{2.5}$ data is much less than for the other species and thus only the period 2008-2017 is investigated.

For the regions with a minimum number of stations during this period, the meteorologically adjusted relative GAM trends indicate a weak decline, most pronounced for the Scandinavian sites. The downward trend is much smaller than for PM_{10}, though. There is a large spread in the Δ values, reflecting a substantial meteorological influence on the wintertime $PM_{2.5}$ concentrations from year to year. For none of these regions, there is a significant linear trend in the meteorology induced anomalies (the Δ values) as estimated by the Mann-Kendall trend test.

Figure 18: Box-whisker plots for the relative GAM meteorologically adjusted trends of mean wintertime $PM_{2.5}$ levels (November-February) for each of the different regions to the left and the corresponding difference between the relative observed levels and the GAM trends to the right.

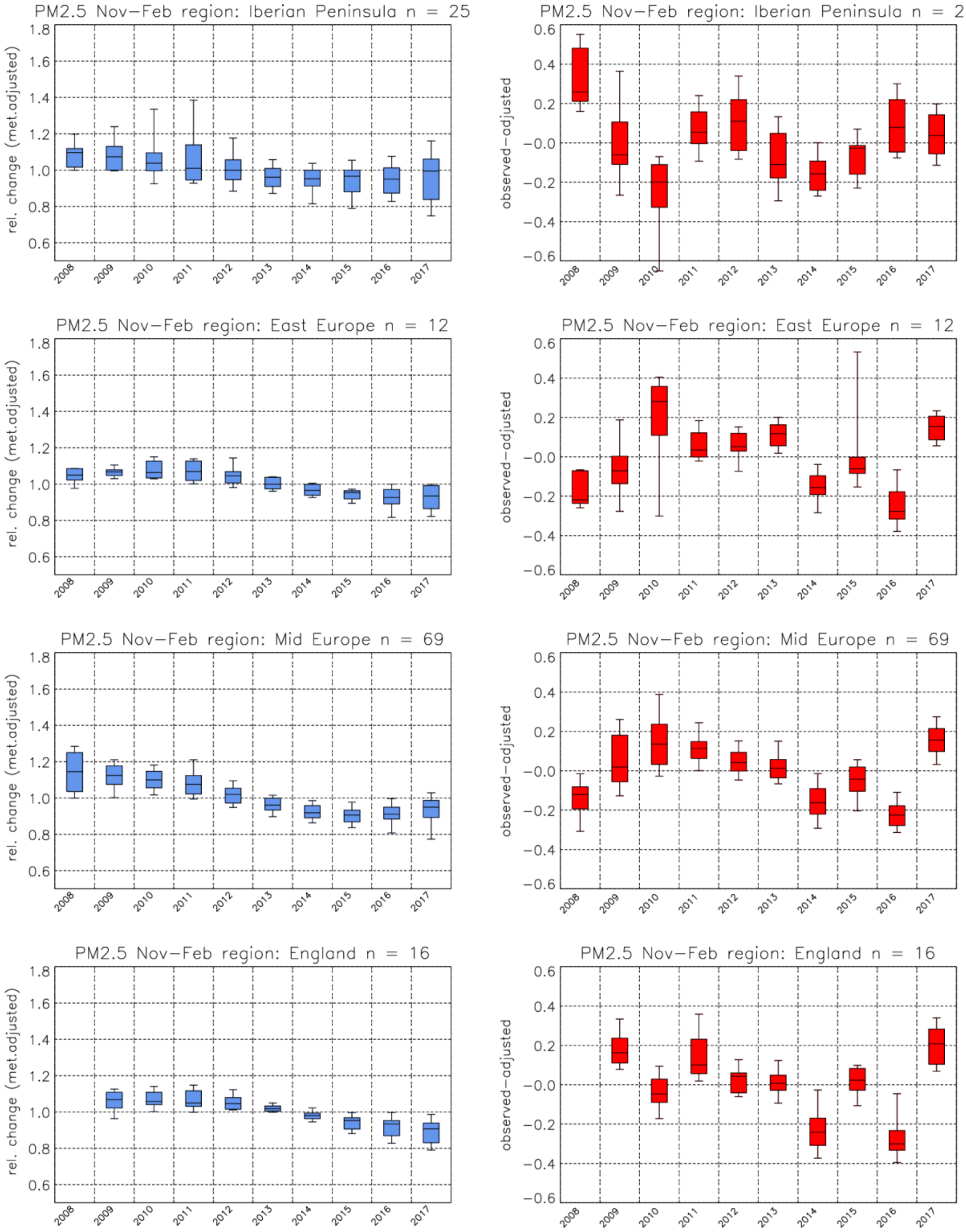
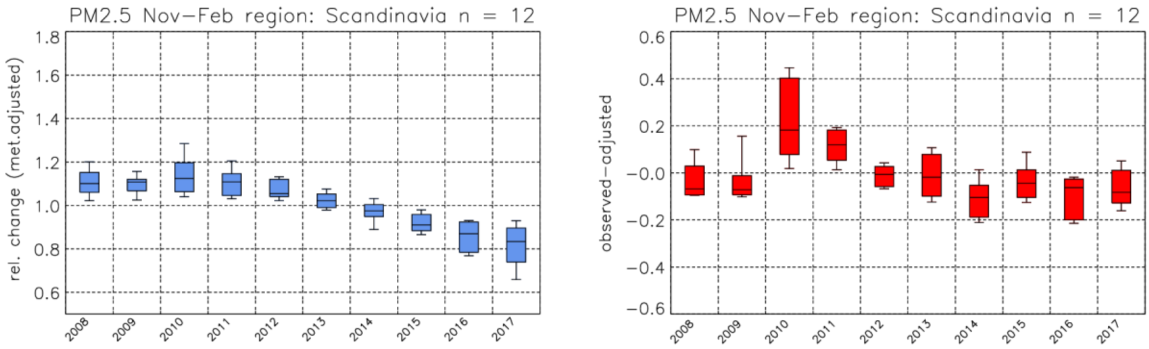


Figure 18 (contd.)



4.5.6 $PM_{2.5}$ April - September

The meteorologically adjusted relative GAM trends and the annual Δ values (the meteorology impact) for the mean summertime $PM_{2.5}$ concentrations for the Prudence regions are summarized in Figure 19.
Six of the nine regions have enough sites to apply the box-whisker plots.

Except for the Iberian Peninsula, the meteorologically adjusted GAM trends for the summertime $PM_{2.5}$ data indicate a more pronounced decline than the data from the wintertime. Furthermore, the spread in the Δ values, reflecting the meteorological induced variability is smaller for the summertime $PM_{2.5}$ data than for wintertime. The Mann-Kendall statistics indicates a significant trend induced by meteorology for the Iberian Peninsula, and a slight increase is estimated by the Theil-Sen slope.

Figure 19: Box-whisker plots for the relative GAM meteorologically adjusted trends of mean summertime $PM_{2.5}$ levels (April-September) for each of the different regions to the left and the corresponding difference between the relative observed levels and the GAM trends to the right.

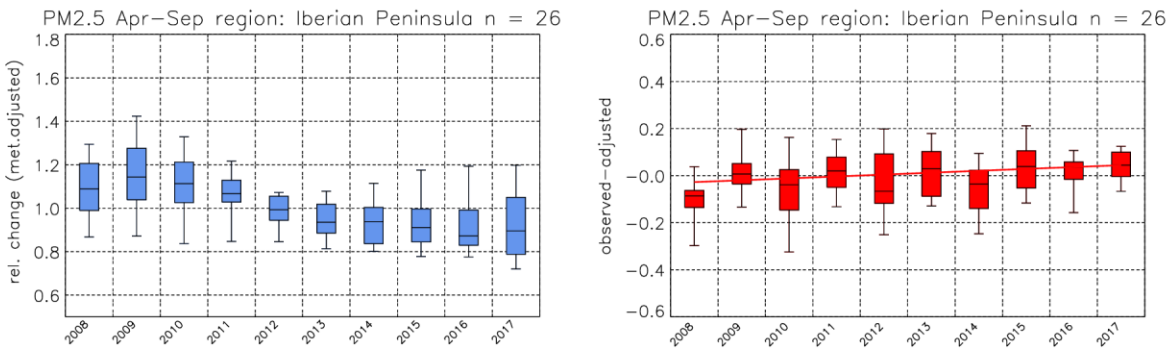


Figure 19 (contd.)

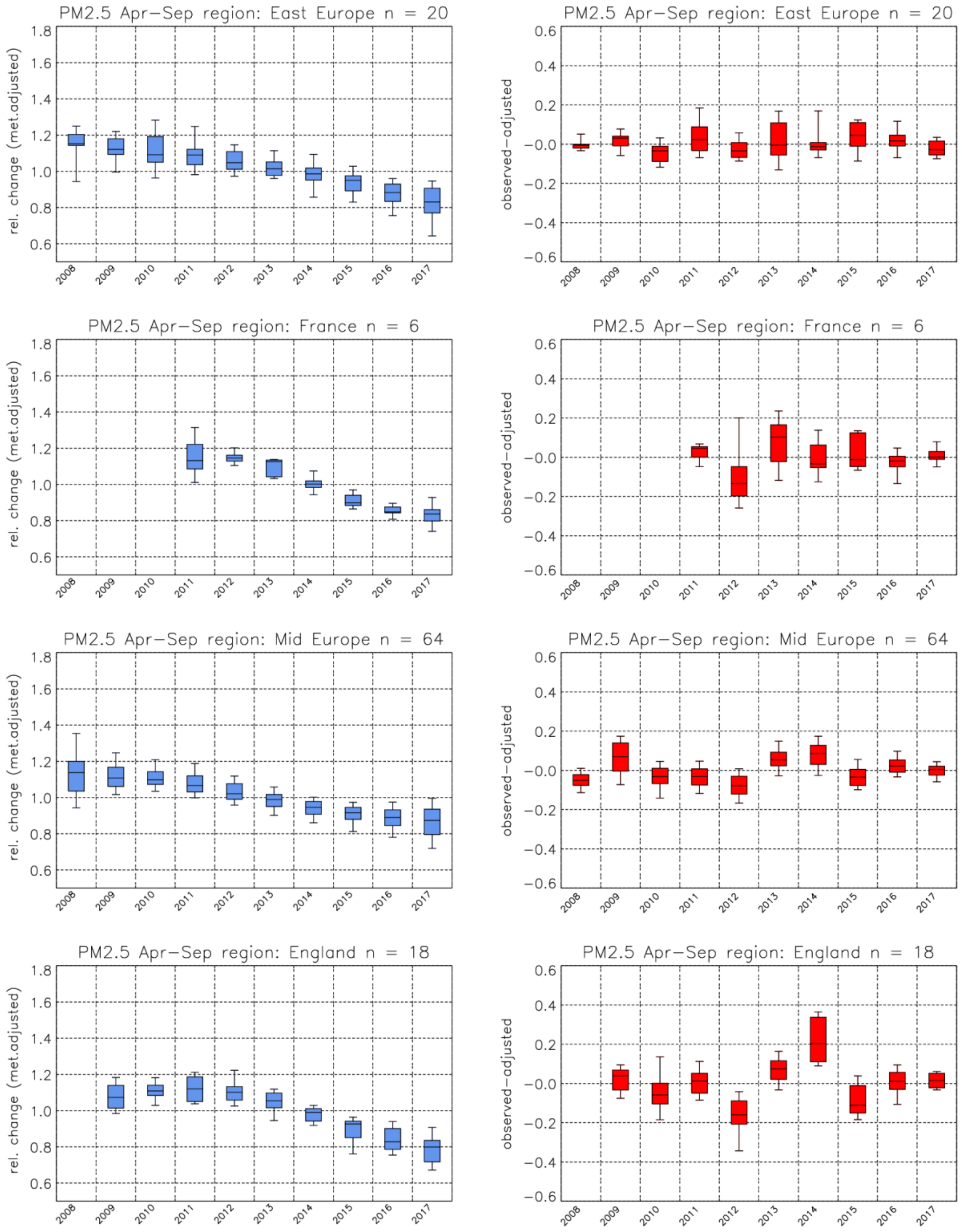
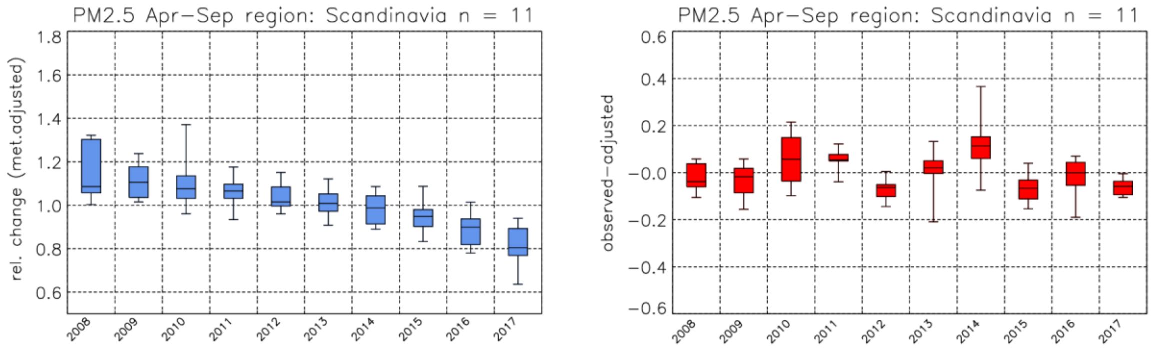


Figure 19 (contd.)



4.6 The impact of meteorology on observed seasonal pollutant levels

In addition to the meteorologically adjusted long-term trends presented above; the GAM method could be used to estimate the meteorological perturbation of the pollutant levels for each year separately. The principle of the GAM method is to separate the influence of meteorology from other effects and thus, it could be used to analyse to what extent e.g. the observed mean NO_2 winter-time concentration could be explained by weather anomalies or due to a gradual change in concentration levels.

Figure 20 shows the observed mean MDA8 (max daily running 8h average) ozone values for the summer half year for a few selected years together with the estimated perturbation due to meteorology for the same years as calculated by the GAM. The meteorology impact in 2003 is very clearly visible in Figure 20 and shows a marked positive perturbation for all of Europe corresponding with the well-known summer heat wave that year. Also, for 2006 and 2015, these results indicate that the weather conditions lead to higher levels than the expected mean, and the positive anomaly was of the order of 5 ppb, or around 10 % of the absolute level.

For 2008 and 2010, meteorology lead to lower ozone levels than the mean in many areas, according to these calculations particularly in the southeast. In 2008 this ozone deficit was widespread whereas in 2010 it was most pronounced in the southeast.

Similar maps for the observed 4-months winter-time mean concentration and the perturbation due to meteorology are shown in Figure 21. Note that the maps for each year are based on the mean of the diurnal mean levels for the months November and December the previous year + January and February the present year.

These maps show that the winter mean NO_2 levels depend strongly on the weather situations from one year to another. 2001, 2007 and 2016 were years in which the meteorology contributed to significantly reduced mean concentrations whereas 2009, 2011 and in 2017 were years where the weather conditions promoted high NO_2 levels over large regions in Europe.

Maps like those presented in Figure 20 and Figure 21 could be provided for any year and any species provided that there are sufficient data available to apply the GAM model and could be used in a routine manner to analyse previous observational data.

Figure 20: Observed mean summer half year MDA8 of ozone (left) and the perturbation due to meteorology (right) for a few selected years. Unit: ppb.

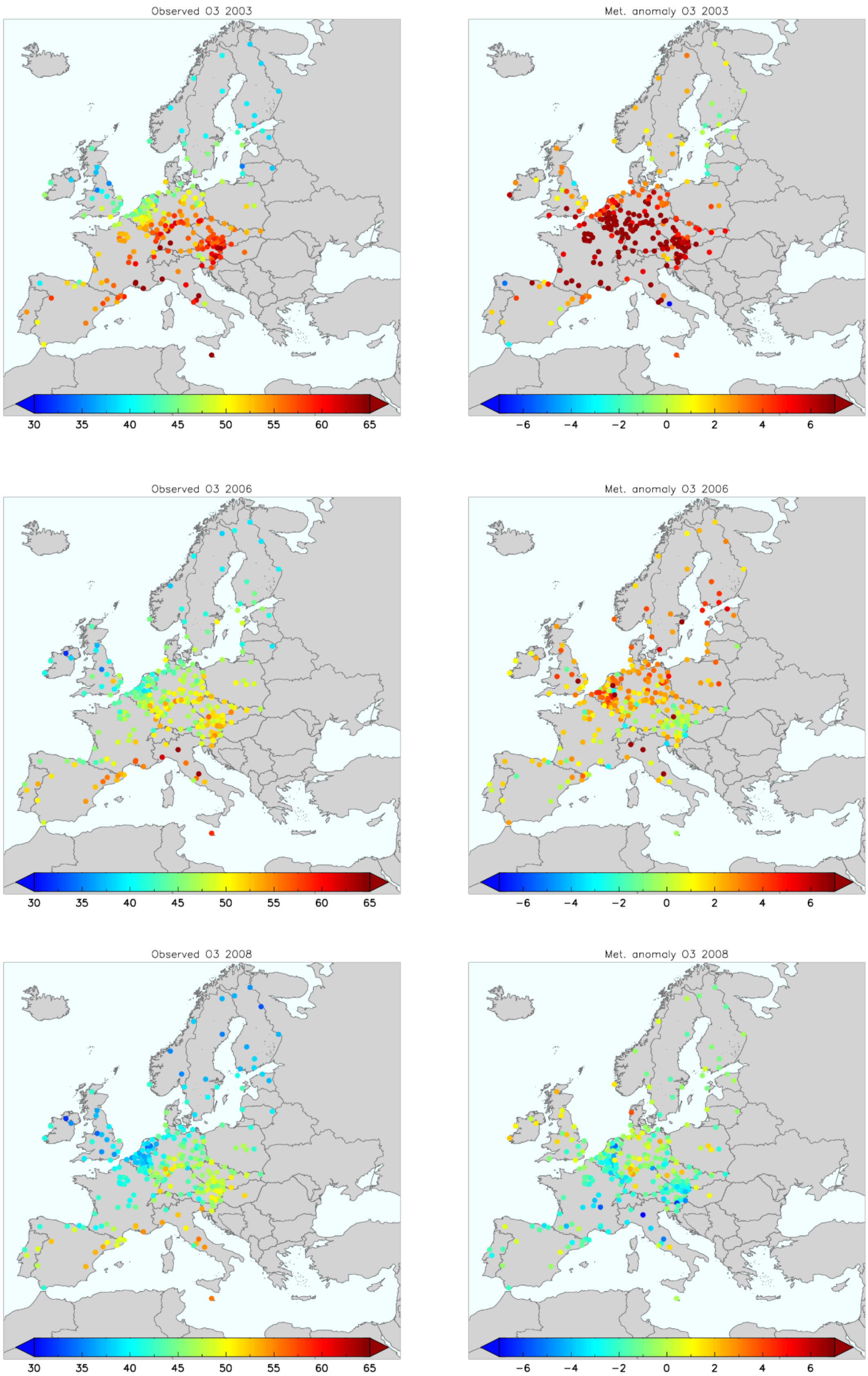


Figure 20 (contd.)

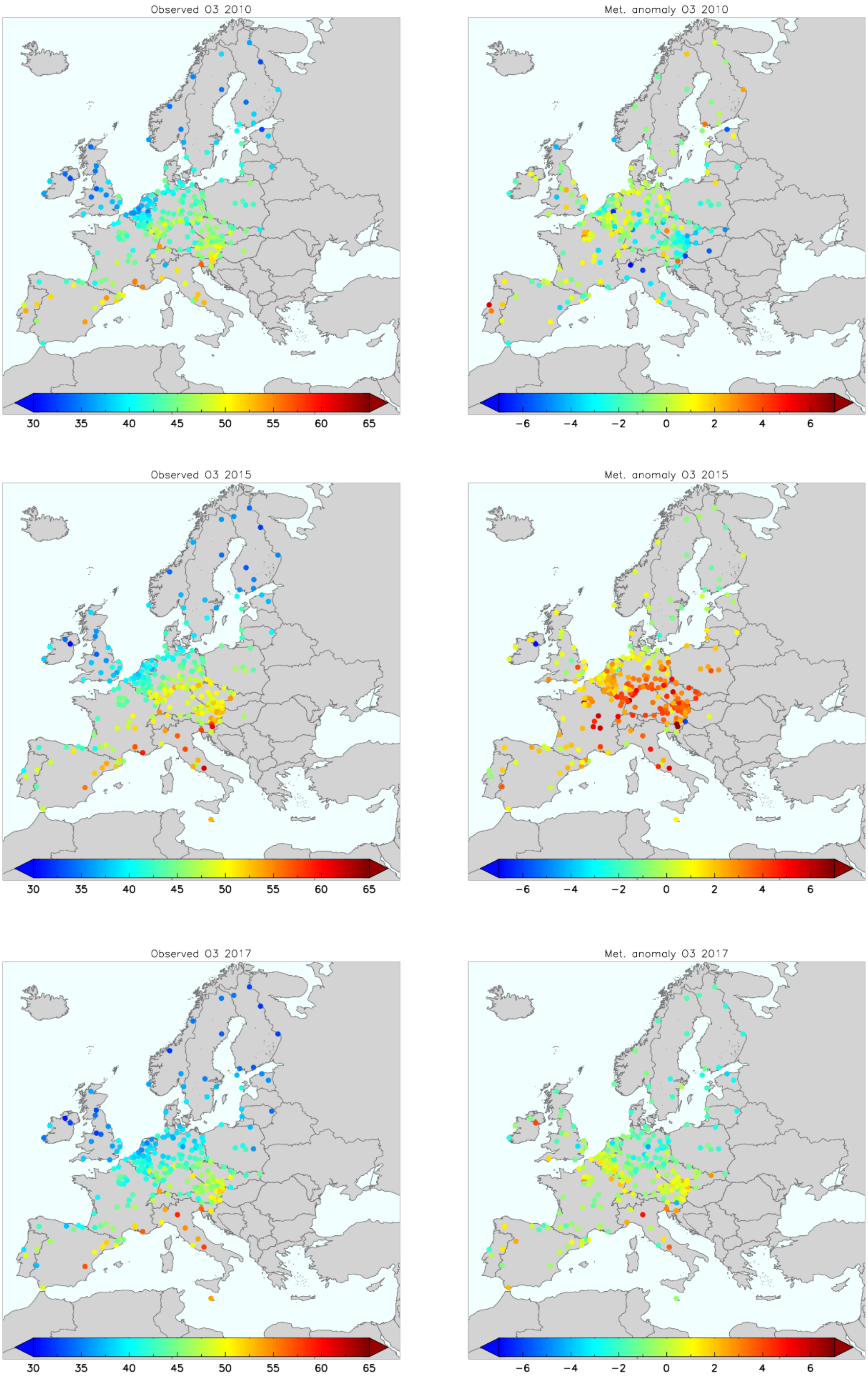


Figure 21: Observed mean winter (Nov-Feb) concentrations of NO_2 (left) and the perturbation due to meteorology (right) for a few selected years. Unit: µg/m^3.

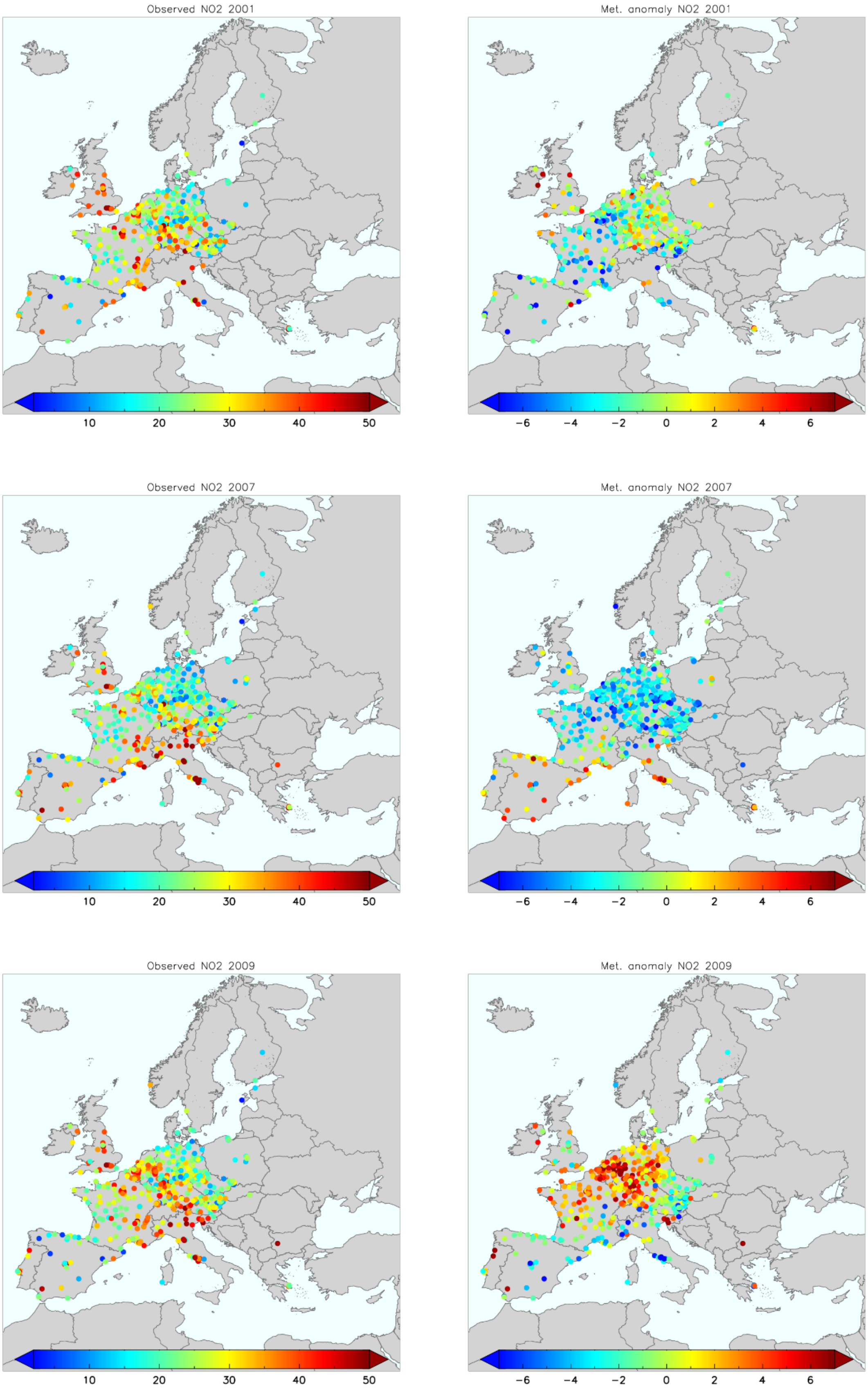
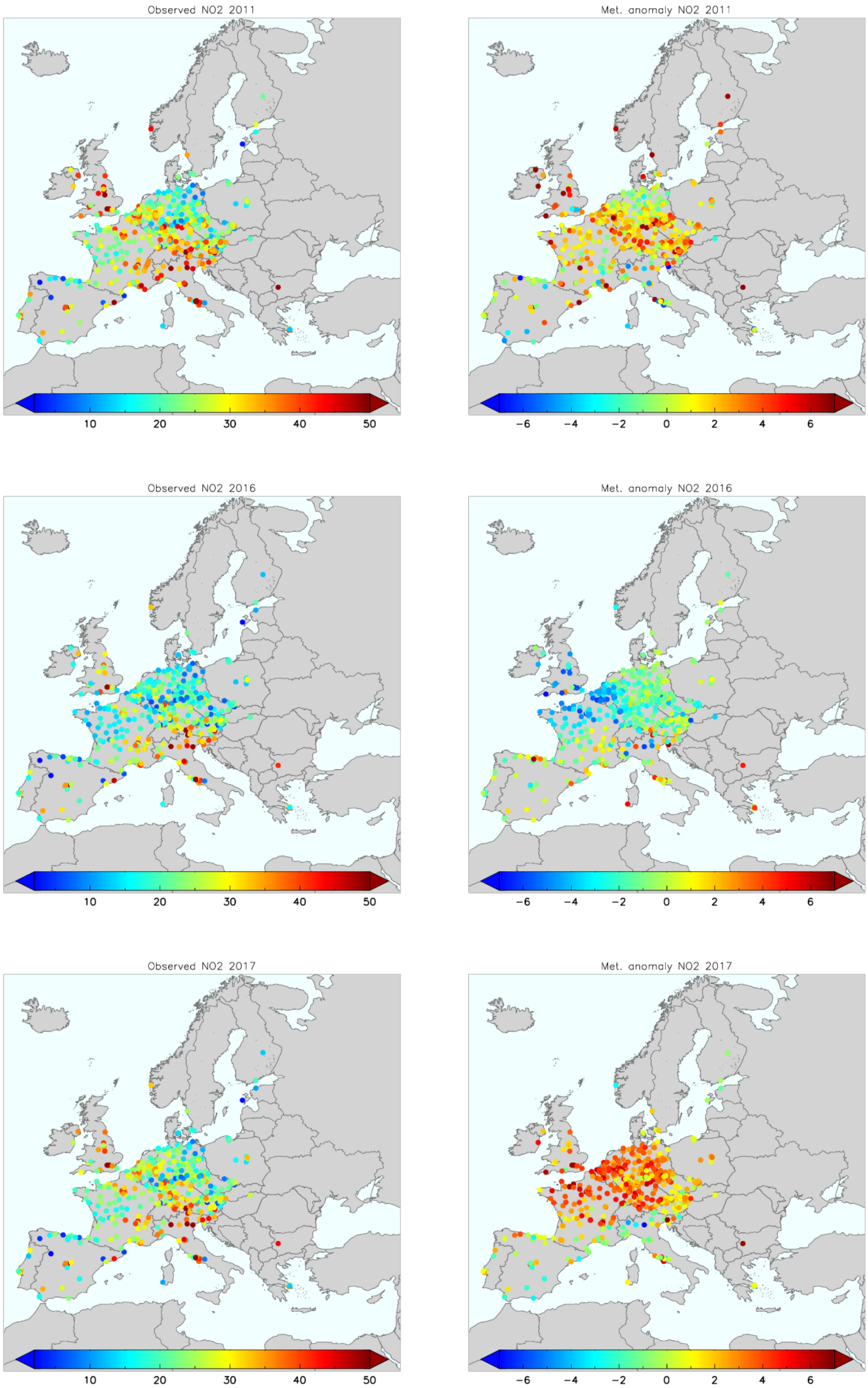


Figure 21 (contd.)



4.7 Outlier time series detected by the GAM model

A spin-off from the GAM model turned out to be the ability to detect what could be called outlier time series, i.e. time series with possible flaws in the data. These were detected either by particularly poor performance of the GAM model overall, or by a strong bias in certain years. We used the linear correlation coefficient, r, and the normalised mean gross error, NMGE, based on the entire time series for each station separately for the screening of data in the present project. Since this work was not part of the planned activity but just a valuable spin-off, we

The time series identified in this way could not be presented in detail in the present report. Instead a list of such time series will be provided to EEA. Some examples of typical flaws are given below.

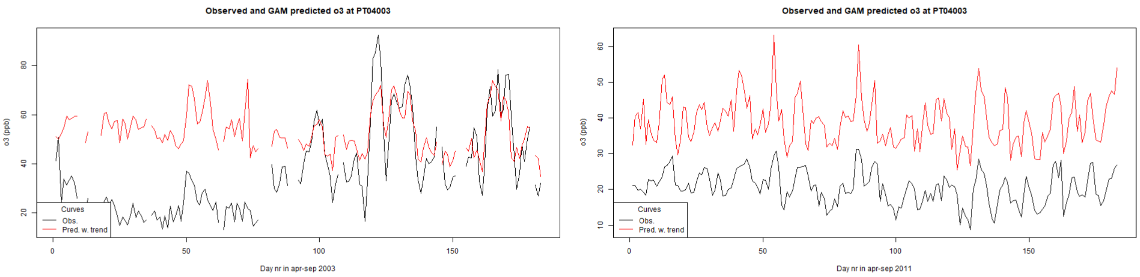
4.7.1 Wrong unit of O_3

The unit of O_3, i.e. whether it is given in ppb or µg/m^3, is a well-known issue of potential confusion. Whereas the UV monitors deliver data in ppb, the air quality guidelines are given in µg/m^3 and thus the data need to be scaled. Normally this is done by a factor of 2.00, i.e. O_3 [µg/m^3] = 2·O_3 [ppb]. Sometimes this issue lead to data in the wrong unit in the databases. This kind of error is very easy to detect as it is just a matter of scaling. With more than thousand ozone monitoring sites there is still a risk that such obvious flaws are hidden in the database. This is apparently the case for data from Irish sites in the e-Reporting database for 2014 and 2015. Figure 22 shows the observed and GAM predicted daily O_3 values at IE0090A for the period 2013-2016 and there is no doubt that the measurement data are wrong by a factor 2. Interestingly, the GAM model predicts the variation from day to day very closely, but with an offset reflecting the multiplicative factor 2.

A similar pattern is found for the site PT04003 as shown in Figure 23. The measured values in the first part of 2003 and all of 2011 are likely given in wrong unit.

Figure 22: Observed (black) and GAM predicted (red) time series of daily ozone levels in the years 2013-2016 for the site IE0090A. The reported data for 2014 and 2015 are wrong by a factor 2.0.

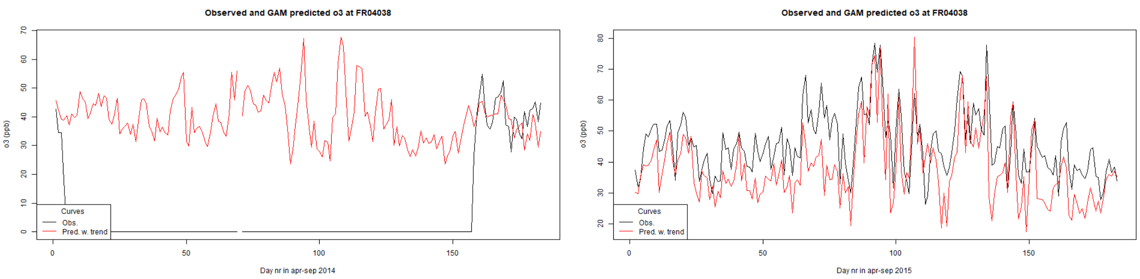
Figure 23: Observed (black) and GAM predicted (red) time series of daily ozone levels in 2003 and 2011 for the site PT04003.



4.7.2 Missing data specified as zero

For some time series, the errors are simply linked to wrong flagging of data. Figure 24 shows the observed and predicted time series of daily ozone in 2014 and 2015 at FR04038. Most of the data in the given period in 2014 are listed as zero while they obviously should be given as missing data. The data from 2015 shows that the GAM predicts the time series to a high degree although underpredicting the levels. Part of this underprediction is probably due to all the zero values in 2014. The flaws in the measurement data thus lead to reduced performance of the GAM model.

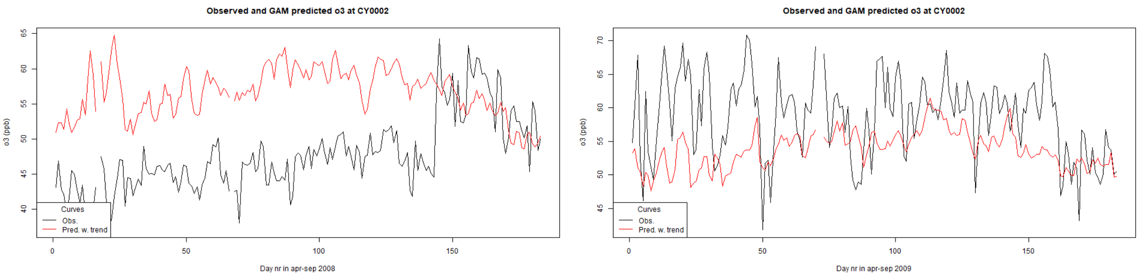
Figure 24: Observed (black) and GAM predicted (red) time series of daily ozone levels in 2014 and 2015 for the site FR04038.



4.7.3 Strong shifts between years

Poor GAM performance as measured by the linear correlation coefficient, r, could be an indication of strong shifts in the measurement data from one year to another. Figure 25 shows observed and predicted ozone levels at CY0002 in 2008 and 2009. The measured data are substantially higher in 2009 compared to 2008 and based on just a visual look at the data it is very likely that the measurement during the first part of 2008 are flawed before a sudden shift in the levels.

Figure 25: Observed (black) and GAM predicted (red) time series of daily ozone levels in 2008 and 2009 for the site CY0002.



The number of time series of NO_2 is considerably larger than for O_3, but it also seems (without doing an objective selection) that the fraction of erroneous or dubious data are larger for the NO_2 measurements. An example of NO_2 time series with a strong shift in the data are shown in Figure 26 for the site CZ0BZNO in 2011 and 2012. The GAM fit in 2011 and most of the other years seem ok, while all measurement data reported for 2012 are substantially lower than the GAM predictions and compared to the observed levels in the other years. It is little doubt that the measurement data from 2012 must be wrong and possibly linked to a scaling issue e.g. reflecting the difference in molecular weight of NO_2 and N.

A similar pattern is seen in Figure 27 for the site ES0584A in 2004 and 2005. In the first year the measured data show recurring peak values of around 30 µg/m^3 whereas in the year after (as well as the year before, not shown) all peak values are of the order of 70-80 µg/m^3. It seems very unlikely that these data reflect the reality except for the uncommon situation that some kind of local road constructions etc changed the overall pollutant levels substantial in one specific year.

Patterns like those shown in Figure 27 and Figure 25 seem to be fairly common in the Airbase/e-reporting data and creates significant challenges for those making use of the data for evaluation purposes.

One last example is given in Figure 26 showing the NO_2 station with the poorest GAM performance of all. Based on a visual inspection of the time series, it seems very likely that the reason for the poor GAM performance is due to flaws in the measurement data.

Figure 26: Observed (black) and GAM predicted (red) time series of daily NO_2 levels in 2011 and 2012 for the site CZ0BZNO.

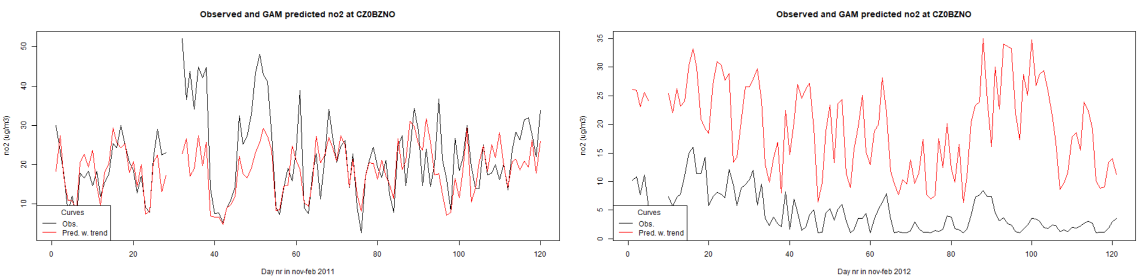


Figure 27: Observed (black) and GAM predicted (red) time series of daily NO_2 levels in 2004 and 2005 for the site ES0584A.

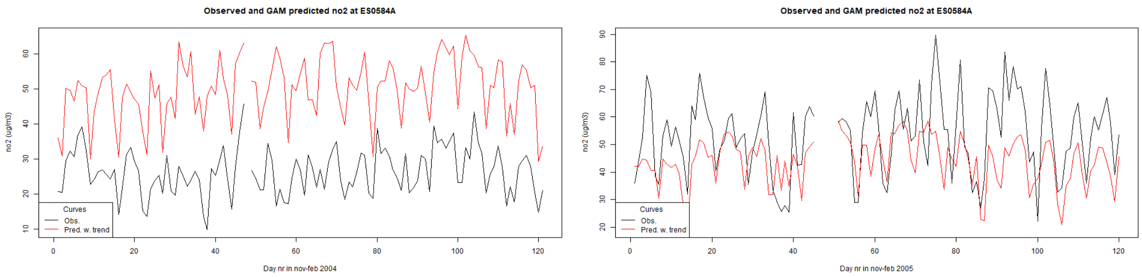
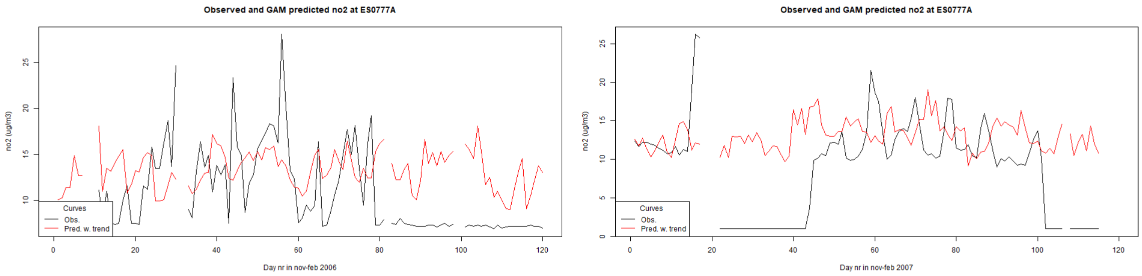


Figure 28: Observed (black) and GAM predicted (red) time series of daily NO_2 levels in 2006 and 2007 for the site ES0777A.



To sum up, a certain fraction of the measurement data shows clear signs of flaws in the data. This creates problems for anyone using the data for evaluation purposes, and screening and «cleaning» of the database implies that substantial extra time and effort is needed. The GAM model has proven very useful for this purpose since it is based on recurring relationships between pollutant concentration levels and meteorology from year to year. Sudden deviations from this pattern, that typically could be seen from model performance indicators, is an indication that something might be wrong and that the measurement data should be inspected more closely. This is particularly useful when the amount of data is large (many sites/years) as has been the case in the present task.

Based on the experience from the current work, which has involved a significant amount of time dedicated to data quality issues, we would strongly recommend that a set of data checking routines are developed to identify these problematic time series. It would probably be a small task to develop procedures that could pick out most of the questionable data. Whether these data then are removed from the database or just flagged as questionable is a political question though. In principle one could also imagine defining a set of particularly strong criteria to select the measurement data suited for trend analyses since such analyses require extra strong emphasis on data quality.

5 Conclusions and recommendations

The previous studies on the link between trends and meteorology has shown that these links could be estimated by a careful design of model setups using CTMs (chemical transport models). Multi-year model scenarios designed in certain ways is thus a very valuable tool for predicting long term trends in pollutant concentration levels and the role of individual physio-chemical processes. There is a scientific consensus that state-of-the-art CTMs provide the best approach for predicting and analysing trends and variabilities of atmospheric pollutants.

Long-term model runs with several model scenarios do, however, require a large amount of computational resources and the set-up of scenarios could be sensitive to the selection of years. Furthermore, the predictions are based on model data only and not tied to the observed pollutant concentrations which could pose a challenge if there are systematic discrepancies between modelled and measured pollutant concentrations.

Although CTMs without doubt constitute state-of-the art for modelling of the atmospheric composition, a large variety of scientific studies have been published the last few years on statistical models linking observed concentration levels to certain input data in a statistical way without any attempt to parameterize the physio-chemical processes. The GAM model that has been developed within the EEA tasks recently is one example of such models. The GAM model could be considered as complementary to the use of CTMs for separating the influence of meteorological variability from other processes.

Results from applying the model, linking observed pollutant concentrations to local meteorological data has been presented. As a model relating meteorology and air quality, the GAM is an efficient method for interpretation of data.

The main limitation of the statistical model is that it contains no parameterisation of the real physio-chemical processes and secondly, that it relies on a local assumption, i.e. that the observed daily concentrations could be estimated based on the local meteorological data. Although the latter assumption in principle is not valid for species like O_3, PM and NO_2, it turns out that the GAM model provides good predictions of the measurements for certain parts of Europe, while performances are more limited elsewhere.

We see three main applications of the GAM method when used on a regular basis:

1. Separate the long-term trend in observed concentration levels from the “noise” induced by meteorological variability and additionally look for any trends induced by meteorology alone.
2. Evaluate to what extent the pollutant levels in one specific year deviates from the expected level due to meteorological anomalies that year
3. Identify possible flaws in the measurement data. Since the statistical model is based on systematic patterns between meteorology and pollutant levels, a poor model performance could indicate errors in the observational data.

We found clear differences in model performance both with respect to geographical area and atmospheric species. In general, the best performance was found for O_3 with gradually lower performance for NO_2, PM_{10} and $PM_{2.5}$ in that order. With respect to area, the model produced the best predictions for Central Europe (Germany, Netherlands, Belgium, France, Austria, Czech Republic) and poorer agreement with observations in southern Europe.

Over the 18 years period studied (2000-2017) we found very few cases for which the meteorology alone caused a statistically significant trend in the data. One exception is the O_3 sites in Mid Europe.

For the sites in this region taken together, we estimate that meteorology alone caused a slight increase in the summer mean MDA8 levels. The general lack of meteorology induced trends in the data reflects the length of the time series in this study. Eighteen years is presumably a sufficiently long period that interannual variations in meteorology is not affecting the trend and still too short for climate change to have a noticeable effect. The advantage of the GAM method is that it can separate the long-term trend from the substantial interannual meteorological variability and thus provide robust trends.

An important spin-off from the GAM model is the ability to identify flaws in the measurement data series. The observational databases apparently contain a certain amount of erroneous data and the GAM model could be used (in combination with other tools) to identify these and thus clean the data.

To sum up, the GAM model is now in a phase that makes it ready for implementation and use on a regular basis. A few minor adjustments and refinements could be considered but else it could be put into operational use. The outcome would be linked to the three main aims listed above: separation of the long-term trends from the interannual variability, evaluation of one year of data compared to the mean and finally, the identification of outlier data.

Several specific action points to develop the GAM further is listed below:

- A detailed screening of the measurement data to filter out obvious flaws are strongly recommended.
- A comparison between data from EMEP and EEA should be carried out in order to identify discrepancies when looking at the same stations and species.
- The GAM model could be extended by extracting the partial dependencies on a daily basis for each input variable. This would allow us to quantify the influence and possible trend from each parameter individually.
- For long-term trend purposes, a subset of monitoring data with the stations showing the most reliable and well-documented data could be prepared.
- The GAM model could be further developed to be used on a region basis instead of for each station individually. This could improve the robustness and skills of the method.
- One could consider to define a set of performance criteria to decide if the GAM model is applicable for a specific station and parameter.

6 References

Berrisford, P., et al., 2011, ‘Atmospheric conservation properties in ERA-Interim’, *Q.J.R. Meteorol. Soc.*, 137, 1381–1399. doi: 10.1002/qj.864.

Camalier, L., et al., 2007, ‘The effects of meteorology on ozone in urban areas and their use in assessing ozone trends’, *Atmos. Environ*., 41, 7127–7137.

Carslaw, D.C. and Ropkins K., 2012, “openair —an R package for air quality data analysis”. *Environmental Modelling & Software,* Volume 27-28, pp. 52–61.

Carslaw, D.C., 2019, *The openair manual — open-source tools for analysing air pollution data. Manual for version 2.6-3*, University of York.

Carslaw, D. C., 2018, *deweather: Remove the influence of weather on air quality data*. R package version 0.5. http://github.com/davidcarslaw/deweather

Chan, E., 2009, ‘Regional ground-level ozone trends in the context of meteorological influences across Canada and the eastern United States from 1997 to 2006’, *J. Geophys. Res*., 114, D05301, doi:10.1029/2008JD010090.

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

Colette, A., et al., 2015, *Air Quality Trends in AIRBASE in the context of the LRTAP Convention*, European Topic Centre on Air pollution and Climate change Mitigation, ETC/ACM Technical Paper 2015/4.

Colette, A., et al., 2016, *Long term air quality trends in Europe Contribution of meteorological variability, natural factors and emissions*, ETC/ACM Technical Paper 2016/7, European Topic Centre on Air pollution and Climate change Mitigation.

Davis, J., et al., 2011, ‘A comparison of CMAQ-based and observation-based statistical models relating ozone to meteorological parameters’, *Atmos. Environ*., 45, 3481–3487.

EEA, 2013, *Air quality in Europe – 2013 report,* EEA Report 9/2013, European Environment Agency.

Fix, M. J., et al., 2018, ‘Observed and predicted sensitivities of extreme surface ozone to meteorological drivers in three US cities’, *Atmos Environ* 176, 292-300, https://doi.org/10.1016/j.atmosenv.2017.12.036.

Grange, S. K., et al., 2018, ‘Random forest meteorological normalisation models for Swiss PM_{10} trend analysis’, *Atmos. Chem. Phys*., 18, 6223–6239.

Grange, S. K. and Carslaw, D. C., 2019 ‘Using meteorological normalisation to detect interventions in air quality time series’, *Science of the Total Environment* 653, 578-588.,

Ordonez, C., et al., 2005, ‘Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003’, *Atmos. Chem. Phys*. 5, 1187-1203. http://dx.doi.org/10.5194/acp-5-1187-2005.

Otero, N., et al., 2018, ‘A multi-model comparison of meteorological drivers of surface ozone over Europe’, *Atmos. Chem. Phys*., 18, 12269–12288, https://doi.org/10.5194/acp-18-12269-2018.

Pernak, R., et al., 2019, ‘Forecasting Surface O_3 in Texas Urban Areas Using Random Forest and Generalized Additive Models’, *Aerosol and Air Quality Research*, 19: 2815–2826, doi: 10.4209/aaqr.2018.12.0464.

Solberg, S., et al., 2015, *Discounting the impact of meteorology to the ozone concentration trends,* ETC/ACM Technical paper 2015/9, European Topic Centre on Air pollution and Climate change Mitigation.

Solberg, S., et al., 2018a, *Discounting the effect of meteorology on trends in surface ozone: Development of statistical tools*, ETC/ACM Technical Paper 2017/15, European Topic Centre on Air pollution and Climate change Mitigation.

Solberg, S., et al., 2018b, *Trends in measured NO_2 and PM. Discounting the effect of meteorology*, ETC/ACM Eionet Report 2018/9, European Topic Centre on Air pollution and Climate change Mitigation.

Thompson, M.L., et al., 2001, 'A review of statistical methods for the meteorological adjustment of tropospheric ozone', *Atmos. Environ*., 35, 617 – 630, doi:10.1016/S1352-2310(00)00261-2.

Vaisala, 2013, *Humidity conversion formulas, calculation formulas for humidity*, Report B210973EN-F. Vaisala Oyj, Helsinki, Finland, [www.vaisala.com.](http://www.vaisala.com)

Wood, S.N., 2017, *Generalized Additive Models. An introduction with R*, CRC Press, Boca Raton, Florida USA. doi:10.1201/9781315370279.

Zheng, J., et al., 2007, 'Interannual variation in meteorologically adjusted ozone levels in the eastern United States: A comparison of two approaches', *Atmos. Environ*., 41(4), 705–716, doi:10.1016/j.atmosenv.2006.09.010.

Annex 1
Statistical method

As in the previous studies, the GAM model (Eq. (1)) is fitted using the GAM library "mgcv" (Wood, 2017) in the statistical modelling system R (R Core Team, 2018) for each station for the period 2000-2017. We use all data in each period in order to estimate the long-term meteorologically adjusted trend function β_9 in Eq. (1) at each station.

New in the current study is that we apply the function "bam" in the "mgcv" library first instead of using the "gam" routine as was used previously. Both routines fit GAM models to data, but "bam" is much faster than "gam", which is important in the current study due to many stations and long period with data. The "bam" routine is therefore now tried first and only if this is found not to converge properly, the "gam" routine is used instead.

The "bam" routine is run with method = "fREML" which is the default numerical solution method in "bam". It stands for fast REML computations. For NO_2, PM_{10} and $PM_{2.5}$, for which we use a log link function, we use discrete = TRUE in "bam" which discretizes covariates for storage and efficiency reasons.

The default fitting method in the "gam" routine is based on Newton-type optimization of the GCV (generalized cross-validation) scores in order to estimate or select the degree of smoothness. This was the method applied in previous years reports (Solberg et al., 2018a, 2018b). This method usually works fine but may sometimes get stuck in local minima close to but not equal to a more correct global minimum. The REML (restricted maximum likelihood) method is less prone to get stuck in such local minima. Another aspect is that the GCV method often tends to overfit and produce too wiggly smooths as its focus is on minimizing the prediction errors. The REML method in "gam" tends to produce less wiggly smooths in this regard and will less easily get stuck in local minima. The consensus in the GAM modelling community now seems to favour the REML method rather than the GCV approach to model fitting of GAMs in general. We have therefore switched to this method in "gam" by setting method = "REML" in the call to this routine.

The "gam" routine generally does a good job of selecting smooth functions of the predictors but does not perform model selection by default. However, by setting "select = TRUE" in the call to "gam" a further penalty for having too many unnecessary predictors in the GAM is introduced. This may lead to one or more of the smooth predictors to be further penalized away to become zero functions. Thus, this setting can be used for more parsimonious models with no unnecessary predictors to be selected automatically by the "gam" routine. We believe this generally to be a good idea also in our application, so we have switched to this added model selection approach by using "select = TRUE" in the call to the "gam" routine.

The same automatic model selection with "select = TRUE" is also used in the call to "bam".
We generally apply thin-plate splines which are the default in bam and gam for all smooth functions except for wind direction where we use circular splines (bs=" cc"). We use standard and default 10 degrees of freedom for all smooth functions except for the time covariates "weekday", "day of season" and "years" where we use 4 degrees of freedom in order to obtain less wiggly functions for these, but still informative with a minimum of residual noise. Note that the "weekday" variable takes discrete values 1, 2,…, 7 but is considered to be a continuous "time of week" 1-7 covariate in our system. Similarly, "day of season" is treated also as a continuous variable.

A check for concurvity of the covariates has been added to the script. Concurvity is the same for GAM as collinearity is for multiple linear regression. It is important that covariates are genuinely different and that no smooth covariate function can be replaced by a combination of smooth functions of the other covariates. It is thus, important to check for this as part of the GAM modelling. We have thus introduced a call to the "concurvity" routine in the "mgcv" library in R as part of the script.

A “rmweather” method has also been added to the script, which is a non-parametric random forest type of approach to non-linear regression (Grange and Carslaw, 2019; Grange et al., 2018). This is an approach like the “deweather” boosted regression tree method (Carslaw, 2018) which was already in our system. The “rmweather” differs, however, from “deweather” in that it can also be used to predict concentrations in whole left-out years, enabling us to check its predictive performance in the same way as we do for GAM. This is not possible to do with “deweather” since only a random sample can be left out to perform testing, not whole years.

These two latter approaches, “deweather” and “rmweather”, represents the two most complex and least interpretable approaches to the trend estimation problem, while a linear regression approach using only a single continuous “years” covariate represents the other end of the spectrum. Our GAM model falls in somewhere between these two extremes.

European Topic Centre on Air pollution,
transport, noise and industrial pollution
c/o NILU – Norwegian Institute for Air Research
P.O. Box 100, NO-2027 Kjeller, Norway
Tel.: +47 63 89 80 00
Email: etc.atni@nilu.no
Web : https://www.eionet.europa.eu/etcs/etc-atni

The European Topic Centre on Air pollution, transport, noise and industrial pollution (ETC/ATNI) is a consortium of European institutes under a framework partnership contract to the European Environment Agency.

European Environment Agency
European Topic Centre on Air pollution,
transport, noise and industrial pollution