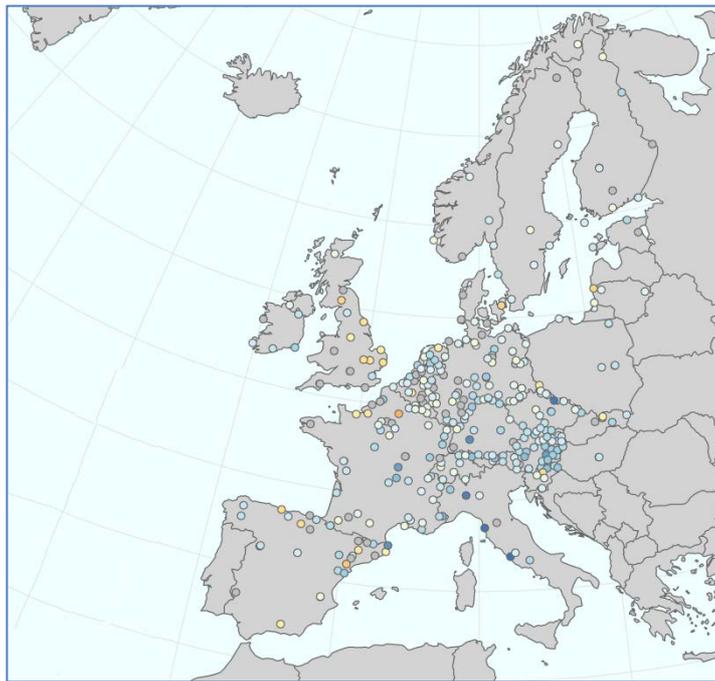


Discounting the effect of meteorology on trends in surface ozone: Development of statistical tools



ETC/ACM Technical Paper 2017/15
August 2018

*Sverre Solberg, Sam-Erik Walker, Philipp Schneider,
Cristina Guerreiro, Augustin Colette*



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

Front page picture:

Meteorologically adjusted ozone trend as calculated by the GAM method for rural EMEP and AirBase sites 2000-2010. See Figure 4.24 (left) of this paper.

Author affiliation:

Sverre Solberg, Sam-Erik Walker, Philipp Schneider, Cristina Guerreiro: NILU - Norwegian Institute for Air Research, NO
Augustin Colette: INERIS - French National Institute for Industrial Environment and Risks, FR

DISCLAIMER

This ETC/ACM Technical Paper has not been subjected to European Environment Agency (EEA) member country review. It does not represent the formal views of the EEA.

© ETC/ACM, 2018
ETC/ACM Technical Paper 2017/15
European Topic Centre on Air Pollution and Climate Change Mitigation
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone +31 30 2748562
Fax +31 30 2744433
Email etcacm@rivm.nl
Website <http://acm.eionet.europa.eu/>

Contents

Executive summary	5
1 Introduction	6
2 Methodology	6
2.1 The US-EPA approach.....	7
2.2 Our approach and comparison with the EPA model.....	8
2.2.1 Difference in meteorological parameters.....	9
2.2.2 Seasonal parameter instead of weekend parameter.....	9
2.2.3 Difference in long-term trend parameter.....	10
2.2.4 Difference in link function	11
2.2.5 Trajectory data not included.....	11
2.3 Use of GAM and openair model software.....	12
3 Input data	13
3.1 Surface ozone monitoring data	13
3.2 Meteorological data.....	13
3.3 Homogeneity and robustness criteria.....	14
3.4 From daily to annual indicators	16
4 Results	17
4.1 Performance of the GAM method.....	17
4.1.1 Results for Waldhof (DE0002).....	18
4.1.2 Results for Harwell (GB0036).....	22
4.1.3 Results for Kårvatn (NO0039)	25
4.1.4 Results for Peyrusse Vieille (FR0013) and La Tardiere (FR0015)	28
4.1.5 Overall statistics	32
4.2 Spatial differences in GAM performance.....	34
4.2.1 The period 2000-2010	34
4.2.2 The period 1990-2000	36
4.3 Importance of the explanatory variables.....	38
4.3.1 The period 2000-2010	41
4.3.2 The period 1990-2000	43
4.4 Long-term trends.....	45
4.4.1 The period 2000-2010	45
4.4.2 The period 1990-2000	48
4.4.3 The influence of meteorological adjustment of trends	50

5	The use of GAM as a data quality checking tool.....	53
6	Comparison with other methods for meteorological adjustments.....	57
7	Summary of findings	58
8	Future developments.....	60
9	Acknowledgement	61
	References.....	61

Executive summary

Trends in ozone concentrations are not only influenced by the changes in anthropogenic emissions of ozone precursors, but also by the interannual variations in meteorology, among other factors. The main aim of this work is to evaluate whether a statistical method applied by the US-EPA (published by Camalier et al., 2007) to adjust for the effect of interannual variations in meteorology on surface ozone trends could be applied for European monitoring sites.

The US-EPA method can in principle be used to estimate long-term ozone trends adjusted for the effect of variations in meteorology, reflecting mostly the ozone trend caused by emission changes alone. Secondly, the method could be used to explain the ozone levels observed in one specific year in terms of meteorology.

The method is a so-called GAM (generalized additive model), which could be regarded an advanced multiple regression method in which the regression coefficients are replaced by smooth functions. Hourly ozone measurements from the EMEP and AirBase databases were used to calculate the maximum daily 8-h running mean ozone (MDA8 O₃) values for the summer season (1 May - 31 August) for two time periods 1990-2000 and 2000-2010. For the meteorological data we used daily gridded values (e.g. of temperature, relative humidity) that were used in the EURODELTA Trends project. The method's assumption is that, although ozone is not determined by the local weather conditions directly, most of the local meteorological variables reflect the large-scale weather situation that is in turn closely linked to the level of ozone.

The performance of the method was evaluated by comparing daily measured and modelled MDA8 O₃ values, giving promising results in certain areas of Europe. The method's performance was good to very good in central Europe and in Germany in particular. In the Nordic countries, the seasonal cycle and the timing of ozone episodes were well reproduced, although with a substantial underestimation of the high peaks. On the other hand, in southern Europe poorer agreement with observed values were seen.

The results showed that the meteorological adjusted trends were systematically less negative than the non-adjusted trends, indicating that meteorology contributed to the downward trend in O₃ seen at most sites for both 1990-2000 and 2000-2010. These results are in accordance with the results from two other methods reported in Colette et al. (2016). The downward trend in ozone during 1990-2010 summers was mainly driven by the emission change.

Additionally, the GAM demonstrated to be useful to detect time series containing errors in the data, such as sudden shifts in the levels. The GAM can also be used to evaluate how O₃ levels in a year deviate from average levels, due to the meteorological variability of that year.

1 Introduction

Long-term air quality trends have been the topic of various ETC tasks the last years (Colette et al., 2015; Solberg et al., 2015; Colette et al., 2016). This work has been linked to ongoing activity within EMEP TFMM and EURODELTA Trends (named EDT hereafter). More specifically, the role of interannual variations in meteorology versus the effect of changes in anthropogenic emissions for the trends has been studied in detail. A variety of methods for isolating these two effects on the long-term trends of ozone and PM have been proposed in the scientific literature, commonly called “meteorological adjustment” of the trends.

The main aim of the work presented here is to evaluate whether a statistical method applied by the US-EPA to adjust for the effect of interannual variations in meteorology on surface ozone trends could be applied for European monitoring sites. The basic method was published by Camalier et al. (2007) and results (maps and time series) are presented at EPA’s web page on an annual basis: <https://www.epa.gov/air-trends/trends-ozone-adjusted-weather-conditions>.

In principle this and similar methods could be used to adjust for the effect of interannual variations in meteorology both with respect to the long-term trends as well as for the evaluation of ozone in a single year.

2 Methodology

The basis for the present study is a statistical method used by US-EPA relating surface ozone levels to meteorological data as presented by Camalier et al. (2007). The method could be regarded as an advanced version of a multiple regression method in which various meteorological parameters are used as input explanatory variables for predicting daily ozone levels as outlined in more detail below.

A major assumption for this kind of statistical method is that the surface ozone level could be related to the local meteorological conditions without consideration of precursor emissions and large-scale dynamics. In the EPA study, the effect of atmospheric long-range transport is included through the trajectory length and direction only.

The assumption of local relationships contrasts with the basic understanding developed the last decades. It is well known that the problem related to surface ozone pollution is indeed not a local phenomenon. Due to its chemical lifetime of one-two weeks or more in the troposphere, ozone is controlled by transboundary and even intercontinental transport in the atmosphere. Hence, the ozone being measured at a rural European site has not been formed locally. On the contrary, the ozone level is the net result of chemical and physical processes taking place during several days prior to the time of measurement.

Thus, it may seem somewhat meaningless to apply a method based on local relationships for predicting ozone. One response to this is to regard the meteorological parameters not as local variables, but rather as indicators of the large-scale dynamical field. As an example, episodes with high ozone levels in summer are typically seen during high-pressure situations

associated with elevated temperatures, gentle winds, dry conditions and a thermally stable boundary layer. In such episodes the link between ozone and local meteorological conditions is reflecting the link between ozone and the large-scale weather situation and should not be considered as indicating local formation of ozone.

2.1 The US-EPA approach

The statistical model applied in the EPA study can be written:

$$g(\mu_i) = \beta_0 + \beta_1(x_{i,1}) + \dots + \beta_j(x_{i,j}) + \dots + \beta_p(x_{i,p}) + W_d + Y_k + \varepsilon_i \quad (1)$$

where

the subscript i denotes the i th day's observation, j indicates the j th independent explanatory variable, $j = 1, \dots, p$, and k denotes the k th year.

$g(\mu_i)$ = the "link" function and μ_i the daily ozone values. A log function ($g(\mu_i) = \log(\mu_i)$) or simply a unity function ($g(\mu_i) = \mu_i$) is typically used for air pollution statistics. In the EPA study they used a log-link function.

μ_i = the maximum daily 8-h running mean ozone (MDA8 O₃) values

β_j = smooth response functions linking explanatory variables $x_{i,j}$ like temperature, humidity etc. to the dependent variable (MDA8 O₃)

$x_{i,j}$ = the value of the explanatory variable j at day i

W_d = the d th day of the week, where $d = 1, 2, \dots, 7$

Y_k = the meteorologically adjusted annual ozone value

ε_i = the error term containing the residuals

A main difference between Eq. (1) and a multiple linear regression (MLR) model, is that in the former β_j are general and smooth functions of the explanatory variables and not simply linear functions $\beta_j x_{i,j}$ as in an MLR.

The EPA study was focused on 39 urban regions in the eastern US over the period 1997-2005 using the months May-September only. The basis of their study was surface ozone monitoring data taken from EPA's database and aggregated into MDA8 O₃ values for a large number of stations. These data were then merged into a single time series for each of the 39 predefined regions by picking the highest MDA8 O₃ value each day for the sites within each urban region.

They then obtained meteorological observational data for a large number of meteorological parameters at nearly 700 stations from the US national databases. These data were also (as for the ozone data) merged into a single time series of daily values for each of the 39 regions by picking the meteorological station data closest to the geographical centre of each urban region.

In addition to ozone and meteorological data, they also calculated 1-day back trajectories for each of the urban regions for the whole period using the HYSPLIT model (Draxler and Hess, 1997).

According to Camalier et al. (2007), they then applied a generalized linear regression model (GLM) to relate the ozone data to the data for meteorology and trajectories for each of the 39 regions, separately. Based on the description in their paper, it is clear though that the model they used was actually a generalized additive model (GAM) and not a GLM.

In the EPA study they started out with a large number of meteorological explanatory variables. By a statistical screening procedure, they reduced this to eight variables that were then used in all their subsequent calculations:

1. Daily maximum temperature.
2. Mid-day average relative humidity.
3. Morning average wind speed.
4. Afternoon average wind speed.
5. Morning surface temperature difference (925 hPa – surface).
6. Deviation of morning temperature at 850 hPa from the 10-year monthly mean.
7. Transport direction (based on Hysplit trajectories).
8. Transport distance (based on Hysplit trajectories).

In their paper, however, the discussion is on the effect of temperature, humidity, transport direction and transport distance whereas the effect of the other four variables are not mentioned, presumably since their importance are small.

2.2 Our approach and comparison with the EPA model

The purpose of our study was to evaluate the performance and validity of the EPA method when applied to European data. Several important differences between our study and the EPA work should be noted, though.

Firstly, since we run the GAM for many hundreds of individual sites and not for 39 merged areas as done by EPA, we have of the order of 10-20 times more time series and regression equations to handle.

Secondly, whereas the EPA study was focused on urban areas, we have run the statistics on rural and suburban data. This is an important difference since the degree of local relationships between meteorology and ozone that such a statistical approach relies on will likely differ for rural vs. urban locations.

Finally, our period of 21 years (1990-2010) is substantially longer than the period studied by EPA (1997-2005) although we split the study into the two 11-year periods 1990-2000 and 2000-2010 by reasons explained in more detail in Chapter 3.3.

We use a slightly revised version of the GAM equation compared to Eq. (1). Our mathematical equation is given in Eq. (2)

$$G(\mu_i) = \beta_0 + \beta_1(x_{i,1}) + \beta_2(x_{i,2}) + \dots + \beta_6(x_{i,6}) + \beta_7 x_{i,7} + \varepsilon_i, \quad (2)$$

with our list of explanatory variables (the $x_{i,j}$) given in Table 2.1. As indicated by Table 2.1 and Eq. (2) there are several differences between our approach and the GAM model used in the EPA study. Each of these differences is explained in detail in the following sub-chapters.

Table 2.1 List of explanatory variables used in the GAM for the present study.

	Associated explanatory variable	Function type
x_1	Daily max temperature	Smooth
x_2	Daily mean relative humidity	Smooth
x_3	Daily mean global radiation	Smooth
x_4	Daily mean 10 m wind speed	Smooth
x_5	Daily mean PBL height	Smooth
x_6	Day number in season (120 = 1 May each year)	Smooth
x_7	Continuous time in fraction of years (0.0 = 1 Jan at start of period). This is the trend term.	Linear

2.2.1 Difference in meteorological parameters

We did not have access to all the meteorological parameters that were used initially in the EPA study, and we don't consider this a major issue since it turned out in their study that just a few of those parameters were statistically significant. Based on the experience from that study and the availability of meteorological data we selected five meteorological explanatory variables as given in Table 2.1: temperature, relative humidity, global radiation, surface wind speed and planetary boundary layer height. These explanatory variables constitute the terms $\beta_1(x_{i,1}), \dots, \beta_5(x_{i,5})$ in Eq. (2).

We did not have access to upper-air meteorological data (925 hPa and 850 hPa) as were used in the EPA study and thus we couldn't use their parameters number 5 and 6 as listed above. However, since these parameters and their possible influence are not discussed at all by Camalier et al. (2007), we assume they are of minor importance.

Next, since we did not have access to hourly meteorological data, we could not calculate averages over parts of the day (like mid-day average, afternoon average etc.) as in the EPA study. This, we admit, could have some effect on the GAM performance, for example regarding relative humidity.

2.2.2 Seasonal parameter instead of weekend parameter

We did not include the weekday term (W_d) as given in Eq. (1) since our monitoring stations were either rural or suburban and thus we expect that the effect of weekend versus week day should be minor compared to the EPA study where they looked only at urban data.

Instead of the weekday parameter, we introduced a term to consider the systematic *seasonal* variation in ozone. This is the term $\beta_6(x_{i,6})$ in Eq. (2). This is similar in type to the weekday term in EPA's method in the sense that it corrects for the mean development in the O_3 levels

within each year. The reason for including this term is that the mean seasonal variation in O_3 at a monitoring site is caused by other processes than the short-term episodes that are driven by meteorology (and emissions). Without separating these two effects in Eq. (2) we would get a more “blurred” GAM model. The seasonal term typically reflects the mean seasonal variation in the hemispheric baseline ozone level, i.e. what could be called the European boundary condition.

2.2.3 Difference in long-term trend parameter

The meteorologically adjusted ozone value each year is a main output of the US-EPA method, as given by the variable Y_k in Eq (1). This value could be considered the seasonal mean ozone level (May-September) that would have been observed if the meteorology (temperature, relative humidity, radiation etc) that year had followed the average seasonal cycle. Thus, ideally, the Y_k variable reflects a mean ozone value that is determined by physio-chemical processes except meteorological fluctuations. This would typically be precursor emissions, mixing from the stratosphere, hemispheric baseline ozone, etc.

It is, however, not completely clear from the Camalier et al. (2007) paper how their Y_k variable was determined from their Eq. (1). They don’t describe this in their paper and it is not obvious how these annual values could be inferred from their GAM model.

To estimate a mean seasonal ozone level when the effect of daily meteorological perturbations on ozone have been subtracted, was a main aim for us as well. This implies that for each day in the season we regard the observed ozone level as a net result of baseline ozone, down-mixing of stratospheric air, photochemical ozone formation from precursors in addition to the effect of meteorological fluctuations on ozone. When looking at a time period of several years, these meteorologically adjusted ozone values constitutes a trend that is not influenced by meteorological perturbations.

We tested various approaches for including the trend term in our GAM model, including a way that seems closest to Eq. (1). In this approach we allowed the Y_k term to be a freely varying variable but with a fixed value each year implying that Y_k was just a fixed annual value independent on the day of season and that Y_k was allowed to vary freely from one year to another, giving us 11 free variables for a period of 2000-2010. This turned out to produce not an optimal solution. The performance of that GAM version was considerably poorer than the GAM model we ended up with (Eq. (2)) and, additionally, produced certain artefacts with Y_k values showing unphysical high variations from one year to the next.

We ended up using an assumption of a linear trend for the MDA8 O_3 summer values. This is reflected in Eq. (2) where $x_{i,7}$ denotes the time given in fraction of years since 1 Jan the first year in the period, and β_7 is forced to be a linear slope.

Thus, the term $\beta_7 x_{i,7}$ in Eq. (2) expresses the long term linear trend in the MDA8 O_3 summer values corrected for the interannual variations in meteorology at each site. The β_7 coefficient is a constant value, representing the slope in this trend. The $\beta_0 + \beta_7 x_{i,7}$ part on the right-hand side of Eq. (2) is then identical to a standard linear regression with intercept β_0 and slope β_7 and where both are constant coefficients. The other terms in Eq. (2) represent the variations due to meteorology or background seasonal cycle, and the $\beta_j(x_{i,j})$ terms are smooth unspecified functions to be estimated from data as indicated in Table 2.1.

The assumption of a linear slope in the meteorologically adjusted ozone levels constitutes a significant restriction of the GAM model as opposed to letting the trend term simply be a smooth unspecified function of time as for the other explanatory variables. It could be argued, however, that for limited periods, like our 11-year periods, this assumption is justified, on the grounds that we are more interested here in the long-term average trend coefficient for these years, rather than any individual variation during this time in the underlying trend curve.

2.2.4 Difference in link function

The term $g(\mu_i)$ on the left-hand side of Eqs. (1) and (2) is called the link function as mentioned above. In the EPA study they used a log-link function, i.e. $g(\mu_i) = \log(\mu_i)$, where μ_i is the value of MDA8 O₃ at day i .

We started out our work with a log-link function as well. However, a closer inspection of the GAM model residuals at the individual sites revealed that such a log transformation in many cases was a too strong transformation to use for our data. Ideally, residuals should be symmetric (without skewness) and not too far from a Gaussian distribution for best performance of GAM models. Based on our evaluation, a unit link function $g(\mu_i) = \mu_i$ in most cases gave a better result than a log-link. A bonus of not using a transformation is also that the interpretation of the results of the GAM model becomes simpler.

2.2.5 Trajectory data not included

The EPA method included air mass back trajectory data as part of the input explanatory variables in their statistical calculations (Eq. 1). They used the trajectory path for the 24 h prior to arrival and converted those data into two explanatory variables: transport length and transport direction.

In our study, the time did not allow for the computation of back trajectories to every station for the 21-year period. However, since pre-calculated trajectories from the Flextra model (Stohl et al., 2001) were already available for certain sites and periods, we looked at the effect of including such data for a few selected sites, namely Illmitz in Austria (AT02), Chaumont in Switzerland (CH02) and Harwell in UK (GB36). For these sites we had 5-days back trajectories for the period 1997-2010 available and for the GAM modelling we used the length of the trajectory path for the last 24 hours as an additional explanatory variable.

The direction or “origin” of the trajectory was not included, though. In technical terms, it’s not clear how to include this in the GAM model since this is a cyclic explanatory variable. In the EPA study they included this as a so-called transport direction going from 0° to 360°. It’s not obvious, however, that this kind of parameter could be included in the same way as the other meteorological data. In a geophysical sense, a value of 350° and 10° for the wind direction would be almost identical, both reflecting a northerly wind. In the GAM model, however, these values would be considered to be very different.

When we included the length of the trajectory path as an additional explanatory variable in the GAM, it turned out that the performance of the method was just slightly improved. For

the three test sites mentioned above, we didn't find any major improvement in performance. This may reflect that the daily average wind speed which already was included in the GAM is so highly correlated with the 24 h trajectory length that the extra benefit of the trajectory data is small. It is possible, though, that the air mass origin would be a more influential parameter to include in the GAM. It is, however, not clear how such a parameter could be included and so this is a topic for further investigation.

2.3 Use of GAM and openair model software

The GAM model (Eq. (2)) was fitted using the GAM library `mgcv` (Wood, 2017) in the statistical modelling system R (R Core Team, 2018) for each station and for each period 1990-2000 and 2000-2010 separately. We used all data in each period in order to estimate the long-term trend coefficient β_7 in Eq. (2) at each station. This was the basis for the estimation of the trends and the meteorology adjustments.

However, in order to see how well the GAM model performed in each separate year we made additional calculations where we excluded the data for that year, i.e. the “target year”, and fitted the GAM model (Eq. (2)) to the rest of the data (the left out years). This means that e.g. when fitting the GAM model to predict the daily ozone levels in 2008, we skipped the input data for 2008, but used the data for the other years in the 2000-2010 period to calculate the β coefficients. Then, these β coefficients were used together with the meteorological data for 2008 to predict the daily MDA8 O₃ values in 2008. It should be noted that these β coefficients would differ slightly from the β coefficients used when all 11 years were included. By this procedure, the GAM predicted MDA8 O₃ levels for any year are completely independent of the actual ozone data for that year. We believe this is a different approach compared to the method presented in the EPA paper.

As output, we calculated the β functions and their significance levels for each station/period as well as various measures of the GAM model evaluation performance such as RMSE, R² etc. For this latter part we used the `openair` library (Carslaw and Ropkins, 2012; Carslaw, 2015). As part of the output we also calculated the GAM fit, i.e. the levels of MDA8 O₃ each day predicted by the GAM as well as the slope of the meteorologically adjusted ozone trend over the period (β_7 in Eq. (2)). The results are presented in plots and tables below.

3 Input data

3.1 Surface ozone monitoring data

The basis for our study was the EDT studies that is focused on the period 1990-2010. Hourly observed ozone data for this period were extracted from the AirBase and EMEP databases. We picked the rural (EMEP and AirBase) and the suburban (AirBase) sites for our study while the urban AirBase sites were considered not well suited for our purpose. Furthermore, as explained in more detail below, we applied the GAM model on two sub-periods separately; 1990-2000 and 2000-2010. In addition, we used data for the period 1 May - 31 August only. This is the time of year with the most marked elevated ozone episodes where we expect to see the clearest dependencies between the meteorological data and ozone.

The number of ozone monitoring stations within the two databases for these sub-periods are given in Table 3.1. The number of stations used in the GAM model vary depending on the period studied since the length and onset of the ozone monitoring varies from station to station. We used the criteria that at least 75 % of the years with 75 % of daily data in the period 1 May – 31 August should be available for the station to be included in the GAM model.

Table 3.1: Databases used in the study with their corresponding number of ozone monitoring stations

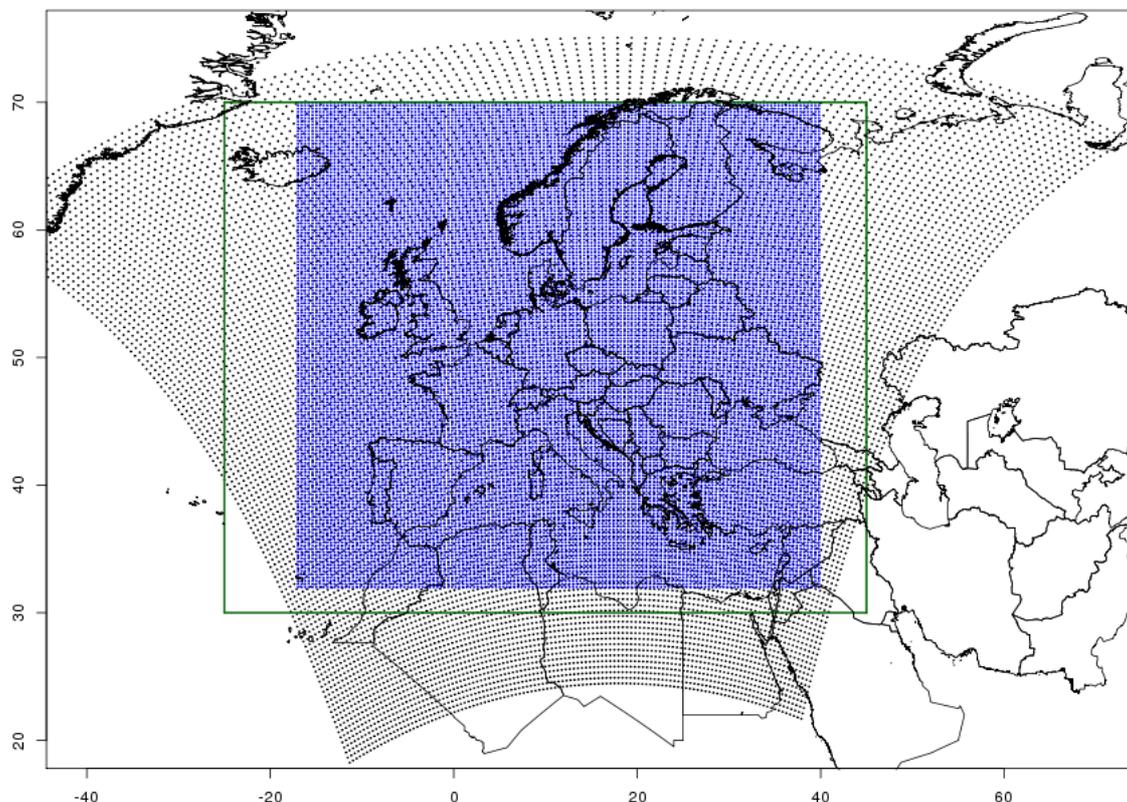
	Total number of sites	Number of sites with data in 75 % of the years within 1990-2000	Number of sites with data in 75 % of the years within 2000-2010
EMEP (rural)	202	97	122
AirBase (rural)	664	163	369
AirBase (suburban)	597	102	288

Based on the time series of hourly observed ozone concentrations we calculated daily values containing the maximum 8-h running mean concentration (MDA8 O₃) at each site and used those daily data as input to the GAM model. For the allocation of 8-h periods to the individual days, we used the same definition as in EU's ozone directive, i.e. we allocated the 8-h period to the day on which the 8-h period ended (EU, 2008). Thus, the period from 18 UT day 1 to 01 UT day 2 was allocated to day 2 and so on. We required that at least 75% of the 8 hours (i.e. 6 hourly values) should contain a real value or else it was defined as missing.

3.2 Meteorological data

Meteorological data were extracted from the meteorology used in the EDT project (Stegehuis et al., 2015) which covers the geographical domain from 30°-70° N and 25°W-45°E in a latitude-longitude grid with a grid spacing of 0.25° latitude and 0.40° longitude as shown in Figure 3.1, corresponding approximately to 25 x 25 km². Details of the set-up and design of the EDT is given by Colette et al. (2017).

Figure 3.1: Model domain in the EDT exercise shown in blue. The grey dots mark the Euro-Cordex domain whereas the green rectangle marks the model domain used for the analysis of campaigns in EURODELTA (Colette et al., 2017).



Based on the gridded fields of hourly meteorological data we calculated annual time series for every monitoring station for the period 1990-2010 containing daily values of the input parameters using either the daily mean or the daily maximum as given in Table 2.1.

In the GAM modelling, the time series of daily meteorological data were paired with the time series of MDA8 O₃ data by selecting the model grid square containing the monitoring site. Thus, no interpolation of neighbouring grid cells was applied to the model data.

3.3 Homogeneity and robustness criteria

The length of the period (i.e. the number of years) that the GAM model is applied for should be determined as a compromise between two basic criteria that could be called the homogeneity and the robustness criteria, respectively.

The robustness criterion states that the more data and the longer time series we use, the more robust the GAM results should be in an overall long-term statistical sense.

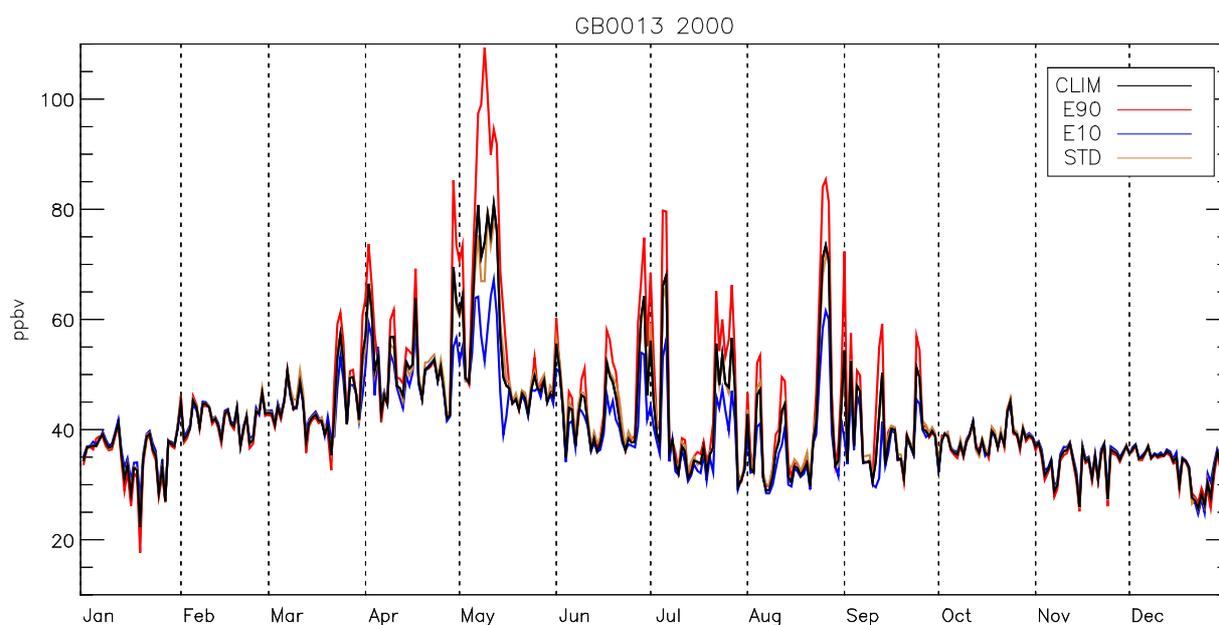
The homogeneity criterion on the other hand, states that the response of ozone with respect to the explanatory variables should be the same, or nearly the same, throughout the period considered. Since this response is determined mainly by the atmospheric levels of NO_x and

VOCs, the GAM should then only be applied to periods with fairly even levels of NO_x and VOC emissions in Europe.

Thus, for a long period such as 1990-2010, a perturbation of e.g. +5°C in the daily maximum temperature will presumably be associated with a stronger ozone response in 1990 than in 2010 due to the decline in European emissions of NO_x and NMVOC during 1990-2010. This effect is illustrated in Figure 3.2 (adopted from Colette et al., 2016) showing the modelled MDA8 O₃ levels for a site in southern England based on meteorology for 2000 and European emissions for 1990 and 2010, respectively. Figure 3.2 shows that the ozone level follows a baseline with marked episodes on top. These episodes are likely caused by photochemical episodes associated with elevated temperatures, solar radiation etc. and would thus be important for the estimation of the β functions in Eq 2.

As seen in Figure 3.2, in the summer season (April-September), the magnitude of the high ozone episodes depends strongly on the emission year. With 1990 emissions, the ozone levels peak at nearly 110 ppbv in the strongest May episode whereas with 2010 emissions this episode peaks at only 65 ppbv. Since the meteorology is identical in these model runs, the difference in calculated ozone is due to the difference in emissions between 1990 and 2010 only. This effect has been extensively documented in various papers discussing the links between climate change and tropospheric ozone levels (e.g. Colette et al., 2013).

Figure 3.2: (Adopted from Colette et al. 2016) Daily modelled MDA8 values at GB0013 (Yarner Wood, UK) in 2000. The various lines show the EMEP modelled ozone levels with 2000 meteorology based on European emissions for 1990 (E90, red line), for 2010 (E10, blue line), for 2000 (brown line, STD) as well as interpolated between E90 and E10 (black line, CLIM).



As a compromise between the homogeneity and robustness criteria, we decided to divide our study into the two periods 1990-2000 and 2000-2010. A further reason to split our study in this way is that numerous studies have revealed contrasting ozone trends in the 1990s versus the 2000s (e.g. Parrish et al., 2014; Logan et al., 2012).

3.4 From daily to annual indicators

The aim of the present work as well as of the EPA study was to use the GAM model to reduce or “subtract” the effect of meteorological variability on annual or seasonal ozone statistics. The EPA study used the GAM to predict seasonal mean MDA8 O₃ values based on the months May-September and in our study, we looked at the mean for the period May-August.

When studying long-term trends, the selection of statistics (i.e. annual mean, annual max, seasonal mean, etc.) is not a trivial choice. As shown by e.g. Simpson et al. (2014) and Lefohn et al. (2017) the long-term trends could differ significantly depending on the choice of the ozone statistics. A main message from these two studies is that changes in anthropogenic emissions lead to changes in the frequency distribution of the ozone levels, typically characterized by a reduction in the highest and lowest concentrations leading to a narrowing of the overall probability density function as a response to a reduction in precursor emissions. These two studies, as well as previous trend studies for EEA (Colette et al., 2016; Colette et al., 2015; Solberg et al., 2015), have shown that the strongest trend signal is seen for the high percentiles of ozone whereas less significant changes are seen for the medium range levels of ozone. Thus, an annual statistic like SOMO35 that is mostly determined by the intermediate ozone levels, will likely show a very different trend than the number of peak episodes (like the exceedance of certain peak thresholds like 180 µg/m³, etc.).

In the present study the aim was to investigate whether the US-EPA method could be applied for European rural and suburban areas. Thus, we kept the methodology as close as possible to the original work, meaning that we looked at the summer (1 May - 31 August) seasonal mean of the MDA8 O₃ values. For looking at trends in high peak values, other statistical methods, more suited than the GAM presented here, are needed.

4 Results

In the following, the main results from the GAM statistical modelling work are presented. The GAM was run for hundreds of individual stations and for several periods and data sets (rural and suburban), thus producing several thousand plots and result files. In the following we provide an overview of the results with geographical maps, examples of time series and some tabulated summary statistics.

Our study raises several questions that are discussed in the following, such as:

1. To what extent is the GAM methodology applicable for adjusting for the year-to-year variability in meteorology?
2. Are there any geographical patterns in the performance of the GAM method?
3. What is the importance of the various explanatory variables?
4. How do the meteorologically adjusted trends look like?
5. Do they differ from the trends computed by simple linear regression?

4.1 Performance of the GAM method

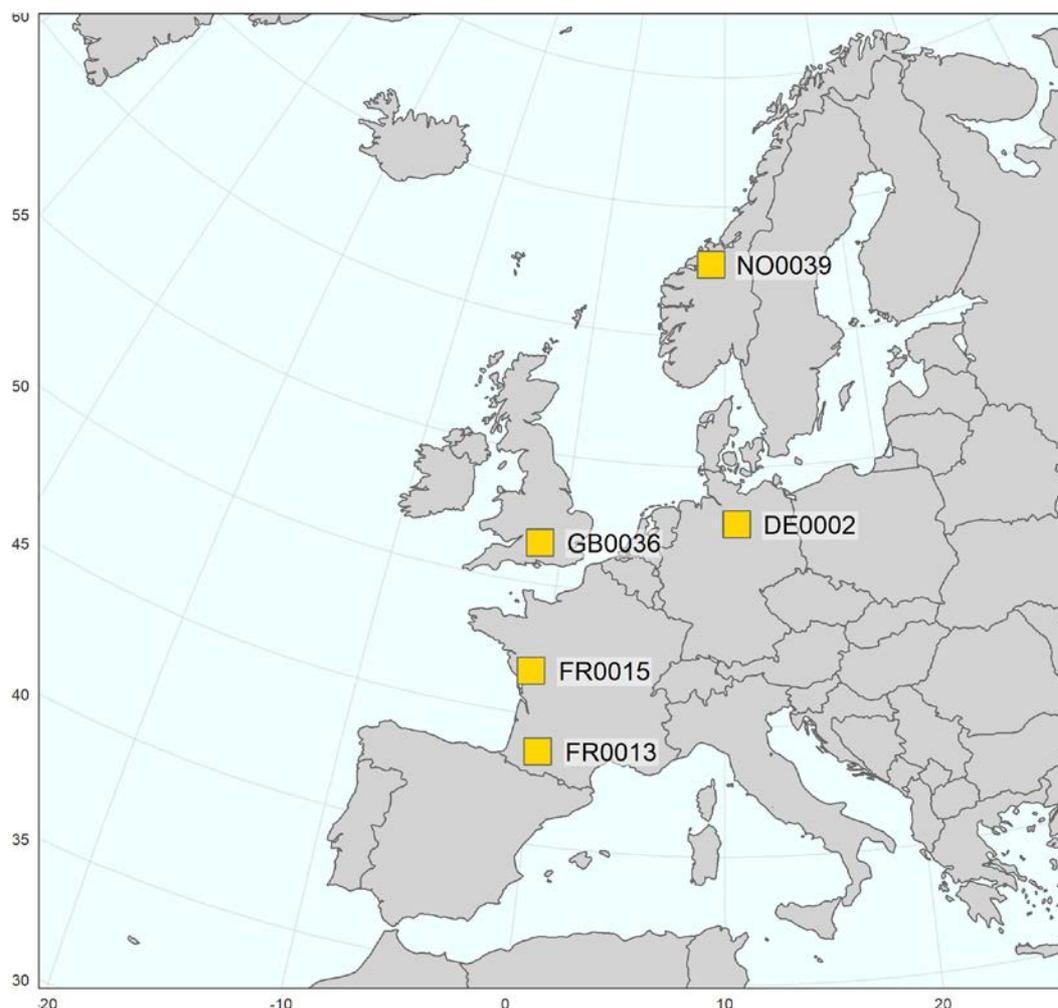
Our main procedure for evaluating the GAM performance is by comparison of the MDA8 O₃ predicted by the GAM method each day with the corresponding measured MDA8 O₃ levels for the summer season defined above, i.e. 1 May - 31 August, station by station. This comparison was done by visual inspection of time series plots and frequency distribution plots, as well as tabulation of standard performance statistics.

Although the main aim of the GAM was to calculate seasonal mean levels (May-August), the performance of the method on a day-to-day basis is crucial for the trust in the method. Without the GAM being able to reproduce the day-to-day variability in MDA8 O₃ to a certain level, the seasonal mean levels would be of little value. The question of the overall applicability of the GAM method could be seen as a response to question 1 above.

Given the large number of stations and long-time series included in our study, we can only show individual results for a few example sites/years in addition to overall tabulated statistics. In the following we show examples of ozone time series and histograms of concentration distributions. The selection of example sites/years will have to be subjectively chosen. We picked well-known monitoring stations representing various geographical and chemical regimes, namely the following as also shown in Figure 4.1:

- Waldhof (DE0002, Northern Germany)
- Harwell (GB0036, Southern England)
- Kårvatn (NO0039, Western Norway)
- Peyrusse Vieille (FR0013, Southwest France)
- La Tardiere (FR0015, Western France)

Figure 4.1: Map showing the location of the example sites discussed below.

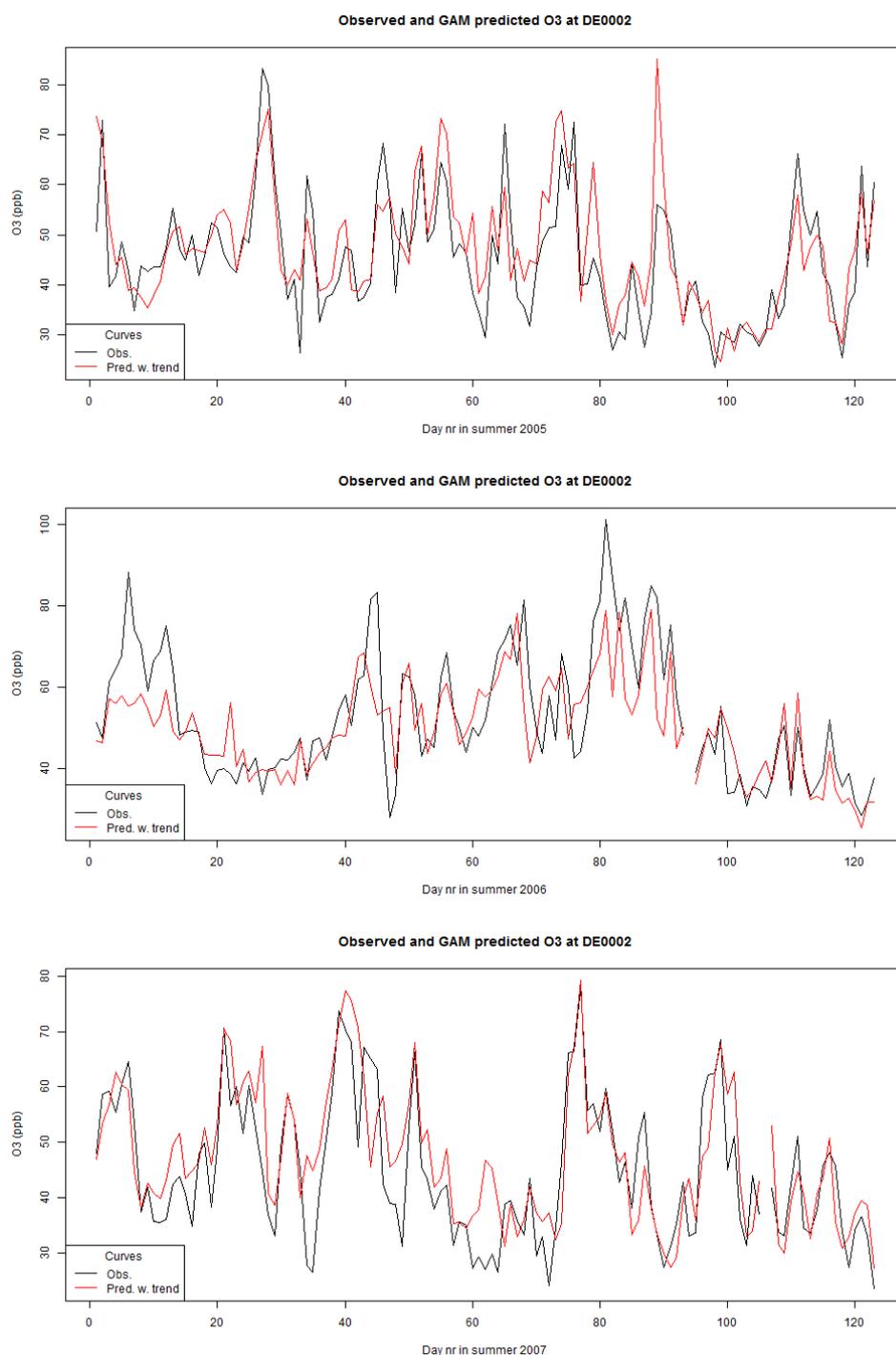


4.1.1 Results for Waldhof (DE0002)

Figure 4.2 shows the measured MDA8 O₃ values together with the levels predicted by the GAM model for the EMEP site Waldhof (DE0002) in Northern Germany for the years 2005-2007. Waldhof is a very well-established monitoring station having records of ozone, VOCs and many other key pollutants for decades. Previous assessments have shown that Waldhof is well suited as a background rural reference site and that the influence from local emissions are minor.

As Figure 4.2 indicates, the GAM method performs very well when compared to the measurements at this station for the given years. Both the general levels as well as the episodes are predicted to a very satisfactory degree.

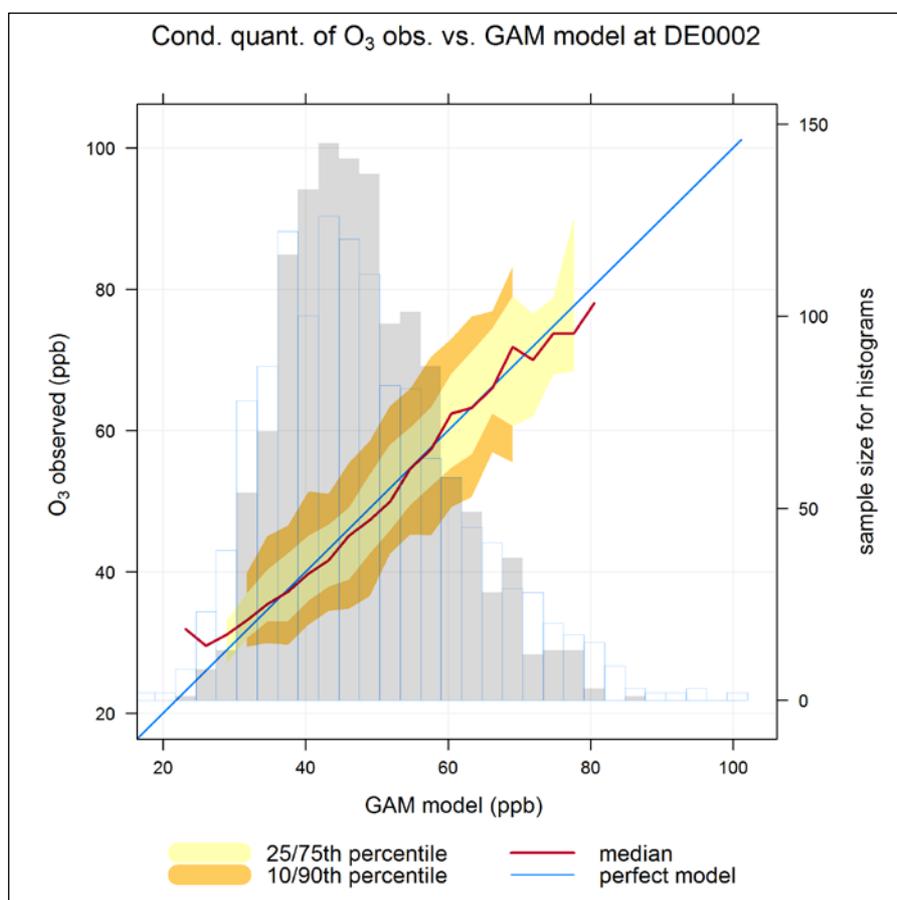
Figure 4.2: Time series of daily max 8-h values of O₃ at Waldhof (DE0002) in Northern Germany as measured (black) and predicted by the GAM method (red). Data for May-August 2005-2007 are shown.



A so-called conditional quantile plot for the period 2000-2010 is shown in Figure 4.3 (Carslaw and Ropkins, 2012; Carslaw, 2015). This plot splits the observations into evenly spaced bins. For each bin, the corresponding values of model predictions are identified and the median and 25/75 percentiles and 10/90 percentiles are calculated for that bin. The data are then plotted to show how these values vary across all bins. For a time-series of observations and predictions that agree perfectly, the median value of the predictions will

equal that of the observations for each bin. Note that the conditional quantile plot differs from the traditional quantile-quantile (Q-Q) plot in that the Q-Q plot considers only quantiles of the overall distributions of observed and predicted values, whereas the conditional quantile plot uses corresponding observations for each interval of predictions. The conditional quantile plot thus shows how well predictions agree with simultaneous observations.

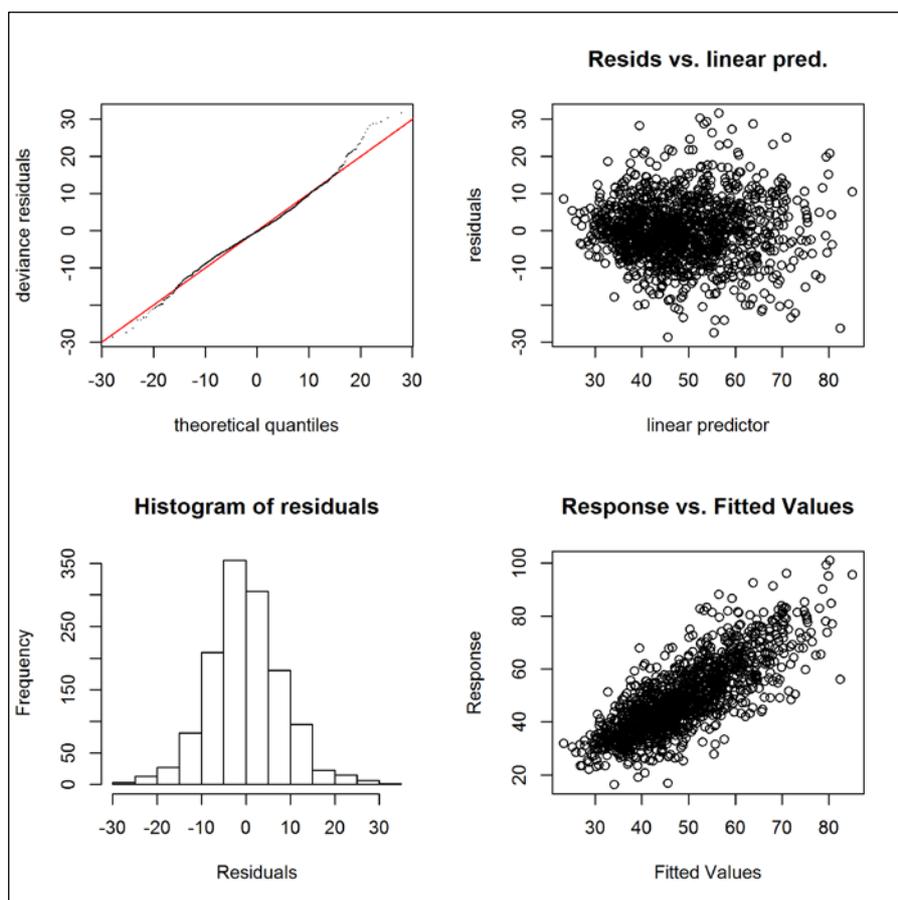
Figure 4.3: Conditional quantile plot for Waldhof (DE0002) for 2000-2010 for the performance of the GAM method. The dark red curve marks the median of the GAM predicted values in each bin of the observed values and the light and dark yellow fields mark the corresponding 25/75th and 10/90th percentiles of the GAM predictions. The 45° line and the underlying histogram in light blue refers to the observed values whereas the grey shaded area marks the histogram of the GAM predicted values. See text for further explanations.



As seen from this figure, the GAM predictions (dark red curve) agrees very well with the observations (light blue 45° line). It seems though (from the histogram) that the highest and lowest peak values are slightly underrepresented by the GAM, and thus the predicted distribution of values show a narrower distribution than the observed values.

The GAM check plot (Figure 4.4) confirms that the GAM method performs very well over the period 2000-2010 at Waldhof. The scatter plot and the histogram of the residuals don't show any clear skewness. This indicates unbiased residuals (the ε_i values in Eq. (2)) in the GAM model.

Figure 4.4: GAM check plot for Waldhof (DE0002) for 2000-2010. The upper left panel shows the GAM deviance residuals against theoretical quantiles of the deviance residual distribution, according to the fitted model. The ideal 1:1 line is given as the red line. The upper right panel shows a scatter plot of the GAM residuals, i.e. the ε_i values in Eq. (2), versus GAM predicted MDA8 O₃ values. Ideally these residuals should not show any systematic dependency on the predicted values. The lower left panel shows the histogram of the residuals. Ideally this should show a perfect symmetric Gaussian like distribution. The lower right panel shows the observed ozone values (y-axis) vs. the values predicted by the GAM (x-axis). Ideally, they should follow a 45° line.



4.1.2 Results for Harwell (GB0036)

Ozone time series from Harwell (GB0036) 1994-1996 are shown in Figure 4.5. Harwell, located in South England, is another long-term running rural monitoring station, somewhat more exposed to nearby emissions than Waldhof, though. Furthermore, this site should presumably be somewhat more dependent on long-range transport and thus harder to predict by the GAM than Waldhof. The measured time series in Figure 4.5 show high peak values up to 100 ppb for these years. Many of the high peaks are reflected by the GAM predictions, although at a generally lower level than observed.

The conditional quantile plot (Figure 4.6) confirms that the highest peak values are underestimated by the GAM model. The occurrence of low ozone levels is also underestimated and thus the GAM predictions show a narrower frequency distribution than observed. The scatter plot of the residuals in Figure 4.7 indicates some bias (residuals not evenly distributed) compared to the observations.

Figure 4.5: Time series of daily max 8-h values of O₃ at Harwell (GB0036) in Southern England as measured (black) and predicted by the GAM method (red). Data for May-August 1994-1996 are shown.

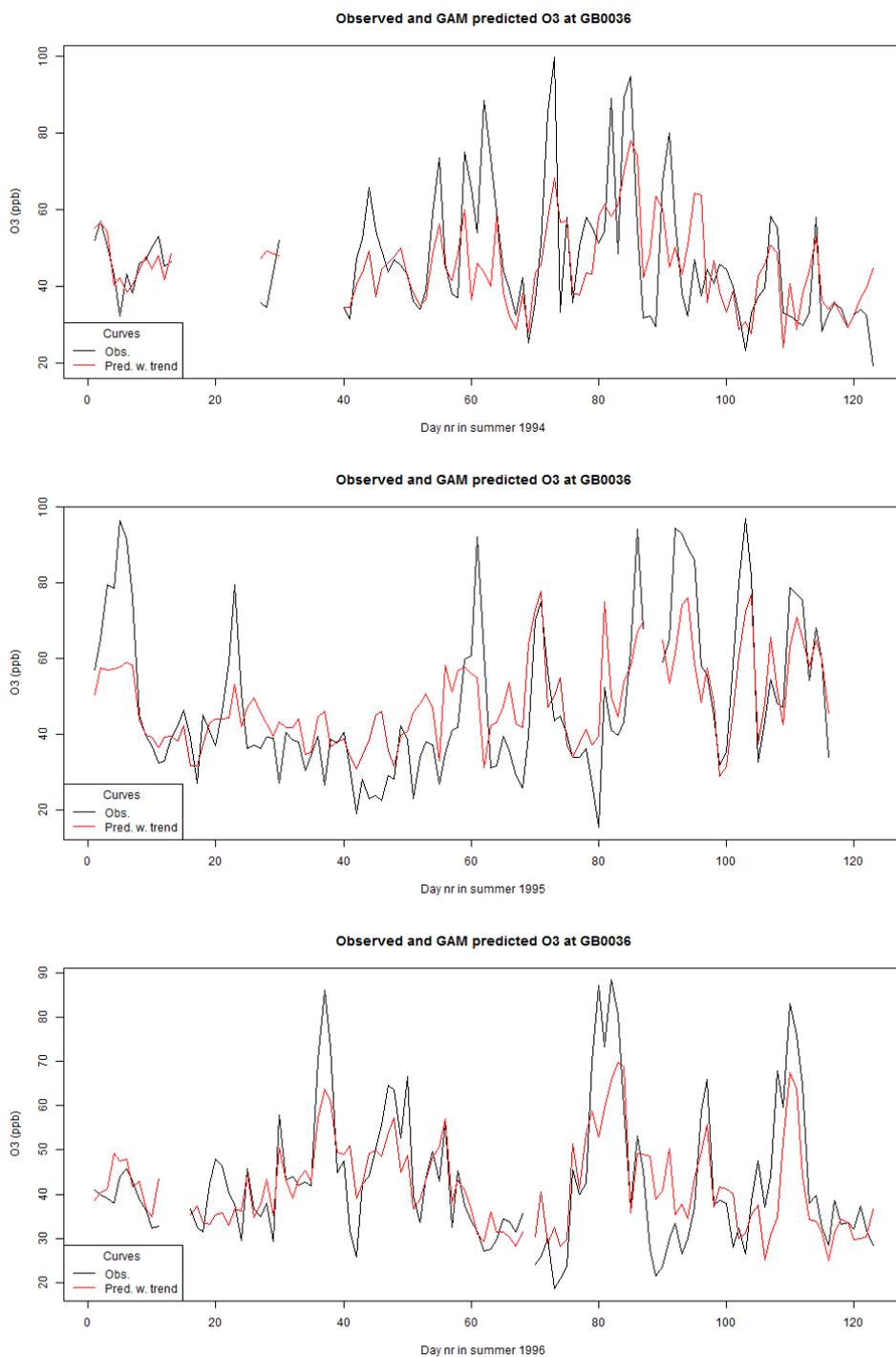


Figure 4.6: Conditional quantile plot for Harwell (G00B36) for 1990-2000 for the performance of the GAM method. See Figure 4.3 and the main text for explanations of the plot.

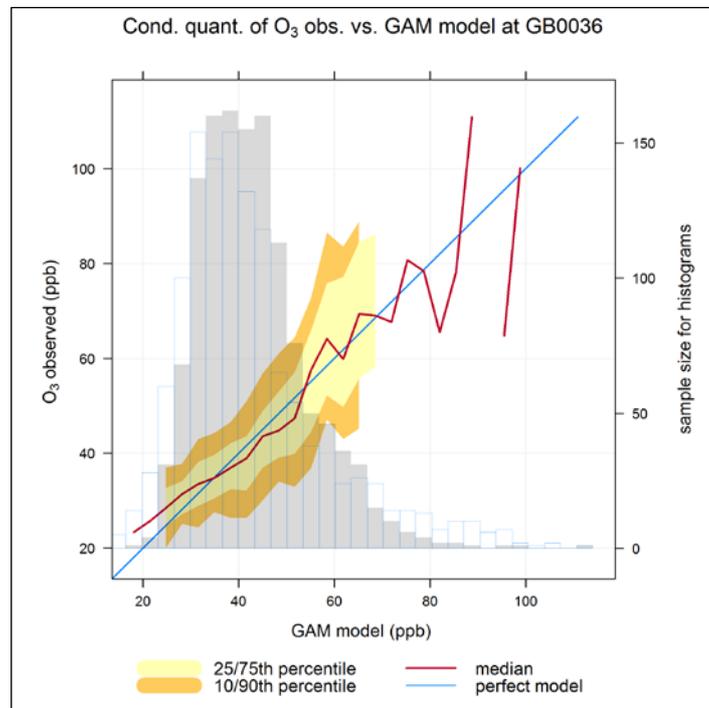
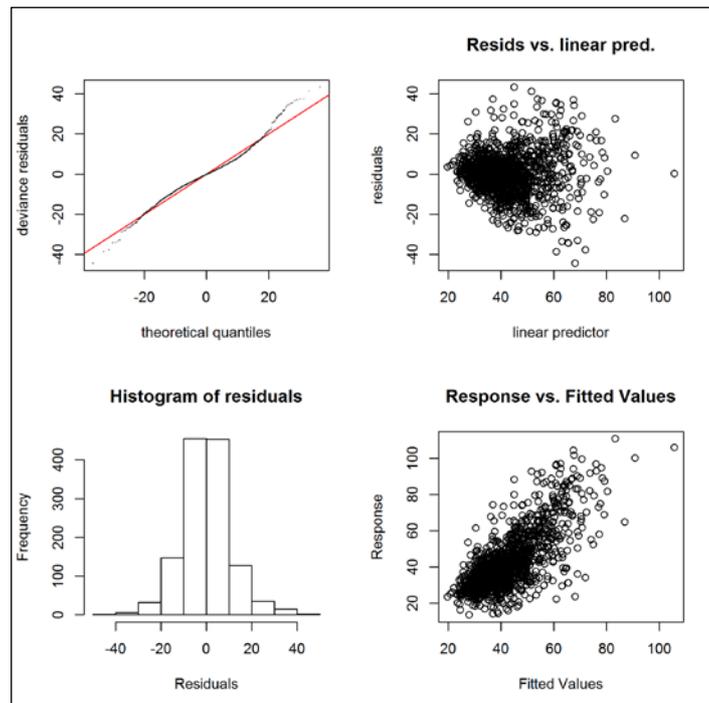


Figure 4.7: GAM check plot for Harwell (GB0036) for 1990-2000. See Figure 4.4 for explanations of the plot.



4.1.3 Results for Kårvatn (NO0039)

An even more challenging test of the GAM method is the application to the site of Kårvatn (NO0039), a background rural site in Western Norway, far away from the main European emission areas. In this region, we expect that the surface ozone levels are completely dependent on the long-range transport of pollutants from the continent and that any links between local meteorology and observed ozone is simply reflecting that high ozone episodes are linked to weather situations with transport of warm air masses from central Europe. The results in Figure 4.8 show that for 1994-1996 the GAM method strongly underestimates the highest peak episodes at this site, while it captures many of the less extreme peaks very well.

The conditional quantile plot (Figure 4.9) indicates that the predicted median levels match the observed ones closely whereas the histogram of the predicted values are narrower than observed, confirming that the GAM model is not able to predict the highest and lowest ozone levels at this site.

This is also confirmed by the GAM check plot (Figure 4.10) showing an underestimation of the high peak levels and an overestimation of the low peak levels.

Figure 4.8: Time series of daily max 8-h values of O₃ at K arvatn (NO0039) in mid-Norway as measured (black) and predicted by the GAM method (red). Data for May-August 1994-1996 are shown.

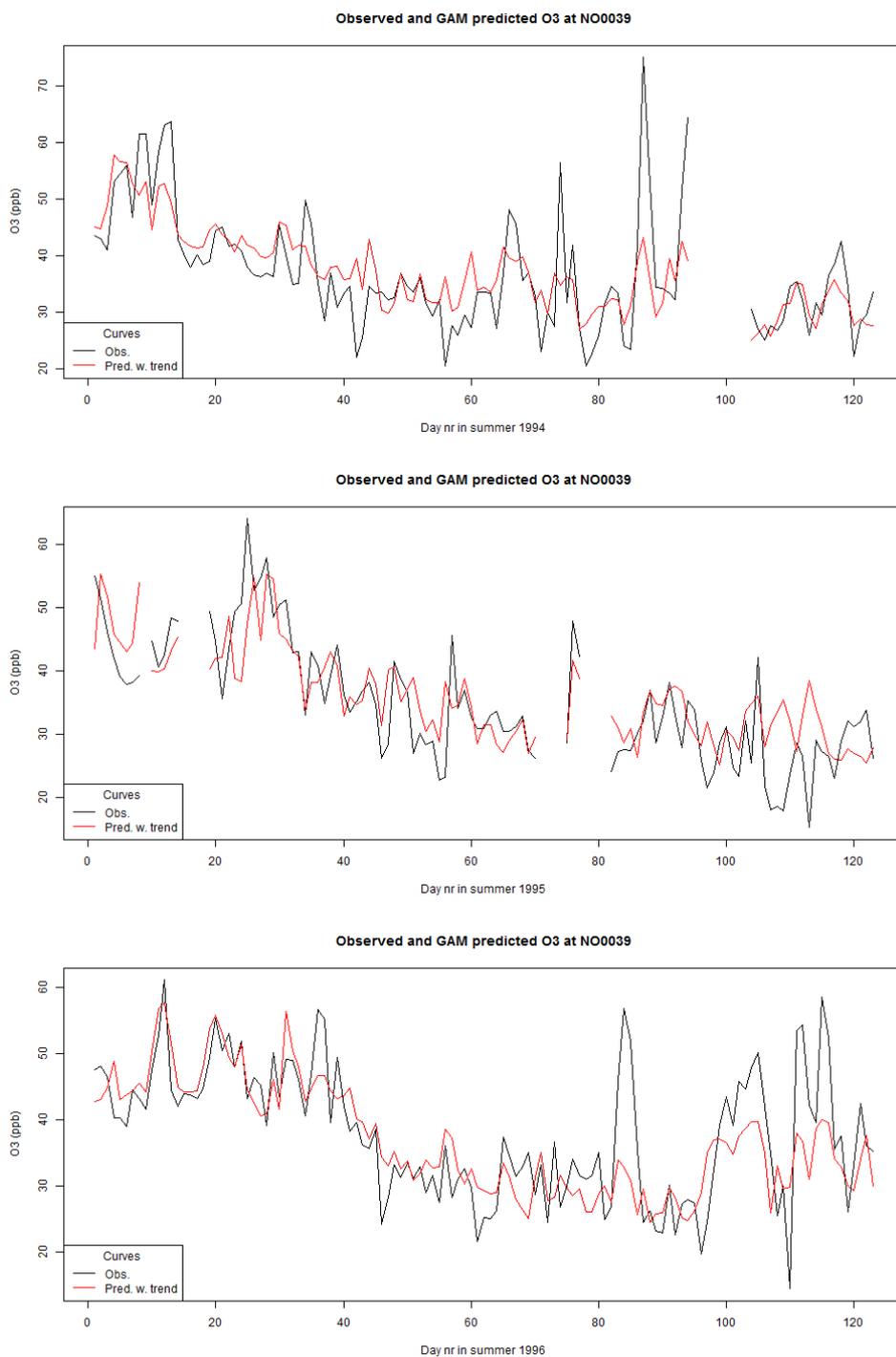


Figure 4.9: Conditional quantile plot for Kårvatn (NO0039) for 1990-2000 for the performance of the GAM method. See Figure 4.3 and the main text for explanations of the plot.

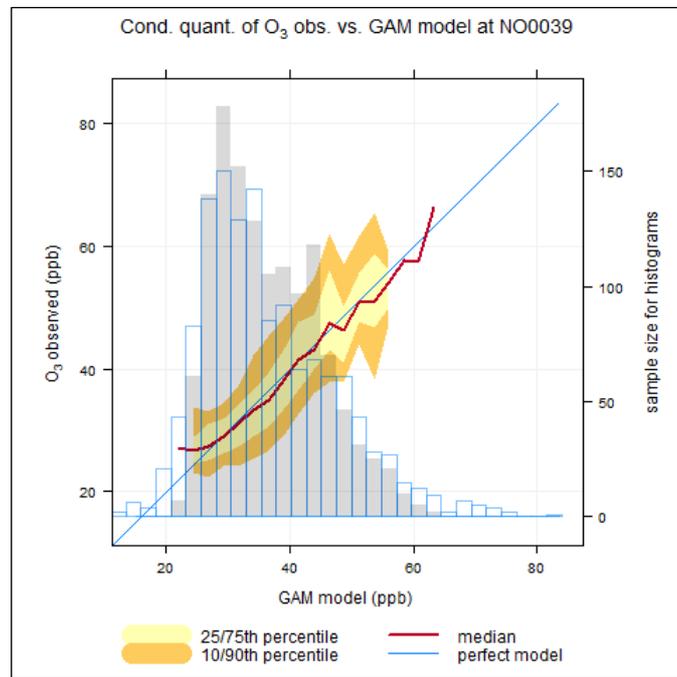
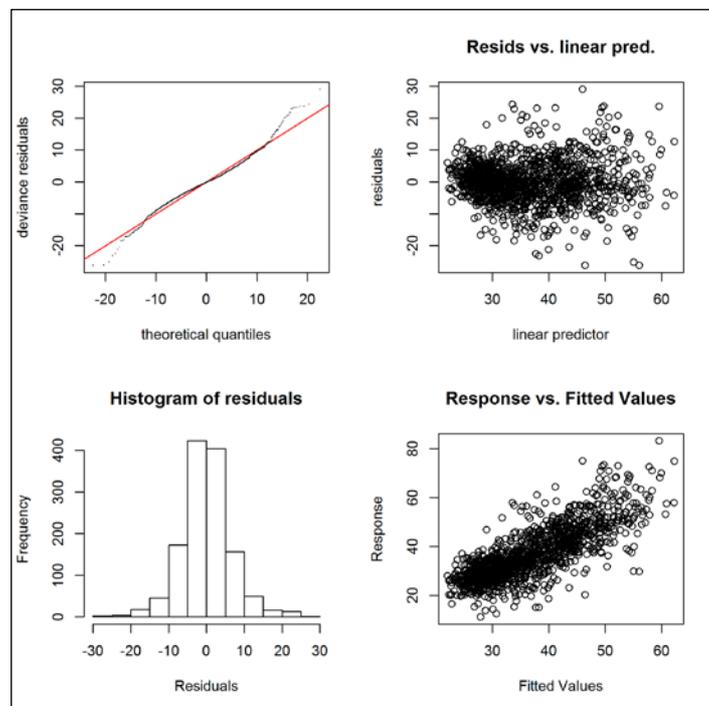


Figure 4.10: GAM check plot for Kårvatn (NO0039) for 1990-2000. See Figure 4.4 for explanations of the plot.



4.1.4 Results for Peyrusse Vieille (FR0013) and La Tardiere (FR0015)

Figure 4.11 and Figure 4.12 shows the time series of daily ozone at two French background rural sites in 2003-2005. Both stations are in Western France some distance inland from the Bay of Biscay: Peyrusse Vieille (FR0013) in the southern part ($43^{\circ} 37' N$) and La Tardiere (FR0015) further north ($46^{\circ} 39' N$). We have included these two sites to highlight the differences in O_3 levels and predictions over a not too long distance (~ 330 km).

These results show clear differences in performance of the GAM method between the sites, with better agreement with the measurements at La Tardiere ($R^2 = 0.56$) compared to Peyrusse Vieille ($R^2 = 0.43$) in the south. Whereas many of the high peaks at La Tardiere are very well predicted by the GAM, they are significantly underestimated at Peyrusse Vieille.

The conditional quantile plots (Figure 4.13) for these two sites confirm a much better agreement between the GAM model and the observations for FR0015 than for FR0013 as seen from the histograms although the median values (red lines) agree fairly well for both sites. This is further seen from the GAM check plots (Figure 4.14).

Figure 4.11: Time series of daily max 8-h values of O₃ at Peyrusse Vieille (FR0013) in southwest France as measured (black) and predicted by the GAM method (red). Data for May-August 2003-2005 are shown.

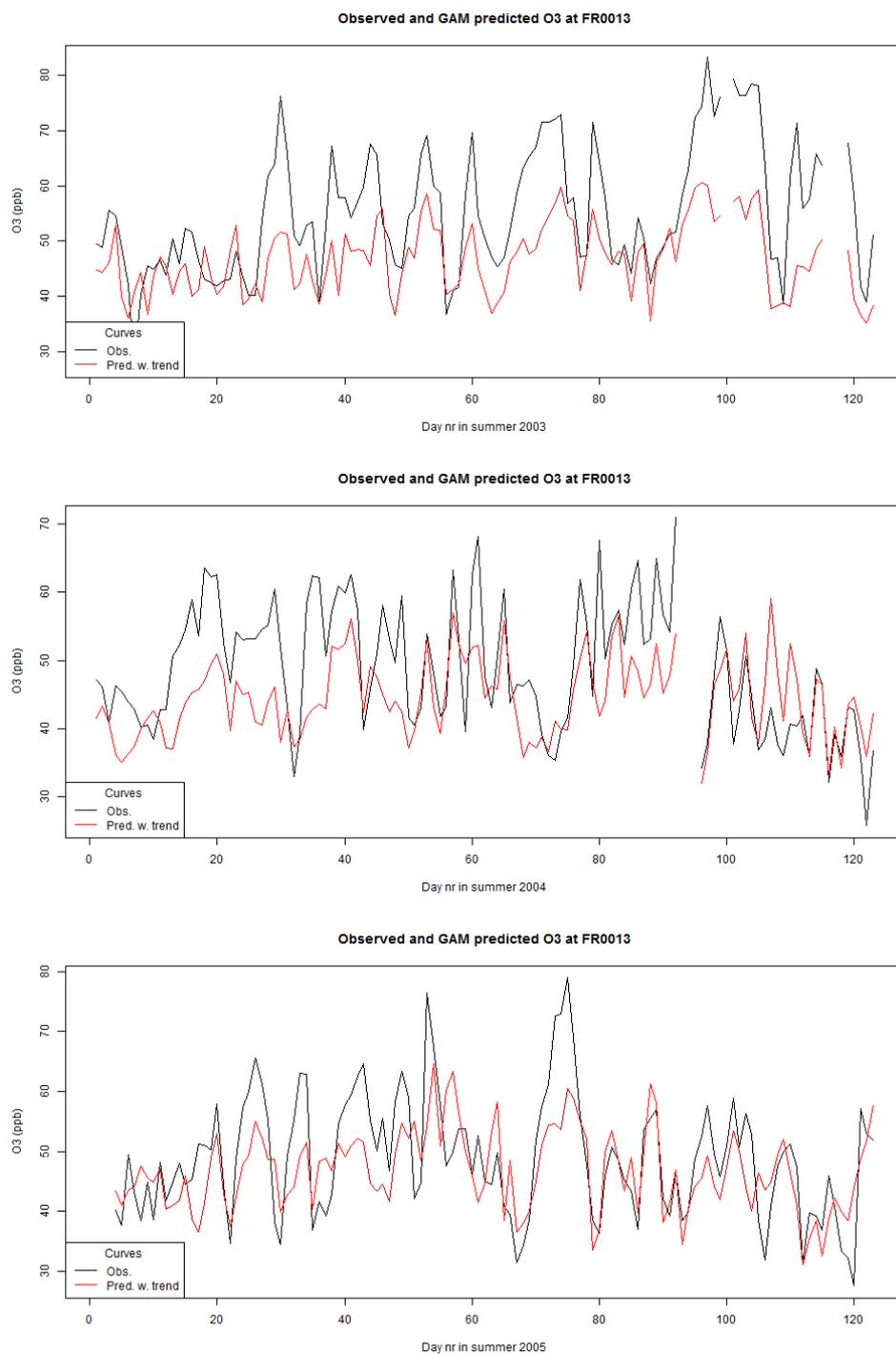


Figure 4.12: Time series of daily max 8-h values of O₃ at La Tardiere (FR0015) in western France as measured and predicted by the GAM method. Data for May-August 2003-2005 are shown.

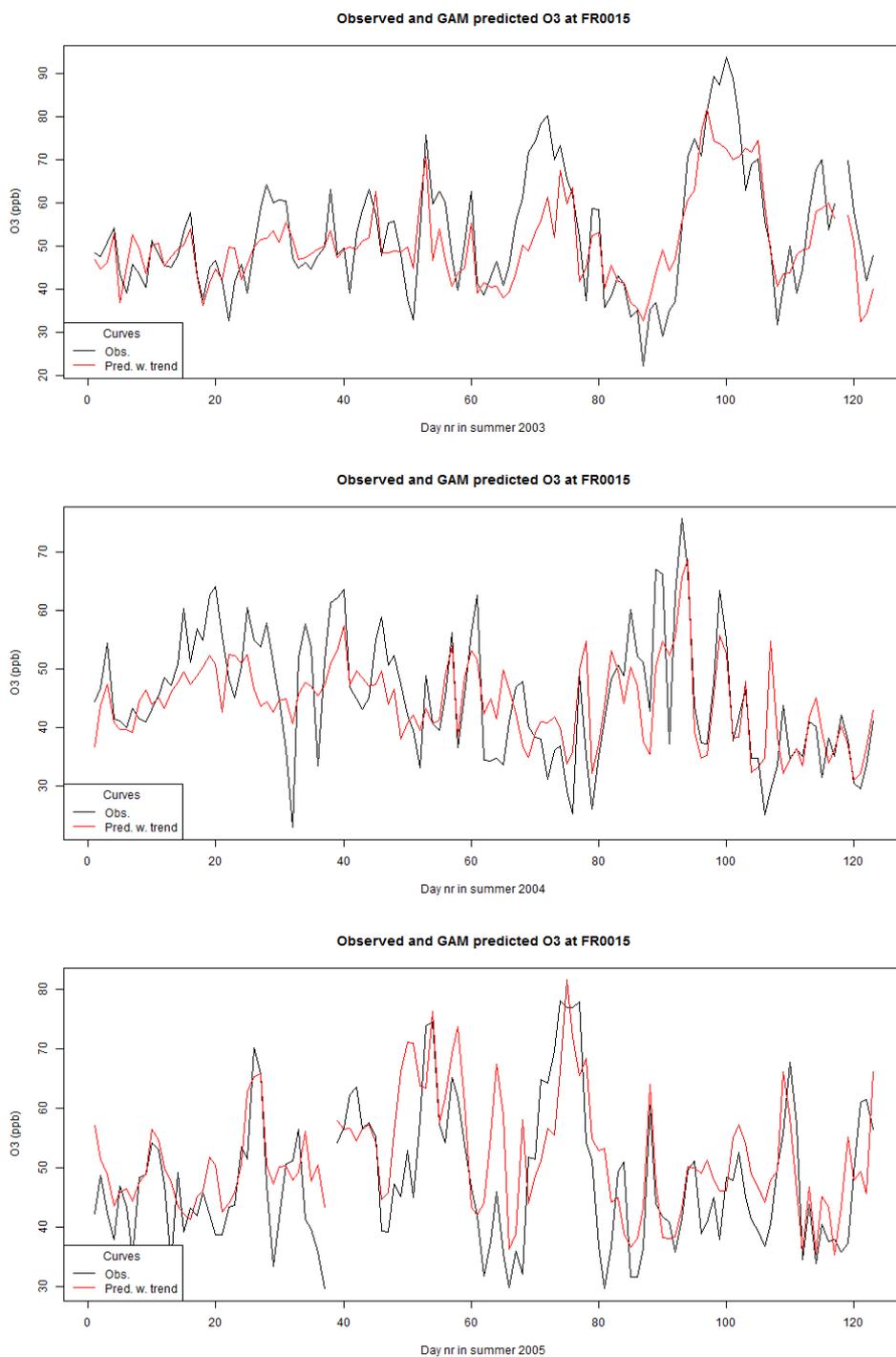


Figure 4.13: Conditional quantile plot for Peyrusse Vieille (FR0013, left) and La Tardiere (FR0015, right) for 2000-2010 for the performance of the GAM method. See Figure 4.3 and the main text for explanations of the plot.

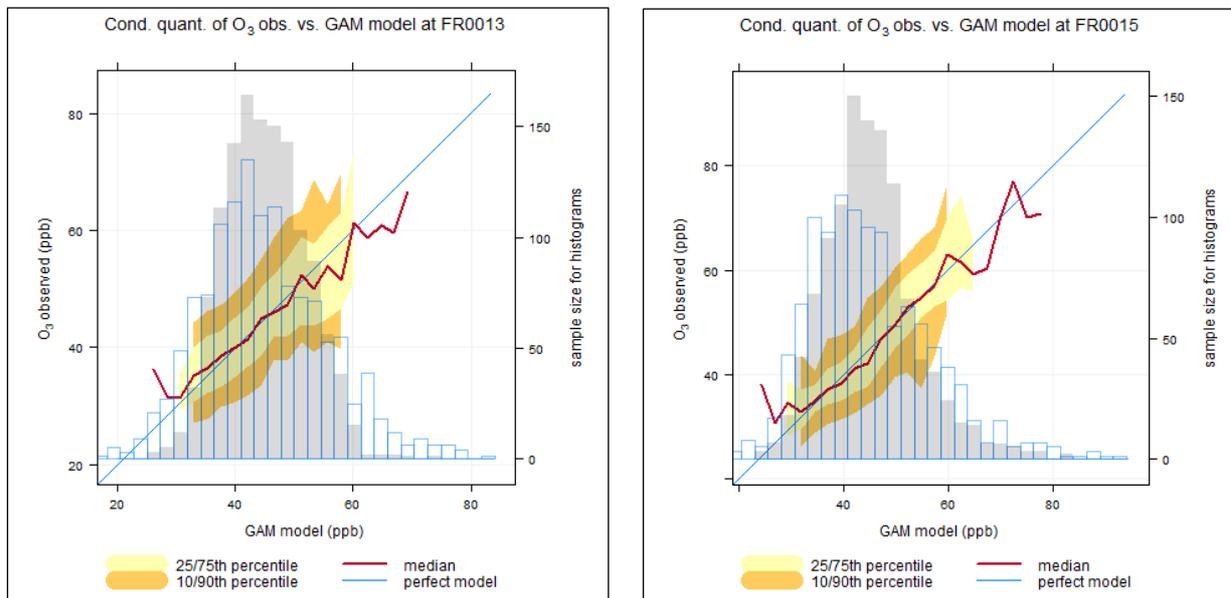
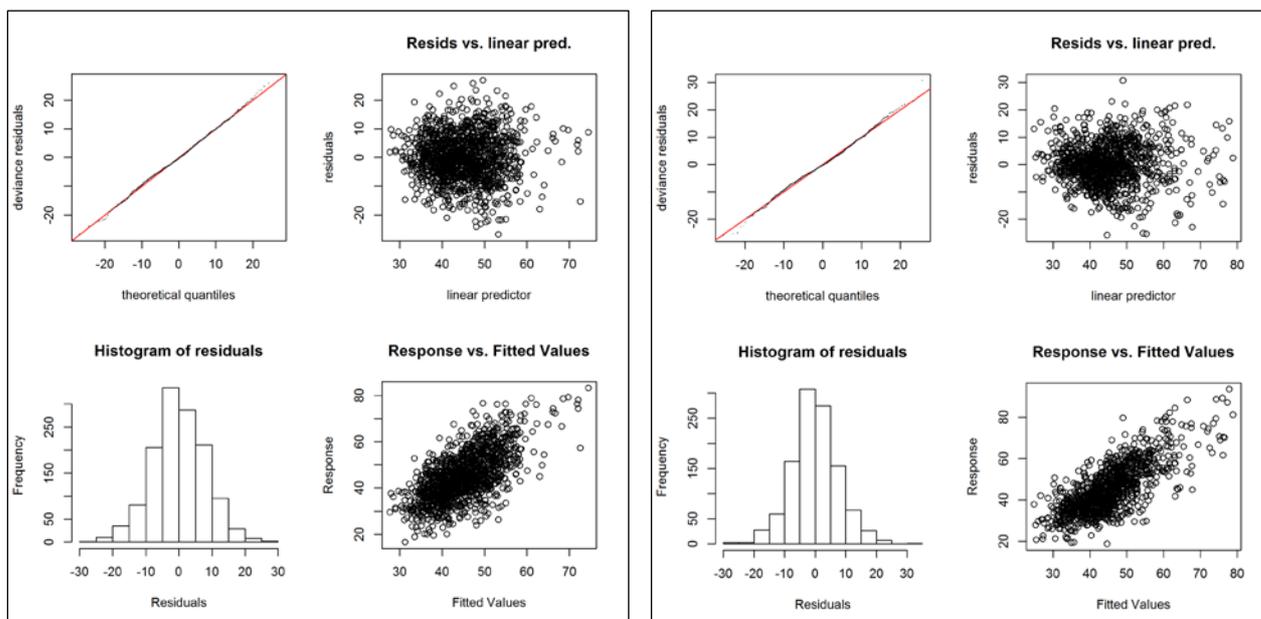


Figure 4.14: GAM check plot for Peyrusse Vieille (FR0013, left) and La Tardiere (FR0015, right) for 2000-2010. See Figure 4.4 for explanations of the plot.



4.1.5 Overall statistics

The discussion above has shown how the GAM method performs for certain sites as examples. To summarize the performance, we calculated various statistics averaged over all monitoring sites for each period, respectively. The list of these statistics is given in Table 4.1. R2 measures the overall predictive power of the GAM method, whereas the other statistics could be used to evaluate the correlation, bias etc.

Table 4.1: GAM model performance statistics

Abbreviation	Explanation
R2	Coefficient of determination (R squared)
RMSE	Root mean square error
COR	Linear correlation
IOA	Index of agreement
MB	Mean bias
NMB	Normalized mean bias

Table 4.2 and Table 4.3 presents these statistics for MDA8 O₃ for the two time-periods (1990-2000 and 1990-2010) for each of the three data sets (EMEP rural, AirBase rural, and AirBase suburban) when averaged over all sites within the individual periods. Note that the AirBase rural and EMEP rural data sets contain many of the same sites.

The results in Table 4.2 and Table 4.3 show that the mean R2 statistic is of the order of 0.50, and that it is slightly higher for the AirBase suburban sites vs. the EMEP and AirBase rural data. This value is considerably lower than what was found for the EPA study for the Eastern US where they calculated an R2 ranging from 0.56 to 0.80. This may reflect that a GAM method with its underlying assumptions of local relationships between O₃ and meteorology is better suited for urban regions in the Eastern US than for rural and suburban locations in Europe. It could also reflect differences in the design of the GAM method in our approach compared to the EPA approach as outlined above (Ch. 2.2). The difference in looking at many individual sites as we have done compared to applying the GAM to merged data for larger areas as done by EPA could be important. Another reason for the higher R2 in the EPA study could be the use of trajectory data in their calculations.

It should be noted that the statistical measures of bias (MB and NMB in Table 4.2 and Table 4.3) are very low, which is a consequence of this kind of statistical method. Since the GAM model is tuned to each site separately, the bias will almost by definition be close to zero.

Table 4.2: Mean values of selected statistics for the performance of the GAM method for the period 1990-2000.

	EMEP rural (n=72)	AirBase rural (n=115)	AirBase suburban (n=82)
R2	0.483	0.510	0.527
RMSE	9.013	10.327	10.921
COR	0.642	0.666	0.684
IOA	0.632	0.643	0.651
MB	0.051	0.024	0.079
NMB	0.001	0.000	0.002

Table 4.3: Same as Table 4.2 for the period 2000-2010

	EMEP rural (n=104)	AirBase rural (n=302)	AirBase suburban (n=210)
R2	0.483	0.518	0.520
RMSE	7.998	8.613	9.445
COR	0.638	0.671	0.671
IOA	0.630	0.644	0.642
MB	0.041	0.028	0.028
NMB	0.001	0.001	0.001

These statistics and the examples of time series shown above indicate that the performance of the GAM model varies. For many sites, we see a good to very good agreement between observed and predicted daily data, while for other sites the GAM fails to reproduce the day-to-day variability. The background seasonal cycle of ozone seems to be well captured through the inclusion of the day-of-season input explanatory variable in Eq. (2). The timing of the high O₃ episodes are mostly reflected in the GAM model, although often at too low levels. Typically, the predicted values show a too narrow distribution meaning that the highest observed levels are underestimated, and the lowest observed levels are overestimated by the GAM method (in other words the number of the highest and lowest levels are both underestimated). This type of results is often seen also in chemical transport models.

Table 4.4 shows how the GAM performs for the seasonal mean values versus for a high percentile, in this case the seasonal 90 percentiles. These results confirm that the GAM

performs clearly better for the seasonal mean values than for the high percentiles. As seen from Table 4.4, the GAM model underestimates the level of the 90 percentiles for both EMEP and AirBase data, and furthermore, the correlation with the observed 90 percentiles are clearly poorer than for the seasonal mean values.

Table 4.4: Overall statistics based on all sites for the period 2000-2010 showing the mean of the observed and GAM predicted seasonal mean MDA8 O₃ (mean obs. and mean GAM); the mean of the observed and GAM predicted seasonal 90 percentile MDA8 O₃ (P₉₀ obs. and P₉₀ GAM); and the linear correlation of these means and P₉₀s (r (mean) and r (P₉₀)). Unit: ppb.

	Mean obs.	Mean GAM	P ₉₀ obs.	P ₉₀ GAM	r (mean)	r (P ₉₀)
EMEP rural	46.2	47.2	59.1	56.4	0.47	0.39
AirBase rural	47.0	48.2	61.7	59.0	0.53	0.44
AirBase suburban	46.6	48.0	62.4	60.0	0.53	0.44

4.2 Spatial differences in GAM performance

The following is a response to the question of geographical differences as raised in question 2 above. We present maps with summary statistics and results and have separated the discussion for the two time-periods 2000-2010 and 1990-2000 below¹.

Furthermore, we have merged the EMEP rural and AirBase rural data. These data sets are to a certain extent identical since many of the same sites are reported to AirBase and to EMEP. To some extent the EMEP rural data could be regarded as a sub-set of the AirBase. Since regular reporting of O₃ data to AirBase didn't start until 1997, however, many of the rural sites are lacking in the first part of the 1990s in AirBase.

4.2.1 The period 2000-2010

Maps of R², RMSE, and mean bias (MB) for the GAM method vs. rural and suburban measurement data for the period 2000-2010 are presented in Figure 4.15 - Figure 4.17. These maps reveal marked spatial differences in the GAM performance. The best performance is found in northern and central Germany, central France and parts of the Nordic countries (high R², MB near zero and low RMSE) in agreement with the time series at Waldhof and La Tardiere shown above. R² is clearly lower in Spain, southern France, Austria, Slovakia and the Baltic countries. RMSE is particularly high in Italy, presumably also reflecting that the average ozone level is higher in that region.

¹ Period 2000-2010 is shown first because there are much more stations and data for this period than for 1990-2000 and thus more relevant results.

Figure 4.15: R2 for the GAM method for EMEP and AirBase rural sites (left) and for the AirBase suburban sites (right) for the 2000-2010 period.

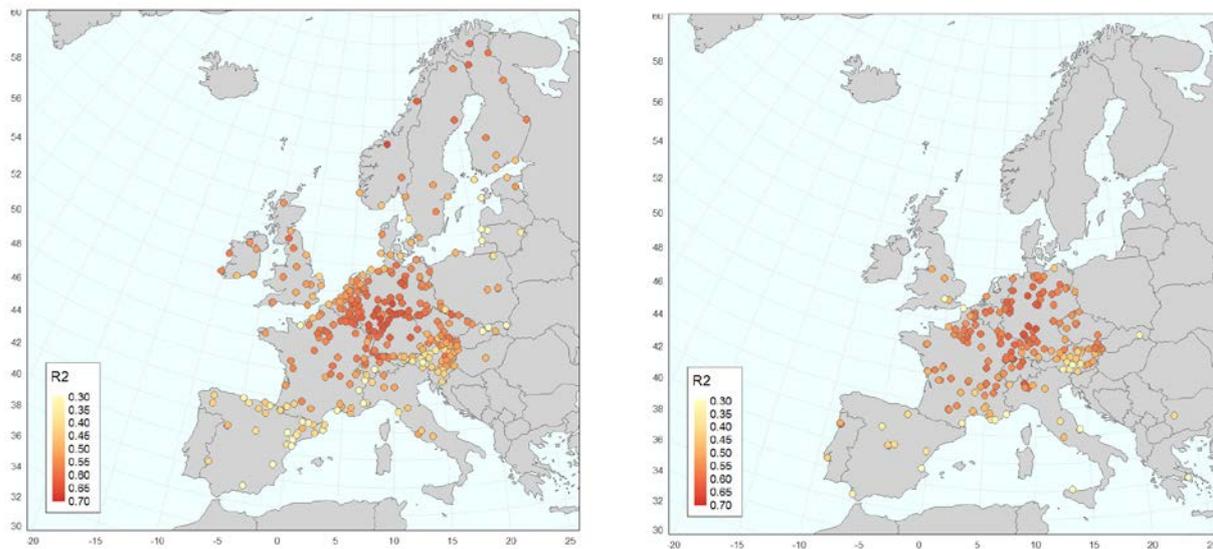


Figure 4.16: RMSE for the GAM method for EMEP and AirBase rural sites (left) and for the AirBase suburban sites (right) for the 2000-2010 period.

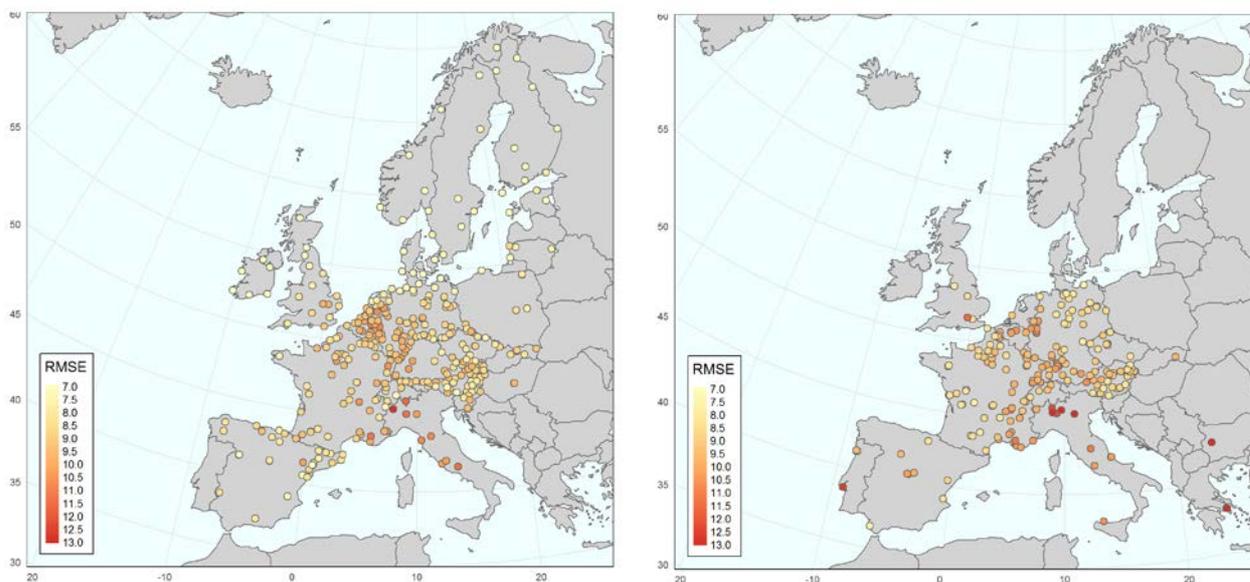
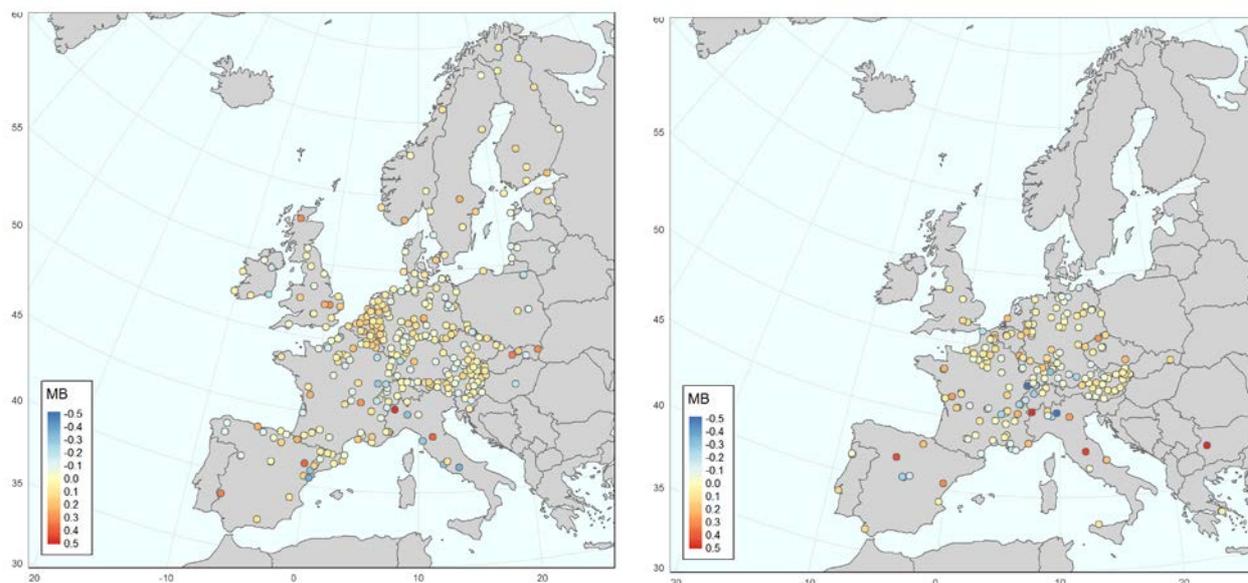


Figure 4.17: Mean bias (MB) for the GAM method for EMEP and AirBase rural sites (left) and for the AirBase suburban sites (right) for the 2000-2010 period.



4.2.2 The period 1990-2000

The number of stations with data for the period 1990-2000 is considerably less than in the later decade, and the spatial coverage is also much poorer. As seen from the maps in Figure 4.18 - Figure 4.20, large parts of the European continent lack monitoring stations fulfilling the criteria of at least 75 % data capture. Anyway, the region with best performance of the GAM method seem to be in similar areas as for 2000-2010, namely Germany, Belgium and parts of the Nordic countries. Lower R2 scores are seen in Austria and parts of Switzerland, pointing to a poorer performance in these regions.

Figure 4.18: R2 for the GAM method for EMEP and AirBase rural sites (left) and AirBase suburban sites (right) for the 1990-2000 period.

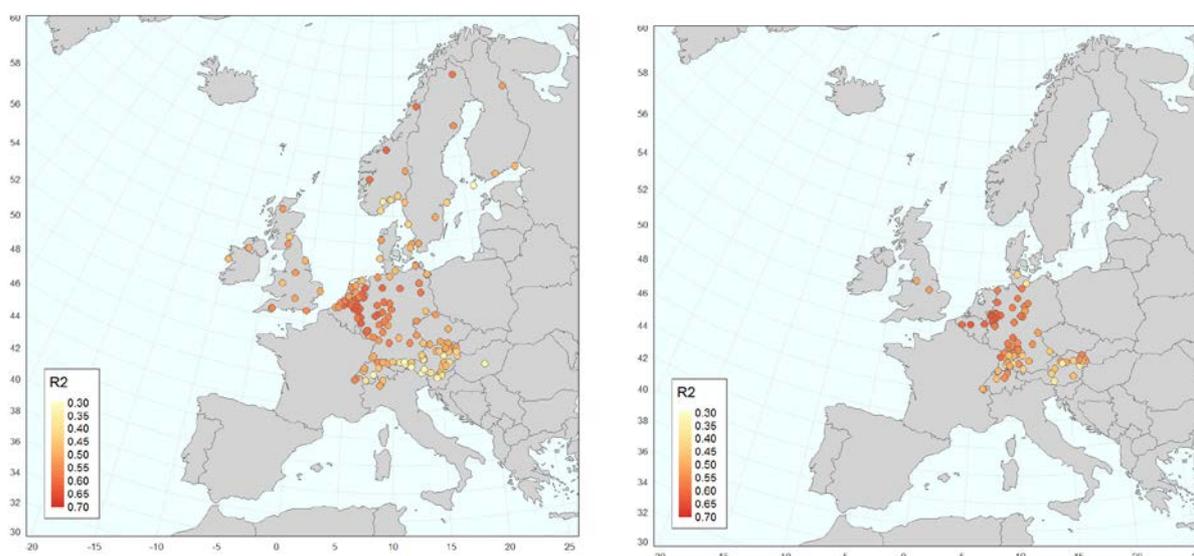


Figure 4.19: RMSE for the GAM method for EMEP and AirBase rural sites (left) and for the AirBase suburban sites (right) for the 1990-2000 period.

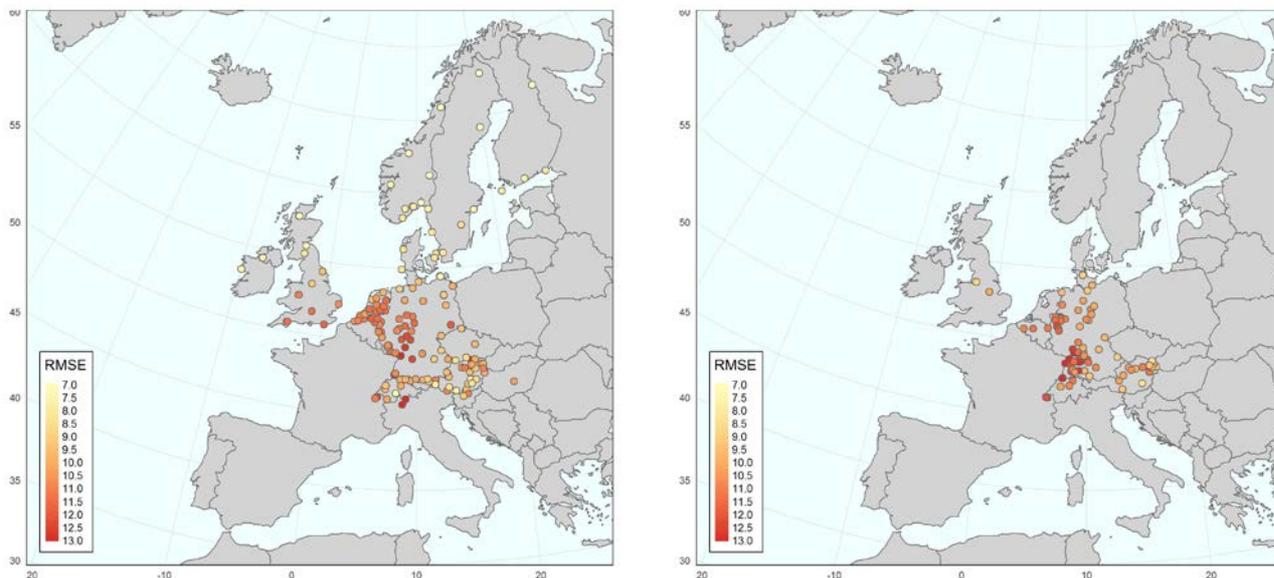
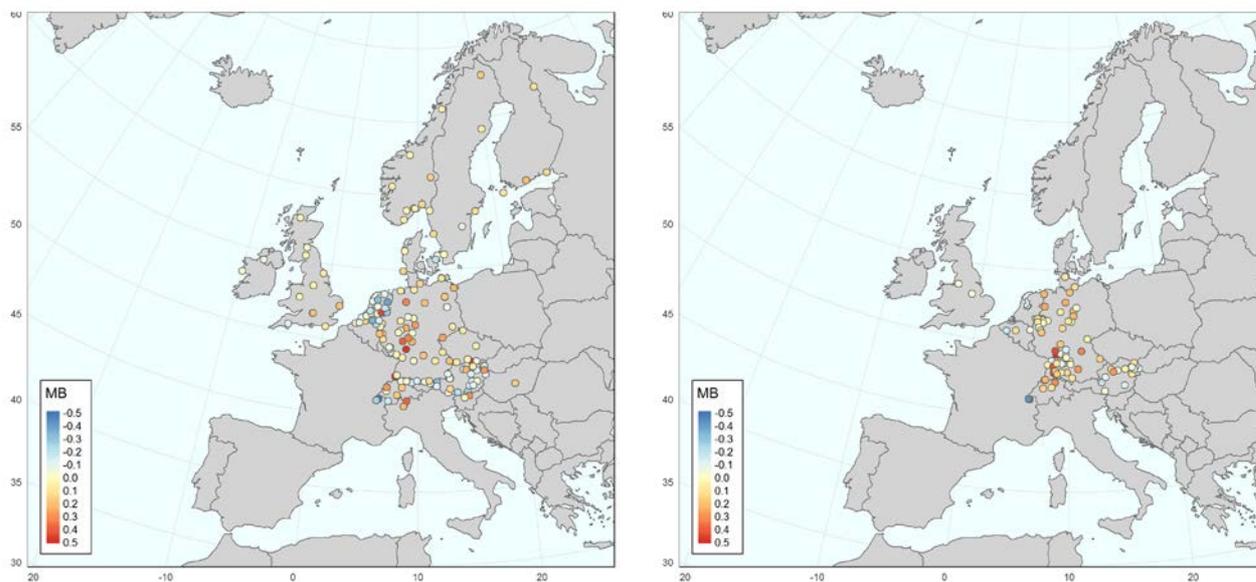


Figure 4.20: Mean bias (MB) for the GAM method for EMEP and AirBase rural sites (left) and for the AirBase suburban sites (right) for the 1990-2000 period.



4.3 Importance of the explanatory variables

As a response to question 3 above, we present statistics for the importance of the various explanatory variables in the following. As above, this is done for the two periods 1990-2000 and 2000-2010, separately. To save space only the combined EMEP and AirBase rural stations are shown. It turned out that the spatial pattern of these relationships was similar for the suburban stations.

As explained above, we used smooth β_j functions for the explanatory variables $j=1, \dots, 6$ in Eq. (2). An example of these functions for Waldhof (DE0002) for the period 2000-2010 is given in Figure 4.21. The curves show the partial responses of ozone on the y-axis to each of the individual explanatory variables on the x-axis based on daily data in the summer season (May-August) for the given period. The lines in the diagrams mark the GAM method's best fit smooth functions, i.e. the β_j functions in Eq. (2). The shaded areas represent uncertainty regions, i.e. ± 2 standard errors confidence regions, for the underlying fitted smooth curves. Note that these plots are normalised, i.e. having an overall mean value of zero for the partial derivatives for each variable. Furthermore, the curves shown in Figure 4.21 represent the normalised perturbations when assuming mean values for all the other variables. One could thus regard these curves as an expression of how each variable influences the ozone levels as a function of time of year keeping all the other variables constant. For example the downward slope of day of season just implies that keeping all other variables constant at their mean value, the day of season tend to lead to an increase in ozone in the start of the period and a reduction in ozone in the last part.

It turned out that daily max temperature (*Tem2*) is the strongest predictor of the explanatory variables, and the shape seen in the example in Figure 4.21 with a marked positive response between *Tem2* and O_3 is seen at many stations. In the example in Figure 4.21, a linear relationship could seem appropriate whereas for many sites the response curve for *Tem2* typically shows a flat (no response) line for the lower temperatures and a very strong increase for higher temperatures, signifying the non-linearity of the ozone-temperature relationship.

The example given in Figure 4.21 shows a negative relationship between O_3 and relative humidity (*sreh*) although at intermediate humidity levels (typically 40-70 %) it seems this relationship is very weak and presumably non-significant.

Short wave radiation (*swrd*) doesn't seem a very strong predictor for the station shown in Figure 4.21. The β function is nearly flat with an increasing spread in data values for higher levels. This turned out to be true for most stations, and thus, the *swrd* parameter is for most sites not statistically significant as discussed more in the following chapters.

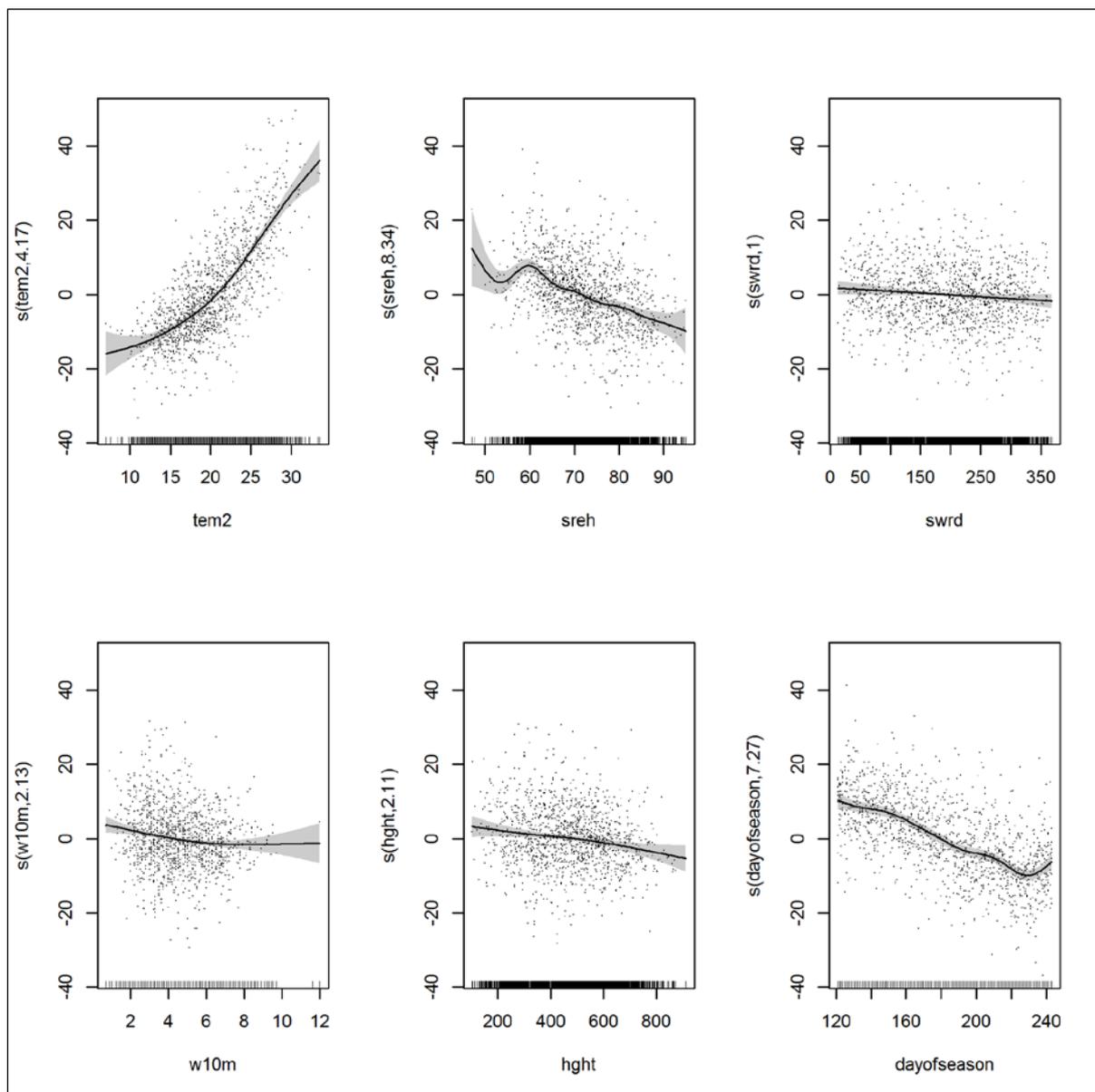
The response plot for wind speed at 10 m (*w10m*) shows a strong spread in data values. Still the GAM found a significant negative relationship, although not very strong in magnitude, between *w10m* and O_3 . This pattern is also typical for many of the other stations. This kind of relationship likely reflects that the high O_3 episodes are linked to weather situations with more gentle winds associated with anticyclones, whereas periods with stronger winds are linked to cyclonic weather conditions when ozone levels are lower.

The response curve for the day of season shows a clear negative response, except for the very last part of the period. This reflects the mean seasonal variation in ozone with highest mean

levels in May followed by decreasing levels until mid-August and then increasing levels again. A similar pattern is seen at many of the other stations in Europe.

The response curves (the β_j functions) discussed above are smooth functions and could therefore not be presented as single values. For mapping and tabulating these functions, we calculated “linearized” β -values using a linear fit to the interior of the data ranges, defined by the area between the 25- and 75-percentiles of these data. This is identical to the EPA approach (see Figure 5 in Camalier et al., 2007). This means that we fit a straight line to the middle part of each of the curves in Figure 4.21, for example the interval approximately 15-25°C for the temperature, 60-80 % for the relative humidity and so on. The slopes of these straight lines are thus just indicative of the overall relationship between the explanatory variables and the O₃ levels. When the interior range of the response curves (Figure 4.21) are approximately linear, these linearized β values are good representatives of the underlying β functions. For highly non-linear responses, however, the linearized values are less informative.

Figure 4.21: Response curves for Waldhof (DE02) for 2000-2010 showing daily values (dots) as well as the calculated beta functions (the smooth functions listed in Table 2.1) and the corresponding perturbation of the ozone value (ppb) on the y-axis for each of the six explanatory variables from top left to bottom right: tem2 = daily max 2 m temperature ($^{\circ}\text{C}$), sreht = daily mean relative humidity (%), swrd = daily mean short wave radiation (W/m^2), w10m = daily mean 10 m wind speed (m s^{-1}), hght = daily mean height of the planetary boundary layer (m) and dayofseason = the day of the season (with day number calculated from 1 Jan, so 1 May = 121). The black curves show the best fit smooth functions (the betas), whereas the shaded grey areas indicates ± 2 standard errors confidence regions.



4.3.1 The period 2000-2010

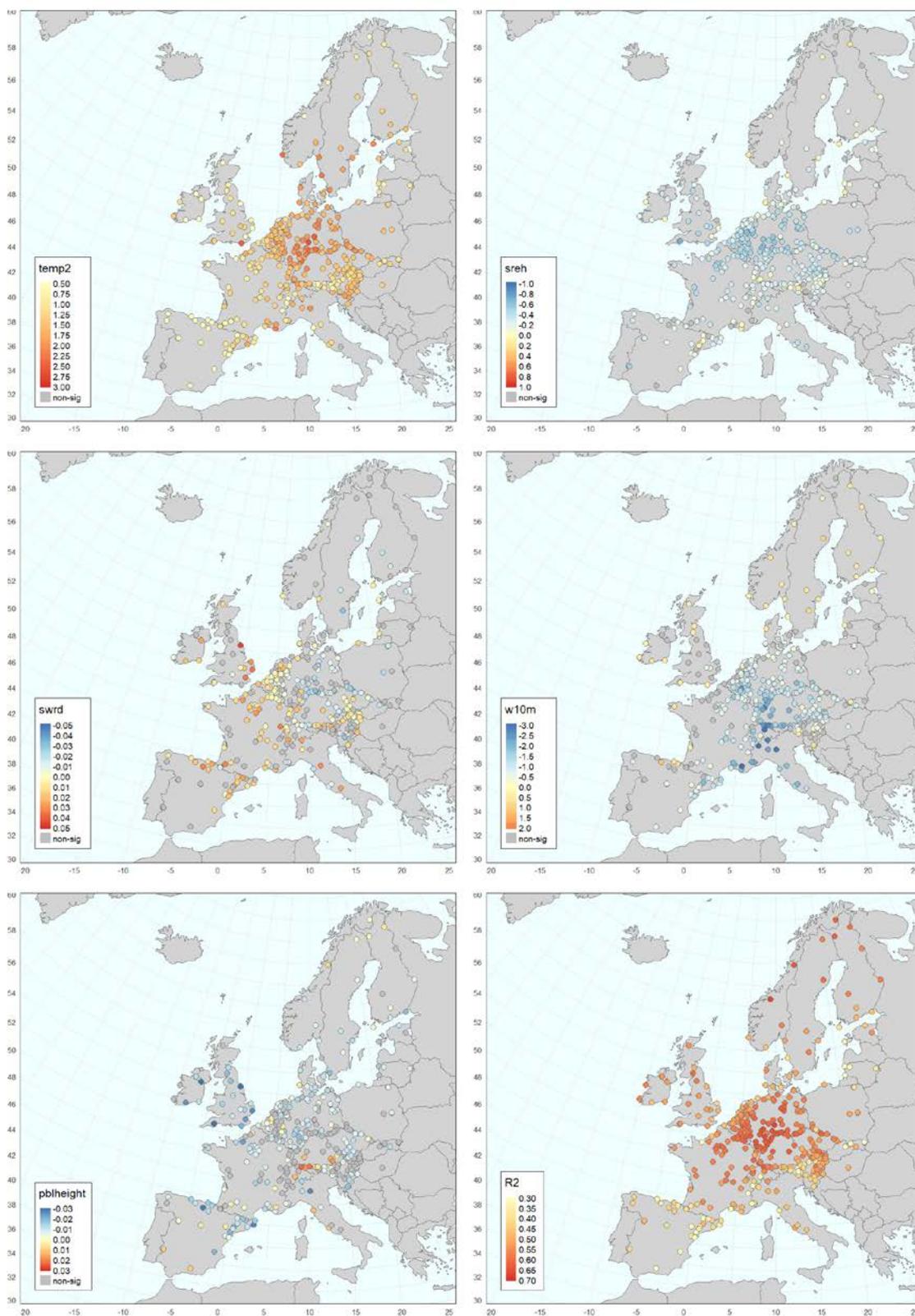
The linearized β values for the rural sites in the 2000-2010 period is mapped in Figure 4.22. Red colours mark positive values, blue colours mark negative values and non-significant values ($p = 0.05$) are given in grey.

Daily maximum temperature (*Tem2*) and daily mean relative humidity (*sreh*) turned out to be the two most important explanatory variables, as also found in the EPA study. Both variables are statistically significant at nearly all sites as seen in Figure 4.22. Furthermore, they show marked opposite responses on ozone – a strong positive response of temperature and a marked negative response of humidity.

More non-significant results are found for the responses of short wave radiation (*swrd*) and wind speed (*w10m*) indicating that these parameters have a generally smaller influence on the GAM results. We find a negative response of ozone with respect to wind speed in central Europe and a positive response in the Nordic countries and the UK. This reflects that high ozone episodes are linked to stagnant anticyclonic weather patterns in central Europe, whereas for the UK and the Nordic countries high ozone levels are dependent on advection of air masses from the emission regions in central Europe.

The linearized β value for the PBL height shows many non-significant values and is apparently not a very important explanatory variable for ozone, although there are some exceptions: For the UK, the PBL height has a clear negative response on ozone reflecting that high ozone episodes are mostly linked to weather situations with a shallow PBL. Thus, for UK the results indicate that elevated levels of ozone are linked to advection of air masses from the continent (as shown above) in a shallow PBL associated with stable atmospheric conditions during high pressure situations. A few sites in southern Germany and Switzerland shows the opposite, i.e. a positive response of PBL on ozone. This could reflect the conditions at mountain sites that are dependent on the upward mixing of high ozone levels from emission areas below and that the sites could be located above the polluted layer in situations with a shallow PBL.

Figure 4.22: The linearized beta values for the individual explanatory variables calculated by the GAM for rural EMEP and AirBase sites 2000-2010. Non-significant results are given in grey. R2 denotes the R2 statistic.



4.3.2 The period 1990-2000

Figure 4.23 shows the mapped linearized β values for the period 1990-2000. The poorer network of monitoring sites in 1990-2000 makes it harder to evaluate the large scale spatial differences for this period. The available network of sites does, however, indicate similar response patterns as for 2000-2010. Certain differences are worth mentioning, though.

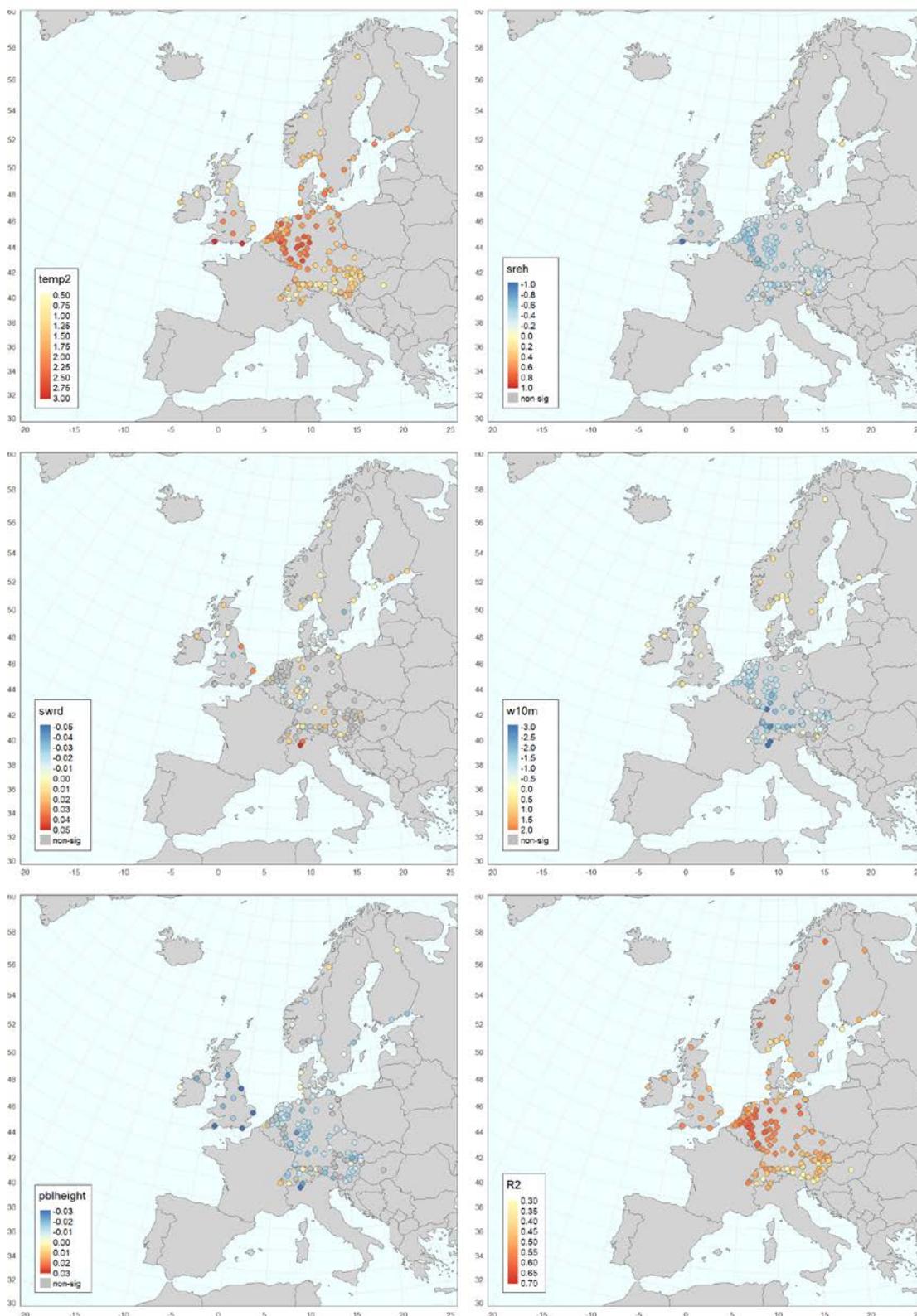
The responses given in Figure 4.23 indicate a stronger response of temperature on ozone for the 1990-2000 period compared to the 2000-2010 period in some areas as e.g. the southern UK, the Netherlands, Belgium and Germany. This likely reflects the reduced emissions of ozone precursors from the first to the second decade. Since, as discussed in the beginning, the temperature response on ozone is determined by the level of precursor concentrations (NO_x and VOC) in the air, we would indeed expect a higher temperature response in the first decade as is also seen from Figure 4.23.

With respect to relative humidity, we see the same type of response in the 1990-2000 period as for the 2000-2010 period. In some areas, the values indicate stronger responses, i.e. a stronger negative response in Germany/UK and stronger positive response in the Nordic countries, in the 1990-2000 period compared to the 2000-2010 period. This could indicate that ozone response with respect to the relative humidity is linked to the atmospheric chemistry as is the ozone response with respect to temperature.

In the same way as for the 2000-2010 period, we find in general a negative response of ozone with respect to the PBL height in the 1990-2000 period. The response in the 1990-2000 period is negative at most of the sites and show stronger absolute values compared to the 2000-2010 period that have more positive values and generally smaller absolute values, presumably of the same reason as for temperature – that the level of ozone precursors was generally higher in the first decade.

Also for the wind speed ($w10m$) most sites show a negative relationship indicating that high ozone levels are associated with low wind speeds in general. As for the 2000-2010 period, the Nordic countries and UK is an exception to this and shows a positive relationship that reflects that the occurrence of elevated ozone episodes in these countries depend on a certain advection of polluted air masses into the countries.

Figure 4.23: The linearized beta values for the individual explanatory variables calculated by the GAM for EMEP and AirBase rural sites 1990-2000. Non-significant results are given in grey. R2 denotes the R2 statistic.



4.4 Long-term trends

One aim of this study was to evaluate to what extent the long-term trends in surface ozone in Europe reflects the change in emissions of ozone precursors and to what extent interannual variations in meteorology could have an influence on the ozone trends.

The preceding chapters have, as we see it, shown that the GAM method can predict the ozone levels to a certain extent and that the performance of the GAM model varies systematically over the continent. The previous discussion has furthermore indicated that the GAM model performs better for medium range O₃ values than for the highest (and lowest) peak values. This is important to keep in mind since several studies (e.g. Lefohn et al., 2017) have shown that the magnitude and significance of long-term trends vary with the selection of ozone metric. High-percentile O₃ metrics are generally showing clearer long-term trends than medium-range ozone metrics.

Anyway, the overall performance of the GAM method documented above, is a first prerequisite for assessing the long-term trends. Without this kind of general trust that the GAM method can predict daily levels of ozone to a certain degree of precision, it would be effortless to use it for trend evaluations.

4.4.1 The period 2000-2010

In Figure 4.24 the meteorologically adjusted trends (β_7) based on the 2000-2010 rural EMEP and AirBase data calculated by the GAM method are compared to a plain linear regression, i.e. with no adjustments for meteorology. Figure 4.25 shows the corresponding results for the AirBase suburban sites.

The comparison of the meteorological adjusted trends vs. the plain linear non-adjusted trends indicate fairly small differences, implying that overall the meteorology did not have a decisive influence on the ozone long-term trends at the rural sites in this decade. Still meteorology could have a very strong impact on the year-to-year variability in ozone, which is also an issue of interest.

Negative (downward) trends are found over most parts of continental Europe by both adjusted and non-adjusted methods. Strongest declines are seen at some Italian sites. A few differences between the plain trends and the meteorologically adjusted trends are apparent, though. For the UK, the meteorologically adjusted trends are calculated to be positive, reflecting increasing ozone levels, whereas the non-adjusted trends are mostly non-significant in the UK. Furthermore, for central Europe the results indicate that the plain non-adjusted trends give stronger reductions in ozone than calculated by the adjusted trends.

The slightly increasing meteorologically adjusted trends in the UK is interesting and somewhat counter-intuitive since we would expect decreasing levels of both ozone precursors and ozone itself during this period also in the UK.

It is essential, though, to be aware of the metric these trends are based on. These trend estimates are based on the MDA8 O₃ levels during the May-August period. Thus, they reflect the trend in the seasonal mean MDA8 O₃, a metric that is more indicative of the medium-range ozone values and less indicative of the high ozone peak values. As documented in

recent papers (e.g. Lefohn et al. 2017), the trends could vary strongly for various parts of the ozone distribution. In general, high peak levels are most responsive to the trends in the precursors (NO_x and VOC) whereas the trend in medium range ozone values are less marked since these values are more influenced by other factors such as the hemispheric baseline level.

Figure 4.24: Meteorologically adjusted ozone trend as calculated by the GAM method (left) and the plain linear trend (right) calculated by a standard regression (i.e. not meteorologically adjusted) for the rural EMEP and AirBase sites 2000-2010. Non-significant trends ($p = 0.05$) are given in grey. Unit: ppb year⁻¹.

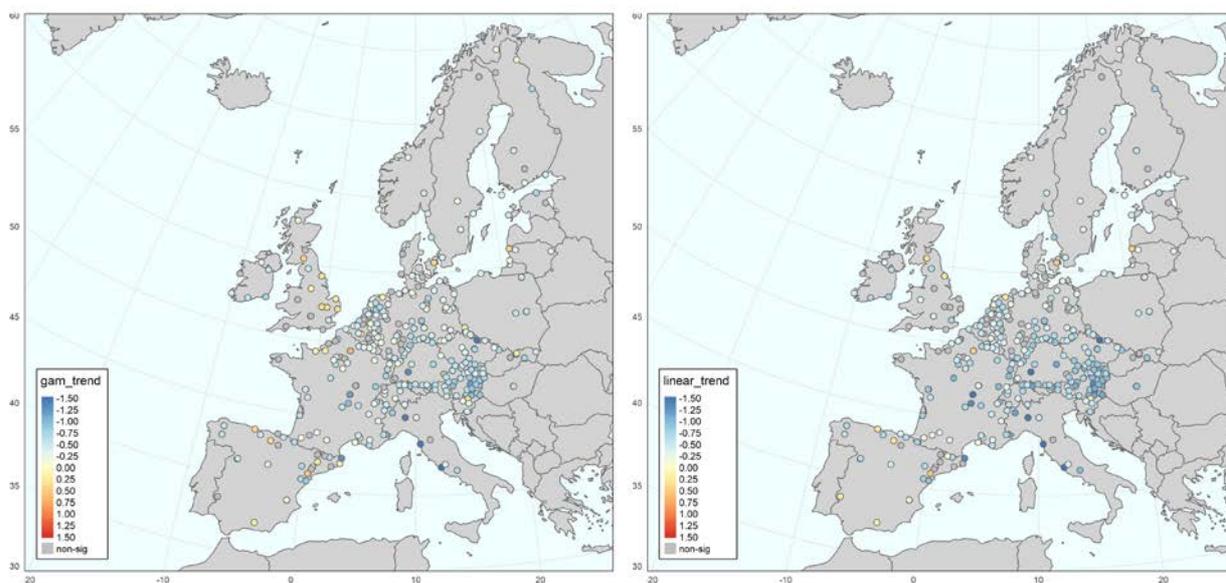
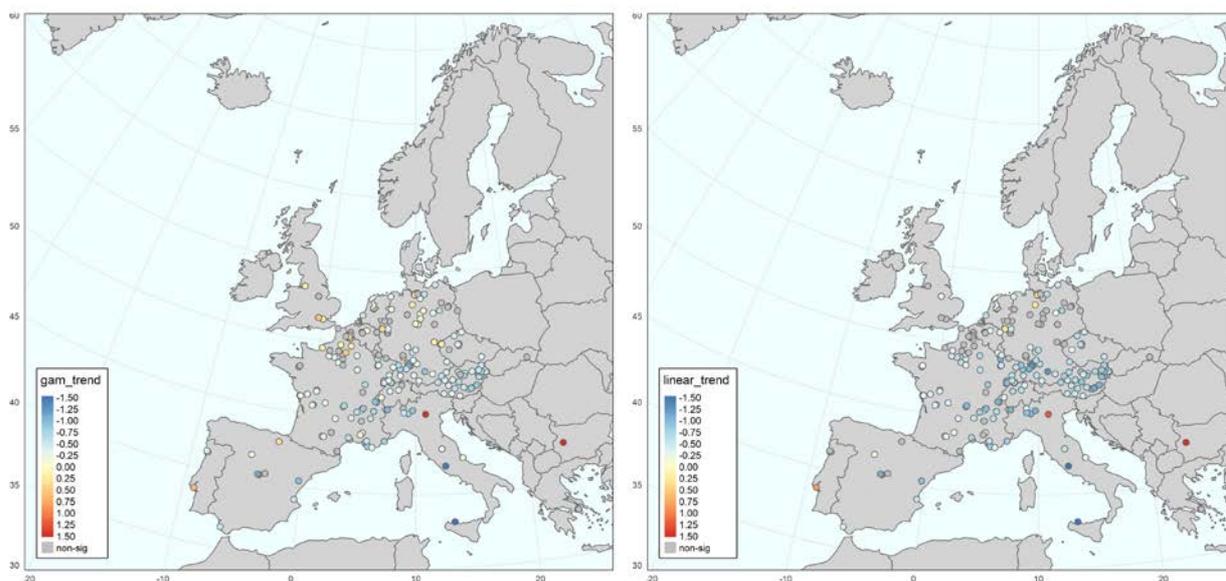


Figure 4.25: Same as Figure 4.24 for the AirBase suburban sites.



In addition to the maps shown above, we have included time series with the observed and predicted seasonal mean MDA8 O₃ values for the same example sites as above, i.e. Waldhof, Harwell, K arvatn, La Tardiere and Peyrusse Vieille. These data are shown in Figure 4.26 and Figure 4.27.

Figure 4.26: Seasonal (May-August) mean MDA8 O₃ levels as observed (black), modelled by the GAM (red), modelled with GAM subtracting the trend (red dashed). The blue straight line shows the linear trend in the meteorologically adjusted data as given by the GAM. Data for Waldhof (top), Harwell (middle) and K arvatn (bottom) are shown.

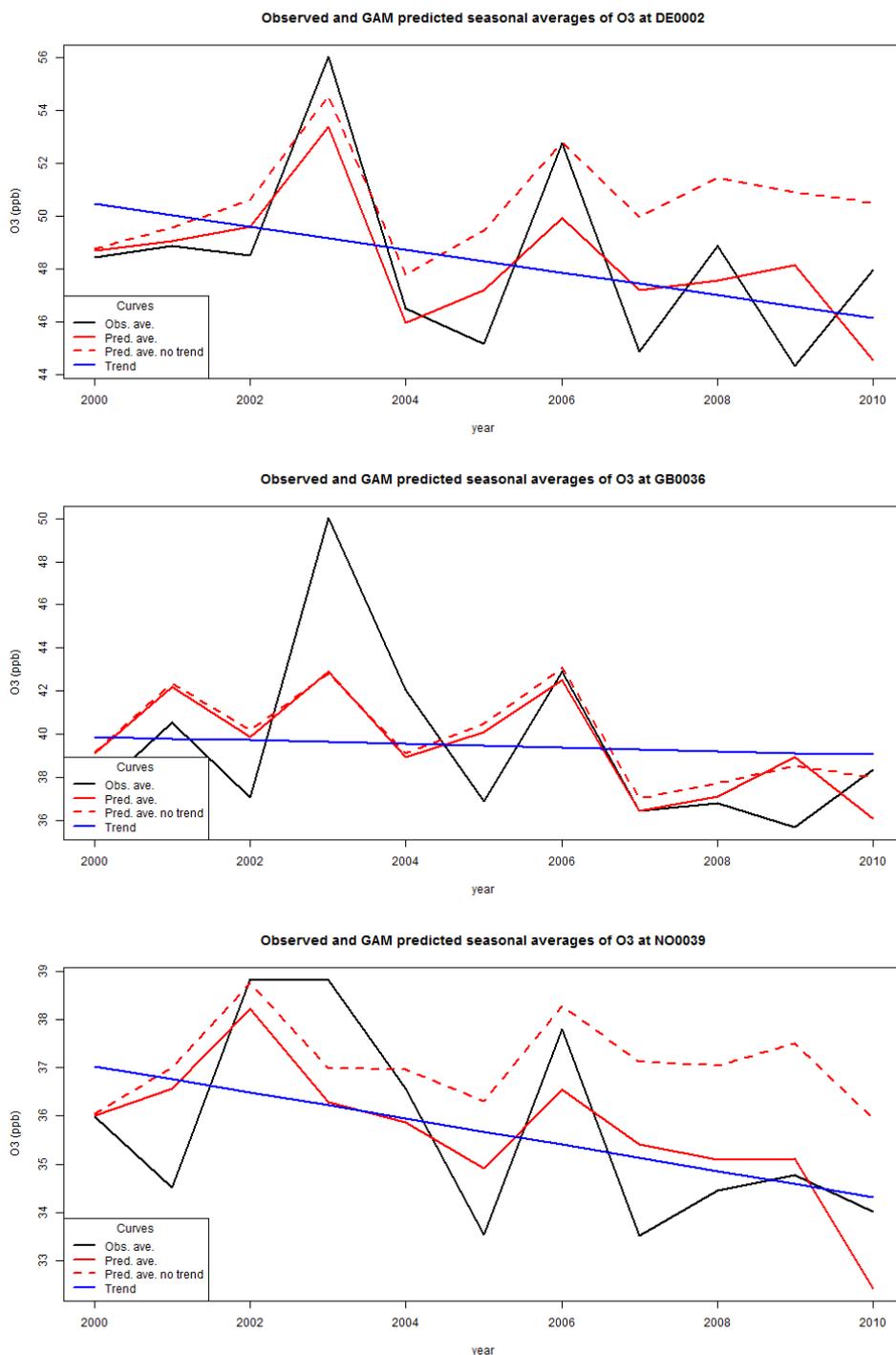
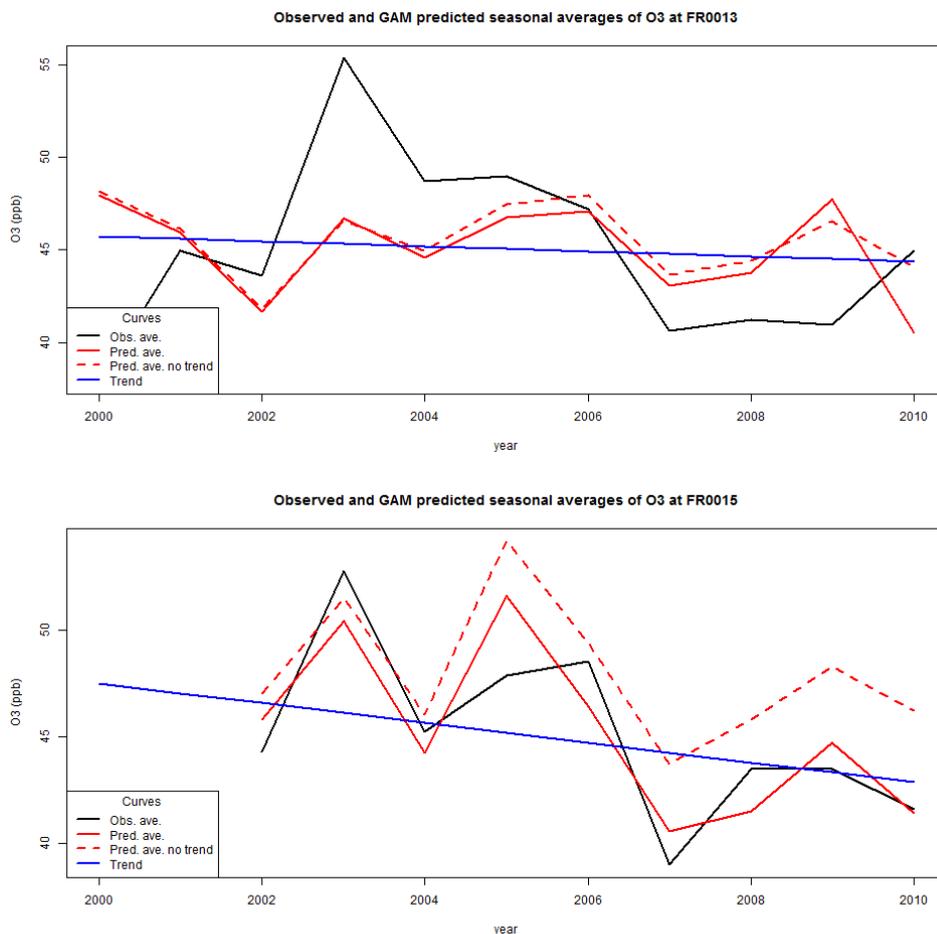


Figure 4.27: Same as Figure 4.26 for Peyrusse Vieille (top) and La Tardiere (bottom).



4.4.2 The period 1990-2000

The meteorologically adjusted and the plain non-adjusted trends for the period 1990-2000 are shown in Figure 4.28 and Figure 4.29 for the rural and suburban sites, respectively. As opposed to the 2000-2010 period, we calculate increasing trends in the mean MDA8 O₃ for many sites in the 1990-2000 period, more marked for Swiss, German and Nordic sites. Furthermore, we find more sites with significant (increasing) trends by the adjusted trends than by the plain non-adjusted trends. Many sites are found to show non-significant trends when they are not adjusted for meteorology (right panel), while they are positive (increasing) when adjusting for meteorology through the GAM routine. This indicates that meteorology in the 1990s counteracted a rising trend in the MDA8 O₃ levels in these regions.

Figure 4.28: Same as Figure 4.24 for the rural EMEP and AirBase sites for the period 1990-2000.

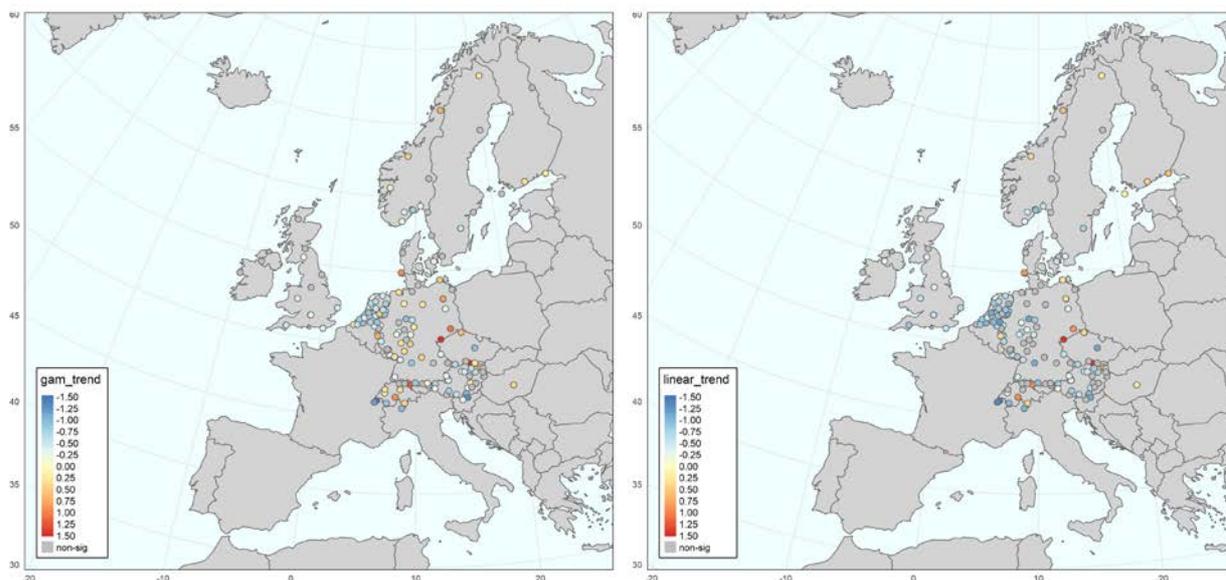
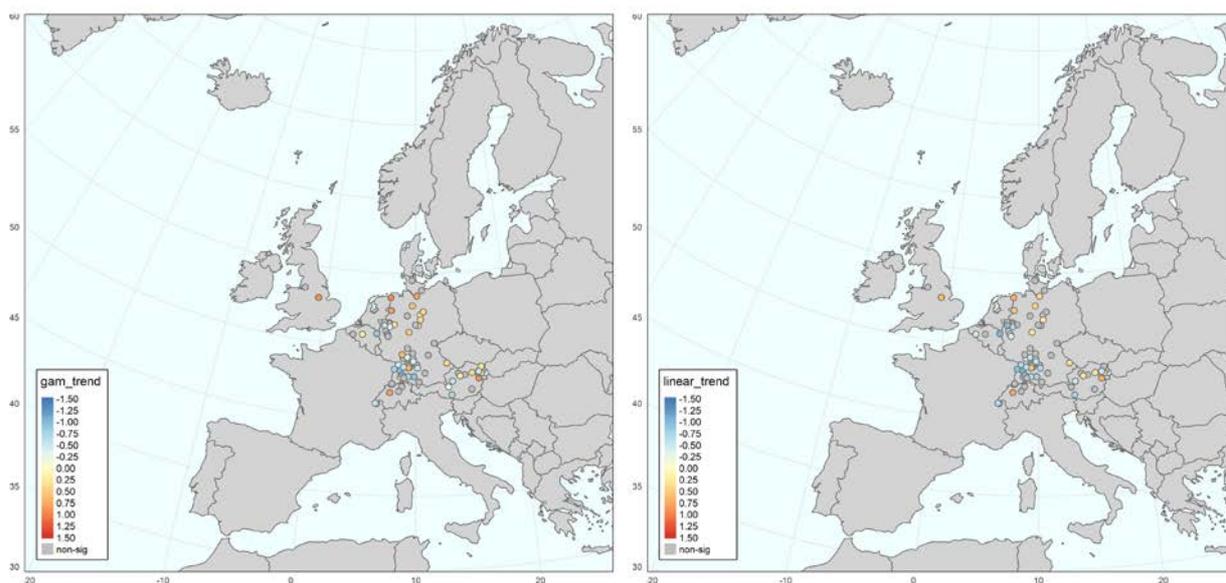


Figure 4.29: Same as Figure 4.24 for the AirBase suburban sites for the period 1990-2000.

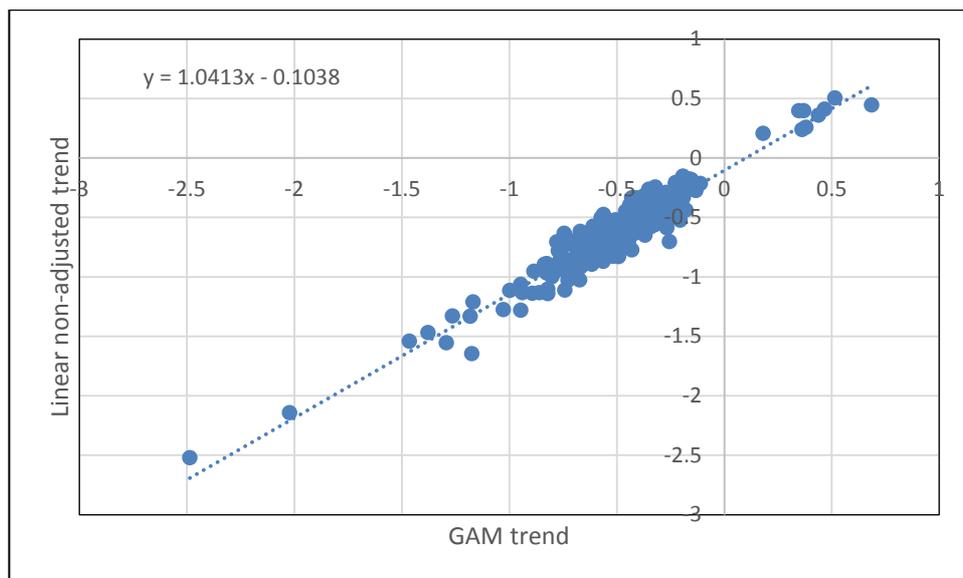


4.4.3 The influence of meteorological adjustment of trends

Based on a visual inspection of the maps presented above, we do not see very clear patterns in the meteorologically adjusted trend values compared to the plain linear non-adjusted trends. In the following some basic comparisons between the adjusted trends and the linear non-adjusted trends are given.

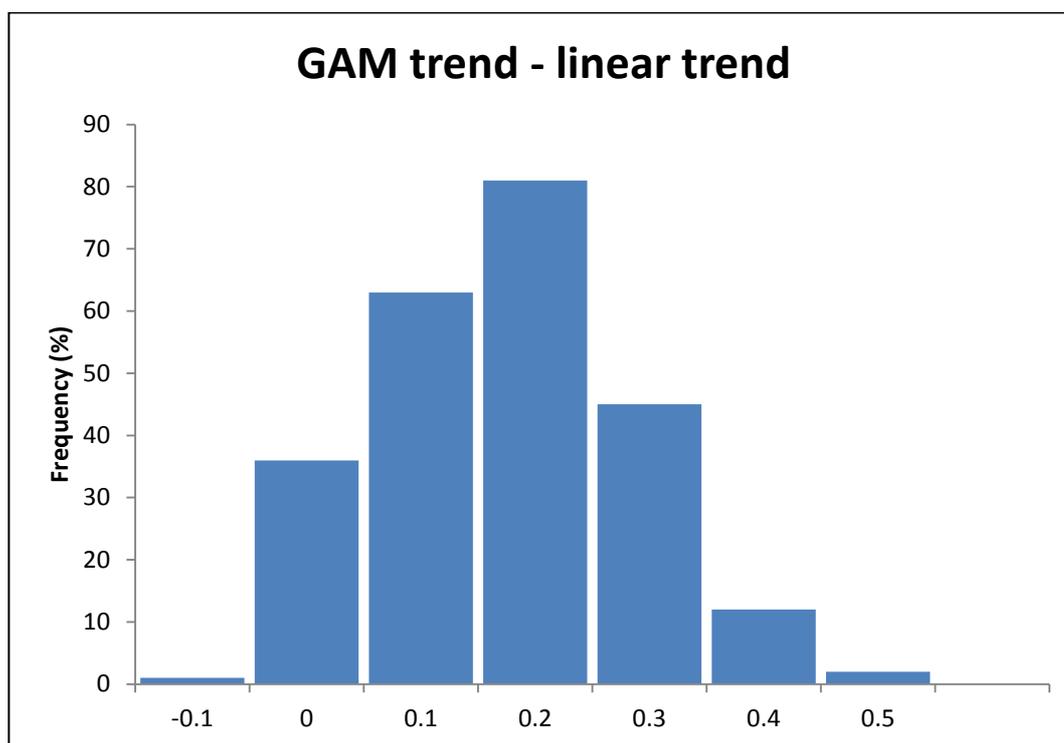
Figure 4.30 shows the scatter plot of the adjusted trends vs. the plain non-adjusted trends based on AirBase rural data for the period 2000-2010. Note that only significant trends ($p=0.05$) from both types of trends are plotted. First, this shows a close agreement between these two types of trend values. The regression between these individual points follow a straight line. Secondly, nearly all the trends are negative implying that both the adjusted and the non-adjusted trends show declining levels of the mean MDA8 O₃ levels in this period for nearly all sites.

Figure 4.30: Scatter plot of the GAM met-adjusted trend (x-axis) vs the plain linear non-adjusted trend (y-axis) for AirBase rural sites 2000-2010. Only significant trend values are plotted. Unit: ppb year⁻¹.



The regression fitted to these points do, however, indicate that the GAM adjusted trend gives somewhat higher values (less negative) in general compared to the non-adjusted linear trends. This is further confirmed by the histogram of all differences, i.e. the GAM adjusted trends minus the linear non-adjusted trends, shown in Figure 4.31. We see that nearly all GAM trends are larger than the linear trends, and the median of the differences are of the order of 0.2 ppb year⁻¹ according to Figure 4.31. This means that the meteorologically adjusted trends calculated by the GAM method are less negative than the linear non-adjusted trends (or that the few positive trends are more positive).

Figure 4.31: Histogram of the GAM met-adjusted trend minus the plain linear non-adjusted trend for AirBase rural sites 2000-2010. Unit: ppb year⁻¹.



If taking the GAM method for “the truth” this implies that the interannual meteorological variability during the 2000-2010 period led to a stronger reduction in the MDA8 O₃ levels than would have been the case with “normal” meteorology. In other words, a part of the decline in MDA8 O₃ experienced during this decade was due to meteorology alone. Interestingly, this is the same finding as both the two methods presented in the 2016 report (Colette et al., 2016) indicated.

That we end up with the same conclusion using very different approaches gives increased confidence to the conclusion that meteorology played a role by intensifying the observed ozone reduction during 2000-2010. It should be said that we still should be careful with making too strict statements. One important point, as discussed above, is that whereas the trend calculations in Colette et al. (2016) were based on high peak values in ozone, the present study has been based on seasonal mean MDA8 O₃ values, and as shown in several papers, these two metrics could show very different trends. Interestingly, a pattern very similar to the 2000-2010 period is found for the 1990-2000 period, as shown in Figure 4.32 and Figure 4.33. That the difference between the GAM trends and the linear non-adjusted trends are so similar in these two periods, adds some concern with respect to the method in general. However, the same kind of relationship between meteorology and ozone for these time periods was found by Colette et al. (2016) with a fully independent method.

Figure 4.32: Same as Figure 4.30 for the period 1990-2000 based on EMEP rural data.

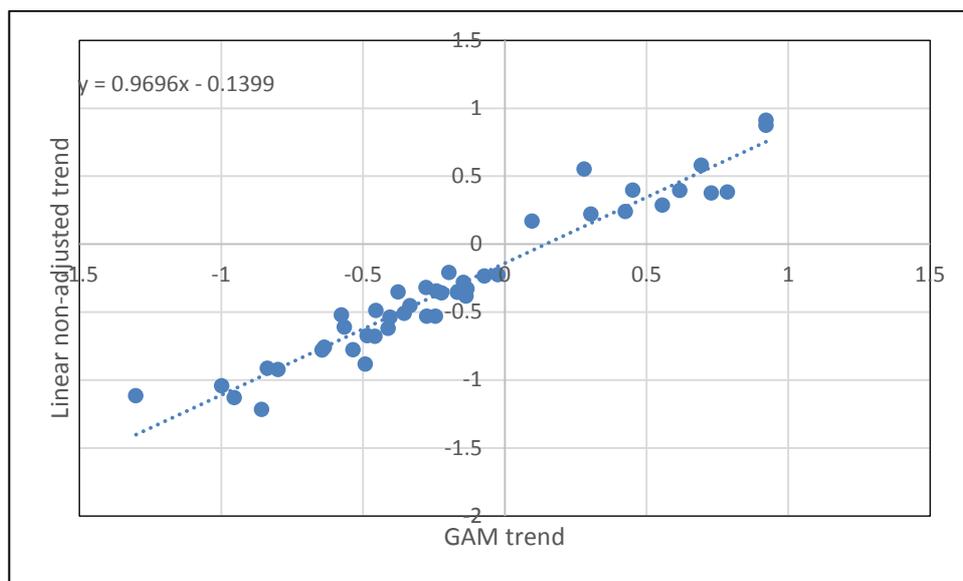
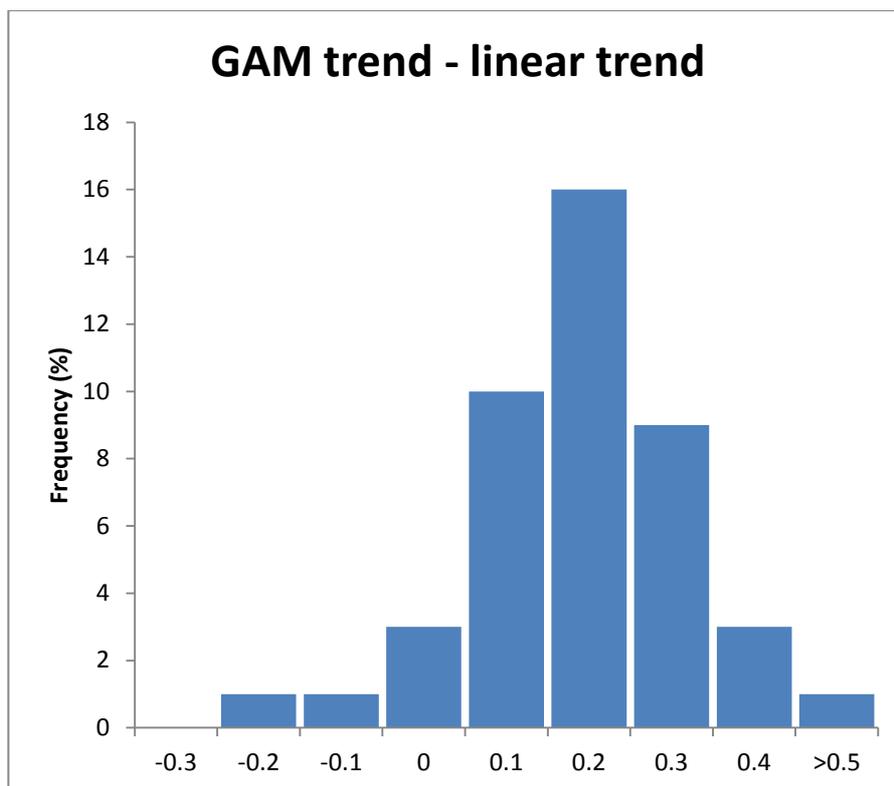


Figure 4.33: Same as Figure 4.31 for the 1990-2000 period for EMEP rural data.



5 The use of GAM as a data quality checking tool

As a side effect of this project work, it turned out that the GAM could be used to identify time series with dubious ozone monitoring data. The R2 statistic, expressing to what extent the GAM was able to predict the observed O3 data typically varied from 0.20 to 0.70 for the individual stations. When inspecting the sites with the lowest R2 scores, it became obvious that some of these time series included data that were obviously wrong.

As an example, Figure 5.1 shows the observed and predicted ozone time series at the Cypriot site Ayia Marina (CY0002) for three years, 2007-2009. The R2 for the GAM based on the 2000-2010 data was particularly low for this site ($R^2 = 0.10$) and the data clearly points to a major offset problem in 2008 when nearly all data are below 50 ppb whereas nearly all values in the neighbouring years, 2007 and 2009 are above this value. Thus, we excluded this site in the further project work, and the site is not included in the map plots and statistics shown above.

A similar problem was found for the Italian AirBase site IT1179A that had an R2 of 0.17 for the period 2000-2010. Inspecting the daily data revealed substantial differences in ozone levels from one year to another, as shown for the years 2004-2006 in Figure 5.2. Whereas the ozone values were of the order of 30-40 ppb in 2004, it shifted to 60-70 ppb in 2005 and then up to 100-150 in 2006.

Another example is given in Figure 5.3 showing time series for the Spanish site Niembro for 2001-2003. In 2002 the observed ozone data were consistently much lower than expected and much lower than the other years. It seems highly unlikely with such low levels over an extended period.

Based on these findings, it seems that as a side effect, the GAM method may be used to identify sites that show problems of major offsets in the data values. We do not believe it could be used in automated way, but rather as part of a manual screening of data quality station by station.

Finally, we mention that the GAM performance score as measured by R2 also give particularly low values for certain sites that do not show obviously erroneous data. The ozone data from Jungfraujoch (CH0001), a key station in the WMO-GAW network being operated by the World calibration centre for surface ozone at EMPA, also give very low R2 scores with the GAM method. This is obviously not linked to data quality issues, but rather that the GAM method fails at this site, located almost 3600 m a.s.l. The likely reason why the GAM fails here is that the ozone levels at this site could not be linked to the local meteorological parameters.

Figure 5.1: Time series of daily O₃ levels for Ayia Marina (CY02) for 2007-2009 as measured (black) and predicted by the GAM method (red). The blue line just shows the results from the GAM model using a normal climatology each day for reference. Unit: ppb.

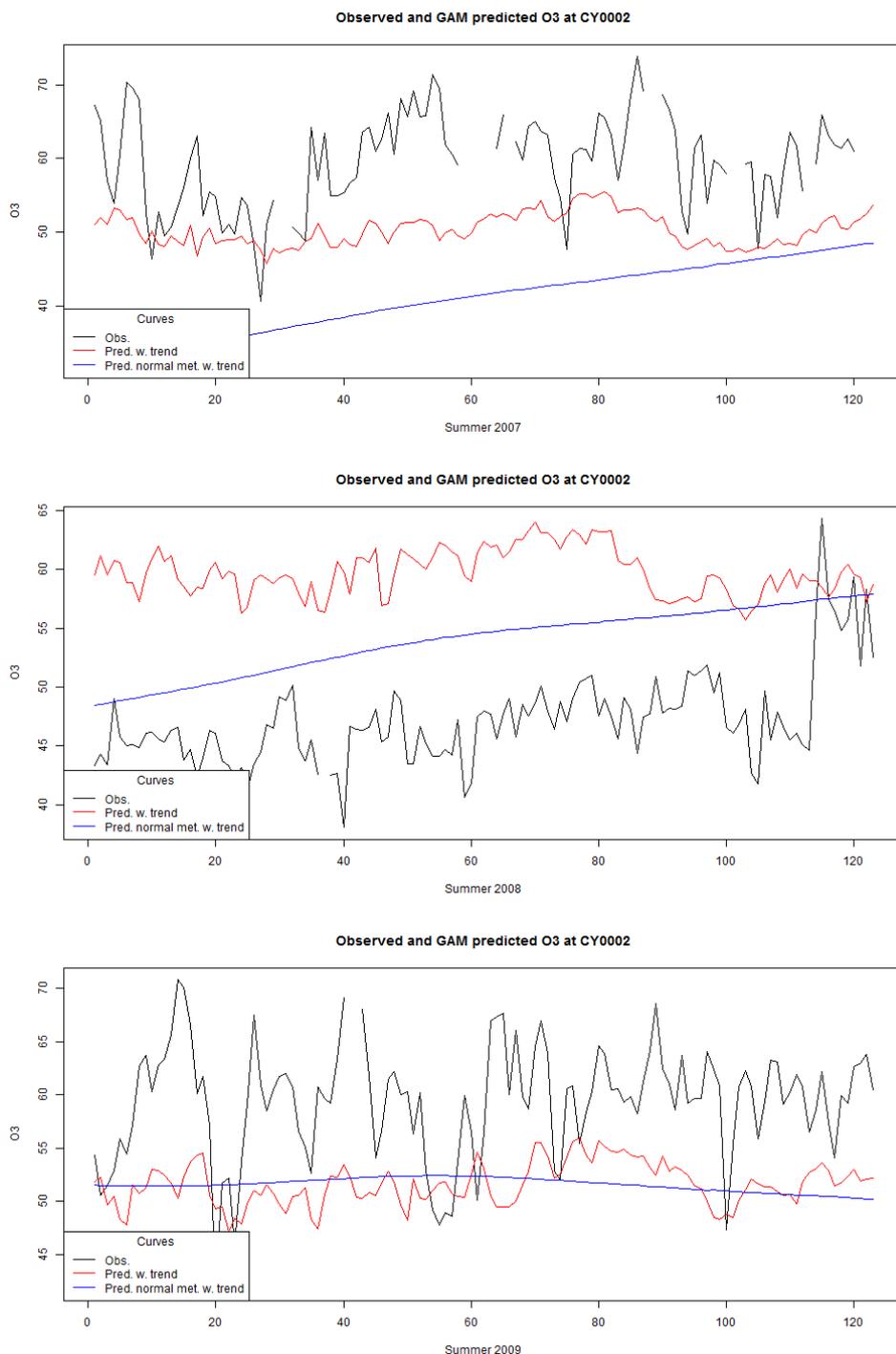


Figure 5.2: Observed and GAM predicted time series for the AirBase site IT1197A for 2004-2006. Unit: ppb.

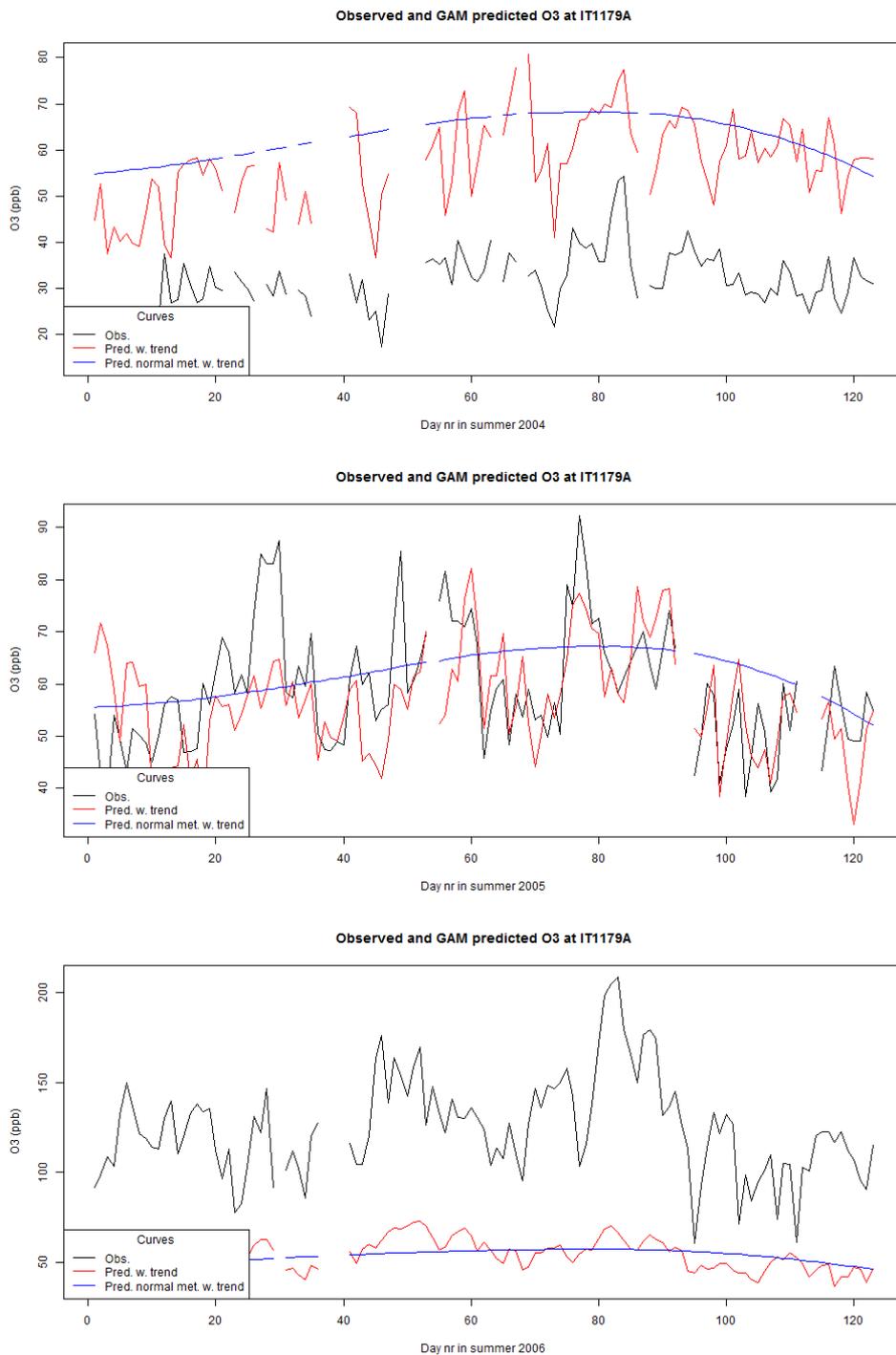
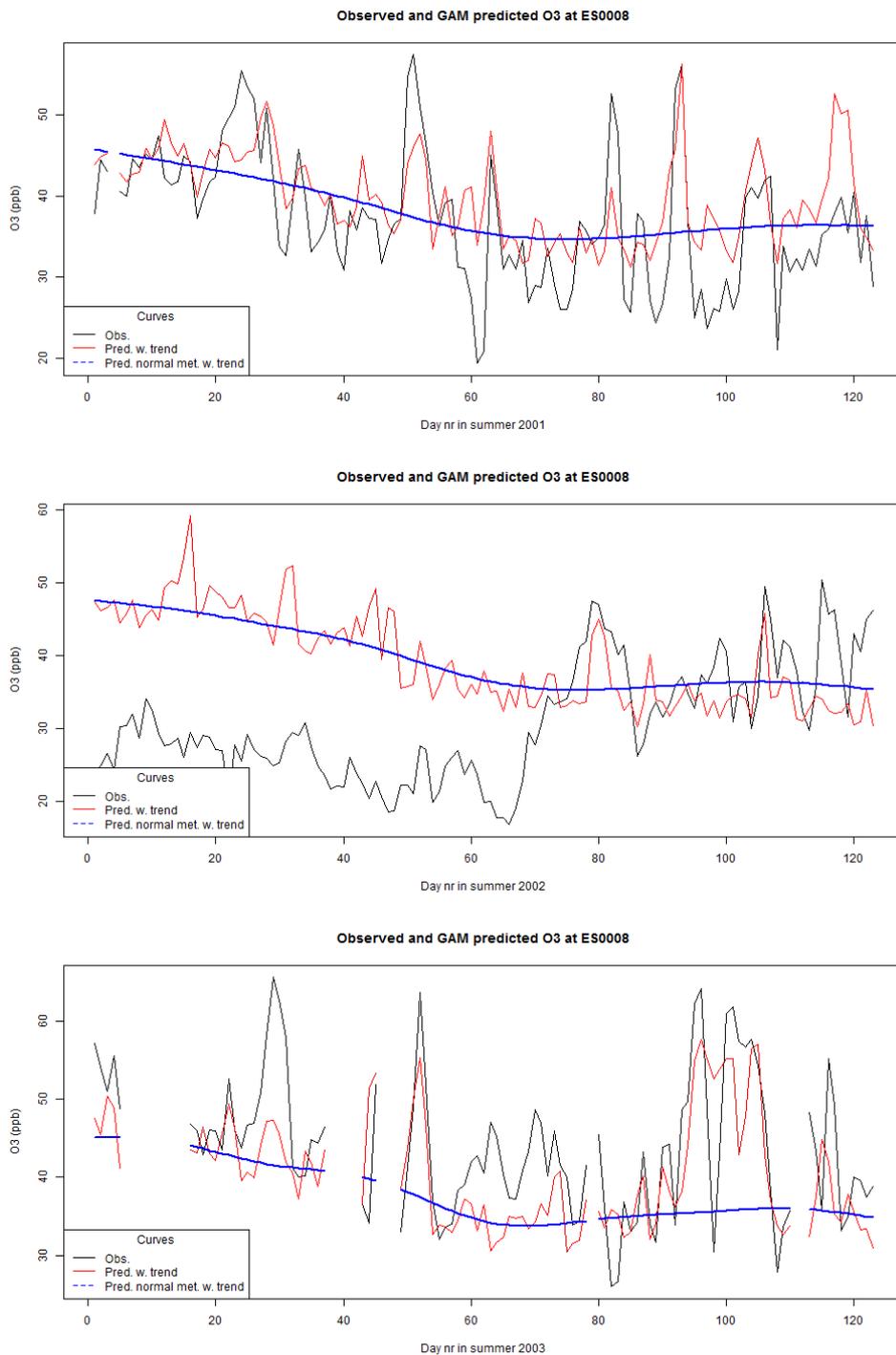


Figure 5.3: Observed and GAM predicted time series for the EMEP site Niembro (ES08) for 2001-2003. Unit: ppb.



6 Comparison with other methods for meteorological adjustments

The study presented in this report was a follow-up on previous studies on surface ozone trends and the question of meteorological influence on these trends. The study in 2016 (Colette et al., 2016) was based on the EDT exercise that included a set of long-term model runs (1990-2010) with various scenarios for emissions, meteorology and boundary conditions. In the 2016 study we found a general improvement in model performance over the years (i.e. a trend in model performance when compared to measurements), and furthermore that the observed decrease in ozone peaks in the 1990s was less than predicted by the models, whereas the agreement was better for the 2000s.

The 2016 report also presented an attribution analysis designed to quantify the relative importance of emissions, meteorological variability, and ozone background levels. That analysis showed the dominating impact of emission changes on observed air quality trends, thereby demonstrating the effectiveness of air quality management strategies. A substantial impact of meteorological variability and background chemistry was also documented, thus showing the need to account for those factors to reconcile the quantitative comparison between trends in emission of precursors and ambient concentrations.

A comparison and evaluation of the GAM approach relative to the methods in the 2016 report is not straight-forward, though. Two different methods for meteorological adjustments were given in the 2016 report: i) a pure model-based approach based on certain combinations of the various EDT scenarios; and ii) a pure model-based approach based on constructing a long-term ozone mean climatology.

Our GAM work presented here differs from the 2016-methods in important ways: Firstly, the GAM method is based on the ozone measurements directly, not on the modelled ozone. Secondly, the GAM predictions and trends are based on the seasonal mean of the MDA8 O₃ values, whereas the trend calculations in the 2016 study was mostly focused on high peak values, typically the 4th highest MDA8 O₃ through the year. As mentioned above, the long-term trends could differ significantly for these two ozone metrics.

We do find, however, that both the present study and the methods applied in the 2016 study point in the same direction, namely that the year-to-year variations in meteorology intensified the observed reductions in ozone metrics. In other words, precursor emission reductions were only responsible for parts of the observed decline in ozone, according to these studies.

7 Summary of findings

This report has presented results from studying surface ozone trends applying a statistical method for calculating so-called meteorologically adjusted trends based on local meteorological parameters. The purpose of this method is to try to reduce or subtract the influence of year-to-year variations in meteorology from the observed trends in ozone. The reason for developing such methods is that long-term trends in surface ozone could be affected by a combination of several physio-chemical processes, of which the two most important ones presumably are i) emissions of precursors; and ii) variations in meteorology. Naturally, policy makers and others would be interested in the possibility of separating the trend driven by emission abatement actions from the other processes.

Strictly speaking, subtracting the influence of one process from the other is not possible since they are all linked together. Many studies have, however, been published on the topic of meteorological adjustment of trends, using a wide variety of methods ranging from pure statistical ones to those based on CTMs.

Here we apply a method originally outlined by US-EPA that we have further developed and adapted for European rural and suburban monitoring sites. The method is called a GAM – Generalized Additive Model – which can be regarded as a sophisticated multiple regression model in which we use smooth response functions instead of linear coefficients for the various explanatory variables. The GAM method is applied for two separate periods: 1990-2000 and 2000-2010, and for three separate datasets; EMEP rural, AirBase rural and AirBase suburban stations.

The calculations were based on the daily MDA8 O₃ values, i.e. the maximum 8-h running daily average concentrations, for the summer season, defined as 1 May - 31 August for the years 1990-2010. As input explanatory variables we used daily gridded meteorological data used in the EDT project extracted from the grid squares containing the respective ozone monitoring site. We used temperature, relative humidity, wind speed, short wave radiation, PBL height and the seasonal day number (counted from 1st May) as explanatory variables together with the long-term trend. For the trend term we assumed a linear dependency, whereas smooth response functions were assumed for the other parameters. The GAM was then applied to each site separately, i.e. of the order of 100-500 stations depending on the decade.

The results showed that the performance of the GAM method varied with region. For some regions, like central Europe and Germany in particular, we found good to very good agreement between the observed and predicted daily MDA8 O₃ levels. For other areas, like the Nordic countries, we found that the seasonal cycle and the timing of episodes were well reproduced, although with a substantial underestimation of many of the high peak O₃ episodes. In southerly regions, particularly in Spain and Italy, significantly poorer agreement with observed values were seen.

Daily maximum temperature, daily mean relative humidity and the seasonal day number were overall the most important explanatory variables. For the period 1990-2000, however, the wind speed was significant at equally many sites as the relative humidity. Shortwave radiation and PBL height showed non-significant contributions to the GAM model at many sites.

The fact that the GAM performance is in general better for the seasonal mean MDA8 O₃ values than for the high peak values sets some limitations on the use of this method in its present form. If the interest is the high peak values, one should probably look for adjustments to the presented GAM methodology.

With respect to the trends, we found close agreement between the plain linear non-adjusted ozone trends calculated by simple regression and those calculated by the GAM. We did find, though, that the GAM's meteorologically adjusted trends were systematically higher (less negative) than the non-adjusted trends. This indicates that meteorology contributed to the downward trend in ozone seen at most sites, i.e. that parts of the reduction in ozone could be explained by meteorology, both for the 1990-2000 and for the 2000-2010 decade. The same result was found by two other independent methods presented in the study by Colette et al. (2016) for other ozone metrics, pointing to a stronger robustness and confidence of this finding.

As a side effect, it turned out that the GAM method could also be used to identify errors in the measurement data. Some stations with particularly low R² scores, indicating very poor performance of the GAM method, turned out to have erroneous data, typically reflecting substantial shifts in the general levels from one year to another.

8 Future developments

As this work has documented that the GAM method, at least for certain sites and regions, is able to predict the daily levels of MDA8 O₃ reasonably accurately, we believe it would be valuable to continue adapting and developing such a tool. It is important, however, to keep in mind what a statistical method like this can do, and what it can't do, as opposed to a chemical transport model (CTM).

For certain regions/stations, our results show that a GAM (or similar methods) may be applied to look at the year-to-year variability in mean ozone statistics and at perturbations in ozone with respect to meteorology. The method can also be used to assess temporal trends in certain ozone metrics to a varying level of accuracy depending on the geographical area. Furthermore, it could be used to aid the interpretation and understanding of the ozone levels for individual years.

More efforts can be made to evaluate the significance of various input parameters, and the possibility of introducing air mass trajectory data should be explored. Furthermore, the performance of the method when applied to individual stations versus when applied to merged data from larger areas should also be investigated. Additionally, it would be of interest to explore the reasons behind the systematic spatial differences in model performance as seen in this study.

For predicting high peak levels of ozone, the GAM method presented here needs adjustments, e.g. by introducing the method of quantile regression as recently documented by Fix et al. (2018) and others. Such methods could also possibly be utilized for assessing trends in peak values of ozone.

Comparisons between CTM performance and GAM performance could be done, and one might also consider including CTM results in the GAM procedure in several ways. It would e.g. be of interest to include the performance of CTMs directly as a dependent variable, i.e. relating this performance, and lack of performance, to certain input explanatory variables in sensitivity analyses.

At the same time, it is important to keep in mind the limitations of a statistical method like this. In contrast to CTMs, a statistical method cannot be used to study the physical links between controlling parameters like emissions and ozone. Next, the homogeneity criterion discussed in this work is important to keep in mind. This criterion states that the relationships established between the ozone response and the input meteorological data is dependent on the atmospheric level of precursors. Thus, with changing emissions, these relationships will also change, and the regression functions for one period is not necessarily applicable for another period with other levels of emissions.

9 Acknowledgement

Meteorological data used in the present analysis were produced in the framework of the EuroDelta-Trends Project initiated by the EMEP Task Force on Measurement and Modelling of the Convention on Long Range Transboundary Air Pollution. EuroDelta-Trends is coordinated by INERIS and involves modelling teams of BSC, CERE, CIEMAT, ENEA, IASS, JRC, MET Norway, TNO, SMHI.

References

- Camalier, L., Cox, W., and Dolwick, P., 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., 2015, The openair manual — open-source tools for analysing air pollution data. Manual for version 1.1-4, King’s College London.
- Colette A., Bessagnet B., Vautard R., Szopa S., Rao S., Schucht S., Klimont Z., Menut L., Clain G., Meleux F., Curci G. and Rouil L., 2013, European atmosphere in 2050, a regional air quality and climate perspective under CMIP5 scenarios, *Atmos. Chem. Phys.* **13**, pp. 7451-7471.
- Colette, A., Beauchamp, M., Malherbe, L., Solberg, S., 2015, Air Quality Trends in AIRBASE in the context of the LRTAP Convention, ETC/ACM Technical Paper 2015/4.
- Colette, A., Solberg, S., Beauchamp, M., Bessagnet, B., Malherbe, L., Guerreiro, C., 2016, Long term air quality trends in Europe Contribution of meteorological variability, natural factors and emissions, ETC/ACM Technical Paper 2016/7.
- Colette, A., Andersson, C., Manders, A., Mar, K., Mircea, M., Pay, M.-T., Raffort, V., Tsyro, S., Cuvelier, C., Adani, M., Bessagnet, B., Bergström, R., Briganti, G., Butler, T., Cappelletti, A., Couvidat, F., D’Isidoro, M., Doumbia, T., Fagerli, H., Granier, C., Heyes, C., Klimont, Z., Ojha, N., Otero, N., Schaap, M., Sindelarova, K., Stegehuis, A. I., Roustan, Y., Vautard, R., van Meijgaard, E., Vivanco, M. G., and Wind, P.: EURODELTA Trends, a multi-model experiment of air quality hindcast in Europe over 1990–2010, 2017, *Geosci. Model Dev.*, 10, 3255–3276, <https://doi.org/10.5194/gmd-10-3255-2017>.
- Draxler, R.R., Hess, G.D., 1997, Description of the HYSPLIT_4 Modeling System. NOAA Technical Memorandum ERL ARL-224. Air Resources Laboratory, Silver Spring, Maryland.
- EU, 2008, Directive 2008/50/EC of the European parliament and of the council of 21 May 2008 on ambient air quality and ozone in ambient air. Official Journal of the European Union.
- Fix, M. J., Cooley, D., Hodzic, A., Gilleland, E., Russell, B. T., Porter, W. C., and Pfister, G. G., 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>.
- Lefohn, A.S., Malley, C.S., Simon, H., Wells, B., Xu, X., Zhang, L., Wang, T., 2017, Responses of human health and vegetation exposure metrics to changes in ozone concentration distributions in the European Union, United States, and China. *Atmos. Environ.* 152, 123–145.

- Logan, J. A., Staehelin, J., Megretskaia, I. A., Cammas, J.-P., Thouret, V., Claude, H., De Backer, H., Steinbacher, M., Scheel, H.-E., Stübi, R., Fröhlich, M., and Derwent, R., 2012, Changes in ozone over Europe: Analysis of ozone measurements from sondes, regular aircraft (MOZAIC) and alpine surface sites, *J. Geophys. Res.*, 117, D09301, doi:10.1029/2011JD016952.
- Parrish, D. D., Lamarque, J.-F., Naik, V., Horowitz, L., Shindell, D. T., Staehelin, J., Derwent, R., Cooper, O. R., Tanimoto, H., Volz-Thomas, A., Gilge, S., Scheel, H.-E., Steinbacher, M., and Fröhlich, M., 2014, Long-term changes in lower tropospheric baseline ozone concentrations: Comparing chemistry-climate models and observations at northern midlatitudes, *J. Geophys. Res. Atmos.*, 119, 5719–5736, doi:10.1002/2013JD021435.
- R Core Team, 2018, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Simpson, D., Arneth, A., Mills, G., Solberg, S., and Uddling, J., 2014, Ozone – the persistent menace: interactions with the N cycle and climate change, *Current opinion in Environ. Sustainability*, 9, 9–19. <https://dx.doi.org/10.1016/j.cosust.2014.07.008>.
- Solberg, S., Colette, A., and Guerreiro, C., 2015, Discounting the impact of meteorology to the ozone concentration trends ETC/ACM, Bilthoven.
- Stegehuis, A. I., Vautard, R., Ciais, P., Teuling, A. J., Miralles, D. G., and Wild, M., 2015, An observation-constrained multi-physics WRF ensemble for simulating European mega heat waves, *Geosci. Model Dev.*, 8, 2285-2298, 10.5194/gmd-8-2285-2015.
- Stohl, A., L. Haimberger, M.P. Scheele, and H. Wernli, 2001, An intercomparison of results from three trajectory models. *Meteorol. Applications* 8, 127-135.
- Wood, S.N., 2017, *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC Press.