**Replies to Referee 1:**

We thank the referee for the thoughtful and careful review. We highly appreciate the comments and suggestions for improving the manuscript. Our replies to the specific comments (in *italic*) are as follows:

*1) Page 14,202, line 4: insert "with emissions" between "were performed" and "for 1990"*

The sentence was revised as suggested.

*2) Page 14,202, lines 9-11: these statements appear to be based on the results shown in Section 3.2. However, as stated in that section, no PM2.5 observations were available that span the entire 1990 – 2005 time period. While the rates of change inferred from more recent measurements may lend some support to the modeled trends, these observations simply are too sparse or for PM10 instead of PM2.5 to be able to draw quantitative conclusions about how well the model captured emissions-driven changes. I recommend rewording this section to more accurately reflect the findings discussed in Section 3.2 and to avoid overstating the accuracy of the model results. As discussed in my general comments, it would be best to perform model simulations for the 2000 – 2010 time period when more observations were available to establish that the modeling system can capture emissions-driven changes.*

The statements about the changes in PM2.5 between 1990 and 2005 are based on PM10 trends for that period and on PM2.5 measurements after 1998. We agree with the referee that there is no direct comparison of modeled and measured PM2.5 for the period between 1990 and 2005. Further simulations for all of the 2000 – 2010 period however, is beyond the scope of this study, therefore we revised the abstract and Section 3.2 to emphasize this issue. In order to strengthen our statement about how well the model captured emissions-driven changes, we added a figure (Fig. S3 in the Supplement) in the revised manuscript to show the measured time series for PM10 together with the modeled values for 1990, 2005 and 2006 in this study, as suggested by the referee in comment 19.

*3) Page 14,202, line 19: insert "average" before "ozone levels in polluted"*

We inserted "average" as suggested.

*4) Page 14203, line 12: please specify which aspects of the ozone distribution were discussed in Wilson et al. (2012), e.g. annual mean hourly values, annual mean daily* maximum values, summertime values, etc.

Wilson et al. (2012) reported that anthropogenic NOx and VOC reductions did not have a substantial effect on observed annual mean ozone trends with the exception of Austria-Hungary. We revised the sentence.

*5) Page 14,205, lines 12-13: please discuss the impact of this choice on the simulated ozone changes and their comparison to observations. In particular,*

*this model configuration does not allow the simulation of stratospheric ozone influences on surface concentrations while such influences may affect the magnitude, interannual variability and potentially trends of the observed concentrations used for model evaluation.*

We used 31 layers going up to 100 hPa in the meteorological model WRF and 14 layers in CAMx. Although CAMx is a chemical transport model for tropospheric simulations, boundary conditions come from a global model. Ordonez et al. (2007) suggested that the positive ozone trends and concentration anomalies in the lower free troposphere over Europe during the 1990s were likely to a large extent due to enhanced stratospheric ozone contributions, particularly in winter–spring. The results published by Cui et al. (2011) on the other hand, do not support the study of Ordonez et al. (2007).

The emphasis in this manuscript is not specifically on long-term trends but on the influence of changes in anthropogenic emissions on surface concentrations using the same meteorology and only changed emissions. We performed therefore the model evaluation only for 2006 using the meteorology and emissions for that particular year.

*6) Page 14,205, line 18: please insert "annual mean" before "ozone"*

Done.

*7) Page 14,205, lines 25-26: Please provide more details on how the 1990 boundary conditions were prepared. Were 5 ppb subtracted from the ozone concentrations in all layers for all boundaries and all hours of the year? How were boundary conditions for other species that may affect ozone (e.g. PAN, CO) prepared? And how was seasonal variation taken into account, and for which species?*

Initial and boundary conditions were obtained from the MOZART global model for 2006. Originally, we kept the boundary conditions constant in all simulations assuming that the background ozone did not change significantly since 1990. Comparisons with observations however, suggested that background ozone might have been lower in 1990. Additional simulations with lower background ozone improved the model results. Changing the background ozone in the boundary conditions was done by reducing each value by 5 ppb (based on Logan et al., 2012) in each layer for all boundaries and for each hour. Ozone values varied seasonally, being highest in summer, and lowest in winter. We revised the sentence in the manuscript to make it clearer. We did not change the concentrations of other species. According to our experience, uncertainties in emissions affect the concentrations more than the boundary conditions in this rather large domain.

*8) Page 14,205 line 27 – page 14,206 line 1: Which year were the ozone column densities extracted for – 2006 only? If so, how sensitive would the 1990 results be towards using ozone column densities that reflected conditions in that year?*

We used the TOMS dataset for extracting the ozone columns. The data however is only available for years after 2004 (not for 1990). Since our emphasis in this project was given to the changes in emissions, we kept all other input parameters and meteorology as in 2006. The changes in ozone column densities might have some effect (depending on how much the change was) on the photolysis rates. This question needs to be handled by much more detailed sensitivity tests and it is one of the topics we are working on in a new project. It is, however, beyond the scope of this study where our emphasis was on emissions-driven changes.

*9) Page 14,206, line 14: Was the replacement of TNO/MACC emissions with Swiss emissions performed for both the coarse and fine grid?*

Yes. We modified the sentence to make it clear in the revised version.

*10) Page 14,207, line 12: Please define SNAP.*

SNAP (Selected Nomenclature for Air Pollution) was defined in the revised version.

*11) Page 14,208, lines 17 – 19: What were the results of this evaluation of meteorological variables? They do not appear to be presented in this manuscript. Also, why was the evaluation of meteorological variables performed only for a small portion of the modeling domain?*

We added some information about the evaluation of meteorological variables in the Figure S1. In general the agreement between measurements and model results was quite good with high correlation coefficients (0.76-0.98) and low mean bias error, MBE (0.00023 for specific humidity, -1.13 for air temperature, 0.57 for wind speed). These values fulfil the desired accuracy suggested by Cox et al. (1998) which is 2 $^{o}$C for temperature, 1 m $s^{-1}$ and 2.5 m $s^{-1}$ for wind speeds < 10 m $s^{-1}$ and > 10 m $s^{-1}$, respectively. The reason of choosing the nested domain for evaluation of model results was simply the higher horizontal resolution (resolution of the fine grid is 3 times higher than that of the coarse grid). The chemical parameters such as ozone and PM2.5, however, were compared with measurements (AirBase) in the coarse domain as well.

*12) Page 14,208 – line 20 – Page 14,210, line 14: It would be good to put the 2006 model performance summarized here in the context of other recent 2006 simulations over Europe, e.g. the AQMEII Phase 1 simulations analyzed in Solazzo et al. (Atmospheric Environment, 2012 a and b)*

Thanks for this suggestion. We revised this part as suggested.

*13) Page 14,209, Lines 1-2: Are these 19 sites equally distributed throughout the modeling domain or concentrated in a particular area? Adding this information would help with the interpretation of the results*

These sites cover a large part of Europe between -8.8 and 27.7 degrees from west to east, and between 37.3 and 60.5 from south to north, including

measurement stations in Great Britain, Sweden, Finland, Portugal, Spain, Belgium, Germany and Czech Republic. We added some information in the text.

*14) Page 14,209, lines 2-3: Was analysis performed to confirm that the the model did not capture the strength these inversions? And if so, how did this affect simulated wintertime ozone concentrations?*

The winter period in 2006 was analyzed in detail in an earlier study with MM5/CAMx models and discussed in Aksoyoglu et al. (2011). There was a 10-day period in January 2006 with very low temperatures and low wind-speed. Although the meteorological model is different in this study (WRF), capturing inversion periods still remains challenging. This leads to underestimation of pollutant concentrations during such periods. It is especially important for PM concentrations.

*15) Page 14,209, lines 5-6: Where is this shown? I suggest adding a time series for ozone similar to the PM2.5 time series shown in Figure 2.*

We agree and added the time series for ozone in Figure 2 as well.

*16) Page 14,209, lines 7-8: Are the results similar for other stations in the modeling domain? Focusing this analysis on two stations only does not allow general conclusions about model performance at rural vs. urban sites.*

The two stations were shown in Fig. 3 as representative for rural and urban sites since the results are similar for other stations.

*17) Page 14,209, lines 10-12: The ozone distributions at Cheaumont are not "very similar", the median of the model distribution is shifted towards the left of the observed distribution and the modeled distribution does not have any of the observed high values above 70 ppb.*

The description of the frequency distributions was probably not clear in the manuscript. What we meant with "very similar" in case of Chaumont was the shape of distribution (higher values in the middle) which was not the case in Zurich (measurements have highest frequency at the lowest concentrations, the model on the other hand, shows highest frequency in the middle). We improved this part in the revised version to make it clear.

*18) Page 14,210, lines 16 – 28: Given that more observations became available starting around 2000, it would have been good to perform simulations for 2000 and 2010 emissions to see how well the model captured particulate matter changes over that time period. This would have been a more direct way of trying to establish the model's credibility in capturing emissions-induced PM changes than the current discussion in this section that is more qualitative because of the lack of observations for the time periods that were modeled. While it may be beyond the scope of the study to perform such additional simulations, the authors may want to add appropriate caveats that a more quantitative analysis to establish the modeling system's ability to capture PM trends is still necessary.*

We agree that the lack of PM measurements in the early nineties is a problem for a quantitative analysis. For such a quantitative evaluation of model data however, more detailed simulations covering the whole period between 2000-2010 using the corresponding input data (meteorological, boundary conditions, emissions) are needed. This is however, beyond the scope of this study. Based on observations between 1998 and 2010, our findings suggest that the model captures the changes in PM reasonably well (see also Fig. S3 plotted for the next comment). We added some discussion about this issue in the conclusions.

*19) Page 14,211, Lines 1-4: It might be good to include a time series showing the 1991 – 2008 observations along with the model values simulated using 1990, 2005, and 2006 emissions (all using 2006 meteorology) to get an idea of interannual variability in the observed trends and how the model predicted emissions-induced change compares to this interannual variability.*

Thanks for the suggestion. The observations between 1991-2008 mentioned in the manuscript were for PM10. We included therefore the modeled PM10 data (black crosses) in the time series for some of the stations in Barmpadimos (2011) which shows both raw data (blue) and data after the adjustment for meteorology (red). The model shows a similar interannual variability as the measurements. (Fig. S3).

*20) Page 14,211, Lines 20 – 23: Please also add a discussion on the impact of the choice of boundary conditions for the 1990s on the results shown in Figure 7. For many areas, the increase between the simulations with 1990 and 2005 emissions appears to be close to the assumed boundary condition increase of 5 ppb.*

We don't think that the increase between 1990 and 2005 is close to the 5 ppb increase in the boundary condition because the absolute change shown in Figure 7 varies spatially between -5 and +18 ppb (1-2 ppb in northern Switzerland, 0-3 ppb in western Europe, 3-6 ppb in northern Germany, 10-15 ppb in eastern Europe, and about -3 ppb in southeast Europe). However, we followed the suggestion of the referee and checked the impact of the choice of boundary conditions for 1990 on results shown in Fig. 7. We performed another simulation with 5 ppb higher boundary conditions for 1990. The results suggested that the changes shown in Fig. 7 would be 1-2 ppb lower. However, long-term measurements show a positive trend of 0.32 ppb/y for the period between 1990 and 2008 (Cui et al., 2011) supporting our choice of 5 ppb increase between 1990 and 2005. We added some discussion about it in the text.

*21) Page 14,211, Line 26 – Page 14,212, Line 7: Why are the results in Tables 3 and 4 only calculated for the Swiss stations? Aren't EMEP/Airbase observations for 1990 and 2005 available to also analyze the observed and modeled ozone changes in other countries?*

We prefer to use the results of the nested domain for any comparison with measurements because of the higher horizontal resolution. In order to

compare the model results with the stations in the other countries, we have to use the results in the coarse domain (3 times coarser horizontal resolution).

*22) Page 14,212, line 27: same comment as above – why are results only shown for the Swiss stations?*

Please see our reply to the previous comment.

*23) Page 14,213, lines 16-17: Please add more discussion on the role of the choice of boundary conditions for the 1990s on the results shown here. It may even be worth performing additional sensitivity simulations with higher and/or lower boundary conditions for the 2005 and/or 1990 scenarios to more fully assess the impact of these choices on changes in AOT40 and SOMO35.*

Both measurements and model results suggest that peak ozone concentrations decreased due to emission reductions between 1990 and 2005 whereas low concentrations (winter, night) increased because of less NO titration especially at polluted areas. Since AOT40 is the sum of ozone concentrations above 40 ppb, changes around this threshold might affect the AOT40 values significantly. As mentioned earlier, we performed another simulation with 5 ppb higher boundary conditions for 1990. Increasing the boundary conditions led to an increase in AOT40 roughly about 5%. Although this indicates the importance of the boundary conditions for the indicators, it needs much more detailed sensitivity studies.

*24) Page 14,215, lines 1-5: Are there any observations these modeled changes could be compared to? If so, this information should be added.*

Unfortunately observations of atmospheric nitrogen deposition are severely restricted in spatial extent and type. Deposition estimates cannot be directly assessed because of lack of measurements, especially of the dry deposition component. We are not aware of any reliable measurement data for total nitrogen deposition to compare with the modeled change between 1990 and 2005.

*25) Page 14,215, lines 20-21: See my previous comment on Page 14,202, lines 9-11*

We think that the additional figure implemented in the Supplement (Fig. S3) made an improvement and we revised this paragraph.

*26) Page 14,216, lines 13-15: Weren't the 2006 boundary conditions obtained from MOZART, not observations? Please clarify this statement.*

Boundary conditions were obtained from MOZART for 2006 and used for 2005, 2006 and 2020 simulations. As we reported in Page 14,205, lines 25-26, however, background ozone mixing ratios were set 5 ppb lower for 1990 simulations following the findings from observations reported by Logan et al. (2012). We modified the statement to make it clearer in the revised manuscript.

**Replies to Referee 2:**

We thank the referee for the review. We improved the description of the methodology regarding the model resolution and lateral boundary conditions in the revised manuscript as suggested. Our answers to specific questions (in *italic*) are as follows:

*- I would suggest to comment on the forecast skill of the adopted methodology for projections of 2020. Is O3 increase attributed to emissions used or meteorology?*

As described in Page 14,205, we calculated the meteorological fields for 2006 and used the same meteorology for all emission scenarios. The changes predicted by the model simulations therefore are attributed to changes in emissions.

- *I would suggest to add a description about the model resolution used for Switzerland and the topographic variability of the area. Why it is considered as adequate? The same is true for the emission inventory used. How about meteorological variability? Is it captured adequately with the model configuration used?*

The horizontal resolution of the coarse domain (European domain) is 0.250deg x 0.125deg. Since Switzerland has a complex terrain, a higher resolution was used for the nested Swiss domain, 0.0833deg x 0.0417deg. We think that the model resolution used in this study is an optimum choice for whole year simulations over complex terrain. The emission inventory used for the European domain has a resolution of 0.125º x 0.0625º. The resolution of the emissions inside the Swiss domain is higher. We improved the description of the model resolution as well as the resolution of the emission inventories in the manuscript. Some statistical evaluation for the meteorological variables was added into the Supplementary (Fig. S3)

- *For clarity of the presentation of results i suggest redraw the figures showing model results for Europe. The color palette they used is too light and the results are not clearly seen.*

We changed the color scale of all the figures showing model results not only for the European but also for the Swiss domain for consistency.

- *The time series of Figure 2 and 4 need a better representation. Too long time series of observational and modeled data. I suggest to break in smaller period.*

The purpose of Figure 2 is to show how the model captures the temporal variation. We therefore showed the whole year. However, in order to simplify the figure we plotted daily average instead of hourly values in the revised manuscript. On the other hand, Figure 4 shows data only for one month (intensive campaign period for aerosol components in June) and we think that splitting it into a smaller period is not necessary.