Response to reviewers – Pre-industrial to end 21st century projections of tropospheric ozone…

We are grateful for the comments of both reviewers, and we hope that our changes and the information below address their comments and concerns. Our responses follow each reviewer comment in blue.

A major improvement to our revised manuscript is the inclusion of some ozone budget data. We revisited the ozone budget output from the ACCMIP models and we have managed to provide data for 5 models. The budgets are not calculated in a consistent manner – so we cannot talk about an ACCMIP mean – but they do provide a generally consistent picture on the increases/decreases of the budget terms relative to the present day.

In addition to any changes in response to reviewer comments, we have addressed some small errors in the text as well as added additional model output as it became available. These are:

- Corrected errors in the mean burden in Table 1
- Addition of RCP2.6 output for CMAM (2030 and 2100 time slices)
- Addition of 2030 time slice data for RCP8.5 for EMAC
- Addition of 2030 time slice data for RCP2.6, 4.5 and 8.5 for HadGEM2
- Small corrections to emissions totals

Responses to reviewer #1
(a) Main concerns:
R1.1. It would be good if anything is said regarding the possible improvements. Where are the main source of errors located? Would a more sophisticate emission dataset lead to large improvement? Although I understand that without a rigorous ozone budgeting this is not quantitatively possible, I still think that even a qualitative assessment will improve the paper. Some of what I would like to see is in some degree already mentioned in the manuscript (see Sect. 6), but it would improve the paper if these interesting interpretations were highlighted in the abstract/conclusions. For example, focusing on the past-future ozone estimations, the stratosphere-troposphere exchange of ozone is shown to be very important (see Page 21640). Is it possible to estimate somehow the relative influence with respect to the other analysed key factor (i.e. methane and NOx)?

A1.1 Regarding the main sources of errors: For the ozone column, we already highlight that there is a high bias in the NH and low bias in the SH, which is generally present in all the models. As the emissions are the only (largely) consistent driver of ozone between the models, this could suggest that the emissions are deficient in some way. However, the biases could also reflect shared deficiencies in transport or chemical schemes.

In general, we would argue that it is not possible to accurately pinpoint the major sources of errors in the models with this kind of study. It is likely that the models and their inputs are deficient in all manner of ways, including – but not limited to – emissions, chemistry schemes, missing processes and limited resolution. The purpose of this study is to present how models perform and where models agree, with respect to present day ozone, and past and future projections. We would suggest that a quantitative and qualitative assessment of model performance would be best conducted using well-constrained sensitivity experiments, perhaps in a future model intercomparison study with those particular goals in mind.

Regarding the strat-trop exchange: At the end of section 5 we state

"….[based on emission changes alone,] we would expect an ozone burden increase of approximately 30 Tg between Hist 2000 and RCP8.5 in 2100. However, at 60 Tg the ACCMIP multi-model mean increase in ozone is double that expected, and **consistent with equal roles for methane increases and the net impacts of climate change**, i.e. through promoting increased influx of stratospheric ozone, changing LNO_x, and impacting reaction rates, through temperature and humidity changes."

The limited ozone budget data that we now have strengthens this conclusion, and we will show an increased stratospheric contribution of 20-140% for RCP8.5 (2100 time slice) compared to the Hist 2000. We will be sure to highlight these findings in our revised abstract and conclusions.

With regard to attribution of the ozone changes to NOx, VOCs, CO and methane, we make clearer reference to the related ACCMIP work by Stevenson et al. in our revised version.

R1.2. I know that this sounds like a race between models, and it is not my intention to suggest it. However any model developer could learn a lot about its own model comparing his results with observations and other models as well (especially the ensemble one, which should give the best results). The deep analysis carried here would be very profitable for all models. Hence I suggest to add in the electronic supplement some comparison based on a selected metric and add a chapter in the text summarizing such results. This is partially present in the tables, and, in theory, this information could be infer from the figures in the supplement. However, I got lost in the hundreds of figures, and an additional summarizing plot (such as Taylor's diagram) is needed. Finally, please compare also quantitatively your results with previous results from other multimodel ensemble: is dry deposition or stratospheric-tropospheric exchange in your ensemble still comparable to what obtained by Stevenson et al. (2006)? In the manuscript I could find only one quantitative comparison to previous results (i.e. Page 21631, line 14-15), while the others were only included in the figures.

A1.2 Tables 2 and 3 provide information on the comparison between the OMI column and the ACCMIP models (including the ensemble mean), which – as stated in the text – gives information consistent with the ozonesonde comparisons. We are not sure that additional information will provide anything new here, and (anecdotally) many people find Taylor diagrams confusing, especially when used for many models. Instead, we propose to include lat/lon plots of the model-OMI biases in the supplementary material, in addition to the multi-model mean plot already in the paper. This way, readers and model developers can see where a given model's bias is the same as (or different to) the ensemble.

Regarding the general comments about the supplementary material: Most of the figures show changes in tropospheric ozone (surface, column and zonal mean) between the time slices for individual models, so that the interested reader can see for themselves the agreement (or not) between the model results. They are mostly not for model evaluation (although see previous paragraph).

Regarding quantitative comparisons: The lack of budget information meant that we could not make comparisons to ozone budget terms calculated by the Stevenson study, although – with the new budget information – this will now be rectified. As for the ozone burden, we already state:

"Values for The mean burden is 337 ± 23 Tg, very close to the 336 ± 27 Tg reported for a subset of the ACCENT models (Stevenson et al. 2006), and the 335 ± 10 Tg estimated from measurement climatologies by Wild (2007), although the latter estimate is from pre-2000 ozone data."

(b) Minor comments

R1.3. Page 21618, line 5. What do you mean with "well"? Could you add a correlation value?

A1.3. This sentence is merely to introduce the subsequent two sentences that discuss model performance. Here, "well" encompasses the full range of sonde and satellite comparisons made in the paper, and is therefore not easily summarised in a single value. We think that this is sufficient for the abstract.

R1.4. Page 21620, line 11. "ozone precursor emissions [..] were high". High with respect to what? Garnaut et al (2008) suggested that the previous emissions are underestimated for greenhouse gases. Could it be the same for ozone precursors as well? Could you also conciliate these different points of view?

A1.4. This text refers to the projected precursor emissions prescribed by the SRES scenarios, as used before the RCPs. Precursor emissions were scaled to the projected changes in GHGs in these earlier scenarios, which, in particular, resulted in high NO_x emissions in the SRES A2 scenario (driving large tropospheric ozone increases). The more recent emissions data allow for changes in GHGs and ozone precursors to be decoupled, as the latter are reduced to mitigate air quality issues. The RCPs allow for this decoupling, but (of course) they might not represent the full range of possible futures.

R1.5. Page 21623, line 8-16. Although tedious I think it is necessary to list the meaning of the acronyms or at least a reference for each model, once mentioned for the first time.

A1.5. We would prefer to leave the reference to the Lamarque et al. GMD paper, which details the ACCMIP study and the models in more detail. All the manuscripts relating to ACCMIP are collated under a joint GMD/ACP special issue, so it should be relatively easy to find this information.

R1.6. Page 21626, line 4-5. The t-Student's distribution converge to the normal distribution only for large number of degree of freedom. In your case you have limited number of models (14 or 15 models). Hence I expect that the 95% confidence level not to be given by the 1.96 times the standard deviation. Please check.

A1.6. The t-distribution converges to the normal distribution for infinite n, but according to Wilks (2006) "even for moderately large values of n…the t distribution is closely approximated by the standard Gaussian distribution". However, we take the point and have changed our practice to quoting 2-sigma uncertainties (approximately equivalent to the 95% confidence interval). This change has effects Figures 9-11, although quite small.

R1.7. Page 21630, line 5. " Figg.2d and 2g show higher ozone levels over source regions". I expect that the authors mean regions where ozone precursors are emitted. Due to page formatting is difficult to see the figures (I suggest to enlarge them in the revised version). However from Figg.2d and 2g the Mediterranean basin has higher ozone concentrations than central Europe. Is this also a sign that precursors are more emitted over the Mediterranean basin than, let's say, France or Germany? Please reformulate this sentence.

A1.7. This sentence has been re-written to emphasize that we mean precursor emissions (changes in bold):

"…also show higher ozone levels over **ozone precursor** source regions, **the** plots also indicate enhanced concentrations downwind of the **these** regions, due to transport of ozone **and** ozone precursors, including…"

Higher concentrations over the Mediterranean reflect the impact of local emissions, transport of non-local ozone/ozone precursors, and the meteorological conditions. In the interests of brevity, we discuss the ozone distribution in very general terms, not highlighting smaller regions. As such, we do not specifically bring out the Mediterranean results in the revised manuscript. (Some regional ozone concentrations are discussed with respect to the validation.)

R1.8. Page 21630, line 17-18. What about different Planetary Boundary Layer Height (PBLH)? Could it also play a role in the variability between model at the surface? If so, how?

A1.8. PBL schemes could indeed be one of the many causes for inter-model differences, impacting the mixing of emissions and transport. It is not clear what the nature of this impact might be, although a recent paper by Menuet et al. (2013, "On the impact of the vertical resolution on chemistry-transport modelling", Atmos. Environ.) touches on aspects of this and concludes that changes in the vertical resolution are not a big impact on their air quality model results.

R1.9. Page 21631, line 4. I am intrigued by the large variability of the ensemble in the Arctic. Could it be related to precursor transport from Europe and its relation with the North Atlantic Oscillation (see Christoudias et al., 2012)? From the introduction I deduce that the model were forced with a climatology of the year 1995-2005. Were the model perhaps presenting a persistent positive NAO index, implying a persistent northward transport of pollutants from Europe?

A1.9. If all the models were in a persistent positive NAO phase then any inter-model variability associated with this should be low. However, we recognise the importance of NAO phase for pollutant transport to the Arctic and it could be that the climatologies from particular models are biased to a given phase of the NAO. Revised text:

"But larger uncertainty for the surface at high latitudes could be related to differences in precursor transport and chemistry from lower latitudes (Shindell et al., 2008), **particularly if there is a spread in the phase of the North Atlantic Oscillation in the climatologies calculated from the models (Eckhardt et al., 2003; Christoudias et al., 2012).**"

R1.10. Page 21631, line 21-25. Is the VOC emissions mostly connected to Isoprene emissions (as the authors suggested in Sec.2.1)? If so, there should be larger standard deviations for future scenario mainly over tropical forest. Is that the case?

A1.10. Where a model includes isoprene emissions (i.e. not HadGEM2 and CMAM), the biogenic VOCs do indeed dominate the total VOC emissions. However, most models (8 out of the remaining 13) do not include interactive isoprene emissions, sticking instead with constant present day emissions. This is now made clear in the revised text, also in response to reviewer 2 (see R2.6).

With regard to the future projections, Fig S6 suggests some spread in the future projections for surface ozone over isoprene emitting regions, which could be related to different isoprene emissions and isoprene chemistry. However, from Fig 11 we also note that the different biogenic VOC emissions do not impact the significance of the surface ozone changes notably. Overall, the diversity in VOC emissions is not a major driver of the inter-model diversity for the projected ozone changes.

R1.11. Page 21633, line 11. From the figure, the ozone levels at 500/750hPa in the NH winter (January) are always overestimated (and outside the standard deviation). This seems to disagree with the text. Please clarify.

A1.11. The text has been revised:

"Both the ACCMIP multi-model mean and median are within the standard deviation of the observations for **most** locations and altitudes**, with the winter NH extratropical comparison being a notable exception. Indeed, c**ompared to the mean observations, the largest relative errors are found for the NH extratropics…"

R1.12. Page 21633, line 17. I like this point very much. Is it possible to guess why there is such improvement?

A1.12. It would indeed be a guess (and therefore we would not add it in the manuscript), but it could due to better representation of seasonally varying emissions (such as wildfires), or transport pathways, including the role of stratospheric ozone. This is a hard thing to pin down without a systematic comparison between a model run under ACCENT versus ACCMIP conditions. Of course, a pessimistic view is that it could also mean several errors cancelling each other out in a different way. See also A1.1.

R1.13. Page 21637, line 5. Please, add a number after "good". Is it $R2 > 0.9$ or > 0.5?

A1.13. The point we were trying to make here is that a model with a higher ozone burden for the present day *generally* has a higher burden for other time slices, but the model with the highest burden for time slice A is not necessarily the model with the highest burden for time slice B. However, we have added that r (not R^2) is > 0.7 for these relationships.

R1.14. Table 1,2, Tab. 3 is exactly the same as Tab.1. You should move Tab.3 to the electronic supplement, as it does not give any additional information.

A1.14. (Presuming that the comment refers to Table 1 and table 4.) While the information in Table 4 can be gleaned from Table 1, we believe that having it available in the main text helps the reader. Furthermore, the entries in the "mean" row for Table 4 are the means of the differences for the given scenario and time slice. I.e., they are not the differences in the ensemble mean burdens in Table 1.

Response to reviewer #2

(a) General comment

R2.1. This is a very nice piece of work, well written and documented, that clearly makes the necessary links to previous similar exercises. However, I do find that the paper suffers from a certain lack of quantitative information about the extent to which different processes contribute to past and future tropospheric ozone (e.g., changes in stratospheric-tropospheric ozone vs. change in lightning NOx, methane, etc.), especially since this is probably the first model intercomparison that includes such a number of models with stratospheric chemistry. In other words, I find it a bit surprising that none of the groups involved in this study have archived the necessary diagnostics to further discuss the past and future ozone budget, and that even if these diagnostics are not entirely consistent throughout the model suite, it is not possible to further discuss the relative contribution of these key processes.

A2.1 We agree that it was disappointing, and clear instructions encouraging consistent budget diagnostics have been included in the document outlining the next major model intercomparison, CCMI (Chemistry-Climate Modeling Iniative). CCMI will have more models than ACCMIP, and many will have a complex representation of the stratosphere.

However, there is reasonably good news regarding the current study. As we noted at the start, we have managed to get ozone budgets from 5 particpating models. Even though these budgets are not consistent (esp. in their definitions of production and loss), they do enable us to make some addition conclusions, especially with respect to the stratospheric influx, e.g. the 20-140% increase in RCP8.5 2100 compared to present day.

We cannot say much about the contribution of the other drivers (e.g. lightning NOx changes), but the Stevenson et al. ACCMIP paper does attribute ozone changes to NOx, VOC and methane changes. We now include a clearer reference to this work.

(b) Minor comments

R2.2. Page 21620, lines 21-24: It is written that the isoprene flux depends on climate but that whether future climate changes will drive isoprene increases or decreases is not clear. Are the Authors only talking about the climate changes or also about changes in CO2 concentrations (which may counteract the changes due to climate change to some extent)?

A2.2. We agree that this is not clear. We meant that climate changes generally drive an increase (through increases in temperature), whereas CO_2 increases could offset this effect to some degree. Changes in land use could result in higher or lower emissions, depending on the nature of the land use change (e.g. afforestation, changing species composition), as we mentioned. The first sentence has now been re-written:

"Its emission flux depends on climate **and (inversely) on CO2 concentration** (Guenther et al., 2006; Arneth et al., 2010), **and whether** future isoprene **emission projections suggest** an increase (Sanderson et al., 2003; Lathière et al., 2005) or decrease (Arneth et al., 2007a; Young et al., 2009) **depends on whether the CO2 dependency is excluded or included** (see also Pacifico et al., 2009)."

R2.3. Page 21623, paragraph starting lines 17: It appears that some models do not include NMVOCs (e.g., isoprene) in their simulations. I wonder to what extent it makes sense to include such models in the assessment, since there is now a clear recognition that NMVOCs do impact tropospheric ozone burden. The global ozone burden from the specific model that does not include NMVOC does not seem to be completely different from the others but I wonder whether there are compensating effects that make the global budget "artificially" correct, and whether this could have an impact on the projections of future ozone changes. I understand that the Authors may not have all the ozone budget data available but could they comment on this?

A2.3. While it is true that the role of NMVOCs for tropospheric ozone is well established, there are a myriad of other key factors that the models do and don't do well and/or comprehensively. To pick a few: increased resolution (vertical and horizontal), comprehensive stratospheric chemistry (for better strat-trop links), interactive photolysis rates, and links with aerosols. For instance, CMAM might have the least complex chemistry scheme (no VOCs), but it is one of the more complex models in terms of its stratospheric physics and chemistry – potentially a key component for future tropospheric ozone, as we highlight.

We describe that there is a correlation between total VOC emission and present day ozone burden (although not 1:1), but we also note that there is not a consistent relationship between the magnitude of the present day ozone burden and the magnitude of the changes in ozone burden. That is, the models respond in different ways to the emission and climate changes, and the present day ozone burden is not a good predictor of these responses. This would suggest that, even if there were good reasons, excluding certain models from this analysis would not substantially limit the uncertainty for the ozone changes.

The whole issue of making "weighted" climate projections (an end to "one model, one vote") is an active research area within the climate modelling community, and we are sure that it is a topic that will be revisited within the context of tropospheric ozone modelling. However, it is not a task that we can undertake with confidence in this study.

R2.4. Page 21624, lines 4-16: I understand that the RCP scenarios are very "popular" and are being used in many studies, but still I think it would be useful to briefly present the underlying hypotheses that were used for deriving these RCP scenarios and result in decreasing ozone precursor emissions. A similar remark could be made for the lines 13-16 pages 21620: why do more recent emission projections include reduced anthropogenic precursor emissions? Do they assume strict air quality regulations?

A2.4. The RCPs are all thoroughly described in a Climatic Change special issue, which we will highlight in our revised version. More recent emissions projections allow for changes in GHGs and ozone precursors to be decoupled, as the latter are reduced to mitigate air quality issues. This was not the case with the SRES emissions (circa IPCC third assessment), which are now believed to have had too strong an increase in ozone precursors, particularly NOx in their "pessimistic" A2 scenario. The RCPs do assume strict air quality regulations, under the assumption that effort to control air pollutant emissions increases as economies grow.

R2.5. Page 21628, lines 9-11 and lines 25-28: It is said that "An increase in LNOx from 2000 to 2100 (RCP8.5; strongest warming) is generally robust across the ACCMIP models, and ranges in magnitude from 10–75%". Another "robust" result appears to be an increase in total VOC emissions for many models because of a climate-driven increase in isoprene. Are those results "robust" because all models include the same parameterization for LNOx and isoprene emissions, or does that say something about the "robustness" of other processes?

A2.5. Only 3 models include online isoprene emissions, which we make clear in our revised version. The models do use a similar parameterisation and the emission projections are similar (all increase; no account of CO2 inhibition), but differences in projected temperatures and (for some models) land cover change will result in model diversity.

For LNOx, the results are indeed "robust" because most models use the same convective mass flux-based parameterisation, with CMAM (using the parameterisation of Allen and Pickering, 2002) being the outlier in terms of the projections. We already state this in the text, as well as the fact that "[c]learly further study is required into the implications of the use of different parameterisations for LNOx, and the different sensitivities across models."

R2.6. On Figure 1, it would also be nice to include an additional plot to illustrate the changes in anthropogenic versus biogenic VOC emissions (similarly to what was done for LNOX), also to give information on whether the changes in VOC are driven by the RCPs or the changes in biogenic emissions.

A2.6. This is indeed a good idea and will be included in the revised version.

R2.7. Page 21633, lines 16-17: Bias and correlation appear to be improved in comparison to the ACCENT mean for the NH tropical mid and upper troposphere. Do the Authors know why there is an improvement upon the ACCENT results?

A2.7. As we said in our answer to reviewer 1 (A1.12), without a systematic comparison with a model running with ACCENT and ACCMIP conditions, anything we say is really just a guess and we would rather not put too much idle conjecture into the manuscript. We hope that future model intercomparison studies will have more data archived to be able to address this kind of comment.

R2.8. Page 21637, lines 8-10: I am not sure I understand what is meant there.

A2.8. This is making a similar point as in our response to your third comment (see A2.3). E.g. just because a given model simulates the largest change in the ozone burden between Hist 1850 and Hist 2000, it does not necessarily simulate the largest change between Hist 2000 and RCP8.5 2100. We have clarified the final clause of that sentence:

"i.e. there are not any models that consistently simulate large (or small) ozone changes **between time slices**…"

R2.9. Typo in the legend of Figure 1: respectively.

A2.9. Fixed. Thanks.