We thank the reviewer for their constructive comments on our paper. Those comments are in italics below, while our replies follow in standard font.

Reviewer 1 (William Collins)

This is an interesting paper that both evaluates a chemistry climate model, and evaluates the impacts of ozone and methane changes on historical and future climate change.

The climate impacts experiments are novel and scientifically important (assuming they are robust). In fact I would have liked to seen these discussed further. The authors do nicely link the evaluation and experimental sections through the radiative forcing. Even so, they might want to consider splitting the paper. I worry that the really interesting later parts get a bit lost amongst all the evaluation.

The evaluation is very comprehensive and covers many aspects. However it is not always easy to tell what the overall message is. It would be useful if the model performance could be related to others too, maybe using some of the CCMVal diagnostics and ACCMIP comparison (e.g. Young et al. submitted). This would help add some perspective. e.g. where a disagreement with observations is found, is this the same for most models or is E2 an outlier? Hopefully over time the community will converge towards some common diagnostics for model evaluation papers. Are there aspects of the evaluation that are particularly new and innovative? If so, maybe these could be emphasised more.

As in our replies to specific comments below regarding this last point, we have added several comparisons with other multi-model analyses.

Section 2.1: I'd suggest removing the description of E2-H as it is hardly used in the evaluation.

The E2-H model is seldom shown, as it typically is very similar to E2-R, but it is relevant to the variability discussion around Figure 13 (and is shown in Figure 12 and Table 2). Hence we feel it's important to introduce it at the start, and it only takes one line.

Section 2.2-2.3: It would be useful to have slightly more description of how the chlorine and bromine compounds are derived from the single CFC tracer.

We have added a description of the assumptions used, which reads "As we use only a single CFC tracer, the ratio of anthropogenic bromine to chlorine is held fixed at year 2000 values, with both released from CFC photolysis in an amount proportional to the total equivalent effective stratospheric chlorine loading in a given year."

Page 23518, line 23. Note HadGEM2 has not submitted any of the data from interactive methane runs.
Thanks for pointing this out. We've rephrased to simplify this and now simply say that all other groups' currently available CMIP5 simulations used the MAGICC-derived concentrations.

Page 23519. In tuning the wetland emissions, the authors are implicitly assuming that all other emissions and loss rates are correct. This should be made clear.

We have added a sentence along these lines to clarify this point at the end of the second paragraph in section 2.4.

Section 3.1.4. The authors could compare the ozone metrics and budgets here against previous studies (particularly multi-model ones).

We appreciate this suggestion, which we have applied broadly. In section 3.1.4, we have added comparison against the multi-model ACCENT burden and budget values and the ACCMIP burden values reported in Young et al (2012) as well as the burden estimated from observations by Wild (2007). We have also added comparison to the multi-model analysis of models against the OMI/MLS tropospheric ozone climatology reported in Young et al (2012). In addition, we added comparison of the correlations analysis with MOPITT CO to prior multi-model results in the last paragraph of the CO section 3.3, and a comparison of ozone trends versus total column observations reported in a recent multi-model analysis in the first paragraph of section 4.1. We hope additional standard metrics can be developed.

Page 23534, lines 1-14: Do the NO2 biases (both low and high) correlate with ozone biases? If so, does the NO2 cause the ozone bias, or do both have a common cause (e.g. transport issues).

A good question. We find that NO2 and ozone biases are correlated in outflow regions, but not elsewhere. We now state this in the last paragraph of section 3.3.1, and describe how biases in both compounds are due to overestimated horizontal transport (as revealed by the autocorrelation analysis).

Page 23534, line 25-28: In the next section the suggestion is that it is VOC oxidation rather than methane that contributes to the upper trop CO. It would be useful if the authors could diagnose the main sources of upper trop CO in their model.

We discussed the production of CO from hydrocarbons, meaning both methane and NMVOCs. However, we've looked at the diagnostics, and have added to the first paragraph of section 3.3.2 that CO production is almost exactly evenly split between production from methane and from NMVOCs, with the latter coming primarily from oxidation of paraffins.

Page 23435, lines 3-6: The NO2 might also be contributing to the ozone bias.

We agree, and have added a phrase stating this.

Page 23538, lines 17-23: It would be useful to see these trend comparisons in a graph rather than a table. Are the results here similar to Lamarque et al. (2010)?

As we already have 28 figures, we are reluctant to add another (and there are so many stations that Lamarque et al (2010) displayed these as two figures). We have, however, added a line comparing with the prior analysis, and to better understand the range in trends due to internal variability in the model, we now give the range across five ensemble members in Table 3. The ensemble information changes the perception somewhat by giving a clearer sense of which model results are truly distinct from the observations (but would make plots very difficult to read, further suggesting we maintain this information as a Table).

Page 23539, line 3: "For comparison with other studies ..." it would be useful if this section presented some comparison with other studies here rather than leaving it as an exercise for the reader.

We have added comparison to the multi-model ACCMIP analysis presented in Young et al (2012).

Page 23539, line 27- page 23540, line 2: Has this difference in variability between
coupled and uncoupled ozone been seen before? If not it may be worth emphasising this more.

We do not believe this difference in variability has been examined before. We’ve added additional discussion of the implication of this result.

Page 23541. How similar are the setups for the GISS and NCAR RTMs? The descriptions read very differently, but I assume they are doing similar calculations?

They are indeed very similar setups, and we have harmonized the two RTM descriptions to clarify that.

Page 23545, lines 4-6. This sentence needs rephrasing. I assume this means that the changes in background state increase the iRF by 0.05 W/m²?

Rephrased to clarify that, as the reviewer assumed, the changes in background add 0.05 W/m².

Page 23547, lines 1-3. Does this RF/iRF ratio also apply separately to the LW and SW? If not, is there a reason to believe the total forcing ratio is more appropriate than the separate LW and SW components?

The GISS RTM calculations did not save LW and SW fluxes separately at the tropopause (only TOA, but as noted in the previous paragraph those can be quite different from tropopause fluxes). While it would be interesting to see if the ratios were the same for the LW and SW components, the net forcing is the end goal in any case, and we see no reason why the RF/iRF ratio for the net forcing due to tropospheric ozone should vary over time (and as the RF is only 6

Page 23548, lines 14-15. I don’t quite understand the relevance of comparing iRFs between the current version and AR4.

Agreed, we have deleted this sentence.

Page 23548, lines 22-24. The cooling might speed up ozone loss in the polar regions (in the absence of circulation changes).

True. We’ve clarified that this refers to extra-polar ozone, which dominates the total stratospheric burden.

Page 23549, lines 20-22. Why are the MAGICC results different for 8.5? Does this imply a different sensitivity of ozone production to methane concentrations?

While it is difficult to determine exactly what relationships between changing temperature, water vapor and constituents were included in MAGICC, it seems likely that the differences in RCP8.5 ozone may stem from differences in the methane concentration itself (see figure 22) as well as potential differences in ozone production chemistry at large methane concentrations.

Page 23549, lines 24-26. I assume the polar superrecoveries are due to increased transport.

That is correct, and we have added a sentence on this to the text.

Page 23556. The heterogeneous and non-local responses discussed here, seem in contrast to the smoother patterns discussed in Shindell et al. 2010. Does this imply that the 2010 conclusions have been superseded? If the ozone RTPs from Shindell and Faluvegi 2009 were used to calculate zonal mean responses, would they agree with the full model calculations (as in Shindell 2012)?

We have now performed the ARTP calculations and compared with the climate model output. A new paragraph discusses these results. In particular, we find that the ARTP works for tropospheric ozone forcing when looking at the tropical and NH mid-latitude responses, but does not work as well for the other regions nor for stratospheric ozone forcing. When discussing the circulation response to stratospheric ozone depletion later in this section, we note that this may be the reason the tropospheric ozone ARTP values are not well-suited to stratospheric ozone forcing. In regard to the 2010 analysis of spatial patterns of forcing and response, we note that that work examined only the
response to aerosol forcing. It may be that ozone, as a short-wave and longwave absorber, causes a different response than aerosols (which are primarily short-wave scatterers). We intend to investigate this in future work comparing the response to ozone and aerosol forcings.

Interactive comment on Atmos. Chem. Phys. Discuss., 12, 23513, 2012.