Authors' comments in reply to the anonymous referee for "Evaluation of ACCMIP ozone simulations using a multi-constituent chemical reanalysis" by K. Miyazaki and K. Bowman

We want to thank the referee for the helpful comments. We have revised the manuscript according to the comments, and hope that the revised version is now suitable for publication. Below are the referee comments in italics, with our replies in normal font.

Reply to Referee #2

Specific Comments:

Abstract: Please mention that a 4-year reanalysis data is used to compare ACCMIP time-slice simulations, and that the evaluation itself can be biased because of this inconsistency.

The abstract has been rewritten to describe the time period of the reanalysis and ACCMIP simulations.

P1L17: Insert anthropogenic between "...important greenhouse.."

Inserted.

P2L1: There are several studies highlighting the use of CTMs/CCMs to assess the radiative impacts of tropospheric ozone prior to Bowman et al (2013). Please acknowledge those.

Several studies are cited in the following sentence.

P2: 2nd and 3rd paragraphs discuss uncertainties in measurements for evaluating chemistry-climate models. I think they can be combined and modified for clarity.

Combined and modified.

P3L12: Please provide motivation for evaluating only tropospheric ozone and not its precursors (e.g., NO2, CO). Presumably biases in ozone are driven by biases in its precursors.

The following sentences have been added:
"Model errors in precursors can also be evaluated using the reanalysis product, and this could help

identify error sources in tropospheric ozone simulations. However, because no other study has shown the potential of reanalysis ozone for model evaluation, this study focuses on tropospheric ozone only."

P4. How is photolysis calculated in the forecast model? Does the model represent methane - prescribed concentrations or emissions? Since the ACCMIP models used different emissions inventory for ozone precursors compared to what is used in this reanalysis product, how would the comparison be affected by this difference?

The following sentences have been added:
"The radiative transfer scheme considers absorption within 37 bands, scattering by gases, aerosols, and clouds, and the effect of surface albedo. Detailed radiation calculations are used for photolysis calculation. Methane concentrations were scaled on the basis of present-day values with reference to the surface concentration."

Because the surface emissions of NOx and CO are optimized using data assimilation in the chemical reanalysis, the difference in the emission inventory should not affect the comparison. In our previous studies (Miyazaki et al., 2012, 2015, 2017), it was confirmed that the a priori emissions do not largely influence the a posteriori emissions through long data assimilation cycles.

P4L13: What is the convection scheme of MIROC-AGCM? Please describe it in a sentence.

The sentence has been rewritten as follows:
"Lightning NOx (LNOx) sources in MIROC-Chem were calculated based on the relationship between lightning activity and cloud top height (Price and Rind, 1992) and using the convection scheme of MIROC-AGCM developed based on the scheme presented by Arakawa and Schubert (1974)."

P5L7: What is DOMINO data? What does the acronym stand for?

DOMINO stands for Dutch OMI NO2. This is described in the revised manuscript.

P5L21: According to Lamarque et al (2013), models were run for several years (up to 10 years) for each time slice. How are the model results for 2000 time-slice handled for comparison with the reanalysis?

The following sentence has been added:
"The number of years that the ACCMIP models simulated for the 2000 decadal simulation mostly varied between 4 and 12 years for each model. Each model simulation was averaged over the simulated years."

To discuss the limitation of the short period of the chemical reanalysis used for the validation, the following sentences have been added in Section 6.1:
"The five-year reanalysis (2005-2009) may cause biases in the estimated model errors in the evaluation of the 2000 decade ACCMIP simulations that used decadal-averaged SST boundary conditions and biomass burning emissions averaged over 1997-2006 (Lamarque et al., 2010). It may neglect the influences of interannual and decadal changes in anthropogenic and biomass emissions and meteorology. Longer-term reanalysis and time-consistent validation are required to obtain more robust error estimations."

P5L29: Young et al. (2013) note that although the models used the same anthropogenic and biomass burning emissions, model-to-model diversity in the implemented chemical scheme resulted in differences in precursor (especially VOCs) emissions across the models. So, the statement "same emissions were used in all the models" is not accurate. Please modify.

Removed.

P6L7: Please provide details on how the ACCMIP model monthly ozone concentrations were interpolated to 2 hour temporal resolution? What diagnostics from the reanalysis were used to compare with observations - monthly or hourly ozone? Some clarification is needed here.

The sentences have been rewritten as:
"The two-hourly reanalysis and forecast model (i.e., control run) fields were linearly interpolated to the time and location of each measurement, with a bin of 25 hPa, and then compared with the measurements. For the ACCMIP models, the monthly model outputs were compared with the measurements at the location of each measurement."

P6L29: Can you please elaborate on the model setting that caused this significant degradation of the representation of ozone in the UTLS?

The clause has been rewritten as follows:

"because of different model settings, such as the upper boundary conditions of NOy, Cly, and Bry"

P7L3: The terminology to refer to model output is somewhat confusing making it difficult to keep track of observations versus model output. Suggest referring to output from the forecast model and ACCMIP as "forecast model output" and "ACCMIP model output", respectively, and observations as "data".

Corrected throughout the paper.

P7L22-24: The simulation of the "wave-1" pattern by ACCMIP models has already been highlighted by Young et al. (2013). This reference needs to be cited.

Cited.

P7L32: It is not clear what "common reported" means here. Please clarify.

Replaced by "Young et al. (2013) consistently revealed the positive bias in the NH and negative bias in the SH using OMI/MLS tropospheric ozone column measurements."

P8L10: Which region is "In this region. . ." referring to?

Replaced by "In the NH extratropics in the lower and middle troposphere".

P8L14-15: These results are consistent with those discussed in Young et al. (2013).

Young et al. (2013) is cited.

P8L24: The differences could also be associated with the way biomass burning emissions are handled across models - whether they are emitted at the surface layer or distributed vertically in the model, as a larger role for biomass burning in the tropical mid to upper troposphere has been suggested recently (Anderson et al., 2016).

The following sentence has been added:
"For instance, biomass burning emissions are handled differently across the models, which may lead to differences in ozone simulations in the tropics (Anderson et al., 2016)."

P8L33-34: Why are results from a specific model highlighted? Is it because model 8 is driving the large model diversity at 200hPa? Please provide a figure to support this statement if this is indeed the case.

The sentence has been removed.

P9L3: It took me a while to understand the meaning of "reanalysis concentrations from the ozonesonde sampling". Please rephrase this to indicate that reanalysis output is sampled at ozonesonde sites instead of averaging the reanalysis at all grid cells.

Replaced.

P929-31: Here again the results are consistent with Young et al. (2013) - see their Figure 4 and its discussion.

Cited.

P10L23-25: Please clarify what is meant by "radiative heating distribution in chemistry-climate simulations are largely uncertain." Are you referring to the radiative heating due to tropospheric ozone? Please elaborate on how the O3 NH/SH ratio provides information on the radiative heating distribution.

The sentence has been replaced by
"The large systematic error in the NH/SH ratio suggests that, for instance, the inter-hemispheric distribution of radiative heating due to tropospheric ozone in chemistry--climate simulations are largely uncertain in most models"

P11L17: replace "completion" with compilation.

Replaced.

P11L19: By coarse resolution, do you mean the coarse horizontal resolution?

Yes. Replaced.

P11L29: Did you mean - "...less variabilities in the SH than in the NH"?

Yes. Corrected.

P12L9-11: Please quantify large in "Large negative model biases..." and larger in "...errors are larger than those. . ."

The sentences have been rewritten as:
"Large negative model biases against the ozonesonde observations have been reported by Young et al. (2013) for 250 hPa (by about -13 % for the NH polar west and -18 % for the NH polar east for the annual mean concentration), whereas results from this study suggest that these errors based on the ozonesonde sampling (by -14 % for the NH polar west and -18 % for the NH polar east in DJF in our estimates) are larger than those from regional and seasonally representative model bias (by -3 % and +5 %, respectively)."

Section 5.1 Please elaborate on how the comparisons discussed here may be influenced by inconsistency in the time period of the reanalysis and the ACCMIP simulations. The reanalysis output is for the 2005-2009 whereas the ACCMIP simulations are representative of 2000. Additionally, the precursor emissions used in ACCMIP simulations were decadal means and not specific to the year of simulation. For example, year 2000 biomass burning emissions were calculated as average over 1997–2006 (Lamarque et al., 2010), so they encompass the high emissions over Southeast Asia in 1998, an El Nino year. One would expect that biases due to sampling in time would occur similar to the biases due to sampling in space. How significant are the biases due to spatial sampling errors (highlighted here) compared to temporal sampling errors. The issue of temporal and spatial sampling was recently highlighted by Lin et al. (2016) in the context of tropospheric ozone trends.

The limitation of the evaluation based on the 2005-2009 reanalysis fields is discussed in the revised manuscript. Please see our reply above.

The spatial and temporal sampling errors are separately evaluated in the revised paper, and the results are presented in Table 7 and discussed in Section 5.1. Lin et al. (2016) is cited in the revised paper.

P14L24: Replace biogenetic with biogenic.

Replaced.

P15L6: I am not sure if the 2005-2009 can be considered a "long-record" Are the authors referring to

the possibility of a long-record reanalysis sometime in the future when observations have accumulated in time.

In this sentence, we are referring to the possibility in the future. The sentence has been replaced by
"A long-record of the reanalysis will allow...."

P15L7-9: From Lamarque et al. (2013): "This averaging was designed to reduce the effect of interannual variability and therefore provide optimal conditions from which average composition changes and associated forcings can be more readily computed." The ACCMIP simulations were designed to remove interannual variability, therefore, it is unjustified to state that "the influence of ENSO was not well-simulated in ACCMIP. . ."

The sentence has been replaced by
"It is noted that the influence of ENSO was not included..."

P33 Figure 5. Please provide statistics such as mean bias and correlation for the comparisons here.

These statistics are provided in Table 4.