Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2016-958-RC1, 2016

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Impact of Intercontinental Pollution Transport on North American Ozone Air Pollution: An HTAP Phase II Multi-model Study" *by* Min Huang et al.

G.S.T Tonnesen (Referee)

tonnesen.gail@epa.gov

Received and published: 12 December 2016

General comments

Most of the paper is focused on comparison of monthly mean model results with very limited evaluation of model performance and no analysis of the causes of the differences among the global model simulations. This analysis is not substantially different from previous HTAP studies and is not informative. I suggest moving most of the text and plots that discuss monthly mean model results to the supplement, and instead, the authors should evaluate and compare model performance on several short term episodes that are more relevant to ozone transport and air quality planning.

The most interesting aspect of the paper is the section that addresses the May 9, 2010 O3 episode. I suggest including a more detailed discussion of this event, including an assessment of the relative contributions of stratospheric O3 and international transport of O3 for this event. Given that all of the global models performed poorly for this event (with the exception of RAQMS with data assimilation), a key finding could be that currently available global models do not perform well for some high ozone events. I also suggest performing additional analysis for at least one other high O3 event during summer 2010 to contrast with the May 9 event. By performing a more detailed evaluation and comparison of the different global models (and the couple STEM/Global model simulations) for specific episodes, the authors can more directly evaluate model performance and the suitability of the individual global models for use as boundary condition data in higher resolution regional models.

I suggest deleting the text that asserts that the use of an ensemble of global models is a preferred approach. The citation (U.S.EPA 2016) is summary of comments at a public meeting and should not be used as citation because the comments were not peer reviewed and do not reflect the consensus of the meeting participants. A better citation would be the EPA whitepaper on background ozone which was reviewed within EPA and is available at https://www.epa.gov/ozone-pollution/background-ozone-workshop-and-information. The whitepaper does not recommend the use of multi-model means to reduce uncertainty. The Li et al. 2016 citation is an analysis of visibility trends and does not evaluate multi-model results. Moreover, there is no valid theoretical basis to assume that the average of poorly performing models will be more accurate than the best performing individual model for key atmospheric processes. While it is possible that, by chance, the average of several poorly performing models will better match observations, the average may still inaccurately represent the individual processes that contribute to O3 and the sensitivity of O3 to emissions reductions. While it might be true that positive and negative bias errors cancel when averaging multiple model results for monthly or seasonal means, this does not necessarily indicate that the multi-model average represents O3 more accurately for episodic events that are of interest to the

air quality planning community. A better approach would be to evaluate and compare models at the process level and specifically for high O3 episodes, and then select the best performing individual model.

Detailed comments

Line 63: Opening sentence is awkward. There is no clear link of the uneven distribution to the health/ecosystem impacts of O3. Also, the uneven distribution of O3 is mostly caused by strong concentration gradients in precursor emissions, but this sentence only identifies the O3 lifetime as a cause of the distribution. Suggest rewriting with a focus on the high mixing ratios, not just the distribution.

Line 73: "to control the emissions of its precursors from these various sources". Not clear what "these various sources" refers to here. The previous sentence identified the stratosphere and local to distant emissions sources, so presumably this sentence is suggesting that there will be benefits of control of both local and international emissions sources, but this sentence then goes on to list precursors categories (VOC, NOx, CO) without reference to local vs international or biogenic vs. anthropogenic. I can infer what the authors mean, but the introductory paragraphs are awkwardly written and potentially confusing to a reader who is not an expert.

Line 75: Also include methane in this list

Line 80: background and baseline are not the same. See Cooper et al. for their definition of baseline, and EPA white paper (link below) for definition of U.S. background ozone. Briefly, baseline O3 (as defined by Cooper et al.) can include contributions from upwind U.S. anthropogenic precursor emissions while U.S. background ozone excludes all U.S. anthropogenic emissions.

Line 82: Given how the authors defined baseline/background O3, this statement is problematic: "below which the air quality standard is not recommended to be set". Baseline O3 can be elevated in some areas because of transport of anthropogenic

precursors and O3 from upwind U.S. states. It is appropriate to set the NAAQS below the baseline O3 level in these areas because the elevated baseline O3 is being addressed by emissions reductions in upwind states. I recommend breaking this very long sentence into several sentences that describe each of the points identified more clearly and more accurately.

Line 90: "It has been revealed" is awkward – "revealed" has other connotations. Suggest "It has been found".

Line 95: "A better understanding of the processes that determine the O3 distributions" Note that the authors have not yet clearly and comprehensively described the processes. They should describe the roles of stratospheric (both routine contributions and discrete intrusion events), biogenic precursors, wildfires, and anthropogenic precursors. We are especially concerned with conditions in which the mixing ratio exceeds the NAAQS, so it is not only the distribution but also the mixing ratio that is important.

Line 96: delete "for recent years". This will be useful for all past years and for future predictions.

Lines 110-112: "Large intermodel diversity was found in the simulated total O3 and the intercontinentally transported pollution for the chosen SR pairs in the northern midlatitudes, indicating the challenges with simulations by any individual model to accurately represent the key atmospheric processes." The conclusion that no individual model performs well is not supported by a finding of inter-model diversity. For example, it is possible that one model performs well while other models do not. The authors need to cite results of the individual model performance evaluations to support the statement that no model perform well.

Lines 113-116: The citation (U.S.EPA 2016) is summary of comments at a public meeting and should not be used as citation because the comments were not peer reviewed and do not reflect the consensus of the meeting participants. A better citation would be the EPA whitepaper on background ozone which did receive review within EPA and

is available at https://www.epa.gov/ozone-pollution/background-ozone-workshop-and-information. The whitepaper does not recommend the use of multi-model means to reduce uncertainty. The Li et al. 2016 citation is an analysis of visibility trends and does not evaluate multi-model results. Moreover, there is no valid theoretical basis to assume that the average of poorly performing models will be more accurate than the best performing individual model for key atmospheric processes. While it is possible that, by chance, the average of several poorly performing models will better match observations, the average may still inaccurately represent the individual processes that contribute to O3 and the sensitivity of O3 to emissions reductions. While it might be true that positive and negative bias errors cancel when averaging multiple model results for seasonal or annual means, this does not necessarily indicate that the multi-model average represents O3 more accurately for episodic events that are of interesting to the air quality planning community. A better approach would be to evaluate and compare models at the process level and for high O3 episodes, and then select the best performing individual model.

Lines 123-125: Note that in certain VOC/NOx chemical regimes the model response to NOx emissions can be strongly non-linear for smaller NOx changes, so the statement that 20% emissions were selected to be "small enough in the assumed near-linear atmospheric chemistry regime" is not consistent with how models respond to NOx emissions and does not provide an explanation for using 20% emissions reductions. A 100% reduction in emissions from a source sector or region is a better approach to evaluate source attribution. Lines 126-130 identify problems with the use of a 20% reduction and this should also be noted in the conclusions. For future work, I recommend 100% reductions when evaluating source contributions.

Line 143: "the necessity of evaluating the extra-regional source impacts on event scale [have] has been emphasized" This is a key point – check to see if this addressed in discussion and conclusions.

Lines 214-216: Biogenic emissions of VOC are larger than anthropogenic VOC glob-

ally, and biogenic and geogenic emissions of NOx, SO2, CO and CH4 are also large and can have a substantial impact on model results. It would have been best to harmonize the natural emissions in addition to anthropogenic emissions, and this approach should be used in future work. For this manuscript, the natural emissions used for each model should be summarized and compared, and, if the natural emissions are significantly different between models, the possible effects on model results should be discussed.

Line 254: Equation 2 is confusing because the labels for the scenarios are confusing. It is not clear what RERER(O3,NAM) represents. Does this represent a percent contribution from local versus non-local sources?

Lines 227-240: The description of the model scenarios and the naming convention is complicated and difficult to understand. In line 231, why is "all" enclosed parentheses? Why is a 20% sensitivity simulation described as "*source region*ALL". It is not clear what "ALL" means, and generally, the approach used to label the scenarios is not intuitive.

Line 266-270: Why would lower than normal temperatures in the western U.S. favor decomposition of transported PAN? Lower temperatures would make PAN more stable.

Lines 287-291: The discussion/conclusions should address the uncertainty introduced by using monthly mean emissions.

Lines 293-294: I doubt that the speciation of VOC emissions in 2005 is substantially different compared to 2010, but if the authors' statement that it is "highly unrealistic" to approximate 2010 using 2005 VOC speciation, this seems to be a significant problem for interpreting the model results.

Line 404: Table 2. The model performance evaluation results in Table 2 are not adequate to evaluate the models. In addition to showing the mean bias for multiple models, the model evaluation should also show the bias and error for each model, and the bias

and error for the highest observed O3 days because these are the days that are most relevant for air quality planning.

Lines 430-432: “Except in the northeastern US, the eight-model ensembles show better agreement with the CASTNET O3 observations than the three boundary condition-model ensemble, suggesting that using a larger number of models in the ensemble calculations may result in better overall model performance.” Given that the goal of this study is to evaluate the contributions of international emissions to O3 transport in different regions of the world, it is critically important to understand the individual performance of each global model. If there are substantial difference among models in the contributions of stratospheric O3, chemical production of O3 from precursors, or transport and dispersion of O3, the effect of averaging multiple models may be to introduce additional error into the analysis. A better approach is to compare each global model at the process level, and select the best performing models. If it is uncertain which model performs best, source response relationship should be evaluated using simulation with each BC from each of the global models to estimate the uncertainty in the SR relationships.

Lines 461-463: Recommend showing the individual model performance results using each global model BC instead of averaging the results for all three simulations.

Table 4: These results are interesting, but to be policy relevant, we need estimates of the contributions on days that exceed the O3 NAAQS. For example, international transport contributions might be highest on days with good dispersion conditions that do not exceed the NAAQS, and lower for days with stagnant dispersion conditions that are more likely to exceed the NAAQS in urban areas. Alternatively, if might be possible that NAAQS exceedances are more likely to occur in rural areas as a result of international transport because of strong mixing from the troposphere to the surface. It is very difficult to interpret the significance of results that are presented as the mean for all days.

Lines 467-469: This is a key uncertainty that the study does not address. If the modeling systems is biased low for international transport and biased high for local O3 production, the results of the SR analysis may not be reliable.

Lines 476-479: Speciation in SAPRC99 is unlikely to be the cause of model overestimates for O3. SAPRC99 underestimates VOC reactivity in chamber experiments, and the most recent updates to SAPRC are more reactive for urban than SAPRC99. For the rural CASTNet sites in this study, it is more likely that overestimates of biogenic VOC in MEGAN and uncertainty in NOx emissions and fate contribute to the positive bias for O3.

Lines 505-510: Note that larger-than-1 RERER values will be less likely to occur if the model results are analyzed for high O3 days. It is not informative to present results for low O3 days on which NO titration is more likely to occur because these days are not relevant to air quality attainment planning.

Lines 516-519: “Comparing to the HTAP I modeling results, the magnitudes of R(O3, EUR, 20%) are smaller by a factor of 2-3, as a result of the substantial improvement in the European air quality over the past decades” The modeling for HTAP II is for 2010 versus 2001 for HTAP I, so any O3 reductions should reflect emissions reductions for 9 years, not for decades. Have European emissions been reduced by a factor of 2 to 3 from 2001 to 2010, or is it possible that other changes in the HTAP II modeling platform are the cause of this change?

Lines 541-545: This text seems to inappropriately discount the significance of international transport and also the possible importance of differences among the global models. For interstate transport EPA uses 1% of the NAAQS as a significant contribution. Thus, differences among global model much less than 5% of the total model O3 can be very important, especially given that the values discussed in the text are based on a 20% emissions sensitivity and that results are reported as the monthly mean. While local emissions will have a larger contribution, it may not be true that local

emissions control programs alone are the most effective way to attain the NAAQS, as the text seems to suggest.

Lines 562-571: It is surprising that the couple STEM/global model predicts large transport contributions than some global models and smaller transport contributions than other global models. The authors provide a list of factors that contribute to model uncertainty as a possible explanation, but it seems like these uncertainties (e.g., terrain, chemistry) should affect in similar ways each of the coupled STEM/global model simulations. More investigation is needed to explain why STEM sometimes shows higher or lower transport contributions compared to the global model.

Lines 607-609: This is an important finding that should be highlighted in the conclusions and abstract.

Lines 620-622: "Therefore, it is important for more HTAP2 participating models to save their outputs hourly in order to conveniently compute the policy-relevant metrics for the O3 sensitivities." I agree with this statement, and moreover, I do not think you can do a meaningful analysis of any models that do not save the hourly outputs (or 3-hour if that is the finest time resolution used), and I would recommend excluding them from this study.

Lines 612-624 and Figure 9: It is obvious that day time O3 is greater than nighttime O3 at surface sites because O3 deposits to surfaces and is destroyed by chemical reactions at night. So the findings in this text that the maximum daily 8-hour average O3 is greater the 24-hour average O3 is self-evident. I suggest deleting this text. I also recommend focusing the analysis on maximum daily 8 hour averages, especially for the highest O3 days, and not showing results for monthly mean O3.

Line 633: "R(MDA8, EAS, 20%) is smaller during the high O3 total days in all subregions." For GEOS-Chem the contribution appears to be the same on high O3 days compared to all days, and the results are very similar for RAQMS. It would be helpful to show more details for this analysis. Is this the mean O3 for all sites for days in which

any monitor was > 70 ppb, or does it only include data for the monitor that was greater than 70 ppb? I suggest performing a more detailed analysis, e.g., show EAS contribution on each day for a few key sites that frequently have high O3, e.g., Great Basin and Canyonlands sites.

Line 655: "We found that the underestimated free tropospheric O3 from the STEM simulations that used any single free-running chemical boundary conditions contributed to the underestimated STEM surface O3 in the high elevation mountain states." Need to edit and clarify meaning of the above sentence. Was this because the global models underestimated stratospheric O3 or international transport?

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-958, 2016.