Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2017-218-AC1, 2017
© Author(s) 2017. This work is distributed under
the Creative Commons Attribution 3.0 License.

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "Coupled Chemistry-Climate Effects from 2050 Projected Aviation Emissions" *by* Andrew Gettelman et al.

Andrew Gettelman et al.

andrew@ucar.edu

Received and published: 20 July 2017

Replies to Reviewers

We sincerely thank the Reviewers for their effort. We appreciate this is a difficult manuscript. We are trying to figure out the appropriate way to present additional work (related to previous work). That requires a bit of repetition, which we have tried to reduce. And we are trying to appropriately present the statistics of our simulations with small perturbations: it is difficult to find a statistically significant signal in surface temperature from aircraft emissions in coupled simulations: it is smaller than the climate noise in most cases. We have however re-run our statistics using the false discovery rate (FDR) approach suggested. This provides we think a more robust result. And

we have now found regionally significant temperature results, and global temperature changes that are barely significant. We thank the reviewers for this suggestion and the pointer to the new method. We think this significantly strengthens the conclusions, and we thank the reviewers for pointing us to these additional tests.

In addition, we have modified the text to try to better discuss model processes and our model comparisons. It is not easy to compare the models, which are very different in construction. We are doing our best because this was not a controlled experiment. The coupled nature of the models is why this is has not been done before. Nonetheless, we have performed tests to bring the models together by testing GATOR-GCMOM assumptions in CESM, and we think this does help understand the processes and assumptions responsible for model differences. This is also a unique aspect to the study.

As noted above, we have redone the statistical tests to be more robust at the suggestion of the reviewers. We have also modified the presentation of the Ozone figures, and added further process discussions of the mechanisms leading to the radiative forcing changes we describe with additional analysis.

We have also rewritten the conclusions for clarity to make them flow better, and better state the revised statistical results.

We think this revised manuscript will answer the reviewers' concerns, and we hope the spirit of this attempt now comes through with better focus in the revised manuscript.

Reply to Dr. Collins

This paper calculates the non-CO2 climate impacts of future aircraft emissions using two very different models. There are some reasonably robust changes in radiative forcing from the CESM model, but the changes in temperature are not found to be robust.

» With slightly more attention to the statistics and a better formulation as described above, we think now we can state some regionally significant results, and we now characterize the global changes as ‘barely significant’ which we think is true (it passes our 95% significance test, but barely).

Overall it is not clear what the new findings coming from this study are. The work is publishable, but more thought needs to be given to the overall messages if this paper is to be of interest to the community.

» Clarified. The main point is to summarize the result of fully coupled experiments. This is now stated explicitly in the introduction, abstract and conclusions. We think we have a more robust and coherent result now, and have rewritten the conclusions significantly

The two models are set up differently which makes it very difficult to draw any useful conclusions from their comparison. In particular the absence of radiative forcing data from GATOR-GCMOM means that the conclusions in section 5 are mostly speculative. Ideally both models would also have run fixed-SST experiments to categorize the rapid responses.

» Both models have run fixed SST and even fixed meteorology experiments to explore the rapid responses. This is now noted in the text more explicitly. We would not say the GATOR-GCMOM results are speculative, rather that they are not significant given the large variability in 5 years of coupled simulation.

In general the text doesn’t flow very well, with many short paragraphs that don’t seem to connect. It would help to understand the messages better if there was a logical chain of argument that could be followed.

» We have now tried to summarize the sections more and provide more discussion of this logical chain. We have also reorganized the text a bit, and tried to remove some of the shorter paragraphs

Specific points: Page 2, line 11: I’m not sure describing the aerosol effects as ‘non-linear’ is a helpful term. None of the aircraft impacts are strictly linear.

» Deleted ‘non-linear’

Page 2, line 23: “and RF increases by : : :”. This clause doesn’t seem to sit with the rest of the sentence.

» Clarified (specific numbers mentioned)

Page 5, line 9: What is meant by “Ensembles are created with a unit temperature perturbation”?

»Clarified. This is the method used to initialize different ensemble members, with a round off level temperature perturbation.

Page 5, line 25: I presume the fluxes are taken at the tropopause because there is no stratospheric adjustment? Does this give equivalent results to a RTM with stratospheric adjustment?

» Yes, there is no stratospheric adjustment with specified dynamics. It is not wise to try to compare to RF estimates with stratospheric adjustment when dealing with forcing in the UTLS around the tropopause, as with aviation.

Page 7, line 32: “non-linear” isn’t the right term.

» Removed non-linear (now removed from the whole manuscript).

Page 8, lines 20-24: Surely the model can tell you whether there are fewer present day contrails, or a higher change for forming contrails?

» There are fewer present day contrails, so there may be higher sensitivity. We have clarified this paragraph.

Page 8, lines 30-34: Presumably the aerosol emission affect the contrails as well, which should be discussed here. » Noted.

Why are the differences between scenarios 2 and 3 not statistically significant? If they are run with specified dynamics the meteorology should be the same and hence no (or very little) variability. I don't understand why the effect of water vapour emissions is so small. According to figure 1A the water vapour alone has a huge forcing.

» Clarified. The 'water vapor' in figure 1 is due to emissions of aviation water vapor causing contrail formation. The difference from Scenario 2 to Scenario 3 is a 5% increase in aviation water vapor emissions, which does not affect things substantially. The contrails mostly contain ambient humidity, not humidity from the engines.

Page 9, line 7. The effect of alternative fuels here doesn't seem the same as the difference between the lines 2050-S1 and 2050-S2 in table 2. In particular in the table the effect on O3-S is 12.0 mW/m2 which seems large. The authors should explain how the changes in sulfur and BC cause such a large change in ozone.

» The reviewer is correct. The paper mis-stated the table value for Short term Ozone effect. This is larger for short term ozone effect. The impact of alternative fuels is $\sim$15%, which is not that large in relative terms. This is probably an 'indirect effect' on ozone that arises from changes in UTLS temperature induced by BC reductions, but the result may not be significant given the small changes in ozone in Figures 3A and B. Noted in the text.

Page 10, lines 1-5. Which scenarios do these forcings come from?

» Clarified (scenarios in Table 1).

Section 3.2.3: Given that it isn't expected that the aerosols affect ozone I suggest this (very short) section isn't needed, nor are figures 3A and B, or 6 A,B,C.

» For completeness, we left the section in. And the figures are necessary for the comments about Ozone RF above.

Section 3.4: Should this be numbered 3.3?

» Yes. Corrected throughout.

Section 3.4.2: The arguments in this section needs to be made clearer.

Figure 2D needs to be the 5-year result from CESM for like-for-like comparison with GATOR.

» It is shown in Figure 5, but we choose to focus the discussion on the 31-50 year period because that is where statistical significance is. Figure 5 is referenced here.

The time evolution needs to be discussed in relation to rapid responses to composition followed by slower responses to SST evolution.

» Noted and clarified.

It is not clear what the message of the second two paragraphs is.

» The second paragraph has been shortened and clarified. It's purpose we think is well described by the topic sentence "the reduction of BC in the Alt fuel scenario (Figure 3C) reduces warming relative to the baseline". The 3rd paragraph has been focused and shortened and it's goal is to note that BC is the most important component in GATOR-GCMOM.

Description of the physics should be moved to section 2, unless the authors are specifically contrasting the different effects of the physics in GATOR and CESM.

» The discussion here of the physics has been shortened.

Page 12, line 10. ": : :contrail radiative forcing dominates: : :" I don't see why this is true, don't scenarios 1 and 2 have similar contrails?

» Clarified. They do have similar contrails. The indirect sulfate aerosol cooling has been removed.

Page 13, line 1. Where does GATOR show a small warming? In figure 4 it cools.

» Clarified.

Page 13, lines 11-13: I didn't see the relevance of this disconnected paragraph on Righi et al.?

» Merged with above paragraph.

Page 13, line 15-16: This disconnected paragraph needs to be moved somewhere else as part of a logical train of argument.

» Removed (discussed in conclusions).

Page 13, lines 20-34: Much of this is model description which could be moved to section 2.

» Actually we think it belongs here. Section 2 references the parameterizations, but here we directly discuss the aspects of the parameterizations that matter for the differences, which is appropriate for the discussion section.

Earlier (page 12, line 30) the BC and sulfate are described as externally mixed in the exhaust, but here they are described as internally mixed.

» Clarified (CESM does treat internal mixtures).

Page 14: These short paragraphs disrupt any flow of argument. What is the message of this section?

» The section is designed to analyze differences in the BC results between the models, and to describe analysis that bring CESM results more in line with GATOR-GCMOM. We have merged several paragraphs and moved the last paragraph to the conclusions. This makes sure the section wraps up with a strong conclusion comparing to previous work.

Page 15, line 11-12: The forcing in GATOR needs to be shown to back this up.

» Radiative forcing was not calculated from GATOR-GCMOM. The sentence has been rephrased.

Page 15, lines 17-20: The difference between the baseline (-0.11K) and the AltFuel (+0.1K) is 0.2K. While this may not be statistically significant due to the length of the

simulations, this is not a negligible difference compared to the 1.5-2.0K Paris recommendations.

» Noted now in the text.

Section 5: The paragraphs in this section tend to be short and unconnected which makes it difficult to pull out the important messages of this study.

» This section has been modified to address this and other concerns. The goal was to try to state each result succinctly, but that perhaps did not work.

Figure 2: A different set of contour levels is needed to show the ozone changes.

» In response to this comment and that of the other reviewers, we have modified the contour intervals on the panels to be relative ozone changes. This does provide a much better picture.

Figures 3, 5, 6: The ozone panels don't add information here.

» As noted, these are now relative changes and they now provide more information.

Reply to Dr. Ponater

ÂăSome comments on A. Gettelman et al. "Coupled chemistry climate effects from 2050 projected aviation emissions"Âă

(This is not meant as a full scale review but rather a collocation of ideas that occurred to me when reading through the paper. Nevertheless, I am confident that the comments express well-founded criticism.)Âă

» They do. Thank you very much for the comments.

The present paper fails to discuss evidence from previous work to an extent that makes it difficult to understand what is actually added here to current knowledge on the subject. Besides other papers I am particularly referring to Huszar et al. (2013), a basic study of future aviation impacts that the authors appear to have overlooked.Âă

» Thank you for highlighting this oversight. We have now noted the Huszar et al 2013 reference where appropriate. They discuss chemistry impacts (NOx) and contrail impacts, but not Aerosols.

1) The results given in section 3.1 all in all look scientifically and statistically plausible, yet they have apparently been presented before (Brasseur et al., 2015; Chen and Gettelman, 2016). However, I notice an inconsistency with re-spect to the contrail cirrus RF estimate for the 2006 baseline scenario between Table 2 (17 mW/m2) and Figure 1A (12 mW/m2). This is rather relevant, as the puzzling evidence that contrail cirrus RF increase more strongly over the years than fuel consumption would vanish, if the Table 2 value were taken as the starting value.Âă

» Corrected. The 17 value is from a slightly different estimate in Gettelman and Chen 2013 that includes effects of water vapor beyond cirrus clouds. 13 mWm-2 is the correct value.

Recently, Forster et al. (2016) came up with a study indicating that radiative flux differences derived from free-running fixed SST simulations (I guess that's what "RESTOM" indicates in Figure 1) should amount to at least 100 mW/m2 in order to reach sufficient statistical significance levels. In case of nudged simulations (resembling the specified dynamics simulations in the present papers) the threshold value may reduce to 10% (Forster et al., 2016, p. 13), which appears to be consistent with the error bars in Figure 1. However, the error bars are clearly overlapping between the different scenarios at all time slices, indicating that the scenarios are statistically indistinguishable.Âă

» Noted in the text.

2) I find the ozone pattern difference presentations from Fig. 2 a,b, Fig. 3 a,b, Fig. 5, Fig. 6, rather pointless. While they suggest large areas of statistical significance (for CESM almost over the whole troposphere), this remains un-convincing as contour lines are largely missing (those that are shown are mainly referring non-significant structures). Figure 8 of Huszar et al. (2013) offers a more satisfactory description,

clearly indicating that patchy stratospheric response patterns are insignificant, despite showing higher concentration difference values compared to the troposphere. It may, hence, be worthwhile to display relative differences, as in many earlier papers (e.g., Grewe et al., 1999) dealing with free running chemistry climate model simulation results.Âă

» These figures do indicate that the lower stratosphere shown is not significant. We have now changed the plots to relative ozone changes as suggested. This does present a much better picture and we think makes the figures more coherent.

3) The severe problems to assess the (statistical and physical) significance of temperature response patterns simulated from aviation effects have been re-ported before (e.g., Rap et al., 2010, Fig. 1a; Huszar et al., 2013, Fig. 10, Fig. 12). A point-by-point hypothesis test suggesting statistical significance in strongly confined regions may well turn out to be unfounded, if spatial correlation is accounted for. (Chaotic negative and positive side-by-side differences, as obvious in Fig. 2 c,d, Fig. 3 c,d, are always raising suspicions in this respect. I notice coherent regions of significance only in Figs. 3d, 4d, and 6f shown here.) Significant temperature response is more easily established for global means (Huszar et al., 2013, Fig. 9), but these are bypassed in the present paper. Sometimes, more sophisticated (multivariate) statistical tests have proved helpful to establish pattern significance (e.g., Sausen et al., 1998).

» We fully agree with your points. As suggested by reviewer #2, we have used the False Discovery Rate approach to better treat field significance (Wilks 2006), and this does eliminate the spuriously significant results (i.e. in the S. Hemisphere). We discussed significance of the global means, but have now made that more explicit with significance tests: the global means for CESM are (barely) significant. We are clear to not overstate the significance (barely significant at the 95% level).

4) In section 4.1.2 much text is devoted to allegedly large effects of aviation black carbon emissions without showing any results. To me this is absolutely unconvincing

as to underpin what is suggested by Figure 1b.Âă

» We have clarified the text here to flow better. We do specifically discuss further analysis here, and trace the effects back to the physical causes. Note that the effects are not in Figure 1B as they are only in the modified version of the model.

5) Given the general lack of statistical significant simulation results, I find large parts of the concluding section to be insufficiently covered by the results. In my opinion, the simulation strategy followed in this paper is only of very limited value for establishing reliable evidence on the relative importance of individual components in forcing a net aviation climate impact. Even in Huszar et al. (2013) statistical noise has made interpretation of their results problematic and I fail to notice any progress on this in the present paper.Âă

» We have looked again at the significance globally and using the false discovery rate approach from Wilks (2006,2016) as suggested. This provides a much better and robust approach for the results. The regional temperature changes are only significant in a few regions of aviation flight corridors, and the resulting global values in CESM are barely significant. We think this better treatment does validate the method of performing long coupled-climate simulations. We have endeavored to rewrite the conclusion section with these concerns in mind and the revised results to address this.

Adding on my main comments, I find the present paper to be written in a rather confusing manner. For example, the description of the simulations is scattered over three different sections (2.1, 2.2, 2.4) and it is not sufficiently recalled in the results section, which of the simulations are actually discussed. A special subsection (3.4.3) is dedicated to alternative fuel effects, yet those are partly addressed already in subsection 3.4.1.Âă

» We have reorganized the paper along the suggested lines. We moved the model sections together, and moved the scenario section, so that at least they are together. We also reference alternative fuel effects in 3.3.1 (only one sentence). We have added

discussion to the results as well that we think make this flow better.

References:Âă Forster, P.M., et al., 2016: Recommendations for diagnosing effective radiative forc-ing from climate models for CMIP6, J. Geophys. Res. 121, 12460-12475.Âă Grewe, V., et al., 1999: Impact of future subsonic aircraft NOx emissions on the at-mospheric composition, Geophys. Res. Lett. 26, 47-50.Âă Huszar, et al., 2013: Modeling the present and future impact of aviation on climate: an AOGCM approach with online coupled chemistry, Atmos. Chem. Phys. 13, 10027-10048.Âă Rap, A., et al., 2010: Estimating the climate impact of linear contrails using the UK Met Office climate model, Geophys. Res. Lett. 37, L20703.Âă Sausen, R., et al., 1998: Climate impact of aircraft induced ozone change, Geophys. Res. Lett. 24, 1203-1206.

Reply to Review #2

The paper presents results from two climate-chemistry-ocean models on the climate impact of future aviation. In particular the paper discusses the climate impact of aviation emissions on climate temperature changes.

Certainly, the topics are interesting and important and deserve careful investigations. An ocean-atmosphere coupled model with a high quality model simulating the fate of the aircraft emissions is essential for computing aviation climate change. Unfortunately, the paper is insufficient in many respects. The material presented, though not irrelevant, does not add enough new insight and results to the existing literature.

» We think there are some important insights in this paper that add to the literature. We think the revised manuscript with better statistical tests and revised conclusions makes a better and clearer contribution to the literature now.

This topic deserves a far deeper investigation and a technically better paper.

» It does deserve a deeper investigation. We have noted this now better in the conclusions and summary. We think these comments and those of the other reviewers have made this a technically better paper and we hope this will answer the reviewers' concerns.

The paper reports quantitative values of the radiative forcing (RF) values for various aviation emissions and effects. The majority of these results are taken from earlier publications and given here with little or no discussion on the ranges of validity and discussions on the differences to results from other studies.

» This is not designed to be a comprehensive review, but is necessary to provide some consistency and a stand alone manuscript that makes sense. We have now explicitly noted where these different previous values come from in the text. We have tried to note previous studies as well to be more comprehensive.

Additions concern specific scenarios.

» Yes, that is one of the important features, and why it is necessary to add the background material.

Then the study reports various "nonlinear effects" from BC, including strong solar radiation absorption by BC from engine exhaust in cirrus, sulfur brightening of low level clouds, regional disturbances, advection effects, and surface temperature response. The simulations were performed with two different models. Here too, a large part of the results was published earlier (in the various cited papers by Jacobson, Chen and Gettelman). In fact, the paper has strong overlap with Brasseur et al. (2016), Jacobson et al. (2013), and Chen and Gettelman (2016), even repeating some of the numerical values in the tables and one of the already published figures.

» We feel it is necessary to add several missing points to previous work in the literature, which is why results are restated. They are also necessary to give the reader a sense of the overall picture. However, this work then goes beyond this significantly to show coupled model simulations and to compare them. This adds significantly by putting two disparate models in different contexts, and highlighting where previous results are confirmed or not in another complex and advanced modeling framework. This signifi-

cantly builds on previous work by both model teams. We focus on our previous work for continuity, and because this is a model analysis, not a review.

The two models disagree in many respects, and the discussion mentions possible reasons, but the discussion remains qualitative and does not present new convincing evidence explaining the differences clearly.

» There is rarely ever a clear answer, but note we have done specific simulations to replicate one model with another for some of the key uncertainties. We try to highlight this better in the discussion and conclusions. We hae added some more quantitative analysis of the mechanisms for the aerosol effects in CESM for example. We naturally were not able to address everything, but we think we have a logical case that goes through the processes responsible for the major differences in sulfate and black carbon emissions between the models. It is difficult to ascertain all the differences between complex model systems. We feel we have gone farther than most. We have tried to bring up in the discussion now what these uncertainties mean. Again: these uncertainties and disagreement between models is a start of understanding of important issues, not the result.

Not surprisingly, the authors did not find statistically reliable results in this respect. This was to be expected, as discussed in other studies, because the disturbances are small compared to the inherent climate noise.

» Actually with some of the better statistics, we do clarify that there are regionally significant responses, and the global values are significant (barely). The revised statistical treatment to use field significance and the false discovery rate approach suggested was very helpful in this regard. We also now reference earlier work with coupled models (there is not much, but Huszar et al 2013).

For this reason, other authors either use enhanced disturbances or follow the idea of climate response models, which are quasi linear in the disturbances, with model parameters (inertia or heat capacities and climate sensitivities) fitted to full climate model studies with enhanced disturbances. The possible inaccuracies of such approaches because of inherent nonlinearities are unavoidable.

» We did NOT want to do this because of these inaccuracies. Too many other studies have tried fitting or assumed linearity with large perturbations and scaling to address these issues, and our intent was to show what can be done with full response models. We have noted this in the conclusions better now. We also are able to better sort out the climate signals with revised statistical tests. We decided to try to run longer simulations (50 years) to better constrain the results.

It would have been interesting to see how linear or nonlinear the model responses are (see Rind et al., 2000).

» To some extent, this can be ascertained from figure 1, which does run simulations over different years.

As the authors mention themselves, none of the climate change results on regional temperature and ozone changes and surface temperature changes is strictly statistically significant.

» We have revised the tests, and we now have higher confidence that the regional responses are statistically significant as the new methodology removes a significant amount of climate noise. This is noted throughout the new results, discussion and conclusions. We now think some of the changes are statistically significant.

Even the global mean surface temperature changes remain in the statistical noise. (This seems to revise some earlier conclusions form the same data; e.g., Jacobson et a. 2013).

» That paper focused on the Arctic, and we do revise the results to note that those regional responses (particularly in the Arctic) are not seen in another model. However, we do find significant regional and global responses using the revised statistical tests. We have tried to make sure we clarify the results against earlier work. But this is an

important revision, and some confirmation to earlier work. The global mean surface temperature changes are just out of the statistical noise now.

The patterns of the simulation results do not look convincing. There is little systematic pattern in the mean responses. Many of the results just look random. Fig. 2b and 4 are examples.

» The revised statistical tests to use field significance to eliminate false discovery help eliminate this problem in the revised draft. We think the results are no longer random. For example, there is now no significance in the S. Hemisphere, and no significance anywhere for 5 years of coupled simulation.

I am not sure, whether the significance tests are reliable because based on local tests; see Wilks, D. S. (2016), "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it, Bull. Amer. Meteorol. Soc., 97, 2263-2273, doi: 10.1175/BAMS-D-15-00267.1. This should be discussed.

» Thanks for the reference. We have not applied this test throughout, and it is very helpful for understanding the results, and we think it removes a lot of the spurious signals, and gives a much better sense of the significance of the results. As noted, we think this improves the paper quite a bit, and we thank the reviewer for this reference. We intend to use this method in the future. It seems quite valuable for reducing climate noise.

The suitability of the models for this study is not sufficiently justified. Since a global model with very coarse horizontal resolution (4 Å∼ 5 degrees) cannot resolve plume dispersion of NOx and neither line-shaped contrail formation nor their merging into contrail cirrus, the results rely highly on the validity of the subgrid scale (SGS) models used. The present method gives no details on these SGS models but refers to previous publications.

»The plume dispersion (contrails) in GATOR-GCMOM is treated explicitly with a SGS model to enable effects to be handled even at low grid scale resolution. You are correct that NOx is not treated this way (this is noted). This is clarified in the text. A few further sentences have been added to describe the SGS and how it works.

When looking to some of the previous publications one finds a lot of adhoc simplifications. I agree that simplifications are unavoidable, but I see a lack of principle justification (e.g., on plume cross-sections), reflection of recent insight and data, and lack of validation of the SGS models with the observations that are now available from various studies.

» We have added further discussion of uncertainties and evaluation of GATOR-GCMOM against detailed models (i.e. a plume model) and CESM (against observations). The assumptions in CESM are analyzed parametrically in previous work, which we reference. Details of the sub-grid scale model in GATOR-GCMOM are also contained in previous work, and as noted above we now refer to them more explicitly.

The paper does not discuss the degree of agreement or disagreement with related contrail studies, e.g., in respect to contrail ice water content, optical depth, life times, cross-sections etc. Hence it is unclear of how good the SGS models are and how much the results change when the SGS model is changed.

» We have added references to evaluation for CESM and GATOR-GCMOM against observations, such as are available.

The assumption that the radiative properties of contrail cirrus are the same as those of normal cirrus clouds is highly questionable and has been overcome in other studies. We know since Minnis et al. (1998) that contrails remain observable at ages larger 10 h. Many further measurement results have been presented on this since then. Many findings indicate that aged contrails differ significantly from other cirrus. In particular they often contain higher concentrations of small ice particles, with impact on optical properties, sedimentation and life times.

» We have evaluated CESM against the Minnis et al data on contrail optical thickness (Pat Minnis was a co-author on the Chen et al 2012 paper). We now cite the validation of method against Minnis work for cloud optical thickness.

A prerequisite for contrail cirrus simulations is the suitability of the simulations of ice supersaturation (and temperature). See Irvine, E. A., and K. P. Shine (2015), Ice supersaturation and the potential for contrail formation in a changing climate, Earth Syst. Dynam., 7, 555–568, doi: 10.5194/esd-6-555-2015. It would be important to show how good the present models resolve temperature and ice supersaturation along aircraft flight tracks, e.g., compared to qualified reanalyzes or to in-situ measurements.

» The evaluation of ice supersaturation in CESM is in Chen et al 2012 as well and we note it explicitly in the revised text.

With respect to NOX and O3, I miss a discussion of the dispersion of NOx etc. from aircraft engines to grid scale which is known to cause nonlinear O3 changes since early studies in the 1990's, depending on the dilution model assumed. See, e.g., Paoli, R., D. Cariolle, and R. Sausen (2011), Review of effective emissions modeling and computation, Geosci. Model Dev., 4, 643-667, doi: 10.5194/gmd-4-643-2011.

» This is now discussed further in the text with references.

The question whether aviation NOx emissions cause a positive or negative or zero RF is still under debate. See Pitari et al. (2016), Radiative forcing from aircraft emissions of NOx: model calculations with CH4 surface flux boundary condition, Meteorol. Z., 23, doi: 10.1127/metz/2016/0776. The CH4 surface boundary condition seems to matter. The present study uses prescribed CH4 at the surface which likely has consequences for the results.

» This is now noted in the text (introduction). The positive RF associated with the short-term NOX-induced O3 production is to some extent offset by the negative RFs associated with aviation NOX emissions leading to a relatively large uncertainty associated with the overall net NOX-induced RF (Holmes et al., 2011). It is noted that this large uncertainty is also in part due to the different spatial and temporal scales associated with the positive and negative forcing (IPCC, 1999; Wuebbles et al., 2010). It is generally reported that the net effect is positive (IPCC, 1999; Holmes et al., 2011). This indicates the importance of accounting for the full suite of aviation NOx-induced RFs when reporting the net aviation NOX-induced RFs which have been often reported considering just the main forcing components (Lee et al., 2009 and Holmes et al., 2011).

Work from other teams is hardly mentioned. Differences in the results between this study and other studies (e.g. for NOx induced O3 and CH4 changes or contrail RF) are neither mentioned nor discussed.

» We discuss contrail estimates with other models further (as noted below) in response to this and other comments. We have added more references. We have also added a paragraph to the section on ozone on page 2 that discusses more fully some further studies on the chemistry and the complexity of the NOx, O3 and CH4 results, and potential differences between models.

As mentioned in a comment by M. Ponater, the studies by Olivié et al. (2012) and Huszar et al. (2013) are related to this work, and should have been discussed.

» They are now mentioned. Huszar does not include aviation aerosols, Olivié has a simple calibrated description of contrails.

The results and conclusions are not always clearly presented. One example: The authors relate (in the abstract and the conclusions) the non-local surface temperature signal from local radiative forcing to advection. The assumption that advection is the reason for nonlocal behavior is reasonable and not fully new (Ponater et al, Ann Geophys., 1996; Shindell and Faluvegi: Climate response to regional radiative forcing during the twentieth century, Nature Geosci., 2, 294-300, doi: 10.1038/NGEO473, 2009; see also Rind et al., 2000).

» Now cited in the text.

However, this paper does not bring any new argument to support this conclusion in this paper, except that the (noisy) temperature patterns exhibits downstream shifts. Is that worth mentioning as a new finding in the abstract?

» Yes, the finding is worth mentioning, because too many studies in the literature assume that regional forcing results in regional response. Noted in the text.

I think, this needs more analysis.

» Noted in the discussion.

The authors claim that heat absorption by BC from aviation is large enough to cause strong warming in contrails. They refer to Liou et al. (2013) in this respect. There is no doubt that BC does change absorption when present in sufficient amount. The quantitative results depend on the assumed BC mass (and effective sizes) of the soot and the mass of ice particles (and their sizes) and the fraction of ice particles containing soot particles. It would be important to check the mass budget of the BC in ice particles and see if this mass budget is consistent with the aviation BC emissions, the lifetime of cirrus and aerosol sinks.

» The mass budget of BC in GATOR-GCMOM has been evaluated by Jacobson et al 2011 against observations, and this has been noted in the text.

For example one could analyze from the model results the total ice mass in the cirrus clouds regionally or globally, convert that to cross-sections, and could compare this with the total BC mass from aviation in the same clouds. I would not be surprised if the mass or area fraction of BC turns out to be by far smaller than that for cirrus ice.

» The BC mass is going to be much smaller than the ice mass in the clouds. The BC contributes to anomalous absorption as described in Jacobson et al 2012 and in the text. We make this clearer in the revised manuscript. We did not mean to imply that BC would be a large mass fraction.

In spite of parallel studies by Righi et al. and Gettelman et al., I am not convinced that aviation sulfur emissions change low level clouds in any significant manner.

» Agreed. However, that is what the results of simulations indicate. We now add a more recent study using very different methodology by Kapadia et al 2016 that indicates the same thing.

I miss a careful and critical discussion of the amount and concentrations of cloud condensation (CCN) particles of reasonable sizes which could be contributed by aviation compared to the many other sources for CCN. Again, one could compute from the model results related statistics. To my understanding, most of the CCN in stratus clouds come from non-sulfur sources.

» We examined the perturbations in more detail and state the percent perturbations to liquid clouds by aviation in 2050 v. no aviation, as well as the locations (more in the sub-tropics). This is now discussed more fully in the text to back up the discussion of the mechanism.

I do not know of any single measurement showing that aircraft could indeed change water clouds by sulfur emissions. So, to me, this effect appears to be purely hypothetical.

» We agree with this and note the extreme uncertainty here. It does seem hypothetical until we can verify it. But the mechanism we propose is self-consistent in the model.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2017-218, 2017.