Response to review of "*Emission metrics for quantifying regional climate impacts of aviation*" by Marianne T. Lund, Borgar Aamaas, Terje Berntsen, Lisa Bock, Ulrike Burkhardt, Jan S. Fuglestvedt and Keith P. Shine

We thank the reviewer for the careful review and useful comments and suggestions. Responses to individual comments are given below.

Anonymous Referee #2
Received and published: 16 March 2017

Major Comments:
142: The RF kernels of Samset and Myhre (2011) are valuable because they are vertically distributed. But they are also limited in their spatial coverage. How do the authors map between the regions in their study and those in Samset and Myhre (2011)? For example, it would seem the latter does not provide any results for the author's SAS and SPO regions. A more complete 2D spatial mapping of aerosol direct radiative forcing efficiencies is provided in Henze et al. (ES&T, dx.doi.org/10.1021/es301993s, 2012). Perhaps results from these two studies could be combined to provide a more complete analysis? Or at least findings from the latter could be used to provide some sense of the uncertainty involved in using only the Samset and Myhre regions as the basis for the present work.
The RF kernels are calculated by applying globally uniform aerosol perturbations in each vertical layer, thereby providing full 3D fields (as already mentioned in the text), not kernels for separate geographical regions (Samset and Myhre 2011 then averages the RF globally and in different regions to illustrate geographical differences). We realize that this may not be clear from the cited literature and have added the following clarification to the methodology section in the current study:

"(..) where the radiative forcing per burden was derived by imposing globally uniform perturbations of given aerosol species at 20 different pressure levels from the surface to 20 hPa."

A comparison of results derived by using the RF kernels and sensitivities from the adjoint modeling of Henze et al. (2012) (as well RF from full radiative transfer modeling) would be an interesting sensitivity study as in both cases RF values depend on the chemistry-transport model used to establish the background aerosol concentrations (OsloCTM2 and GEOS-chem). However, we feel that this is outside the scope of our study.

180-185: This argument feels a bit thin, given that some aspects of aerosol cloud interactions are at least better known that others. At the very least, could uncertainties owing to these processes be carried through the calculation, so that we know when uncertainties in these effect may alter the sign of the next outcome?
While we agree that the understanding of interactions between anthropogenic aerosols and liquid clouds has improved, we still maintain that the impacts of aerosols on ice clouds remain highly uncertain. Furthermore, estimates of sector specific aerosol-cloud RF remain scarce. Studies of the potential effects of aviation BC on large scale cirrus clouds have yet to agree even on the sign the radiative forcing, and the magnitude of the impact depend heavily on assumptions in the

models, ranging from -350 mW/m^2 to + 90 mW/m^2 even in a single study (Zhou and Penner 2014). Of course, if either of these number represent the actual impact, this effect would dominate the climate effect of aviation.

To our knowledge, only three studies have presented estimates of the impact of global aviation aerosols on liquid clouds, with results ranging from -46 to -15 mW/m2. We do not have the resources to quantify aerosol-cloud interactions of regional aviation emissions, but these three studies at least agree on the negative sign the global mean forcing, which could offset a considerable fraction of the warming from other components in the short term. So we have rewritten and added more detail:

"Moreover, our results do not include effects of aerosol-cloud interactions, which is an important caveat. Studies have suggested a potential impact of aviation BC on large scale cirrus clouds, but have yet to agree even on the sign of the radiative forcing (Zhou & Penner, 2014). A few studies have investigated effects of aviation emissions on liquid clouds, with global mean RF estimates ranging from -46 to -15 mW/m^2 (Gettelman & Chen, 2013; Kapadia et al., 2016; Righi et al., 2016), i.e., a cooling that could offset a considerable fraction of the positive RF of contrail-cirrus and ozone on a global scale. However, at present uncertainties in these estimates are also very large, and we consider that their inclusion here would be premature."

For Fig 4: why compare ARTP(20) and ARTP(100), when a more direct and fundamental comparison would be to just consider the RCS's? The RCS is what other people will need, if they are to use the results from this study themselves to calculate ARTPs. At the very least it would be quite useful to compare the RCS values in addition to the existing figures using ARTP in particular years.
While it is correct that the RCSs are needed as input if other people are to repeat the ARTP calculations, for instance with updated RF estimates, the ARTPs are what is needed in order to be able to make first-order estimates of the regional temperature impact of given emissions (the core application of the emission metrics). The RCS also do not have a temporal resolution, but are constant factors to distribute the impacts regionally. So, for instance the behavior of the net NOx impact over time would not be illustrated by the RCSs.

Furthermore, we have not estimated new RCSs in this study (Fig. 4 shows a comparison against temperature response to global aviation – not the response to perturbations in individual latitude bands – and only for NOx). The RCSs have been tabulated in previous literature, but to further aid the reader, we have added a summary table in the supplementary material.

General: For other people to make use of these results, it is useful to provide more information on the aviation emissions used in this study. The authors should provide a table of emissions by species and region, and they should provide separate total for emission by takeoff vs cruise altitudes. While it would be great if they could provide metrics broken down by the later category as well, I'd guess that would involve repeating a lot of calculations. But at least providing the details of the inventory they used would allow future users to be able to scale evaluations of the climate impacts of their own inventories accordingly, given some knowledge of how the authors' inventory was distributed vertically.

A table with total aviation emissions by species and region is already included in the supplementary material.

To separate metrics by cruise and landing/takeoff (LTO) operations we would indeed need to repeat our model simulations, which require additional resources and time that are not available. Guidance on how to access and use the AEDT emissions in atmospheric models is provided in a technical note by Barrett et al. (2010), including how to define and estimate emissions during LTO, allowing users to for instance use our metrics with emissions broken down by category. We have added the following:
"Guidance on how to access and use the AEDT emissions in atmospheric models is provided by Barrett et al. (2010). For input to the OsloCTM3, emissions are interpolated to the model's horizontal and vertical resolution, and averaged monthly."

Minor comments:
35-38: True, but this is evident from the fact that the RCS's in e.g. Shindell 2012 are not uniform. So it is a bit odd to place this in the abstract, although I agree the application does being attention to the issue.
The lack of one-to-one relationship between regional forcing and temperature response is one of the key features that can be emphasized by the use of sub-global, temperature-based emission metric such as the ARTP. It also points to added value of moving beyond RF-based emission metrics, such as the GWP. While this may be clear to the scientific community, it is not necessarily obvious to decision makers. We therefore feel that is an important point to highlight.

40-41: The feels a bit obvious (biggest emissions have the biggest impact) âˇAˇT would discussing the impact per emission be of more interest?
We do not feel that it is obvious that the bigger emissions lead to a largest net warming impact "in all latitude bands" (which is what we write in the abstract) as that is a result that emerges from our analysis. Since the reviewer flags this as a minor comment, we prefer to keep this as it is. To summarise the relationship between each emitting region and each response region (as shown in Figure 2) in a short sentence in the abstract would be very challenging.

66: This statement is missing references.
There are several studies of how regional emissions affect atmosphere and climate; we have added to examples – one general and one aviation specific: Berntsen et al. (2005, doi: 10.1007/s10584-006-0433-4) and Stevenson and Derwent (2009, doi: 10.1029/2009gl039422).

78: The phrase "in a grid cell" is vague (we don't know yet how big your model grid cells are) and also ambiguous with regards to whether you are referring to grid-scale changes in temperature or grid-scale changes in emissions.
The wording on line 78 is in fact "at a grid point level", which is meant to be general so as not to make the link to a specific model or resolution. The sentence also states "temperature response and other climate variables", thereby not referring to emissions. However, to clarify have rephrased to:
"(…) very detailed spatial scales (e.g., grid point level)"

83-84: Can the authors reference any in particular?
Specific measures to reduce emissions are implemented at the sectoral (as well as regional/local) level. To assess their effectiveness in terms of reducing the climate impact, one needs to know how the sector contributes to climate change to begin with. We feel that this is a very generic statement that does not require specific references.

91-93: How much uncertainty / error can this aggregation lead to?
This is not easily quantified in a general way, as it depends on, among other things, which measure, which impact and which driver is considered. For instance, in Shine et al. (2005) it was found that the net temperature response to NOx emissions was positive in the Northern Hemisphere and negative in the Southern (due to different relative importance of ozone production and methane reduction), resulting in a very small on global-mean temperature impact. Lund et al. (2012) found a factor 2-7 higher impact of aviation NOx when assuming a non-linear impact function compared with one based on global global-mean input. In terms of RF, Burkhardt and Kärcher (2011) estimated a global-mean contrail RF of 37 mW/m2, but regionally values were up to 300 mW/m2.

105-110: See also Sand et al., Nature Climate, doi:10.1038/nclimate2880, 2015.
Yes, that is a relevant reference and has been included.

183: I don't understand what mechanisms this refers to. Please be more specific and provide references.
This refers to the existing uncertainties also for contrail-cirrus, NOx and aerosol effects considered in our study. We agree that the current wording should be improved. The sentence also fits better after the first two sentences in this paragraph and now reads:
"It should be noted that there is a broad range in the estimates of RF caused by the various aviation emissions reported in the literature (e.g., Brasseur et al. (2016); Lee et al. (2009)) and such uncertainties in RF will propagate to the emissions metrics."

Table 1: could you list NOx in the first half at the bottom the list of species, so that it is easier to compare these numbers to the results for NOx from other studies listed below?
Good point. We have changed the order (and for consistency, also in Table 2).

Table 1: I must be missing something âˇAˇT the GTP and GWP metrics are computed by emitted species (i.e., SO2 instead of sulfate), yet the authors report values for nitrate(Table 1), and these are reported separately from NOx emissions, even though nitrate is formed secondarily in the atmospheric from NOx. Can the authors please explain this more?
We thank the reviewer for pointing this out, the tables are not labeled correctly. The NOx entry includes the impact of NOx on ozone and methane, while the nitrate label is the effect of NOx-induced nitrate formation. To make this clear, the table should read NOx-nitrate and NOx-ozone-methane. We separated out the nitrate contribution since to our knowledge no previous estimates exist, making it difficult to compare the NOx metrics with previous literature if it was included. However, we realize that this may be unclear in further application of the metrics, since these two NOx values would first have to be added. So, we have combined to one metric for the net NOx effect. Tables 1 and 2 have been changed and the text clarified:

"Our estimates also includes the cooling effect from NOx-induced formation of nitrate aerosols, which has to our knowledge not been accounted for in any previous GWP and GTP estimates."

Fig 2: Please explain the difference between the color bars vs star points in panels E and F, and define O3 vs O3PM in the figure caption itself.
This has been included in the caption.

381: Also Lacey et al. (PNAS, doi:10.1073/pnas.1612430114, 2017) used ARTPs to investigate this for cookstove emissions.
This new study of cook stove emission is very interesting, but as far as we understand it focuses on the impact of regional emissions on global temporal temperature change, determined using regional climate sensitivities, but not explicitly presenting results to support the current sentence.

132: Is this perturbation positive or negative? Does it make a difference, for SO2 and NOx?
The perturbation is negative, i.e., we remove 20% of emissions. However, the difference between the reference and perturbed run is chosen so as to determine the impact of the aviation sector emissions (rather than the impact of a specific emission increase or decrease), so in that sense it will not matter. It is possible that non-linearities in the chemistry would result in differences compared to a positive perturbation. The impact of such non-linearities when applying different size perturbations with the same magnitude has been found to be relatively small (e.g., Hoor et al. 2009; Lund et al. 2014; Myhre et al. 2011).

Fig 2: Given the factor of 0.5 in Eq 2, why isn't O3PM 50% that of the CH4 in panels E and F? Is this a consequence of the spatial re-scaling from Fry 2012? If so, I would have expected it to be less than 50% in some regions and greater than 50% in others.
This is because CH_4 includes the RF of the methane-induced stratospheric water vapor change as described in Sect. 2, while the RF O_3PM is estimated as 0.5 of the "pure" RF CH_4. We have added a clarification in the figure caption. We also discovered a very small error in a couple of the O3PM RF. These have been corrected, but do not affect our results.

222: Why not use the RCS for sulfate for sulfate, rather than for the mean of CO2 and sulfate? What RCS is used for nitrate (although not clear how nitrate is treated anyways)?
The same RCS is applied for nitrate as for sulfate and OC. This has been clarified in the text. We follow the approach of previous studies when adopting the mean sulfate/CO2 RCSs (Collins et al. 2013; Shindell and Faluvgi 2010). There has been one application with the sulfate-only RCS as well (Shindell 2012). In fact, when comparing the RCS, they are very similar (Shindell 2012 and Shindell and Faluvegi 2010), and choosing one over the other are not likely to significantly affect our findings.

492-497: This statement is a convolution of two issues that could be separated, which are that the RF of O3 per ppb is horizontally and vertically variable, and that the climate response to this RF is also variable.
A good point. We have added:
"(…) which in turn also depends on altitude (e.g., Olsen et al. (2013))."

536-538 and 550-551: Fig 4 only shows the normalized results, so it is hard to know how much of an overestimate the authors are talking about here. Can they also provide the absolute results?
The lines in question discuss the possible overestimation in more general terms, based on the existing literature on BC forcing-response. We do not currently have sufficient information on the vertical sensitivity in the BC forcing-response in latitude bands other than the Arctic to say how large such an overestimation could be. The regional distribution in Fig. 4B gives an indication, but it is only one climate model, and given the uncertainty in the magnitude in temperature response to BC in current global climate model we have some reservations against discussing the absolute magnitudes. This is further strengthened by slight differences in the experimental setup and input data in the HadSM3/ECHAM versus CESM/RTP, which, as described in the text, means that differences in absolute magnitude cannot be entirely attributed to model differences. All in all and given the scope of our study, we feel that a focus on geographical distributions is preferable.

Did the authors consider using ARTPs for the land-only response from Shindell 2012?
The reviewer brings up an interesting suggestion. A comparison of the response estimated with the land-only regional climate sensitivities against the corresponding temperature response simulated by other climate models could be an interesting part of a more detailed evaluation/sensitivity study. However, here we are limited by the availability of output from the HadSM3 and ECHAM models.

Technical corrections:
87: as a bridge
133: sulfer dioxide
137: each region are
303: in the present analysis we
Corrected.