

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. Fuglestedt 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 #1

Received and published: 11 February 2017

This manuscript attempts to calculate Global Warming Potentials (GWP), Global Temperature Potentials (GTP) and Absolute Regional Temperature change Potential (ARTP) for aviation emissions. This is a difficult paper to put together since there is essentially I think an important negative conclusion that this method does not work that well for aviation emissions. I would like to see a bit more uncertainty analysis to quantify where the problems are. I think a bit more discussion of the physics of the climate system would help the reader understand the problems with this method. This paper could be publishable with some important revisions.

Generally, it looks like the method described is highly model dependent, which is not really obvious until the end of the paper, and there is not really quantification of the uncertainty. I think this could be better woven throughout. It would be nice to comment a bit more on how Stohl et al 2015 and Shindell 2012 apply to these results: I am pretty convinced that for small perturbations and for non-gaseous species like contrails that these results may not apply.

I think that this work could also use a better focus on some of the physical aspects of the climate system and the feedbacks that might or might not operate. And at a minimum I would like to see quantification of uncertainty in the different parts of the ARTP terms.

Some of the figures (e.g. Fig 3 and S1) could use some clarification as noted below.

The general points are addressed through responses to the following specific comments, where the issues are raised again.

Specific comments are below.

Page 4, L108: how does any of this work account for effects on the coupled climate system? Regional forcing can show up very differently in surface temperature depending on whether the energy goes into sensible heat, latent heat, or evaporation, and will depend on surface type and coupled modes of the climate system. For example: cooling the N. Atlantic deep water formation region with contrails may simply cool deep water and not be felt at the surface. Or it may alter the formation of deep water itself. Don't you need a coupled climate model to do this?

The ARTPs are indeed calculated using input from a coupled climate model. The basis for the ARTP is the regional climate sensitivities for four latitude bands, i.e., relationships between the pattern of RF caused by a certain forcing mechanism and the consequent surface temperature in a

given latitude region. These relationships are established from simulations with the GISS climate model and hence incorporate the response of the coupled climate system to the different forcing mechanisms and regions.

To fully account for the detailed regional temperature responses, one would naturally need a coupled climate model, for instance to capture potential changes in regional responses in a climate differing considerably from the present state. Currently, assessing the impact of individual sectors or mitigation measures is challenging and costly given the large natural variability of the global climate models, and requires significant resources and high technical skill. Emission metrics cannot replace the detailed information from climate models, but they serve as very useful tools to provide first-order estimates of the climate impact of various emissions, and they are based on information from complex models. Until recently, emission metrics have used globally averaged input to produce globally average impacts. Being able to include spatial information is an important step, in particular for the short-lived climate forcers that produce more heterogeneous RF than CO₂.

Although we believe we stated clearly the coupled-climate model origin of our ARTP values, at lines 198-200, to further emphasize this point, we have added the following in the introduction: *“The ARTP uses a set of regional climate sensitivities to provide time-varying surface temperature response in four latitude bands to emissions, accounting for the regional RF caused by the emissions. These sensitivities are derived from simulations with a coupled climate model and express the relationship between the pattern of a radiative forcing and the consequent surface temperature in a given latitude band.”*

Page 4, L110: I’m not sure you can do this with a CTM. The circulation doesn’t adjust.

We realize that this sentence is unclear. As the reviewer points out, the emission metrics cannot be calculated directly from the combination of CTM simulations and radiative transfer modeling, but must include a representation of the climate (temperature) response to the given forcing. The CTM plus radiative transfer calculations allows us to capture detailed atmospheric chemistry and forcing. For emission metrics, the climate response is represented by an impulse response function. This is a simplified, global mean function, but is based on more complex climate models. In our analysis, we also use regional climate sensitivities which are based on coupled climate model simulations in order to capture information about the spatial distribution of the temperature response depending on the forcing mechanism and location.

To clarify, we have rephrased:

“In this study we calculate GWP, GTP and ARTP for global and regional aviation emissions. We consider a broad set of forcing mechanisms and emissions in six separate source regions. Input of the aviation-induced radiative forcing of ozone and aerosols are obtained from simulations with the chemistry-transport model OsloCTM3 (Søvde et al., 2012) and radiative transfer calculations, while the radiative forcing from the formation contrail-cirrus is simulated with the ECHAM5-CCMod (Bock & Burkhardt, 2016b; a). Calculating global and regional emission metrics allows us to capture (...)”

Page 5, L114: again: surface temperature response depends on the surface type. Is that accounted for?

The regional climate sensitivities that form the basis for the ARTP calculations are derived from simulations with a couple climate model, which includes a dependence on the surface type. See also response to the comment above.

Page 5, L117: how uncertain are the regional climate sensitivities?

This is a very good question. A full set of regional climate sensitivities (RCS) have so far only been estimated using one climate model by Shindell and Faluvegi (2009). Standard deviations over the 80 year model integrations show quite some variability in the response to certain species, at least in some regions. Two studies have repeated the experiments from Shindell and Faluvegi (2009) for BC with similar findings in terms of spatial distribution of forcing-response (Sand et al. 2013; Flanner 2013). The coefficients also seem fairly robust compared to the response to historical aerosol forcing in several other climate models (Shindell 2012). However, a formal quantification of their uncertainty is currently not possible and repeating the experiments to establish sensitivities with additional climate models is beyond the scope of this study, and would likely require a dedicated intercomparison exercise. Baker et al. (2015) showed that the zonal mean temperature response to BC emissions can differ significantly between different climate models both in magnitude and spatial distribution, while the models agreed better for SO₂ emissions. Some degree of model dependence in the RCSs cannot be excluded, in particular for some species.

To make the potential model dependence clear earlier in the manuscript, we have specified in the final paragraph of the introduction that the sensitivities have so far only been derived by one climate model. We have also incorporated some of the information above in the text.

Page 7, L173: I do not think you can throw out this term: it is a compensating effect or feedback. You could include the reduction in cirrus as a separate forcing, but now you are overstating the contrail impact.

We agree with the reviewer here and have included the 20% reduction in the RF of contrail-cirrus, assuming it is spatially constant, i.e., that the same impact occurs in each of our emission source regions.

Page 7, L184: What are these uncertainties?

We have clarified this sentence and added references to studies showing how there is a large range in aviation RF estimates reported in the existing literature. The sentence also fits better after the first two sentences in this paragraph and has been moved up. It 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.”

Page 7, L197: Does previous work show that this represents the physical response?

I have a hard time believing it works for non-uniform emissions in the vertical, or for aviation emissions.

The regional climate sensitivities are based on coupled simulations with the GISS model. One can argue whether current climate models represent the physical response, but they are at present the only tools we have.

As described later in the manuscript, the GISS simulations involve perturbing the RF resulting from anthropogenic emissions from all sources. Since, as the reviewer points out, the temperature response per unit RF can depend on the altitude of the RF for some species and regions, applying the sensitivities to situations which differ significantly from that originally used to derive them does introduce uncertainties. This is precisely why we take the analysis one step further to evaluate the application and point to where more research is needed in order to improve the general applicability of the ARTP.

Page 7, L200: Are these done for contrails? This is a cloud forcing, which is very different than other forcings. For example: contrail forcing is a SW and LW forcing that depends strongly on surface albedo. Do these sensitivities take account of that? If not, I'm not sure the method is valid.

We feel that the text starting at line 221 made clear what RTPs are used for the contrail-cirrus case, and the limitations of the assumptions we made.

The contrail-cirrus RF is quantified through detailed simulations with the ECHAM5-CCMod and takes the surface type and other cloudiness into account.

In terms of temperature response, we do not have regional sensitivities that have been derived from climate model simulations specifically for contrails (as we noted at line 221 to 223). Currently, such information does not exist in the literature, and we recognize that this introduces additional uncertainties in our analysis. To represent the regional response to a regional contrail-cirrus RF, we use the sensitivities based on the forcing-response relationship for CO₂ and sulfate. This includes trapping of long-wave radiation and reflection in the short-wave and the sulfate forcing likewise, as is well established, depends strongly on surface albedo. As described above, the sensitivities are based on coupled climate model simulations, which includes the dependence on surface type – for the CO₂ and sulfate forcing.

As we noted at line 226, the only two available studies of the climate response to contrails have shown that the global efficacy of line shaped contrails could be significantly lower than one, and the reasons for this may well be as a result of the distinct nature of the contrail forcing which the reviewer mentions. This lower surface temperature response per unit forcing compared to CO₂ could be accounted in our (and future) calculations by scaling the temperature response (as we discuss in Sect. 3.3). However, we do not have any information about how the efficacy could vary regionally, as we noted at line 229.

Page 8, L202: What does the statement in parentheses mean? What is a component? If the time is longer than the lifetime for a pulse emission, isn't the forcing gone?

This statement has been corrected, clarified and moved to the beginning of the next paragraph: *“Equation 4 can be used for short-lived species where H is significantly longer than the lifetime of the species (typically days to weeks).”*

Page 8, L207: This presumes that all RF is created equal with respect to sensitivity. I think this is valid for GHG in the IR. I'm not sure if it is valid for SW effects, and probably not for clouds.

This is an important point. We already touch upon this issue in the following paragraph on efficacy, but we agree that a more detailed discussion would be valuable. Efficacies can be included in the application of metrics by adding a scaling factor, and will act to reduce or increase the temperature change to emissions. However, only a couple of studies have investigated the (global-mean) efficacy for aviation-induced forcing mechanisms and are limited to line shaped contrails and ozone, and we therefore include efficacy only as a point for discussion in Sect. 3.3.

The concept of efficacy, or climate sensitivity parameters, also relates to the RCS (as the reviewer touches upon in the next comment). These are in themselves a type of climate sensitivity parameters for the four forcing agents considered in the Shindell and Faluvegi (2009) study (BC, ozone, sulfate and CO₂), expressing differences arising from the forcing distribution in the horizontal.

We have therefore the discussion and added:

“The temperature response per unit RF can differ between different forcing mechanisms, i.e., mechanisms can have their specific climate sensitivity parameter. This is often expressed as efficacy, defined as the ratio of the climate sensitivity parameter for a given forcing agent to that for specified changes in CO₂ (Hansen et al., 2005). Efficacies can be included in the metric application as a scaling factor. However, presently only a few studies have investigated the efficacy of selected aviation-induced forcing mechanisms. The efficacy depends primarily on the spatial distribution of the RF, both in the horizontal and vertical. The $RCS_{i,l,m}$ implicitly include differences in efficacy of individual components arising from the horizontal forcing distribution (to the extent that the driving processes are accounted for the underlying climate model simulations). The $RCS_{i,l,m}$ are established for the four forcing agents BC, ozone, sulfate and CO₂. Contrail-specific regional sensitivities do so far not exist. By using the average sensitivities of sulfate and CO₂ to calculate the contrail ARTPs, we include both a longwave absorption and a shortwave scattering component. The efficacy of scattering aerosols and greenhouse gases is also likely less dependent on altitude than for absorbing aerosols. However, studies have indicated that the contrail efficacy may be as low as 0.3-0.6 (Ponater et al., 2005; Rap et al., 2010). Furthermore, little or no information about the dependence of the climate sensitivity parameter of contrails on the horizontal forcing distribution exist. It should also be noted that efficacies from the small sector-specific forcings can currently only be derived by scaling them to produce a forcing large enough to get a significant response in the model. This add an additional uncertainty to deriving reliable $RCS_{i,l,m}$. Using the $RCS_{i,l,m}$ for a different forcing agent to approximate the response to contrail-cirrus allows us to include a broader set of aviation-induced forcing mechanisms in our analysis, but is also an important caveat. As discussed in more detail in Sect. 3.4, the climate sensitivity parameter can also depend on altitude, in particular for absorbing aerosols and ozone in certain regions. We explore potential uncertainties in our analysis arising from such vertical differences by comparing the temperature responses estimated using the $RCS_{i,l,m}$ with temperature simulated by three additional climate models (see Sect. 2.3).”

Page 9, L229: So you are trying to put these uncertainties into 'efficacy', which would affect the contrails by a factor of 2, but then ignore it? I don't think that is appropriate. Efficacy might be a way to address some of the issues I am referring to with contrails in this context.

The issue of efficacy is not ignored, but brought up again in Sect. 3.3, where we discuss how a lower contrail efficacy may affect the regional temperature response to present-day aviation emissions estimated with the ARTPs. However, currently, only two studies have estimated the efficacy of contrails and estimates of efficacy of other aviation-induced forcing mechanisms are very scarce (one study looked at aviation ozone). We emphasize again that these two studies provide information on the global-mean efficacy but we have no information about how this would influence regional climate sensitivities. Furthermore, in response to the question above, we have added a more detailed discussion about efficacy in general.

Page 9, L241: There are emissions scenarios for aviation that specifically address the non-uniform growth of emissions and the efficiency improvements. Why not use them? There are already RF estimates in the literature with these scenarios.

Here we try to relate RF to an emission and discuss whether to relate contrail RF to flown distance or to fuel use. Since fuel efficiency changes with time, the two metrics would give different results. Simulations for future air traffic do not help with the decision how to formulate the metric.

An assessment of the contrail formation and RF from regional aviation emissions under a different emission scenario and climate would of course be an interesting study on its own. However, repeating the regional model runs with a different emission inventory requires resources beyond what we have available. Our global emission metric calculations could be repeated using input of RF from studies that the reviewer alludes to and could make an interesting sensitivity study. However, since we focus primarily on the regional metrics here, we leave this to further applications of our framework.

Page 10, L270: I do not think the climate system is linear enough to small perturbations to make this a valid assumption. Can you provide a reference demonstrating that multiplying a forcing by an order of magnitude or more yields a linear scaling to the response to the small perturbation in a coupled climate model? I see now there is a reference below. Is it valid?

Here, the reviewer seems to question the approach and requests a citation, but then (without any justification) questions the validity of our citation upon discovering that we already have one.

Applying a scaling factor (or performing very long model integrations) is a common approach that makes it possible to use climate models to study the response to specific sectors or perturbations. Because of their small size, the use of unscaled forcings to derive reliable responses is very challenging and costly due to the climate model's internal variability.

The Shine et al. (2012) study illustrates that non-linearities may indeed arise from using different scaling factors in the specific case of BC perturbations from road transport. In contrast, the experiments carried out by Mahajan et al. (2013) (BC x 1, 2, 5 and 10 in the CAM4) actually show a quite linear relationship between the magnitude of the BC perturbation and global temperature response. Nonetheless, this is why we encourage the reader to keep the potential

uncertainties in mind when interpreting the absolute magnitudes of the simulated temperature response (see response to next comment).

Page 10, L274: But isn't the temperature response the heart of this paper? Yet it is to be used 'with care'? I think you might need to quantify the potential uncertainties here.

At the core of the evaluation of our metrics against the additional climate model simulations is the geographical pattern of temperature response. The sentence in question aims to caution the reader against potential uncertainties in *absolute magnitude* that may arise from applying a scaling factor in the climate model (as the reviewer also points out in the comment above). A quantification of uncertainties arising from the use of different magnitude scaling factors require a large set of additional model simulations and is beyond the scope of the study. However, we agree that this can be made clearer and have rephrased:

"(...) and potential uncertainties should be kept in mind when considering the absolute magnitude of temperature responses."

Page 10, L278: I do not think figure S1 is correct. I do not see how a 2ppb ozone perturbation causes a 1K surface temperature change in Figure S1. Please clarify or justify that.

Fig. S1A shows the ozone perturbation caused by present-day aviation NO_x emissions simulated by the OsloCTM3 before scaling for input to the CESM. We thank the reviewer for noticing that this was not specified in the figure caption. It has now been added:

"Figure SI 1: (A) Annual mean ozone concentration change from OsloCTM3 caused by global aviation NO_x emissions and (B) annual mean surface temperature response to the aviation ozone perturbation scaled by a factor 40 as simulated by CESM1.2. Hatching indicates statistical significance at the 0.05 level."

The text in the manuscript has also been corrected accordingly:

"Figure SI 1A) shows the zonal annual mean ozone concentration change caused by global aviation NO_x emissions from the OsloCTM3 (i.e., before scaling), while Fig. SI 1B) shows the annual mean CESM2.1 temperature response to the scaled ozone perturbation (hatching indicates statistical significance at the 0.05 level)."

Page 11, L287: But I thought above the forcing was an impulse, I.e. Non- constant?

Yes, the emission metrics ARTP(H) are pulse-based. As described in Shindell (2012), the RCSs can also be used to estimate the approximate equilibrium temperature response to a constant forcing by multiplying the sum of RCS-weighted regional RFs by the equilibrium climate sensitivity (in our study the ECS inherent in the IRF from Boucher and Reddy (2008)). We have modified the text and hope this makes it clearer:

"For comparison with climate model results, we use the regional climate sensitivities to estimate the regional equilibrium temperature response ($\Delta T_{i,r,m}$) to a constant forcing following Eq. 6 of Shindell (2012)

$$\Delta T_{i,r,m} = \sum_l RF_{l,r} \cdot RCS_{i,l,m} \cdot ECS \quad (8)$$

and adopting the equilibrium climate sensitivity (ECS) inherent in the IRF from Boucher and Reddy (2008) described above."

Page 11, L294: This is confusing regarding emission units: since they are rationed to CO₂, shouldn't they be relative to aviation CO₂, then the comparison with aviation CO₂ is explicit. Pardon me if that is not the appropriate definition, but I'm not sure how to interpret the very large BC GWP number.

We adopt the standard method of reporting GWPs as being relative to CO₂ and since the radiative efficiency of CO₂ (which goes into the emission metric calculations) is the same for all sources of CO₂ the reviewer's comment is not relevant. The unitless GWP and GTP of individual species is the AGWP and AGTP normalized by the AGWP_{CO₂} and AGTP_{CO₂}. However, for CO₂, the radiative efficiency is the same regardless of the source. Of course, if one wants to for instance estimate the temperature response to aviation CO₂ using the AGTP_{CO₂}, one needs to multiply by the CO₂ emissions from this sector.

Page 11, L305: But does this treat compensating LW and SW effects which are a function of surface type?

Yes it does.

Page 12, L323: Dry conditions in S. Asia? Really? In some seasons. What about the summer monsoon?

We agree with the reviewer that this may be a confusing sentence to some readers. We are referring to the high altitudes where the majority of aviation emissions occur in this region, as is noted in the first part of the sentence in question. However, drier conditions at these altitudes are not limited to this region. Furthermore, this region covers both parts of the Middle East and Arabian Peninsula, in addition to India. We have rephrased for clarification:

"(...) dominated by emissions at high altitudes (i.e., few flights landing or departing within the region), where conditions are drier (i.e., less wet scavenging of the aerosols)."

Page 13, L345: I'm not sure I agree with this. How do you deal with non local energy balances in each region?

As we noted above, the regional climate sensitivities are based on coupled climate model simulations where the perturbation is imposed separately in the four individual latitude band and the consequent temperature response averaged over each of these. Hence, these non-local energy budget terms are included.

Page 13, L364: Is it valid to apply RCS from one model to CTM results from another model? They might be vastly inconsistent, particularly in remote regions.

There are of course differences between the different chemistry-transport schemes, such as in vertical profiles of trace gases and particles or polar transport. However, we are not aware of literature that support that current models are vastly inconsistent. The reviewer's contention is not supported by any references and seems speculative.

Page 13, L370: I'm not sure the remote effects being larger make physical sense. See next comment.

Our results are supported by the existing literature. BC aerosols from aviation that are emitted in the 60-90°N region, or transported in there from lower latitudes, are located at high altitudes. Several studies have found that the efficacy of BC decreases with altitude and that in the Arctic, high altitude BC even causes a surface cooling (Flanner 2013; Ban-Weiss et al. 2011; Samset

and Myhre 2015). In the case of aviation, the contribution to temperature response in the 60-90°N region caused by within-region BC is cooling. In contrast, BC at latitudes further south warms the atmosphere locally, which in turn gives a warming impact on the Arctic transport of energy, not transport of aerosols. Since the Shindell and Faluvegi (2009), two other studies have found the same feature (Flanner 2013; Sand et al. 2013).

Studies have also shown that the relative importance of local and remote BC for the Arctic temperature response dependence on the source and/or region: BC located at lower latitudes in the Arctic have a warming impact on surface temperatures. Hence, for emissions from sectors/regions closer to or in the Arctic, the local contribution is more important (e.g., Lund et al. 2014).

While the scientific community's understanding of the climate response to BC forcing may not yet be complete, we are not aware of published literature that disagrees with the findings of the above mentioned studies.

Page 14, L378: Here it looks like the GISS results are convolving the effect of transport with the effect of aerosols: they co-vary, but the aerosols do not cause warming. They just come in with warmer air. I think this highlights some of the problems with this methodology. I think you draw the wrong conclusion here.

If we understand this comment correctly, we think this is precisely the point behind what we (and previous analyses) argue: The remote impact is not caused by the direct warming by the aerosols, but by the transport of the air that has been warmed. See also response to comment above.

Page 14, L390: Unless emissions occur in the stratosphere, in which case contrail formation drops.

Yes, good point. We have specified this in the text.

Page 15, L406: what about significant changes in meteorology: warming reduces contrails, and also changes the tropopause height.

Yes, this is a relevant point, but it is one we addressed through the citation in the very next sentence. To clarify further we elaborate:

“Furthermore, future changes in climate may alter the meteorological and dynamical conditions, and hence affect the potential for contrail-cirrus formation in a given region (Irvine and Shine, 2015)”

We also note that the reviewer is being too simplistic – the change in tropopause height leads to a moistening of the air at a given altitude in the extratropics, and likely leads to more contrails. This point is too detailed for this paper, but our sentence deliberately refers to changes rather than specifying signs.

Page 16, L454: I'm not sure exactly what is being plotted in figure 3: please explain further and in a more detailed caption. Temperature change in each region by different emissions in each region? Are the asterix just a label for the bars or does their vertical position mean anything.

The following has been added to the caption:

“Figure 3: Regional and global mean temperature change by species and source region after A) 20 years and B) 100 years following a one-year pulse of emission from the present-day aviation sector in each source region. The asterisk shows the net temperature response in the respective latitude band for each emission source region, while the bars show the contribution from each species to the net.”

And the text has been slightly modified for clarification:

“Figure 3 shows the temperature change (net and contribution from each species) in each latitude band 20 and 100 years after a one-year pulse of present-day aviation emissions in each source region.”

Page 17, L484: If previous studies only did surface emissions (e.g. Stohl et al 2015 and Shindell 2012), then how will some of the terms from these models be valid (e.g. GISS study with RCS I think it is)?

The Stohl et al. 2015 study used the same approach as in our analysis and calculated the ARTPs for a set of components, emission source regions and seasons. They did not derive additional, new regional climate sensitivities, but compared the ARTPs with results from transient climate model runs (similar to our approach, but we consider the simulated equilibrium temperature responses). However, their emission source regions included all anthropogenic sources, i.e., mainly surface sources, and it is unclear if the conclusions from their evaluation of the ARTP can be extrapolated to the aviation sector.

Similarly, Shindell (2012) did not derive new RCS, but used those from the Shindell and Faluvegi (2009) paper and evaluated their robustness compared to historical aerosol forcing simulated by several other models. Again, total anthropogenic aerosol forcing is largely determined by emissions from surface sources, with only a small contribution from aviation.

Hence, while the application of the RCS have been evaluated in previous studies, our analysis is novel because it considers a specific sector. Neither of the abovementioned studies derived new RCS or include new results that could be directly incorporated in the current study. Furthermore, discussing the potential uncertainties in their analyses or the validity of the results is beyond the scope of our study.

Page 18, L511: I could also argue using this figure that the ARTP concept only works within the main emissions region, and is off by 50% or more relative to a physical calculation.

There are of course several ways these findings could be phrased and we could arguably explicitly say that the 28°S-28°N and 28-60°N latitude bands, where we find the best agreement, are the main emission regions. Hence, we feel that this has already been covered in the previous sections of the paper. However, since the ARTP concept is based on coupled climate model simulations, we do not agree with the interpretation in the last part of the reviewer’s argument. And that final part aside, the rest of the argument is just a different way to phrase the same findings.

Page 19, L528: I think this whole analysis highlights that the ARTP concept is very model dependent and not a general concept because of strong dependencies you have identified here.

We think it is important to separate the concept from the input data here. While the ARTP is a useful tool, and is the only emission metric to provide information at a sub-global scale, there is a certainly a need for additional studies to investigate the robustness of the regional climate sensitivities. It is also a relatively new concept, and presently we do not know how model dependent the sensitivities are. As discussed in response to one of the comments above, they seem to seem fairly robust compared to the response to historical aerosol forcing. However, to further examine the model dependence it is necessary to repeat the original simulations with additional climate models, which is of course a major undertaking. What we have shown in our analysis is that there are indeed uncertainties, as well as a need for improved knowledge of vertical dependence in the forcing-response relationships, not only globally, but also in different regions. Our study presents a framework that can be expanded and improved in the future when/if new, more robust and detailed estimates of regional climate sensitivities becomes available.

The ARTP has already been used in many assessments and we believe analyses such as ours are important to highlight the potential applications and associated uncertainties and point to where future studies could contribute. We also conclude that further work is needed for a more robust and general application of the concept, in agreement with the reviewer's argument.

Page 19, L546: I think you are going to have the same problem with contrails, which also have strong vertical effects, and regional forcing dependent on the climate system itself.

We agree that there may be some strong vertical effects for contrails, but at this stage this can only be speculation – unlike the case for BC where a number of pre-cursor papers already demonstrate the effect. Also the situation is quite different for BC and contrails. There is only a limited vertical domain over which contrails can physically exist, whereas there is no similar physical constraint on where the BC can be.

Page 19, L553: Can you do an uncertainty analysis here? Where are the largest uncertainties and model dependencies in the analysis.

We agree with the reviewer that a better quantification of the uncertainties is valuable. Uncertainties in ARTP arise from a number of factors, including emissions, RF and climate sensitivity. Currently, there is insufficient information to quantify the contribution from the regional climate sensitivities. Similarly, uncertainties in the aviation emission inventory are not available. We have however included an analysis of the impact of uncertainties in the different RF mechanisms and the global climate sensitivity on the global mean temperature response to the regional aviation emissions. We have also expanded the discussion on efficacy and aerosol indirect effects.

Page 21, L589: Where do you demonstrate that the response is stronger than the forcing? And how does this analysis include effects of the large scale circulation and feedbacks? You are breaking most of those feedbacks with use of a CTM and unitless response functions from another model.

This can be seen by comparing the latitudinal distribution of RF values given in Table SI2 with the distribution of the ARTPs (although the absolute magnitudes are not directly comparable since the ARTPs are given per unit emissions and as a function of time). Since we do not present the RF in separate figures and additional calculations are needed for a direct comparison of absolute values, we agree that this should be made clearer and have added (in Sect. 3.2):

“This can be seen by comparing the latitudinal distribution of the RF values given in Table SI2 with that of the ARTPs (note that absolute magnitudes are not directly comparable since the ARTPs are given per unit emissions and as a function of time).”

The effects of large-scale circulation and feedbacks are included through the use of the regional climate sensitivities (forcing-temperature response relationships), which are derived from coupled climate model simulations. See also response to previous comments.

Page 21, L606: Add to this some scaling by a factor of 40 for emissions in here somewhere, and I think this is problematic.

To clarify, it is not the emissions that are scaled by a factor 40, but the concentrations. Furthermore, this is not done in the ARTP calculations, but in the climate model simulation used for evaluation (since otherwise there is a difficulty in detecting the signal above the model-generated variability).

Page 21, L608: At least I agree with the negative conclusion here: it is not clear to me that this concept has a lot of utility for aviation. Especially since you have left out aerosol indirect effects. Anything that involves clouds I think is highly problematic using this method. Would like to see a bit more uncertainty analysis to quantify where the uncertainty lies.

Concerning the first sentence of the reviewer’s comments, we note again that it is very important to separate the concept from the characteristics of the input data we have had to use here. We believe that the concept has great utility and hopefully studies like ours will stimulate the development of more “bespoke” input data.

An important application of emissions metrics is the assessment and comparison of the impact of various components under different emissions scenarios or of emission changes following mitigation measures. This in turn provides a basis for comparing and evaluating different mitigation strategies or policies. For instance, if a mitigation measure has different effects on NO_x, CO₂ or contrails, the subsequent climate impact of these species over time can be compared using our metrics. So while we agree that our frameworks has limitations when it comes to for calculating the total climate impact of aviation since we do not include indirect aerosol effects, we do not agree that there is no utility for the aviation industry. Furthermore, once established, our framework can be expanded and improved as science progresses and the knowledge of the aviation-induced forcing mechanisms is improved.

An uncertainty analysis has been added - see response to comment Page 19, L553 above.

Page 28, L882: What are the units of BC? Per unit of fuel or something else? Not clear from the caption.

BC, OC and SO₂ are per unit BC, OC and SO₂, respectively. This has been specified in the captions of both Table 1 and 2.

Page 32, L922: What are the units here? Mili-kelvin? Over what area? Do the regional colors (asterisk) correspond to the emissions colors? I’m not sure what is being plotted here.

The caption has been modified to include a more detailed description of symbols and colors (see also response to comment above). The unit (milli kelvin) is given on the y-axis and the region/area is given on the x-axis (and now specified in the caption).