Reply to the referee comments (comments in black / reply in **blue** / text changes in **red**):

Referee #2:

This GMDD paper describes a massive research effort to combine various classes of models that collectively evaluate the impact of an aircraft flight routes on climate. The mix of atmospheric chemistry and transport, aircraft simulators, and radiation-climate codes is impressive, clearly a result of extensive collaboration of the co-authors on the REACT4C FP7 project. The connection across models is new and highly valuable to the community. In my view the paper should proceed to GMD, but there are a number of worries that I really wished the authors had dealt with. Some are trivial fixes, but some would require much more extensive (and intensive) analysis of this modelling system that may not be possible – I leave that decision up to the editor. Either way, it would be helpful to address the major issues of closure, conservation, time scales of perturbations, and uncertainties.

Reply: We appreciate the positive feedback. We are grateful for all the remarks supporting a better understanding of our paper. Indeed our modelling approach is very extensive and includes a wide range of processes. That is why we think a detailed model description is necessary. We plan to first publish the model description, including a first sanity check. That is, to our understanding, the objective of GMD, and then to publish results concerning the climate cost function and a traffic analysis, elsewhere. We feel that what we have achieved is pioneering, but certainly is not the last word in how such a modelling chain can be constructed. Of course, within the confines of a project with a finite duration, it is impossible to achieve all the closure and analysis that one would ideally wish to do. Nevertheless, we believe that we have a credible, novel, system.

Can GMDD or the authors please use continuous line numbering as it is almost impossible to keep track of and type in the endless page number here?

Reply: This is standard GMD layout, where we do not have any influence on. At least every line is clearly identified by a page and line number.

4346-3 I thought that most persistent (ie, important) contrails were formed by the cold snap in temperatures in the wake vortex and NOT by the added H2O. In fact the added H2O does nothing unless we are supersaturated w.r.t. ice.

Reply: Right, but indeed both are necessary: Supersaturation with respect to ice and water vapour emissions (the Schmidt-Appleman criterion includes explicitly the water vapour emission index). Without water vapour emissions only aerodynamic contrails are formed. But this is a different process. And in fact the same is true for the chemistry. One needs nitrogen oxide emissions and regions with active chemistry (not the lowermost stratosphere during night) in order to produce ozone. This is described in detail in Sec. 3.3.2. First sentence in the abstract is revised.

4346-23ff The sensitivity analysis and sanity checks have failed here ot convince me of closure (ie, that the entire along-flight impacts are correctly integrated) and that we have learned anything about the uncertainty of the results. Since we are talking about small differentials in climate per alternate routes, it seems that the uncertainty is critical and must be estimated. While some is systematic and may cancel across routes, others are specific to different altitude processes or different chemistry (O3 vs CH4) and hence do not cancel when comparing routes.

Reply: We already have formulated our statement in a somewhat imprecise way and talk about sanity checks rather than validation, because we find it difficult to compare our approach with

either measurement data or other modelling studies. This is discussed in detail in Sec. 4. However, this does not negate the need for from sensitivity studies in the application of the climate-cost functions. On the contrary, it shows the need of detailed uncertainty analysis in the application of this modelling approach in future studies. This is now added in the introduction to Sec. 4 (Verification) (see red text below). In addition, we have included a new analysis of the whole fleet and compare this with an AirClim simulation to provide a closure of the approach and a direct intercomparison with an existing model with the assumption that a 1 day simulation can be compared to an annual mean simulation (see red text and figure below).

The limited possibilities to directly compare the results of our modelling approach to either observational data or other model results, emphasises the need to included sensitivity studies in future investigations on the changes in air traffic routing when optimising with respect to climate impact. This can be achieved by targeted manipulation of the climate cost functions with respect to, e.g., the ratio of the impact individual compounds, their variability or pattern and their consequences on the climate optimal routing changes.

To obtain an overall assessment of the metric results (here P-AGWP20), we take the trans-Atlantic air traffic emissions, calculated with the SAAM model for the minimum economic cost, and compare these results with an AirClim simulation based on the same emission data. AirClim is a fast climate-chemistry response model (Grewe and Stenke, 2008; Grewe and Dahlmann, 2012), which takes into account annual mean emissions and their regional different effects (basically latitudes and altitudes), based on a number of precalculated cases with complex chemistry-climate modelling. Note that in REACT4C the climate response is taken the specific weather situation into account, whereas in the case of the AirClim simulation, the emissions are assumed to occur every day at the same place, i.e. identical for all weather situations throughout a year. The results (Fig. 14) show almost identical values for all emission components, except for contrail-cirrus, which is reasonable, since the day-to-day variability of the contrail effects are highly variable and a one-day simulation cannot be expected to be representative for the whole year. However, the comparison shows that the overall results are comparable in magnitude.

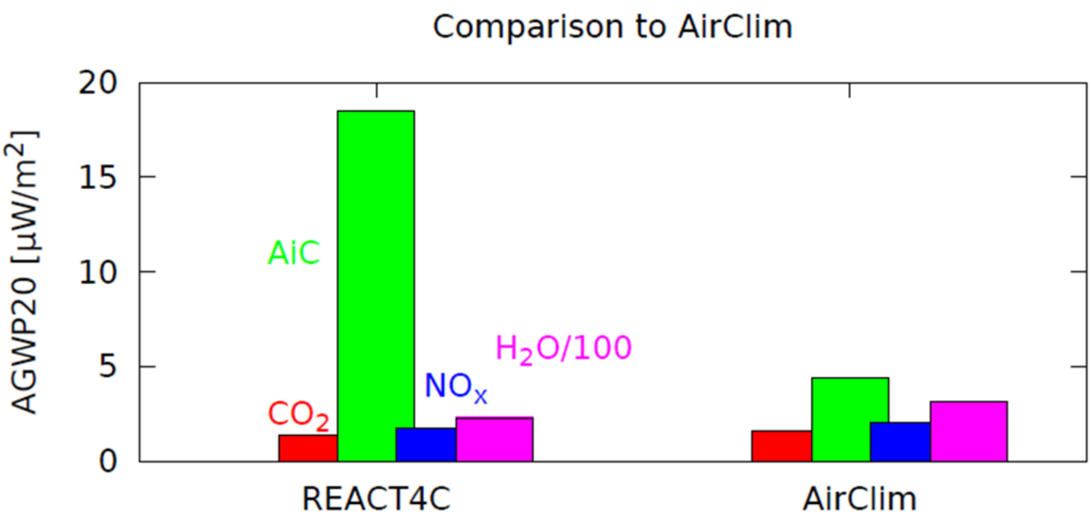

Fig. 14. Comparison of the pulse absolute global warming potential (P-AGWP20) in [$\mu W/m^2$] for emissions of CO_2 (red), contrail-cirrus (green), NO_x (blue), and H_2O (magenta). Note that the H_2O values are multiplied by 100 for presentational purpose. The emissions from the minimum economic cost 1-day trans-Atlantic air traffic, which are calculated with SAAM are multiplied with the climate cost functions (REACT4C, left) and taken as annual mean emissions for the climate-chemistry response model AirClim (right).

4347-8ff It would seem appropriate to reference the IPCC 1999 chapters on chemical modeling and climate in this intro since they really started the major EU initiatives.

Reply: Good point; reference added twice.

4349- It seems the community is moving to "effective RF" – can this be harmonized? If not, OK< but why not?

Reply: There are indeed different definitions of radiative forcing. While the basic concepts behind "effective radiative forcing" (if not the name) have been known for some years, it is not simple to apply. For example, for each forcing one would have to apply a method (such as the Gregory et al. (2002) regression method) using a comprehensive climate model. For our methodology, with so many individual forcings, this would be a truly formidable undertaking. Worse than that, the effective forcing can only really be computed for large radiative forcings, where the signal in surface temperature change exceeds the "noise" in the climate system. The only way around this would be to scale emissions so that they cause a large forcing, but this would then open up many issues concerning the linearity of the response. We have added some text to explain this in the text. Perhaps robust and rapid methods for calculating effective radiative forcing will become available in time, but they are not available now.

4346- Very nice point on the ozone production efficiency

Reply: Thank you!

4350-2 This discussion of a sanity check was very disappointing. There are many such checks that should be done and were not even discussed. For one, there are Oslo co-authors who could calculate aviation effects independently (not optimized with the complete system, but at least for some test cases to compare two alternate routes). The EMAC model could have been run to completion itself with different flight routes to compare with the climate RF calculated with the Langrangian and limited squares analysis. The lack of a 6 month or longer integration to get the CO perturbations correct and thus the O3 and CH4 perturbations is most worrisome. For example the studies is excellent, but it fails to address the major questions about closure and consistency.

Reply: As mentioned in the text (see also Fig. 1), we indeed thought about such a closure experiment and we are still working on it. We have included a closure experiment with the AirClim model, which includes the same effects as in our model approach (CO_2, contrail-cirrus, ozone, methane, PMO, water vapour). Please see the comments above.

Are these partitioned, model-handoff-to-model with totally different scale and scope conserving and integrating the impacts properly? Why not just run EMAC and compare the chemical perturbations – that is what looks most worrisome here. Are the total effects being accounted for. The current paper does not give confidence in this matter.

Reply: The individual models, which we use, are just EMAC and the metrics formulas. There might be a misunderstanding in the model layout. EMAC is modular system with a very advanced interface (MESSy) and a lot of standard procedures. It includes a variety of submodels, such as ATTILA. The procedures to extract data and hand it over to other routines are standardised subroutines well tested and published in 2005 (Jöckel, P., Sander, R., Kerkweg, A., Tost, H., & Lelieveld, J.: *Technical Note: The Modular Earth Submodel System (MESSy) - a new approach towards Earth System Modeling*, Atmospheric Chemistry and Physics, 5, 433–444, doi: 10.5194/acp-5-433-2005, URL http://www.atmos-chem-phys.net/5/433/2005/ (2005)).
The data exchange between individual EMAC submodels is a standard procedure, well tested, and we even haven't considered to mention this. Though, we have now added a sentence in Section 2.1 at the end of the first paragraph.

The second point mentioned suggests a different layout of the experimental design. Actually, Section 3.1. is dealing with this question in great detail. We have decided to use a Lagrangian approach with a tagging scheme for two reasons:

1) **Computational efficiency: For the perturbation approach suggested by the reviewer, one simulation for every time-region grid point has to be performed whereas 15 time-region grid points can be calculated in one simulation using our approach. Since the calculation is enormously CPU-time consuming an order of magnitude reduction in computing time is important.**
2) **Contribution calculation: We decided to concentrate on contribution calculations. And take into account the right impact on production efficiencies, which was appreciated by the reviewer (see above)** "4346- Very nice point on the ozone production efficiency"

4404-Fig6 Cannot figure out what is important here – it just repeats the chemistry with different colors? What's up?

Reply: The point is that one air parcel includes information of all considered time-region grid points (here 15), which is also shown in the graphic and which enables a computationally efficient

simulation (and also diffusion processes in addition to the background air parcels). This is an important issue and also raised by the reviewer in his/her last comment.
We have added this in the figure caption.

4406-Fig8 This set of perturbations is confusing. Is this global? Is the CH4 feedback consistently done within EMAC and is it similar to the results from other published studies, including those from Oslo CTM ? What is conserved here and why/why not? Where are the perturbations occurring – otherwise the odd timings make no sense. When was the pulse? Dec 20? Strange timing.

Reply: Yes, as stated in the caption, these are global atmospheric masses. The emission is on 23rd December 6:00. (Date added now to the caption – was indeed missing - apologies).
The timing looks good to us and we do not know why the reviewer refers to 'strange timing'. The NO_x pulse is largely reduced by 4th January. During this time ozone is built up by NO_x and afterwards ozone is destroyed again. Methane is depleted due to the reactions $NO+HO_2$->$OH + NO_2$ and $O(^1D)+H_2O$->2 OH, which means that methane is reduced as long as NO_x and ozone is available. Lifetimes and the strength of the changes are analysed in Sec. 4.1 and Fig. 14 and agree well with previous findings.

4411-Fig13- This figure makes little sense as it would seem that the RF from absolute H2O would depend entirely on WHERE that H2O was. It would certainly not be linearly if the injected water were below 500 hPa where it would be swamped by the background. So scaling to total H2O perturbations makes little sense.

Reply: This is an important point and hints at a misunderstanding because we were not sufficiently clear as to what is being shown. We totally agree with the referee. The Figure does NOT show the relation of an arbitrary change in mass to the RF. Instead, we consider a constant emission at different points in the atmosphere. This emission leads to two things

1) **shorter lifetime and lower atmospheric mass**
2) **a lower RF**

if we compare an emission at 500 hPa to 200 hPa. What the Figure says is that for a given emission of H_2O we can estimate the RF, if we calculate the change in the atmospheric mass. Hence the relationship between where the emission occurs, and what mass change (and the RF) it causes, is built into the plot. It is not possible to decouple the two points. One would need to set-up a different experiment to extract point 2) the relation between a mass change at various atmospheric locations and the resulting change in RF.

We have added a paragraph in Sec. 3.4.5 and in the Figure caption