Reply to

RC1: ['Comment on jecats-2026-3'](), Anonymous Referee #1, 19 Mar 2026

We thank the Reviewer for the thoughtful and insightful evaluation of our manuscript. We highly appreciate the constructive comments, which have helped us to improve both the presentation and the scientific clarity of our work. We have carefully considered all remarks and revised the manuscript accordingly.

Our detailed point-by-point responses are provided below, providing review comments (**in black**) and our responses (**in blue**).

General Comments

This manuscript describes the D-KULT project, a national scale initiative aimed at demonstrating the operational feasibility of eco-efficient flight routing by building an end-to-end information chain connecting numerical weather prediction, climate response modeling, flight planning, and air traffic control. The topic is timely and societally important. Furthermore, I commend the authors for leading such a complex project across multiple domains.

I find several of the results interesting but many of the details on why such results are observed or sensitivity analyses are deferred to future forthcoming publications.

My high-level concern is that this paper in its current structure ends up downplaying the scientific insights that were gleaned from the project. I therefore recommend the authors to revise the paper with the primary goal being a structural reorientation of the paper to make the new scientific nuggets be highlighted and perhaps sections that primarily describe the consortium workflow (e.g., section 3.1) can be moved to an appendix or supplementary material.

Specific comments

This paper covers a breadth of topics and I realize that it both serves as a "preview" of other publications in preparation and a summary of prior published articles so I keep my specific comments focused on only a few of the issues that I think struck me the most. I also realize that this paper is already fairly long and some of my suggestions ask for more quantitative results, but perhaps moving some of the programatic descriptions of the activities and coordination to a supplementary material might help free space to present some quantitative results. I leave this to the authors.

Reply: Thank you for this thoughtful feedback on scope and length of the manuscript. We realized that initial wording when describing objective of our paper might have been in parts misleading. The revised abstract now better reflects the focus of the paper, which is to report on how to establish a collaborative workflow and the progress achieved, but also challenges identified when working towards enabling climate-optimized trajectories in the current ATM system. We understand that a paper that presents more quantitative results would also be of high interest to the reader. However, this paper is not having a focus about quantitative results but about forging the information chain, testing it and lessons-learnt, which opens the view into necessary future work.

This additional aspect would make this paper simply too long, as such quantitative assessment requires, e.g. to select a set of dedicated flights, traffic inventories or route network that would be analyzed. More quantitative analysis of contrail-avoidance flights which ultimately results in an assessment of climate benefits and damages is under preparation as separate manuscripts. This required more space holds in particular true, as the analysis concept that we suggest relies on a comprehensive approach integrating uncertainties, as proposed in the FlyATM4E project pointing towards robust decision making (e.g. Matthes et al. 2021). Such an approach requires a systematic characterization of individual sources of prevailing uncertainties, which ultimately are combined in an error-propagation approach relying on Monte-Carlo-simulations (e.g. D1.1 FlyATM4E). From within D-KULT there is one quantitative study which explores sensitivities on the forecast uncertainty by providing an overview on a sub-sample of the TF-100 flights, that addresses very specifically sensitivities of estimated climate benefits when applying various MET input data (Peter et al., 2026, submitted). For an understanding how results of applying numerical tools could look like, we have included a visualization of climate estimates for flight segments, informing the reader which non-CO_2 effects would need to be considered, however without showing any uncertainty assessment with it.

- **The two-moment cloud ice microphysics scheme**

In my initial reading I thought this was one of the most important scientific results in the paper, but I realized that Figure 8a is actually a subset of lines (RHi > 100%, 110%, 120%) from Hanst et al 2025 (which additionally has RHi > 105%). Any new insight on this figure beyond Hanst et al 2025, should be really highlighted in this section. Additionally, the paper does not also show the precision curve from the earlier Hanst et al paper which suggests that even in the "3/10 ensemble members agree" case the precision was ~40% so, of the predicted ISSR regions less than half were actually confirmed by radiosondes. Which I think is the central challenge given the <10% prevalence rate. Is this correct? If so I think this needs to be highlighted and further discussion on this would be warranted, especially if new insights since the publication of Hanst et al 2025 have been gleaned.

Reply: We thank the reviewer for his thoughtful assessment of the insights presented in the paper. We have followed the reviewer's suggestion to structurally reorganize the manuscript, resulting in a more coherent and rigorous presentation of our results. The primary scientific contribution of this work is the presentation of methods developed that facilitate interdisciplinary collaboration and enable the collaborative workflow presented herein.

It will also be useful to discuss what the spatial coverage of the radiosonde data was in this comparison.

The paper also states that the 10-member ensemble was initialized from the operational one-moment analysis which "has practically no supersaturation". This results in the spin-up period. I would be interested to know what the resulting ISSR distribution and its evolution look like compared and if it reflects the measured atmospheric state. Additionally did the spin-up period present any real operational hurdles or was this largely immaterial?

The forecast skill analysis (Section 4.1.2) shows ETS declining from 0.7 at 6-hour lead time to 0.4 at 24–30 hours. Crucially, the paper states that "the forecast skill is insensitive to the choice of microphysics scheme". This I think is a result worth expanding on. If the primary source of uncertainty is the underlying linearities in the atmosphere rather than the microphysics parametrization, this has important implications - improving the microphysics scheme alone cannot substantially extend the forecast horizon for PPC prediction. I encourage the authors to discuss this more explicitly and connect it to the practical planning timelines discussed in the other sections.

Reply: Indeed, the insensitivity of the ETS to the choice of the microphysics scheme was a result obtained in the cited paper by von Bonhorst et al. (2025), based on a Master's Thesis. The result may be characterized as this: with the two-moment cloud scheme it is possible to get ice supersaturation and more realistic cirrus clouds in the forecast, but not necessarily at the right time and the correct location. The latter is not a result of the cloud microphysics but rather of the atmospheric circulation and synoptic scale dynamics with its non-linear dynamics. The logical implication is that more cruise-level humidity measurements are needed for data assimilation to keep the modelled atmospheric flow close to measured reality. Such additional upper tropospheric humidity data could be integrated by way of assimilation in NWP forecast models, to better describe the initial state.

Meanwhile, another Master's Thesis is underway which tries a characterization of days with relatively robust contrail forecast and days with non-robust forecasts. This thesis uses the data from the one-moment scheme. A paper has been submitted (von Koslowski and Gierens, MetZ, 2026). It clearly shows that situations with non-robust contrail forecasts are characterized by small length-scales (e.g. autocorrelation of the vertical velocity), probably caused by convective activity and a smaller meridional temperature gradient (weaker and thus less stable dynamics) than in situations with robust predictions. The new paper and its main result are now mentioned in the text. It is noteworthy that both papers point to atmospheric dynamics as an important player in the game.

- **The operational tests**

Section 4.2.3 reports that 9 flights were selected for active contrail avoidance and 16 used the tool in shadow mode. However essentially no quantitative information on these flights are presented - fuel burn penalties, comparison of predicted vs observed contrails etc. The authors state that "evaluation of the changing climate effect of the test flights is ongoing and the results will be published elsewhere". This presents a real problem from a review perspective because it makes it difficult to evaluate the claims that the program accomplished the implementation of eco-efficient flights without referring to future publications. In the current state these sections describe the activities and workflow. I think it will benefit the paper if at least some minimum quantitative results are presented.

Reply: Thanks for the thoughts, whe primary objective of the work presented in this manuscript, is a scientific evaluation of real-world flight trials which were demonstrating the operational feasibility of avoiding regions marked as climate-sensitive by trajectory optimization. Based on these alternative routings, a statistical analysis of satellite data was performed, in order to deliver an independent observation of contrail segments, which is still not a direct quantification of the climate effects or an desired climate benefit. The optimization results generated by FPO Cloud are forecast- and simulation-based, whereas the executed Lufthansa flights were subject to operational influences preventing a scientifically robust one-to-one comparison for which a much larger number of flights would have been necessary to arrive at statistically significant conclusions.

In addition, a validation of predicted contrail occurrence requires a dedicated satellite- or observation-based analysis, which is still ongoing for selected flights. Simulation studies showed that the effectiveness strongly depends on the meteorological situation and forecast quality: for example, the case shown in Figure 5 forecasted an almost complete cruise contrail avoidance with a fuel penalty of about 130 kg, approximately 4 minutes additional flight time, and a reduction of the temperature change metric by 5015 pK. However, other situations would have required trajectory deviations that lead to operational and economical challenges that could have resulted in unacceptable magnitudes in additional time and fuel as well as unacceptable slot delays and capacity restrictions. Consequently, the presented results should mainly be interpreted as a conceptual and operational proof-of-feasibility.

In lines 589 presents an interesting finding about the average thickness of the avoidance regions being 4000ft while the assumed reference value was 2000ft. This is an important finding that I wish had more details on the characteristics of the PPC regions. Was this an artifact of the ICON 2-moment model and it's spin up period? What implications does this have for the operational constraints of contrail avoidance?

Reply: This large vertical extent was statistically examined by DFS based on Clima-1 data of year 2023 over the region EBG Ost. It is not an artefact based on ICON weather model. It is based on determining an average overall propulsion efficiency and the amount of underestimation of relative humidity over ice due to the 1-moment scheme. That is, the combination of both the low 93% RHi threshold for treating a region as supersaturated, combined with the relatively large overall propulsion efficiency of 0.365 enhances the possibility to diagnose contrail persistence (in the model). In our experiment, this led to surprises (or problems) at Air Traffic Control: Increasing dimension of PPC-areas has significant influence on reduction of airspace capacity. Note that this is a model result and one should remain sceptic. To our view, (modern) radiosonde data are still the most important reference to gauge the vertical extension of ISSRs and PPC=1 regions. In order to avoid misunderstanding, the revised version clearly states that the 4000 ft thickness is probably an artefact of the two mentioned conditions and that the true thickness of ISSRs can only be obtained from radiosonde measurements. Hence our analysis also identified the importance of the vertical resolution in data and of the accuracy of such forecasts.

- **Climate response function comparisons**

The finding that contrail aCCF peak values in Clima-3s are up to 10–20 times higher than in Clima-1(s) in some regions despite identical underlying meteorology is an interesting result and potentially represents a major source of uncertainty in the eco-efficient routing optimization. The authors attribute this primarily to differences in "calculation steps". It unclear what exactly this means from physically, and there is no justification provided. Is this a numerical artifact of a regridding step described as a best-practice recommendation in using the aCCFs (lines 511-512)? Or is this because of the differences between the aCCFs and the CoCiP representation. The paper acknowledges the impact this might have on the optimized trajectories but further analysis or discussion would be useful.

Reply: The answer to this question contains two major elements. First, during revision of the manuscript, it was found, that the contrail climate response of Clima-3s as implemented in the experimental operational model simulations needed to be multiplied with 0.244 to receive F-ATR100 to be directly comparable to Clima-1s contrail climate responses. Hence the difference between the contrail climate responses is brought down to a factor of 5. Second,

remaining differences are caused from on the one hand side different methodologies, and on the other hand side different contrail properties assumed during individual "calculation steps". Clima3s relies on a plume modelling on one single Lagrangian airparcel trajectory which is released in the ensemble mean synoptic weather situation. It is expected that single airparcel trajectories cause a higher variability resulting also in the occurrence of higher individual values. Clima1s relies on a statistical regression approach, with training data from 3 different historic weather situations. Within these distinct different concepts, furthermore, contrail properties are chosen quite differently, comprising lifetime, optical depth, contrail depth, width, etc. We state this more clearly in the revised text.
The difference in evaluating individual forecast ensemble weather situations, affects mainly the horizontal extent of the PPC areas. Clima-3s data set is different to zero, as soon as there is one single member showing supersaturation. Clima-1s data set is only showing supersaturation, if the mean atmospheric state is equivalent to ISSR conditions. For consistency, this difference is considered through masking of Clima-3s (with Clima-1s). We point out this more clearly in the text. Regridding can have an effect on individual trajectories and their climate effect, this part was moved to the conclusions section.

It is unclear if the results in Figure 9, where a trajectory optimized using Clima-1s fields yields higher contrail climate estimates when evaluated with Clima-1 is a result of the same issue with differences in calculation steps or if it is due to other reasons. Unfortunately the readers are directed to forthcoming paper for an explanation.

Reply: More detailed information on the differences between the optimized trajectories was added.

Technical corrections

Title - Would it be more accurate to replace "routine" with "operational prototype of" or similar perhaps.
Reply: We agree that the current state of affairs does not yet allow a routine application of the information chain for eco-efficient routing. As we explain in our manuscript, there are still problems to be solved, in particular the amount of manual work is still preventing a routine operation. However, it is the goal of our work to be working towards a routine operation, not to approach prototypes. Therefore, we prefer to keep the expression "routine".

Abstract - The abstract is structured more as an executive summary of a project report than as a scientific abstract. It does not state the main quantitative results of the study.
Reply: We reformulated the abstract to make the objective of the current manuscript clearer. Quantitative results will be published in other papers.

Figure 3 caption and Section 3.2.3 - The term "masked" (as in "Clima-3s masked by ensemble mean PPC") appears in Figure 3 without definition in the main text
Reply: We added this information in the text.

Section 4.2.1 numbering - There are two subsections both labelled "4.2.1" (one on sensitivity to weighting, one on strategic flight planning).

Reply: Corrected.

Line 501 - There is no section 2.2.3 in the paper, this must be a typo.
Reply: Corrected.

F-ATR vs ATR - I believe this terminology is follows from Megill et al 2024 but the "F" is never defined or explained in this work.
Reply: We added this information to the text.

Figure 9 - is this per flight distance/ flight or for all the flights that were optimized?
Reply: The information was added to the text.

Lines 621 - 622 - "54% of the planned aircraft were kept out of the PPC area" - worth clarifying if this is 54% of all the aircraft considered in the simulation or 54% of those whose planned routes intersected the PPC positive regions.

Reply: Here, only 54% of the aircraft planned to cross the PPC area were kept out of the PPC area.