Response to Reviewer 2

We would like to thank the Reviewer for their careful review. We found the recommendations to be very relevant, making the readers more aware of the current state of the science. Below is our response to the comments on a point-by-point basis. The following convention for text fonts is used:

- Review comments in black color
- Our answers in blue color
- *Pieces of text taken from the revised manuscript in red color and in italic font*

Review

This paper introduces the concept of "risk-aware" contrail avoidance strategies. In such a strategy, a trajectory is optimized against a contrail forecast, and the resulting climate impact is re-evaluated by ensemble members to create a risk score that captures the likelihood that such a re-routing will result in a climate benefit or harm. Additionally, the authors present a “risk-optimized” approach, where multiple trajectories are created based on the differing ensemble members, allowing an operator to select the resulting trajectory that achieves the desired climate benefit at an appropriate risk level. The contributions in this manuscript are novel, appear scientifically and technically sound. An exploration of robust optimization strategies for contrail avoidance is currently limited in the literature on contrail mitigation, and thus I see this paper as a valuable and timely piece of work. In addition, the paper is very well written and organized. I recommend the paper be accepted, and only have very minor comments as noted below.

We thank the Reviewer for their overall positive assessment of the study. Replies to the comments are below.

- The measure of risk used in this work is related to the fraction of ensemble forecast members that, when used as input to a contrail model, cause climate harm rather than benefit. The authors attempt to also capture model uncertainty by also adjusting model parameters and measuring how frequently these adjustments result in climate damage. While this presents a valid metric, the authors do not present any evidence that this risk score is well calibrated. For example, seems like it should be possible for no forecast ensembles to predict a climate harm, but for a reanalysis to later show a climate harm. Calibrating this risk score ultimately seems like a very difficult task, because it is difficult for us to directly measure the climate impact of an individual contrail, and ultimately, the risk should be calibrated against observations rather than reanalysis products and models. For this reason, I find it acceptable for such a calibration to not be within the scope of this study. However, I suggest that the authors place a discussion of this important limitation within either Section 1 or 2 of this paper.

The Reviewer correctly points out that for our method to be valid, the actual climate benefit of rerouting should be reliably predicted. Such a condition can be verified using e.g., rank histograms. However, the scarcity of contrail impact observations hinders this verification process, which is left for future work. Following the Reviewer’s recommendation, we added a small discussion about this issue in Section 2: *Moreover, the effectiveness of the proposed strategies relies on the actual climate impact of the rerouted flight being reliably predicted during the planning process. Such a condition can be verified using e.g., rank histograms (Bröcker and Ben Bouallègue, 2020). Whether this condition is met cannot be verified at present, as direct observations of the climate impact of individual contrails remain scarce. Future work should focus on verifying that this condition is met.*

Note that a discussion of this limitation is already included in the conclusion (which was slightly reworked in the revised paper), thus we did not extend too much the new discussion in Section 2: "In particular, in order for our risk-aware decision-making to be valid, the actual climate benefit of a rerouting should be reliably predicted. This condition can be verified using for example rank histograms, that are commonly used to assess the reliability of ensemble forecasting systems (Bröcker and Ben Bouallègue, 2020). This is still an open question that needs to be addressed. We strongly advocate for additional research in evaluating and verifying CoCiP and similar models against observations, before they are used for operational contrail avoidance. Until then, a first step would be to assess whether the climate benefit estimated using reanalysed meteorological data falls within the estimated variability from ensemble weather forecasts, which will be the subject of future work."

- On lines 315-316, the authors state that, for Flight B, the climate impact is highly sensitive to the parameter controlling the enhancement of nvPM emissions, but the same is not true for Flight A. The authors should report the engine types assumed for each of the two flights. It should be noted that recent work has shown that CoCiP is less sensitive to nvPM emission indices when accounting for the activation of vPM emissions, especially in newer lean-burn engines. Further, more recent versions of pyContrails have models of vPM activation as experimental parameters. I do not see it necessary for the authors to re-run their simulations with these experimental parameters, but I do recommend the authors include a reference to Ponsonby et al (https://acp.copernicus.org/articles/25/18617/2025/), and include a comment, possibly in Section 3.3, noting the above and that research in this space is evolving.

Following the Reviewer's comment, we've added the following text to Section 3.3: *Recent work showed that in addition to nvPM emissions, ice crystal formation is dependent on volatile particulate matter (vPM) emissions, especially in new lean-burn engines (Ponsonby et al., 2025). While work is underway to include such findings in CoCiP, we chose not to include these experimental features in our study, but they could be included in the CoCiP parametric uncertainty estimation in future work.*

We also extended the discussion on the role of each parameter in the second paragraph of Section 4.2, explicitly stating the assumed engine types: *Part of the difference in the role of the enhancement factor of nvPM emissions is due to the different emission index of nvPM for both flights, as expected from different engines. The aircraft flying flight A (resp. B) is assumed to be equipped with a GE90-115B engine (resp. CF6-80C2B6 engine) for the estimation of nvPM emissions, leading to an average nvPM emission index of about 2.8 10^{14} kg^{-1} (resp. 7.8 10^{14} kg^{-1}). As the nvPM emission index is higher for flight B, the enhancement factor may have a stronger relative impact for this flight. A detailed attribution of the differing sources of variability between the two flights is beyond the scope of this study, owing to the complexity and the strongly nonlinear nature of the processes represented in CoCiP.*

- In Section 6, the authors explore the concept of risk-optimized avoidance, where flights are optimized individually against a number of ensemble members. This concept is introduced in Section 2, where is it compared to by Simorgh et al, where risk is directly incorporated into the objective function of the optimization process. The authors' approach to risk optimization is valid and has the advantage of being far easier to implement. However, missing from both Section 2 and Section 6 is a comment on the limitation of the author's alternative approach. Namely, that by jointly optimizing against multiple ensemble members simultaneously, it may be possible to construct trajectories that achieve lower risk scores with similar operational costs. It is not clear a priori what advantage such a scheme would achieve, if any, and so this is topic that would warrant further research in a future study.

We added a discussion in Section 2, where the risk-optimised strategy is described, to explain the main differences between the two approaches: *However, each candidate trajectory in our approach is optimised against a single ensemble member rather than jointly against the full ensemble. While our approach is simpler to implement, it may be possible to construct trajectories that achieve lower risk scores with similar operational costs using a single optimisation in which all ensemble members are considered simultaneously, as done by Simorgh et al. (2024).*

We emphasise that no matter the approach, the idea of the risk-optimised strategy is to process ensemble members directly in the flight planning system, rather than after an alternative route is made available.

The following comment is entirely editorial. The authors may consider this comment in the revision of this work at their discretion:

- Section 5 of your paper shows that the majority of contrail warming may be avoided through trajectory optimization at a relatively low risk level. This is somewhat a corollary of the main results of this paper. I have a concern that the presentation of the current paper may lead some readers to reach a different conclusion. Specifically, in Section 4, the authors show an example of a flight with a relatively high risk level. Based on the statistics in Section 5, it appears that this higher risk flight is somewhat of an outlier. These examples are still highly useful to illustrate the merit of the authors proposed risk mitigation strategy. To alleviate this concern, in Section 4, the authors may consider quantifying how likely it is to encounter flights like Flight A and Flight B. Further, the authors may consider switching the order of Sections 4 and 5, which would help provide more context to the reader for how typical Flights A and B are.

We thank the Reviewer for sharing this concern. Indeed, flights A and B are not randomly picked flights, but were chosen to nicely illustrate the objective of the risk-aware strategies. Following their recommendation, we added at the end of Section 4 a short analysis of the occurrence of flights A and B in the subset of flights: *However, flights A and B are not representative of the entire flight subset. For the 137 flights for which the nominal benefit is higher than 100 tCO_2e, the average estimated risk is 5%, with 71 reroutings for which the estimated risk is 0% and 9 for which the estimated risk is higher than 30%.*

However, we chose to remain as pedagogical as possible, going from the example to the general case, so we decided not to switch the order of Sections 4 and 5.