AC: We thank the reviewer for their time and thorough comments, which we think have lead to a significant improvement of the manuscript. We have revised the manuscript following both reviewers' suggestions, which were generally in agreement with each other. Below we respond to the individual reviewer comments and suggestions (RC) with either author comments (AC), manuscript changes (MC), or both, as appropriate.

Overall evaluation:

In this manuscript, the authors present a novel algorithm capable of producing synthetic contrail observations that can be used as benchmark for evaluating contrail attribution algorithms, as well as an enhanced contrail-to-flight attribution algorithm which is highly-scalable and shows significant improvement over previous ones when evaluated using the aforementioned benchmark. The topic is interesting and of good novelty. Meanwhile, the algorithms that are developed are of good scientific and technical value. The only major shortcoming of this manuscript as a journal publication is the structuring of contents, with so many appendices separating some critical information from the main contents that add to the readers' difficulty in understanding the already novel algorithms. Therefore, I would recommend a minor revision from the authors to make this manuscript better reader-friendly to the general peers in and outside the field before it gets published.

General comments:

RC1. The synthetic contrail dataset generated as SynthOpenContrails, as mentioned by the authors, can be used for benchmarking contrail attribution algorithms. This is a more general concept comparing to an enhanced attribution algorithm, as the latter should also be evaluated by the former to demonstrate its superiority. Therefore, I would suggest the authors put the relevant contents describing SynthOpenContrails in front of those on CoAtSaC. This would also correspond to the order of the two algorithms' appearances in the current title and abstract, which is more logically reasonable.

AC1. Thank you, we have adopted this suggestion and we believe that this has substantially improved the flow of the paper.

MC1. Swapped the order of the synthetic dataset section and the attribution algorithm section, and then divided up the subsections of the "Tuning and Benchmarking" section into the other sections, as appropriate.

RC2. There are too many appendices in the manuscript, making the article look like more of a script of program codes with multiple subroutines responsible for different functions, rather than a journal article in the field of geoscience. While it is somewhat common to have appendices attached as 'extensions' from specific contents in the main text, they should not include critical information so that it would not affect the understanding of paper even when removed from manuscript. For the current manuscript, however, some appendices can definitely be merged into the main text such as A1~A3, and A5, while some are beyond the analysis (A22 and part of

A27) that should be removed and potentially restructured as a new paper if the authors would like to. Therefore, it is strongly recommended that the authors consider restructure the appendices and merge those vital information into the manuscript as subsections.

AC2. We have integrated appendices A1, A5, A19, A20, A21, A24, A26, and part of A27 directly into the manuscript body. We feel that A2 and A3 should remain in the appendix, as including them in the manuscript body might cause the reader to have a hard time understanding what we did because there is too much interspersed discussion of why we did it.

We have dropped appendix A22 entirely, as we agree that it is perhaps better suited to a followup paper. We replaced appendix A23 with a short summary of it in the main text.

The remaining appendices have been reorganized into logical sections.

MC2. We integrated appendices A1, A5, A19, A20, A21, A24, A26, and part of A27 directly into the manuscript body. We have removed Appendix A22 and its references. We replaced appendix A23 with a short summary of it in the main text. Reorganized remaining appendices into logical sections.

Specific comments:

RC3. For subsections 2.2.1 to 2.2.3, the contents within different subsections are kind of mixed up and duplicate, not fully corresponding to the titles of each subsection. Take 2.2.1 as an example, the details of training and validating procedure, such as data splitting, shouldn't appear in the 'Data' subsection, but in the subsection 2.4 or an independent subsection. Also discussions on the error sources of the flight advection simulation algorithm is not related to 'data' but an independent subsection.

MC3. Restructured much of this section. Flight advection is now its own subsection, outside of Data. The dataset splits are now discussed in the "Tuning the Synthetic Dataset Parameters" section.

RC4. I would recommended that the authors merge Appendix A5 into subsection 2.2.3 and incorporate it into the main contents, as both parts are associated with enhancements over the single-frame attribution algorithm by Geraedts et al. (2024).

MC4. Moved Appendix A5 into the main text.

RC5. Fig. 5(b), could you provide more details on how did you separate the mixed attributions into different groups? From the figure there are a lot of points around W=0 line, which are difficult to separate from others from Fig. 5(a) but are ultimately divided into different groups.

AC5. Thank you for pointing out that this is confusing. Each attribution is associated with a range of advected waypoints that ostensibly formed the detected contrail. Flights can of course

form contrails multiple times along the flight path. If we imagine a scenario where a flight formed 2 contrails an hour apart from each other, and there was no advection error, the contrail age x W plot would put all attributions along the W=0 axis, and it would be impossible to separate them. What we therefore do is divide up the attributions into groups such that there are no common waypoints attributed to multiple groups. In the example given, this would now give us 2 groups, each with some subset of the attributions along the W=0 axis, and we can then trivially produce the correct fits. As this point was confusing to both reviewers, and it is essentially an implementation detail, we have removed the discussion of the separation into groups, and instead just stated that we start from sets of single-frame attributions with overlapping waypoints. We also then removed Fig 5(a) and the other groups from Fig 5(b).

MC5. Simplified the process of producing the groups of attributions for running the Fitting stage to just say “gathering all single-frame contrail attributions that are attributed to overlapping sets of waypoints for the same flight.” Removed Fig 5(a) and the subplots from Fig 5(b) that showed other groups. Updated the caption for Fig 5.

RC6. Equation (4), please clarify what impact does SSC, or the slope have in the fit score? Do you expect the slope as low as possible to gain more confidence in the fitting? What’s the relative impact of slope compared to the intercept, like a high SSCwith a low |b| against a low SSC with a high |b|.

AC6. This was a term that the tuning algorithm had the option to set to 0, but did not. Given that it's black-box tuning it's hard to know exactly why it did that. We can speculate that it might come into play in scenes with a large number of short-lived contrails (maybe similar to Fig 4(g)), where the supremacy of the number of inliers term during the fit generation phase might still produce a fit with a relatively large slope that joins detections of many different physical contrails. This term allows the S_fit for such fits to be high, and then ideally be rejected in the “Rejecting” phase.

MC6. Added this speculation to the text.

RC7. Line 338, is the threshold for Sfit constant for the algorithm, or is to be customized when scaled to different parts of the world and global usage?

AC7. The threshold value of 3 is not tuned and was chosen for consistency with (Geraedts et al. 2024) to make downstream analysis easier. The tuning and resulting performance would have been the same if we had tuned with any other threshold, but the resulting parameter values would change. Further research is required to determine whether the tunable parameter values will need to change in other regions of the world. The most likely regional difference that would cause re-tuning to be needed is if there is consistently higher flight and contrail density, which we are likely to find in parts of Europe. The performance fall-off of all 3 algorithms with increased contrail density, as shown in Fig 9(a) and elsewhere, is, to our knowledge, the first time that this density effect has been quantified, and we hope that this will motivate future research into attribution methods suitable for high density regions.

RC8. This is only a suggestion. For Algorithm 1, it's more common in my experience to use a flowchart rather than a pseudocode describing the work and logic flow. The current display resembles a technical report or a User's Guide, but not as reader-friendly as a journal article in the field of geophysics, especially to readers from different backgrounds.

AC8. We thank the reviewer for the suggestion. The algorithm pseudocode has been replaced by a flow chart. We further added a flow chart describing the CoAtSaC algorithm to help better structure that section of the manuscript.

MC8. Replaced algorithm with a flow chart. Added flow chart for the CoAtSaC algorithm.

RC9. Subsection 3.1.1, I would suggest the authors merge Appendix A27 into this subsection in the main content. As the comparison of validation results among the algorithms is pretty obvious, the contents can be summarized and shortened with the main properties and relevant comparisons.

MC9. Integrated parts of Appendix A27 into the main text.

RC10. For subsection 3.2, it's hard to illustrate if there is improvement given that the 'truth' labels are also generated from the algorithm, as is mentioned by the authors, rather than stand-alone observations. Also, this is not really an analysis or result of the SynthOpenContrails, but some application prospects. Therefore, I would suggest this part of information to be relocated to the discussion part in the section of conclusion rather than an independent subsection. If the authors insist on keeping it an independent subsection, some further elaboration and analysis should be given on the model trained by Sonabend et al. as well as the difference in forecasts from using different labels for training.

MC10. Removed this section and added a line about it in the Conclusions section.