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:

1. 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.

2. 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.

Specific comments:

1. 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.

2. 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).

3. 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.

4. Equation (4), please clarify what impact does S_{SC}, 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 S_{SC} with a low $|b|$ against a low S_{SC} with a high $|b|$.

5. Line 338, is the threshold for S_{fit} constant for the algorithm, or is to be customized when scaled to different parts of the world and global usage?

6. 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.

7. 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.

8. 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.