General response

We thank both reviewers for their detailed and constructive comments. Your insights have helped us improve the clarity and quality of our manuscript. Several comments address related issues, so we summarize the main changes here; details are given in the point-by-point replies.

Setting and scope of the algorithm: Our method is designed for MSG/SEVIRI over Europe, which is more demanding than the GOES/ABI setting of previous studies: contrails appear delayed after formation and are more difficult to detect at the coarser spatial resolution of MSG/SEVIRI (3 km nadir) compared to GOES/ABI (2 km nadir); the pixels grow with latitude (Europe is further North than US); no wind information is used; air traffic density is high; and surrounding air traffic is not considered.

These characteristics are explicitly considered in the design of our method, which is self-contained (satellite data and flight track for the flight under investigation only), exploits the contrail life cycle instead of using wind data, considers thick ice clouds along the flight path to quantify limited contrail visibility, and gives a confidence score for each surrounding contrail instead of a threshold decision.

Its purpose is to assess the confidence with which detected contrails in MSG/SEVIRI can be attributed to a given single specific flight as their originating instance and thus distinguish flights with and without contrails, which is the key information for avoidance trials. It is not meant to assign all contrails to their producing flights, and its results are not directly comparable with those of advection-based studies.

Main changes:

1. *Evaluation extended* (R1-C1, R2-C3): seven case studies instead of two, an analysis of the ERA5 wind along 38 flights, and a visual evaluation of 212 contrail attributions for these flights. All these aspects are addressed in the section now called “Evaluation”, where additional subsections have been added.

2. *Shell size* (R1-C4, R2-C9): the maximum advection speed assumed to define the contrail search areas, originally given at the sub-satellite point, is now additionally given for the spatial resolution over Europe, is evaluated with ERA5 winds, and its limits are acknowledged.

3. *Comparison with previous studies* (R1-C2, R1-C5, R1-C6, R2-C15): advantages and limitations of both approaches are now stated and claims about performance and scalability were toned down.

4. *Clarity* (R2-C4 to R2-C14, R2-C16 to R2-C19): sections, figures, acronyms and references were revised.

5. *Renaming:* the title is now "An automated flight-specific contrail attribution using Meteosat Second Generation satellite data". We no longer use "matching", which suggested an exclusive assignment of contrails to flights, and "matching score" is now "confidence score" (S).
6. *Abstract:* We have revised the abstract to better reflect the content of the paper and more accurately capture the goal of the algorithm.

Point-by-Point Replies

REVIEW #1

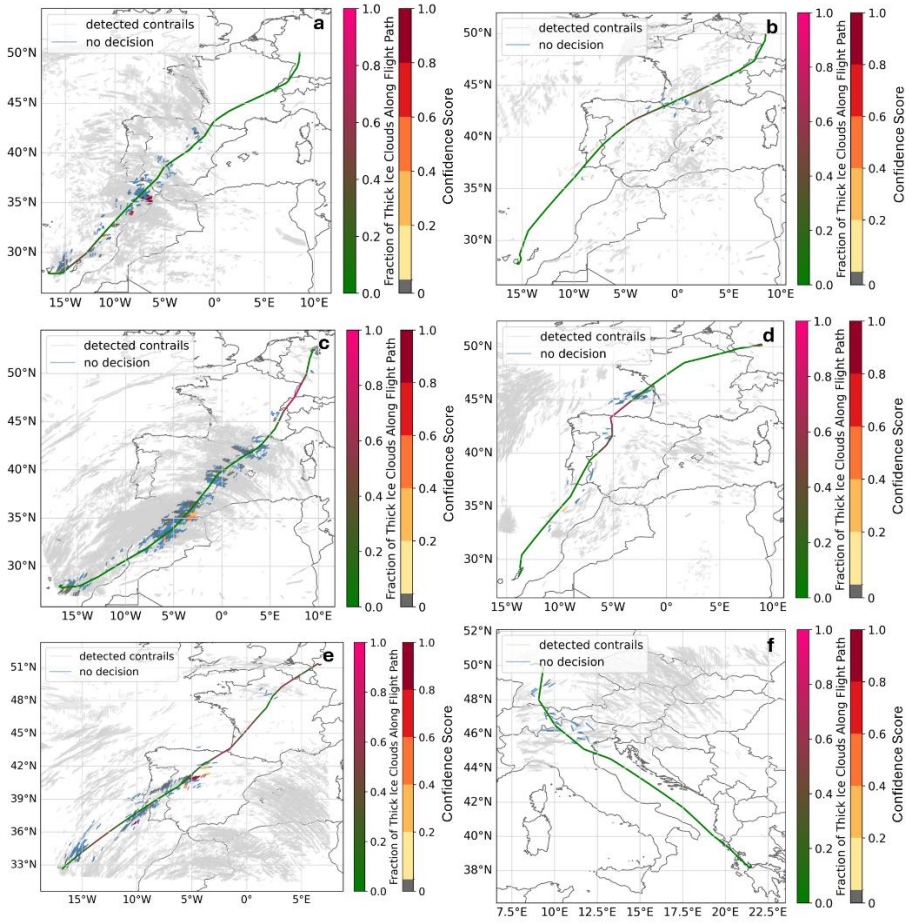
Comment 1: *I have one major criticism of this paper which is that they do comparably little analysis of how their algorithm actually performs. They make the statement "The method demonstrates high potential for automated, observation-based verification of contrail mitigation strategies in practice" but I don't think they've done enough validation to justify this statement. Producing validation as thorough as that in Sarna 2025 might not be feasible, but the authors should do better than just the 4 case studies presented in this work. Even increasing the number of case studies to 10 would be an improvement. Another approach, similar to Chevalier/Geraedts, would be to run their system on a few hundred flights and check if aggregate statistics of things like contrail age, diurnal/seasonal patterns, fraction of flights making contrails, are plausible. Ideally upon resubmission the authors should include some more efforts towards validating their methods. Failing that they should not stress much more strongly that they do not have a good estimate of the model's performance.*

Response: As stated in Sarna et al. (2025) all contrail-to-flight attribution algorithms suffer from missing validation because there is hardly no "ground truth" to compare with. Geraedts et al. (2024) have evaluated contrail detection in the context of their attribution algorithm, while Sarna et al. (2025) have constructed a synthetic dataset to test exactly such algorithms. Unfortunately, the dataset is based on 10-min GOES data and for the application of our algorithm a key aspect is the availability of 5-min satellite observations to apply the tracking algorithm ACTA. Thus, we have decided to focus on various aspects of our algorithm and we have extended our evaluation substantially. The evaluation contains now the following parts:

Evaluation	*Where*	*Sample*	*Purpose*	
Automated vs. manual detection	Sect. 4.1, Fig. 5, Appendix (Figs. D1–D3)	4 flight segments	Impact of detection uncertainty on contrail attribution results	
Selection step sensitivity (shell and orientation +10 %)	Sect. 4.3	4 flights	Effect of selection step parameters choice on contrail attribution results	
Score perturbations (+/- 0.2 in all combinations)	Sect. 4.3, Fig. 6	4 flights	Sensitivity/Robustness of the consistency score S to the criteria S1-S5	
Case studies	Sect. 3.6 (Fig. 4), Appendix (Fig. C1)	1 + 6 flights, with and without attributions	Behavior and illustration of the method	New
ERA5 wind along flight paths	Sect. 4.3	38 flights	Shell size adequacy	New
Visual verification of attributions	Sect. 4.4	212 contrail attributions, 38 flights	Share of correct attributions	New

In the following, we go into the details of the single aspects listed above:

- We added 5 case studies, so 7 are shown in total, including flights with and without contrail attributions. One can see that the method is able to reduce thousands of contrail detections to those ones that are spatially, temporally and geometrically consistent with the flight of interest. Furthermore, one can observe the response of the method in different meteorological situations with few or abundant surrounding contrails. The examples also show the ability of the method to assign confidence scores S>0 to few or various contrails, but also its ability to exclude all surrounding contrails as originating from the given flight. This is the basis for applying the method to evaluate contrail avoidance experiments.



“To complement the methods application, Fig. C1 presents six additional flight case studies. Similar to the results in Sect. 3.6, the algorithm filters out thousands of contrail detections and isolates only those that meet to some degree the attribution criteria. Panels on the left show examples with successfully attributed contrails (S > 0). The three examples on the left show a medium-to-strong occurrence of contrails (Fig. C1a and c) and a low occurrence of contrails (Fig. C1e) along the flight track. In all cases, contrail attributions could be achieved. For these flights, the typical West-East flow induces attribution only East of the flight track. As usually expected, no mixed attributions (East and West) are found. Furthermore, confidence scores in all bins are produced by the algorithm, from dark red to light yellow (in Fig. C1c). The frequent occurrence of blue contrails ('no decision') is largely related to the identification of false alarms produced by COCOS through the attribution algorithm. In flight sections with a high fraction of thick ice clouds (pink flight path), few contrails are detected. This does not prove their absence, as contrails may be hidden below or within the ice clouds (Sect. 3.2). Panels on the right display flights where contrail detections do not sufficiently fulfil the attribution criteria (S = 0) or no decision could be made.”

- In addition, ERA5 winds are shown in Fig. 4 to show that the contrail attributions are consistent with the model wind. Additionally, we analyzed the ERA5 wind along 38 flight paths to test the shell size:

  > "To assess the choice of the shell size, we analysed the ERA5 wind at 250 hPa along the flight paths of 38 flights. The flights exhibit highly similar operational patterns: most of them operated between Germany and the Canary Islands or between Germany and Greece, and took place from the afternoon into the night. The wind speed is below 67 km/h at 48.0 % of the waypoints and below 110 km/h at 79.4 %. The shell is thus large enough for about half of the waypoints in the east--west direction and for about 80 % in diagonal direction. As v_max is a free parameter, it could be increased for strong-wind situations in future work."

- Most importantly, we performed a visual verification of 212 contrail attributions found in 38 flights:

  > "Inspired by the manual evaluation of Geraedts et al. (2024), who let human labelers assess whether automatically attributed flight segments corresponded to an observed contrail, we perform a visual verification of all 212 contrails ($S > 0$) obtained for the 38 flights mentioned above. These attributions are not evenly distributed across flights: 23 flights have at least one high-confidence match ($S>0.5$) while the remaining 15 flights only have matches with scores $S \leq 0.5$. The verification can only remove attributions; no new attributions are added. It therefore yields a measure of precision but not of recall. This evaluation focuses on two primary criteria: the identification of the detection as a true contrail (vs. a false alarm, although many of them have already been excluded by the method itself at a previous stage) and the consistency of its motion relative to the flight trajectory and surrounding cloud movements. Evaluation of further criteria, such as the contrail formation time, would be subjective. In addition, the verification cannot evaluate the confidence scores themselves but only a yes/no decision per attribution. The manual verification revealed that 68.4 % of the attributions were correctly assigned. False alarms accounted for 16.0 %, while 15.6 % exhibited motion inconsistent with the trajectory (inconsistent motion directions were already identified as a source of uncertainty in the automated tracking in Sect. 4.1). Notably, the false alarm rate (16.0 %) is substantially lower than the general false alarm rate of the underlying automated detection algorithm (85.5 %). This confirms the findings already observed above, with the larger dataset indicating that the attribution process acts as an efficient false alarm filter. For orientation, in the benchmark of Sarna et al. (2025) the contrail precision of the algorithms of Geraedts et al. (2024), Chevallier et al. (2023) and Sarna et al. (2025) lies between about 40 and 73 %, which they caution to interpret as relative metrics, as it was obtained for synthetic data. The share of 68.4 % of confirmed attributions found here is

within this range. However, these numbers are not directly comparable to ours because of different sensors, reference definitions and because our verification only covers attributed contrails and therefore yields no recall.

A comparison of the confidence score distributions with and without manual verification shows similar patterns. The effect of the manual verification is a slight reduction in the total number of attributions. The influence on this subset of flights is moderate: the fraction of flights with at least one high-confidence attribution ($S > 0.5$) decreases from 61 % (23 flights) to 53 % (20 flights) after manual verification. This reinforces the confidence in the automated attribution results."

- In this paper we focus on single flights and their attributed contrails. Aggregate statistics such as the fraction of flights forming contrails for the avoidance trial will be the subject of future research (as noted in Sects. 1 and 5).
- Caution about performance: We replaced the statement criticized by the reviewer and now state the limits explicitly (Sect. 5):

  "These results are limited by the small number of cases and by the subjectivity of the visual verification. They therefore do not provide a full estimate of precision and recall. A synthetic benchmark adapted to MSG/SEVIRI or ground-based observations of contrail formation combined with high-resolution satellite observations to bridge the visibility gap between formation and appearance in MSG/SEVIRI could provide such an estimate in the future."

Comment 2: *From what little validation the authors do, their method doesn't appear to do very well. When compared to manual labels, their method only agrees on one out of four contrails, implying a precision and recall of 0.25. This is far less than what is achieved by Chevalier 2023, Geraedts 2024 or Sarna 2025 (as evaluated in Sarna 2025) [it's a small sample size and not an apples-to-apples comparison, but still I think this is a bad sign]. This does not mean that the work does not merit publication. Even if the performance does not match the best algorithms already in the literature, sharing how different design decisions affect performance is useful to inform future algorithm development. Furthermore the poor performance could in part be because of the poor resolution of SEVIRI. But I think the authors should at least call out this disparity, the statement ""The method demonstrates high potential for automated, observation-based verification of contrail mitigation strategies in practice" does not seem justified. It would be good if people reading this paper who might be interested in using the algorithm should at least be cautioned that it might not give the same results as some other methods in the literature. (For example, the authors plan to use this algorithm to validate a trial. If such a study gives a*

negative result, I would not necessarily say this implies the trial did not work, it might just be that the algorithm is not good enough to see the impact)

Response: Thank you for this important observation. To address this point, we have adjusted the text as follows:

- We compare the performance to existing literature and contextualize the results (Sect. 4.1.):
  "While this level of agreement is lower than the detection performance reported in high-resolution benchmark literature (e.g., 30 % false alarms in Geraedts et al. (2024)), these differences must be interpreted with caution: the evaluation is based on a small sample, and the lower resolution of SEVIRI, which decreases further towards higher latitudes, limits the sensitivity to thin or narrow contrails. Yet, those developments are important to investigate contrails above the European/African Airspace."
- In the test mentioned by the reviewer, the outcome of the attribution algorithm was tested using a visual contrail detection instead of automated contrail detection. This is of particular relevance for MSG/SEVIRI due to the challenges in contrail detection related to its moderate spatial resolution. This was not thought as an evaluation of the method as a whole. In that test, the automated and the manual detections led to the same number of attributions with S>0 (four each), and the attribution with the highest score coincided. The remaining differences are explained by contrails missed by the detection or by the subjectivity of manual labelling. Despite the detection challenges, the automated contrail detection provides meaningful / similar results in numbers (but not always with respect to the single contrail attributions).
- Next to the sensitivity of the attribution result to contrail detection, we performed also a larger verification of 212 attributions to 38 flights (Sect. 4.4.), where 68.4 % of them were confirmed. We rather regard this new evaluation as an evaluation of the method itself. We compared our results with previous studies:
  "For orientation, in the benchmark of Sarna et al. (2025) the contrail precision of the algorithms of Chevallier et al. (2023), Geraedts et al. (2024) and Sarna et al. (2025) lies between about 40 and 73 %, which they caution to interpret as relative metrics, as it was obtained for synthetic data. The share of 68.4 % of confirmed attributions found here is within this range. However, these numbers are not directly comparable to ours because of different sensors, reference definitions and because our verification only covers attributed contrails and therefore yields no recall."

Comment 3: *The authors claim that their method is scalable to large numbers of flights. This claim is somewhat undercut by the fact that the authors ran the algorithm on a total of 4 flights. (c.f. Geraedts 2024, which also claims a scalable algorithm and runs on >250 000 flights). The*

authors should prove that the algorithm is scalable by running on at least 100 flights. If this is too big a burden, then the algorithm is not actually scalable.

Response: Thank you for raising this point regarding the scalability claim. We acknowledge that testing the algorithm on a small set of case studies in this work does not constitute a proof of large-scale performance. We have addressed this comment by refining our statements and adding key context:

Sect. 3.6: We softened the text to clarify that further application across larger flight datasets remains an essential topic for future research. We noted that given the computational runtime required for our processing pipeline, scaling to orders of magnitude around 100-500 flights (rather than hundreds of thousands like *Geraedts et al., 2024*) is the realistic operational target for this approach:

> "The integrated attribution pipeline is demonstrated here on individual case studies. For flights taking place at similar times in the same region then COCOS application is due only once and this helps saving time since then only ACTA has to be applied. Scaling the attribution framework to larger datasets is realistically targeted at the order of hundreds of flights or requires further computational optimization, especially in the contrail detection with COCOS and contrail tracking with ACTA."

The sentence "supports efficient large-scale analysis" and "inherently scalable" was deleted (Sect. 5). Beyond the case studies, the pipeline has now been applied to 38 flights (Sect. 4), and the application to an avoidance trial is presented in the companion paper.

Comment 4: *The authors made a design decision to limit the maximum advection speed to 50 km/h. The authors claim this is a fine decision, but I don't agree. They point out that 50 km/h is higher than the mean, which may be true, but that's not the right metric. The authors should instead compute what fraction of the time the wind is higher than 50km/h. The authors later try a larger wind error. They find in total 26 matching contrails instead of 17. They claim this is a small change, I don't agree. I think ideally the authors would recompute all their main results with a larger wind threshold. Failing that they should at least change their claim that 50 km/h is sufficient. I would like to see increasing this threshold as a possible direction for future work.*

Response: We thank the reviewer for this point that was not clearly explained and changed the following:

- We clarify in Sect. 3.2 that 50 km/h refers to the 3 km pixel size at the sub-satellite point. Because the shell is defined in pixels (1.39 pixels per 5 min), it corresponds to larger speeds over Europe:

"As the shell is defined in pixels, its extent in km depends on the local pixel size. The current implementation of contrail-flight attribution focuses on Europe and North Africa and the Eastern part of the North Atlantic. In this region, the pixels are about 4-6 km wide in north-south and 4 km in east-west direction (EUMETSAT, 2017). The same 1.39 pixels per 5 min therefore correspond to about 67-100 km/h in north-south and 67 km/h in east-west direction. Along the pixel diagonal, which is longer than both pixel sides (about 5-7 km), the shell reaches up to about 85-110 km/h. The effective maximum advection speed assumed in the algorithm thus depends on direction and location of the investigated flights and ranges from about 67 to 110 km/h over Europe."

- We computed the fraction of waypoints with wind speeds smaller than the assumed maximum wind speed used for the shell size definition, using the ERA5 wind at 250 hPa along the paths of 38 flights (Sect. 4.3). This showed that most of the possible contrail attributions are covered by this shell size. Finally, we noted that the size could be easily adjusted for strong-wind situations (see our reply to Comment 1 of Review #1).
- The sensitivity test (shell and orientation tolerance +10 %, Sect. 4.3) shows that the number of attributions increases mainly in scenes with high contrail density. We no longer claim that the increase from 17 to 26 is negligible per se, but note that the relevant information for avoidance trials is whether a flight forms contrails at all. That classification changed for one of four flights.

  "However, for the intended application, the relevant information is whether a flight formed contrails or not, and not the exact number of attributed contrails. In this respect, the increase from 17 to 26 attributions in the two dense cases is considerable, but both numbers are small compared to the order of 1000 detected contrails in these scenes, and both flights remain classified as flights with contrail attributions. Only for the flight with an increase from 0 to 1 does the classification change."

Comment 5: *Another design decision the authors make is to not advect the flights. They claim this is a strength of the paper, I think instead it is a weakness that likely contributes to the relatively poor performance of their method compared to existing literature. The authors correctly point out that wind data is uncertain. But having no data at all about the wind is more uncertain! For example the authors consider all contrails within 50 km/h of the flight (which I claim is not big enough). The wind uncertainty is only 20km/h, so if they used the advected flight instead they would need to consider a lot less flights. They have a similar issue with angle. In the introduction they authors claim that sedimentation + wind shear is a reason not to use advection, but then in their algorithm they assume a constant advection speed, which would not be a reasonable assumption of sedimentation + wind shear was really a problem. I think the authors' claims that not advecting is a virtue need to be toned down, and I think they should say that adding advection to their algorithm is a possible direction for future work.*

Response: Our approach is different from previous advection-based methods and we have toned down the claim that not using wind data is a strength. The Introduction now states advantages and limitations of both routes:

> "Both routes have advantages and limitations. Not using wind data avoids the errors of modelled winds and of the altitude assumed for advection, but it requires a wider initial search area and assumes a constant advection speed during the non-visible phase of the contrail. Advection-based methods narrow down the search area and can resolve the competition between flights, but they inherit the wind uncertainty, which grows with advection time (Sarna et al. (2025)). Our method does not decide which of several flights formed a contrail and is therefore complementary to methods that account for the full air traffic."

Our results are not directly comparable to those studies because of the different approach, the different satellite with stronger challenges related to the lower spatial resolution, and the different (but still related) scope. For avoidance trials, flights without contrail attributions are as important as flights with contrail attributions, and we can evaluate them without the surrounding air traffic. To do this, avoidance flights and a set of reference flights with predicted contrail formation are compared with respect to the number of contrail attributions and corresponding confidence scores.

A hybrid approach is proposed as future work (Sect. 5): "using numerical wind data to constrain the search space during the initial time gap before contrail visibility, followed by contrail tracking in satellite data".

Comment 6: *The authors method does not consider multiple flights that might match the same contrail, and the authors correctly highlight this as a drawback of their work as well as discuss why it is hard. A reader might be left with the impression that such consideration is therefore infeasible, but in fact Chevalier, Geraedts and Sarna all do it. The authors should inform the reader of this.*

Response: We never wanted to insinuate that multiple flights consideration is unfeasible in general. But this is not the goal of the method presented here. We have informed the reader accordingly. Sect. 4.2 now states:

> "Frameworks for taking into account the surrounding air traffic have been demonstrated in recent advection-based studies (Chevallier et al., 2023; Geraedts et al., 2024; Sarna et al., 2025). Beyond the computational overhead, incorporating general air traffic into the current algorithm requires access to the respective flight trajectories as well as a strategy to define a relevant spatio-temporal search domain where the present algorithm can be applied to the concurrent flights in order to possibly attribute contrails with high

confidence score with respect to one aircraft to one of the surrounding flights. This is possible but remains a subject for future investigations."

The Introduction adds that advection-based methods "can resolve the competition between flights" and that our method is "complementary". The flight-specific design is deliberate: it makes the method independent of (complete or incomplete) air traffic databases and concentrates on the potential for contrail formation of single flights.

Comment 7: *In some of the case study figures I find the label 'no decision' confusing. I think the authors mean that these are flights that were within the max velocity of the flight but got a score of 0. Please be more explicit about these.*

Response: Thank you for pointing this out. To avoid any confusion, we have clarified the definition of the "no decision" status in both the text and the relevant figure captions.

Sect. 3.3, Figure 4 and C1: We explicitly clarified that "no decision" does not mean a score of 0, but applies to contrail clusters that do not meet the physical minimum requirements for a meaningful advection assessment (i.e., fewer than 2 contrails or a spatial extent of ≤ 3 pixels such that advection properties cannot be defined).

> "All contrail clusters that do not satisfy these requirements are assigned the status "no decision" because the temporal evolution of the contrails during their (early) life cycle cannot be assessed. Consequently, this status does not represent a confidence score of zero for a flight, but rather marks candidate clusters that are structurally insufficient for further attribution evaluation. This situation occurs predominantly for the large number of false alarms and is examined in detail in Sect. 4."

REVIEW #2

Comment 1: *The authors present a fully automated, completely observation-based, multi-step framework that matches contrails to individual flights. The precise attribution of contrails to individual flights is relevant because it allows evaluation of climate-optimized flight trajectories. Such re-routed flights aim to avoid regions prone to the formation of (persistent) contrails, one strategy - among others - to achieve climate-neutral aviation by 2050. The framework uses rapid scans (5-minute temporal resolution) from MSG-SEVIRI which cover a latitude range from about 15° to 70°N at 3 km by 3 km spatial resolution at nadir. The fully automated framework is new and in comparison to existing algorithms it does not rely on external wind data to determine contrail advection. The authors aim to prevent the two primary sources of uncertainty (i) inherent inaccuracies in modeled wind fields, and (ii) the altitude-dependent advection errors caused by contrail settling. Instead, the assignment of contrails to flights is based on geometric*

criteria as well as several metrics that are combined into an overall value for the assignment.

The framework is built on different steps (i) collocation of flight segments to rapid scans, (ii) contrail detection and filtering, (iii) contrail-to-flight matching, (iv) contrail tracking, and (v) calculation of a matching score based on multiple criteria.

In MSG-SEVIRI rapid scans, contrails are detected using COCOS. The algorithm is briefly introduced, sufficient detail about its underlying machine learning model and training is missing. It may also be of interest to the reader how the COCOS algorithm performs in comparison to existing literature, such as Meijer et al. (2022) or Ortiz et al. (2024, 2025). It is also known that brightness-temperature differences are sensitive to rivers such as the Nile, leading to false detections; therefore, it would be interesting to learn how COCOS handles this issue.

Response: We have expanded Sect. 2.3 with the following description:

> "It is a classical image processing algorithm rather than a machine learning model: candidates are extracted from an input image using the infrared channels at 8.7 µm, 10.8 µm and 12.0 µm by adaptive thresholding, separated into line-shaped objects using a probabilistic Hough line detection, and scored by pixel-wise, morphological and object-wise BTD checks. The confidence functions behind these scores are derived from data, using human-labelled MSG scenes. The assigned confidence scores indicate detection certainty to each contrail."

We also added the treatment of linear surface features and the comparison with other algorithms:

> "Compared to the other contrail detection algorithms available for MSG (Mannstein et al., 1999, 2010), COCOS reaches more than three times higher precision at similar recall. Linear surface features such as coastlines and rivers, which can also show high BTD values, are suppressed by the gradient, morphological and object-wise BTD checks of COCOS. Detection algorithms for other sensors, domains and evaluation settings exist t (Meijer et al., 2022; Gryspeerdt et al., 2024; Ng et al., 2024; Ortiz et al., 2024; Ortiz et al., 2025); a quantitative comparison requires a common reference dataset and is planned for future work."

Comment 2: *Detected contrails need to be selected as 'potential matches with the flight'. To account for horizontal advection, the authors assume a maximum wind speed at flight altitude (50 km/h) that defines an oval-shaped region with a radius that is defined as the product of the assumed wind speed and the elapsed time between flight passage and contrail detection.*

Contrail tracking, then, relies on an existing algorithm, ACTA, which was translated from IDL to Python and for improvement slightly modified.

Potential matches are analyzed using clusters and the contrail life cycle is evaluated based on five matching criteria. A cluster is a group of contrails in consecutive rapid scans that correspond to the same physical contrail (based on position, orientation, and length).

The combination of the five matching-criteria scores gives a matching score that assesses the consistency of the cluster with the flight segment.

Finally, the study presents 4 case studies that illustrate the framework's performance.

Advanced cloud-tracking algorithms already exist and those reliably track clouds in geostationary satellite data, cf. TOBAC (Heikenfeld et al., 2019) or LIMOTTo (Seelig et al., 2025). These object-based tracking algorithms apply cross-correlation to calculate a displacement map - due to horizontal wind and/or advection - which is then used to forecast the position of cloud objects in the next time step.

The development of a method using a different Ansatz in detecting clouds (in the present case contrails) in MSG-SEVIRI data and tracking them in time, and additionally assigning a matching score for contrail-to-flight attribution is a valuable contribution.

Response: Thank you for pointing out tobac and LIMOTTo. We now state in Sect. 2.5:

> "Cloud tracking algorithms such as tobac (Heikenfeld et al., 2019) or LIMOTTo (Müller et al., 2025) already exist and reliably track cloud objects in geostationary satellite data. We do not use these algorithms here, because contrails are thin, elongated objects whose tracking benefits from contrail-specific properties such as their linearity. For this reason, the Automated Contrail Tracking Algorithm (ACTA, Vazquez-Navarro et al., 2010), which was designed specifically for contrails and exploits these properties, is used in this study. "

In addition, ACTA is designed to start with a contrail (detected in MODIS data in the original paper) and subsequently look for a similar contrail structure in the previous and subsequent satellite image, without any additional contrail detection information. ACTA is also designed to weaken linearity criteria as a function of time and thus allow to detect older contrails whose shape is no longer strictly linear as directly after formation. This is particularly important for MSG/SEVIRI since due to its moderate spatial resolution the contrails appear in the satellite observation when they are already 20-120 min old.

Comment 3: *However, because there is no 'ground-truth' contrail-to-flight attribution available, the reader would have at least expected a set of a few hundred or thousand cases, especially*

while the authors state that 'through batch processing multiple flights can be evaluated simultaneously'. This would allow for a thorough and robust verification of their framework. Furthermore, aggregate statistics can be calculated - i.e., contrail age, formation time, visibility, young/old contrail coverage etc - that can be compared to existing literature. The framework could also be tested with the synthetic model-based reference dataset for GOES (Sarna et al., 2025).

Response (Same as in Review #1 Comment 1): As stated in Sarna et al. (2025) all contrail-to-flight attribution algorithms suffer from missing validation because there is hardly no "ground truth" to compare with. Geraedts et al. (2024) have evaluated contrail detection in the context of their attribution algorithm, while Sarna et al. (2025) have constructed a synthetic dataset to test exactly such algorithms. Unfortunately, the dataset is based on 10-min GOES data and for the application of our algorithm a key aspect is the availability of 5-min satellite observations to apply the tracking algorithm ACTA. Thus, we have decided to focus on various aspects of our algorithm and we have extended our evaluation substantially. The evaluation contains now the following parts:

Evaluation	*Where*	*Sample*	*Purpose*	
Automated vs. manual detection	Sect. 4.1, Fig. 5, Appendix (Figs. D1–D3)	4 flight segments	Impact of detection uncertainty on contrail attribution results	
Selection step sensitivity (shell and orientation +10 %)	Sect. 4.3	4 flights	Effect of selection step parameters choice on contrail attribution results	
Score perturbations (+/- 0.2 in all combinations)	Sect. 4.3, Fig. 6	4 flights	Sensitivity/Robustness of the consistency score S to the criteria S1-S5	
Case studies	Sect. 3.6 (Fig. 4), Appendix (Fig. C1)	1 + 6 flights, with and without attributions	Behavior and illustration of the method	New
ERA5 wind along flight paths	Sect. 4.3	38 flights	Shell size adequacy	New
Visual verification of attributions	Sect. 4.4	212 contrail attributions, 38 flights	Share of correct attributions	New

In the following, we go into the details of the single aspects listed above:

- We added 5 case studies, so 7 are shown in total, including flights with and without contrail attributions. One can see that the method is able to reduce thousands of contrail detections to those ones that are spatially, temporally and geometrically consistent with the flight of interest. Furthermore, one can observe the response of the method in different meteorological situations with few or abundant surrounding contrails. The examples also show the ability of the method to assign confidence scores S>0 to few or various contrails, but also its ability to exclude all surrounding contrails as originating from the given flight. This is the basis for applying the method to evaluate contrail avoidance experiments.

  > "To complement the methods application, Fig. C1 presents six additional flight case studies. Similar to the results in Sect. 3.6, the algorithm filters out thousands of contrail detections and isolates only those that meet to some degree the attribution criteria. Panels on the left show examples with successfully attributed contrails (S > 0). The three examples on the left show a medium-to-strong occurrence of contrails (Fig. C1a and c) and a low occurrence of contrails (Fig. C1e) along the flight track. In all cases, contrail attributions could be achieved. For these flights, the typical West-East flow induces attribution only East of the flight track. As usually expected, no mixed attributions (East and West) are found. Furthermore, confidence scores in all bins are produced by the algorithm, from dark red to light yellow (in Fig. C1c). The frequent occurrence of blue contrails ('no decision') is largely related to the identification of false alarms produced by COCOS through the attribution algorithm. In flight sections with a high fraction of thick ice clouds (pink flight path), few contrails are detected. This does not prove their absence, as contrails may be hidden below or within the ice clouds (Sect. 3.2). Panels on the right display flights where contrail detections do not sufficiently fulfil the attribution criteria (S = 0) or no decision could be made."

- In addition, ERA5 winds are shown in Fig. 4 to show that the contrail attributions are consistent with the model wind. Additionally, we analyzed the ERA5 wind along 38 flight paths to test the shell size:

  > "To assess the choice of the shell size, we analysed the ERA5 wind at 250 hPa along the flight paths of 38 flights. The flights exhibit highly similar operational patterns: most of them operated between Germany and the Canary Islands or between Germany and Greece, and took place from the afternoon into the night. The wind speed is below 67 km/h at 48.0 % of the waypoints and below 110 km/h at 79.4 %. The shell is thus large enough for about half of the waypoints in the east--west direction and for about 80 % in diagonal direction. As v_max is a free parameter, it could be increased for strong-wind situations in future work."

- Most importantly, we performed a visual verification of 212 contrail attributions found in 38 flights:

  “Inspired by the manual evaluation of Geraedts et al. (2024), who let human labelers assess whether automatically attributed flight segments corresponded to an observed contrail, we perform a visual verification of all 212 contrails ($S > 0$) obtained for the 38 flights mentioned above. These attributions are not evenly distributed across flights: 23 flights have at least one high-confidence match ($S>0.5$) while the remaining 15 flights only have matches with scores $S \leq 0.5$. The verification can only remove attributions; no new attributions are added. It therefore yields a measure of precision but not of recall. This evaluation focuses on two primary criteria: the identification of the detection as a true contrail (vs. a false alarm, although many of them have already been excluded by the method itself at a previous stage) and the consistency of its motion relative to the flight trajectory and surrounding cloud movements. Evaluation of further criteria, such as the contrail formation time, would be subjective. In addition, the verification cannot evaluate the confidence scores themselves but only a yes/no decision per attribution. The manual verification revealed that 68.4 % of the attributions were correctly assigned. False alarms accounted for 16.0 %, while 15.6 % exhibited motion inconsistent with the trajectory (inconsistent motion directions were already identified as a source of uncertainty in the automated tracking in Sect. 4.1). Notably, the false alarm rate (16.0 %) is substantially lower than the general false alarm rate of the underlying automated detection algorithm (85.5 %). This confirms the findings already observed above, with the larger dataset indicating that the attribution process acts as an efficient false alarm filter. For orientation, in the benchmark of Sarna et al. (2025) the contrail precision of the algorithms of Geraedts et al. (2024), Chevallier et al. (2023) and Sarna et al. (2025) lies between about 40 and 73 %, which they caution to interpret as relative metrics, as it was obtained for synthetic data. The share of 68.4 % of confirmed attributions found here is within this range. However, these numbers are not directly comparable to ours because of different sensors, reference definitions and because our verification only covers attributed contrails and therefore yields no recall.

  A comparison of the confidence score distributions with and without manual verification shows similar patterns. The effect of the manual verification is a slight reduction in the total number of attributions. The influence on this subset of flights is moderate: the fraction of flights with at least one high-confidence attribution ($S > 0.5$) decreases from 61 % (23 flights) to 53 % (20 flights) after manual verification. This reinforces the confidence in the automated attribution results.”

- In this paper we focus on single flights and their attributed contrails. Aggregate statistics such as the fraction of flights forming contrails for the avoidance trial will be the subject of future research (as noted in Sects. 1 and 5).
- Caution about performance: We replaced the statement criticized by the reviewer and now state the limits explicitly (Sect. 5):

  > "These results are limited by the small number of cases and by the subjectivity of the visual verification. They therefore do not provide a full estimate of precision and recall. A synthetic benchmark adapted to MSG/SEVIRI or ground-based observations of contrail formation combined with high-resolution satellite observations to bridge the visibility gap between formation and appearance in MSG/SEVIRI could provide such an estimate in the future."

Comment 4: *The manuscript provides sufficient detail in the methods section. However, sections could be improved.*

For example Section 2.1, the authors introduce all the MSG satellites, which is not necessary, only name the one(s) used for the study. Precisely describe the region and the period used for the study. Next, the authors introduce which channels and combinations are relevant, and that the Ash RGB composite "is also used to highlight contrails". However, information is missing regarding exactly which channels and combinations of brightness-temperature differences are used. The data level (1, 1.5, ...) is also important to mention. The discussion about how complicated it is to observe contrails in geostationary satellite data could be dispensed with. Lines 111 - 118 belong to Section 4.

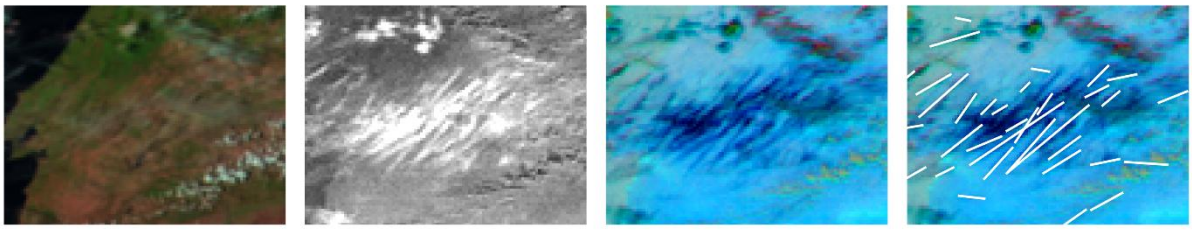
Response: We have revised Sect. 2.1 as follows:

- Only the satellite used is named (MSG4 at 9.5°E).
- Region, period, data level and channels are specified:

  > "We use these Rapid Scan data for the European and North Atlantic region (30° W to 35° E, 15° N to 55° N) and for the period of the analysed flights in May to August 2024 as High Rate Information Transmission (HRIT) Level 1.5 data, i.e. geolocated and radiometrically calibrated images.
  >
  > For contrail remote sensing, we use the infrared channels at 8.7 µm, 10.8 µm and 12.0 µm and the brightness temperature differences (BTDs) 10.8-12.0 µm and 8.7-10.8 µm."

- The discussion of the difficulty of contrail observation was shortened to one sentence, and the content of the previous lines 111–118 was moved to the introduction of Sect. 4.

Comment 5: *The reader might be interested in why Figure 1 shows 3 different channel combinations and which connection the subfigures have to the methodology.*



Response: We thank the reviewer for this comment. Figure 1 is intended to give the reader an impression of what contrails look like in MSG/SEVIRI imagery and how they appear in different channel combinations. We have clarified this in Sect. 2.1 and we now state how the panels relate to the methodology. The Ash RGB is the basis for contrail detection with COCOS. The 10.8-12.0 µm BTD image is used for contrail tracking with ACTA.

Comment 6: *Section 2.2 is a discussion about the contrail-appearance period in geostationary satellite data and could be shortened and moved to Section 3.*

Response: We thank the reviewer for this suggestion. We now state at the beginning of Sect. 2.2 that it provides the basis for the assumed contrail visibility window used in the attribution (Sect. 3). We have kept the section in the methods of the manuscript because the assumption of 20 to 120 min between contrail formation and first visibility is a prerequisite for the attribution procedure and is derived from the characteristics of MSG/SEVIRI and the literature on contrail spreading. The topic of visibility fits of course the method design, but we think it fits best in Sect. 2 where we describe contrail appearance in MSG (Sect. 2.1), contrail first visibility (2.2), contrail detection (2.3), contrail observability within cirrus clouds (2.4) and contrail tracking (2.5). Furthermore, due to its length, moving this paragraph into Sect. 3 would make it more difficult to follow the description of the attribution algorithm.

Comment 7: *Section 2.4 is a mixture of introduction, discussion, and methodology. The authors might sort out the parts into the corresponding sections.*

Response: We thank the reviewer for this observation. While Section 2.4 does indeed combine contextual background, methodological description, and brief interpretive elements, we have chosen to keep it as a consolidated section. Scattering these aspects across the Introduction, Methods, and Discussion sections would fragment the narrative and make it harder for the reader to follow the self-contained rationale of this specific step and of Section 2 in general (see answer above).

Comment 8: *Surprisingly, the description of ACTA functionality and its modifications is outsourced to Appendix A, although it is an important part of the methodology. Lines 473 - 474 belong to Section 2.1. The thresholds could be mentioned explicitly in line 499.*

Response: We thank the reviewer for this helpful feedback. To address your concerns and make the manuscript more self-contained:

- We have moved the sentences previously found in lines 473–474 directly into Section 2.1, as suggested.
- We have added a concise summary of ACTA's two-step core mechanism (position estimation via linear regression and shape determination using spatial, edge, and adaptive threshold masks) directly to Section 2.5.

  "ACTA starts from a detected contrail (originally identified using a polar orbiting satellite sensor like MODIS, Vazquez-Navarro et al., 2010; Vázquez-Navarro et al., 2015) and operates iteratively on MSG rapid scan observations in two main steps: first, it localizes the contrail position in the subsequent satellite image using linear regression; second, it determines the contrail shape by combining spatial search regions, edge detection via the Laplacian of Gaussian operator, and an adaptive threshold mask to capture low-contrast features effectively. "
- We have explicitly included the specific thresholds used for the adaptive thresholding (9×9 neighborhood and constant offset of 4).

However, we have kept the full, detailed technical explanation of ACTA in Appendix A to keep the manuscript length manageable and maintain a fluid reading flow in the main methodology section. Notably, to accommodate other structural adjustments (such as moving the detailed explanation of the derivation of contrail formation parameters from the appendix into the main text, as suggested during the revision process), keeping the extensive implementation details of ACTA in the appendix helps balance the overall length and focus of the paper.

Comment 9: *In Section 1, the authors describe how the application of wind-advection methods significantly degrades the attribution accuracy. Therefore, it is somewhat surprising to read in Section 3.2 that at cruising altitude a fixed wind speed of 50 km/h is assumed for horizontal advection. In dense air-traffic areas with contrail outbreaks and assuming a lower horizontal wind speed than 50 km/h this method would overestimate contrails near flight. And assuming higher wind speed at cruising altitude, the method misses contrails near flight. Can the authors comment on that?*

Response: Thank you for this comment. The shell is not an assumption about contrail motion but a candidate-screening step with an upper bound, and the actual advection speed is derived from contrail tracking (Sect. 3.3). The following points are now in the manuscript (Sect. 3.2):

> "This selection step serves as a broad candidate-screening procedure designed to capture all potential contrail-flight pairs despite wind variability. Subsequent filtering steps are then applied to refine these candidates and eliminate false positives."

Furthermore, we explain now the advantages as well as the limitations of this method compared to advection-based methods in the Introduction (Sect. 1).

> "Both routes have advantages and limitations. Not using wind data avoids the errors of modelled winds and of the altitude assumed for advection, but it requires a wider initial search area and assumes a constant advection speed during the non-visible phase of the contrail. Advection-based methods narrow down the search area and can resolve the competition between flights, but they inherit the wind uncertainty, which grows with advection time (Sarna et al., 2025). Our method does not decide which of several flights formed a contrail and is therefore complementary to methods that account for the full air traffic. "

Our results are not directly comparable to those studies because of the different approach, the different satellite with stronger challenges related to the lower spatial resolution, and the different (but still related) scope. For avoidance trials, flights without contrail attributions are as important as flights with contrail attributions, and we can evaluate them without the surrounding air traffic. To do this, avoidance flights and a set of reference flights with predicted contrail formation are compared with respect to the number of contrail attributions and corresponding confidence scores.

Furthermore, we now explain in Sect. 3.2 that the maximum advection speed of 50 km/h refers to the 3 km spatial resolution of MSG/SEVIRI at its sub-satellite point. Over the evaluated regions this corresponds to 67 – 110 km/h depending on the location and the wind direction. To test the sensitivity of the attribution results to the shell size, we performed two evaluations (Sect. 4.3). First, we increased the shell size by 10 %.

> "As expected, more contrails are initially identified, but after subsequent processing, the number of attributions with a score S>0.5 -- a threshold for ensuring a high degree of confidence in contrail attribution -- remained 0 for one flight and increased from 0 to 1 for another. For the two flights with very high contrail density, the number of S>0.5 attributions rose from 6 to 9 and from 11 to 16, respectively. These results indicate that the attribution outcomes are sensitive to selection parameters primarily under high contrail density, where the contrail attribution is most challenging. However, for the intended application, the relevant information is whether a flight formed contrails or not,

and not the exact number of attributed contrails. In this respect, the increase from 17 to 26 attributions in the two dense cases is considerable, but both numbers are small compared to the order of 1000 detected contrails in these scenes, and both flights remain classified as flights with contrail attributions. Only for the flight with an increase from 0 to 1 does the classification change."

Second, we analyzed the ERA5 wind in 250 hPa along 38 flights:

> "To assess the choice of the shell size, we analysed the ERA5 wind at 250 hPa along the flight paths of 38 flights. The flights exhibit highly similar operational patterns: most of them operated between Germany and the Canary Islands or between Germany and Greece, and took place from the afternoon into the night. The wind speed is below 67 km/h at 48.0 % of the waypoints and below 110 km/h at 79.4 %. The shell is thus large enough for about half of the waypoints in the east--west direction and for about 80 % in diagonal direction. As v_max is a free parameter, it could be increased for strong-wind situations in future work."

A hybrid approach is proposed as future work (Sect. 5): "using numerical wind data to constrain the search space during the initial time gap before contrail visibility, followed by contrail tracking in satellite data".

Comment 10: *Section 3.3 line 244-245: Clustering efficiently captures a part of the contrail life cycle, which is "complemented by ACTA... particularly during early formation stages when they are not yet ... detected." Can the authors comment on that?*

Response: We thank the reviewer for this question. Our empirical experience showed that ACTA has a higher sensitivity towards earlier and older contrails or those exhibiting partial non-linearity, which can be missed by the general COCOS detection algorithm. We now state in Sect. 3.3:

> "It is complemented by ACTA, which tracks less linear or weakly detectable contrails, particularly during early formation stages when they are not yet fully detected, as well as for older or partially non-linear contrails towards the end of their life cycle that can be missed by the detection with COCOS."

We also explain in Sect. 3.3 why this is important:

> "ACTA increases the number of contrails per cluster, providing a more complete representation of the contrail life cycle without changing the total number of clusters. By reconstructing a larger portion of the contrail life cycle, more precise information on location and time of formation, movement direction and speed is obtained. For example, the young contrails, which might be missed by clustering alone, may lie outside the

allowed time window, and clusters with few contrail detections yield less reliable motion estimates."

Furthermore, ACTA performs a tracking on single contrails and we consider it more reliable (but more CPU time consuming) than the clustering alone. Since our estimates of the initial part of the contrail life cycle are crucial in the attribution method, we want to rely on ACTA for this.

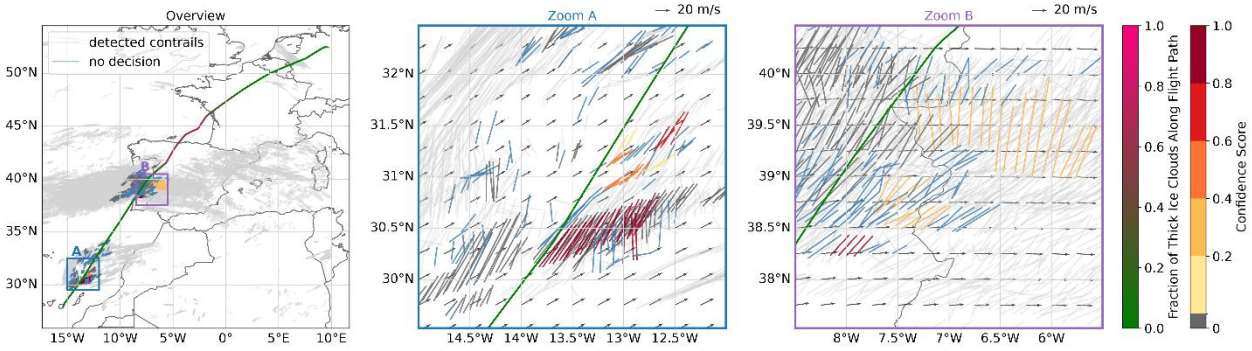
Comment 11: *Section 3.3 line 257: Surprisingly, the description of the contrail-formation parameters is outsourced to Appendix B. Lines 530 to 545 belong to Section 4.*

Response: We moved the description of the contrail formation parameters from the Appendix into Sect. 3.3 (flight position and time at contrail formation, advection velocity, advection-derived contrail formation time, movement direction). The uncertainty discussion that the reviewer refers to was moved into Sect. 4.3.

Comment 12: *Section 3.4 line 277: Start the sentence on a new line - "The Proximity Score ...". Section 3.4 line 287: Start the sentence on a new line - "The Advection Consistency Score ..."*

Response: Done. Each of the five scores now starts on a new line in Sect. 3.4. Thank you for the suggestion!

Comment 13: *Figure 4: It is understandable that an example with and without matching contrails is shown. However, the example with matching contrails makes the music and it is quite busy; why do the authors not neglect the example without contrail matches and, instead, show the two regions where the contrail matches appear as separate subfigures? It would also be interesting to see wind barbs at cruising altitude - perhaps ERA5 winds. This illustrates that ACTA chose the right direction for contrail advection.*



Response: We redesigned Fig. 4 as suggested. It now shows an overview panel and two zoomed panels (A and B) of the regions where contrails are attributed to the flight. The example without

attributions was moved to the Appendix (Fig. C1), which now contains six case studies. We kept example flights without contrail attributions, since for avoidance trials, flights without contrails are as important as flights with contrails, and we can evaluate them without the surrounding air traffic.

We added ERA5 winds at 250 hPa from the hourly field closest in time to the respective observation as arrows with a reference arrow (20 m/s) in Fig. 4. The wind direction is consistent with the advection direction found by the contrail tracking. The confidence score colour scale is now discrete (see Comment 14).

Comment 14: *Is Figure 5 based on the flight segment in Figure 4 (left)? If so, the authors should describe it. Insets A to E appear, but they are not mentioned in the main text or in the caption. Clustering - a colorbar would increase readability. Matching - the authors chose a continuous colorbar, however, a discrete colorbar (same color for bin 0 - 0.2, same color for bin 0.2 -0.4, and so on ...) would greatly increase visualization of the matching score.*

Response:

- Relation to Fig. 4: The caption now states that Fig. 5 shows "a segment of the flight shown in Fig. 4".
- Insets A–E: They are discussed in Sect. 4.1.
- Clustering colorbar: We did not add one because each cluster has its own colour only to separate clusters, and the colour has no quantitative meaning (this is stated in the caption: "each cluster in a distinct colour").
- Discrete colorbar: We implemented it in all figures that show the score (Figs. 4, 5, C1, D1-D3): five bins of width 0.2. Contrails with a score of 0 are shown in dark grey and have their own small bin in the colorbar.

Comment 15: *Section 5 - The authors state in lines 440 - 442 that their framework eliminates the source of meteorological uncertainty. However, they should fairly mention that their approach introduces uncertainty, i.e., due to (i) digital image processing, (ii) assuming a fixed maximum wind speed, and/or (iii) relying on geometrical parameters for connecting contrails within time steps. Line 448, lines 451 - 452, and line 461 ("... inherently scalable ...") - these statements still need to be demonstrated. Lines 91 - 92 belong to this section.*

Response: We have revised the Conclusions:

- Uncertainties: The statement that the method eliminates a source of uncertainty has been replaced. It now names the uncertainties of the approach.

  > “In addition, while methods relying on wind-driven flight advection suffer from uncertainties that compound over time due to cumulative model wind errors, our approach does not inherit the wind uncertainty; its uncertainties are related to challenges in contrail detection and tracking due to the moderate spatial resolution of MSG/SEVIRI and the assumption of a constant advection speed during the non-visible phase of the contrail.”

  Furthermore, a hybrid approach could be an idea for future work:

  > “In the future, combining this observation-based approach with wind-driven flight advection represents a highly promising possibility to increase attribution accuracy. In particular, a hybrid approach -- using numerical wind data to constrain the search space during the initial time gap before contrail visibility, followed by contrail tracking in satellite data -- could significantly reduce attribution uncertainties.”

Scalability: We deleted "inherently scalable" and describe the scalability honestly in Sect. 3.6: We softened the text to clarify that further application across larger flight datasets remains an essential topic for future research. We noted that given the computational runtime required for our processing pipeline, scaling to orders of magnitude around 100-500 flights (rather than hundreds of thousands like *Geraedts et al., 2024*) is the realistic operational target for this approach:

> “The integrated attribution pipeline is demonstrated here on individual case studies. For flights taking place at similar times in the same region then COCOS application is due only once and this helps saving time since then only ACTA has to be applied. Scaling the attribution framework to larger datasets is realistically targeted at the order of hundreds of flights or requires further computational optimization, especially in the contrail detection with COCOS and contrail tracking with ACTA.”

- Companion paper: The application is described in the last paragraph of Sect. 5:

  > “While this study focuses on the development and evaluation of the contrail-to-flight attribution methodology, the method will be applied to a contrail avoidance flight trial over Europe to quantitatively evaluate how pre-tactical flight trajectory optimizations can reduce contrail formation in real-world operations.”

Comment 16: *Check acronyms and abbreviations for first appearance/redundancy. For example,*

line 22: write out COVID; otherwise introduce COVID if it is used several times in the following text,
line 59: NWP not yet defined,
lines 80 - 81: MSG-SEVIRI already defined in the abstract,
line 99: SEVIRI already defined,
line 137: The acronym COCOS should be introduced here,
line 335: see line 22, etc pp.

Response: We thank the reviewer for carefully checking the acronyms and abbreviations. We have thoroughly reviewed the entire manuscript and ensured that all abbreviations and acronyms are handled in accordance with the journal's guidelines:

- They are defined upon their first appearance in the abstract.
- They are defined again at their first appearance in the main text.
- All subsequent mentions use the acronym consistently, and redundancies or undefined terms have been corrected.

Comment 17: *The authors have not yet made their code publicly available, which would allow interested users to reproduce the study and apply the toolbox.*

Response: We thank the reviewer for this suggestion. While the source code is not yet hosted in a public repository, we are very open to sharing our code for reproducibility and collaborative research. Interested researchers are warmly invited to contact the corresponding author directly to obtain access.

Comment 18: *Check titles for capitalization, i.e., lines 576 - 577, 582, 587, 589, 598, etc pp.*

Response: Thank you for pointing this out. We have checked all titles, including section and subsection headings, appendix headings, table and figure captions, and the titles in the reference list, and changed them to sentence-style capitalization. This includes the lines mentioned (576-577, 582, 587, 589, 598) and all other affected instances throughout the manuscript.

Comment 19: *References:*

Meijer, V. R. et al., (2022) "Contrail coverage over the United States before and during the COVID-19 pandemic" Environmental Research Letters, 17(3):034039, DOI 10.1088/1748-9326/ac26f0

Ortiz, I. et al., (2024) "Satellite-based quantification of contrail radiative forcing over Europe: A two-Week analysis of aviation-induced climate effects", Submitted to SESAR Innovation Days 2024, pre-print DOI: 10.48550/arXiv.2409.10166

Ortiz, I. et al., (2025) "Robust evaluation of neural networks trained on the OpenContrails dataset", IEEE Transactions on Geoscience and Remote Sensing, DOI: 10.1109/TGRS.2025.3629628

Heikenfeld, M. et al., (2019) "tobac 1.2: towards a flexible framework for tracking and analysis of clouds in diverse datasets", Geoscientific Model Development, 12 (11), 4551–4570, DOI: 10.5194/gmd-12-4551-2019

Seelig, T., Müller, F., and Tesche, M., (2025) "LIMOTTo (Leipzig Institute for Meteorology Object Tracking TOolbox)", Zenodo. DOI: 10.5281/ZENODO2347.15076642

Response: We have added the suggested references: Meijer et al. (2022) and Ortiz et al. (2024, 2025) in Sect. 2.3, and Heikenfeld et al. (2019) and the LIMOTTo description (Müller et al., 2025) in Sect. 2.5. They are in the reference list.