

Response to referees

Manuscript ID: <https://doi.org/10.5194/egusphere-2026-1171>

Title: *Observing formation and early evolution of contrails formed by IAGOS aircraft using high-resolution LEO satellite imagery*

We thank the reviewers for their thorough and constructive review. Each comment is addressed individually below, with references to the relevant changes made in the revised manuscript.

Shortly after the pre-print was posted for discussion, we noticed an error in Figure 7a, in which ERA5 RH_i data was plotted instead of IAGOS RH_i data. We have corrected this in the revised manuscript. As the corrected figure is qualitatively the same as the original, neither the discussion nor the conclusions are affected.

Response to Referee #1

General comments

G1. The authors develop and apply a new method to observe contrail formation and their early evolution with LEO satellites. This is a welcome step forward, since contrail formation and early stages cannot be observed with geostationary satellites due to their low spatial resolution. Almost two decades ago, Vazquez-Navarro et al. (2009) used LEOS to detect contrails that were then tracked in image series from a geostationary satellite. This might be mentioned in the paper.

Response: We thank the reviewer for pointing us to this work. It is not entirely clear whether this refers to the PhD dissertation from 2009, the journal paper (Vazquez-Navarro et al., 2010), or the 2015 paper added to the review (Vázquez-Navarro et al., 2015). All three works, however, report a methodology in which contrails are detected in MODIS/AVHRR imagery and subsequently tracked in MSG/SEVIRI images using the Automatic Contrail Tracking Algorithm (ACTA). We agree that this work is relevant and have added it to the introduction.

We have updated the introduction as follows: *LL 47–49: For example, the relatively coarse (1–2 km spatial resolution at nadir) geostationary satellite images can miss thinner contrails (Vazquez-Navarro et al., 2010; Driver et al., 2025; Euchenhofer et al., 2025).*

G2. Overall, I like the description of the method [for] how the contrails are detected and geolocated which seems to be surprisingly difficult. However, the conclusion of this paper and sketch of further work (text after L 593) is not something that I would rate as a very important topic. I think it is not necessary to check the SAC again, since this has been done in extenso many years ago (there is a figure in the 1999 special IPCC report). There are other interesting problems that I would see as more important, in particular the strange fact that sometimes contrails survive for a while in subsaturated or even quite subsaturated cases. How is this possible? There are some examples in this study and a thorough investigation of these cases would be a better follow-on project.

Response: We agree that the SAC has been evaluated in the past (e.g. [Kärcher et al. \(1998\)](#), the source of the figure in the IPCC report we believe you refer to), but we do still believe that it is important to expand the current study to aircraft beyond the ones present in the IAGOS fleet. Even if we were to use such an expanded dataset only to compare predictions by the SAC with observations in satellite imagery, we still believe this could be valuable. This is because discrepancies between the predictions and the observations may even occur when the SAC is correct (e.g. the overall propulsion efficiency is underestimated): a better understanding of the frequency and the sources of such errors is in itself valuable for performing contrail formation predictions in practice. Throughout the manuscript, we have tried to emphasize that when speaking of “the performance of the SAC”, we refer to the entire prediction pipeline: encompassing meteorological input data, aircraft performance parameters and the resulting predictions by the SAC. Given your comment, we have gone through the text again to ensure that this distinction is made consistently (see also our reply to comment S4, which describes the exact changes).

Additionally, we would like to point out that recent work utilizing computational fluid dynamics (CFD) suggests that the presence of the bypass stream in turbofan engines causes the “mixing trajectory” to deviate from the mixing line predicted by the SAC (e.g., [Ferreira et al. \(2025\)](#)). We therefore consider it to be interesting to see whether such simulation results would be noticeable in an observational dataset constructed using the presented methodology.

We also note that a recent study using ground-based observations of contrails (or their absence) to evaluate the performance of predictions by the SAC reported lower agreement between the two than can plausibly be attributed to NWP input alone ([Gierens et al., 2020](#)). Until now it has not been possible to determine unambiguously whether such disagreement reflects a shortcoming of the SAC or a limitation of the observations. Our work addresses part of that question by examining the effect of using higher-fidelity meteorological data and observations, and our results suggest that a substantial part of the disagreement is observational in origin. They cannot, however, settle the question for modern aircraft, because the IAGOS fleet consists of older engine types with bypass ratios of approximately 5–6.4. We therefore choose to keep a recommendation for further work in this direction within the manuscript.

Subsaturated contrail survival. We agree that this is an interesting feature of the dataset and that it requires further investigation. Comparable behaviour has been reported by [Li et al. \(2023\)](#) and [Petzold et al. \(2025\)](#), and has recently been discussed at EGU ([Driver et al., 2026](#)) as an opportunity to improve upper tropospheric humidity estimates. Because we have shown that high-resolution satellite imagery detects contrails that remain below the detection threshold of some ground-based cameras, an extended dataset of this kind is well suited to examining this behaviour in more detail.

G3. Another possibility could be to make this method available for further tracking with geostationary images. You should think about it.

Response: We thank the reviewer for this suggestion. We agree that the high-resolution detections presented here could serve as a starting point for tracking the same contrails through subsequent geostationary imagery, which would extend the observed lifetime well beyond the single overpass available to us. In this case, the contrail observations could also help flight attribution by providing intermediate labels with which to update the advection estimates. We have added this to the recommendations for future work in the conclusions.

LL 613–616: Secondly, because the LEO observations used in this work resolve contrails well

before they become detectable in geostationary imagery, they provide an intermediate ground truth for the position of a contrail and for its advection during this interval. Such constraints could support the attribution of geostationary-detected contrails to individual flights.

Major issue

M1. Engine efficiency: This misleading expression is used throughout the paper. It is misleading since it indicates that it is a property of the engines. This is not correct. It is rather a property of the whole system Aircraft. In Schumann’s (1996) original paper it is termed “overall propulsion efficiency” and defined as $FV/(Q\dot{m}_f)$, that is, it is the driving power of the aircraft divided by the energy per second necessary to maintain this power. Of course, the engine is an important part of the system, but the flying body itself with its drag is important as well. Later in the paper the problem appears that BADA and PS models give differing results for descending aircraft. Perhaps η is no longer well defined in such a situation, since both F (thrust) and $\dot{m}_f$ (fuel flow rate) approach zero (I don’t know this, however).

Response: We thank the reviewer for this correction. [Schumann \(1996\)](#) indeed uses the term “overall propulsion efficiency”, and we have replaced the term throughout the manuscript.

The differing efficiencies obtained for descending aircraft can be explained by the methodologies used within BADA 4.2 and The Poll–Schumann model. When using BADA 4.2, the efficiency is not a model output but is derived from its definition $\eta = FV/(Q\dot{m}_f)$. At engine idle (e.g. during descent), the drag may exceed the thrust, so that the net propulsive force can be negative and the computed efficiency is negative as well. Within the implementation of pycontrails, a negative efficiency is subsequently clipped to 0.0.

The Poll–Schumann model instead computes the efficiency from the coefficients η_1 and η_2 , which represent engine characteristics (Eqs. 32–37 in [Poll and Schumann, 2021b](#)), and which do not return zero at engine idle.

The difference between the two models during descent is therefore explainable when referring to their methodologies, although we cannot say which of the two is closer to the true value in this case. Of the 543 collocations in our dataset, only 34 occur during descent, and, as reported in Sect. 3.2.1, all incorrect predictions occur during climb or cruise. The uncertainty in the computed efficiency during descent therefore does not affect our main conclusions, and we leave this open for further research focusing specifically on contrails formed during descent.

Specific comments

S1. L 26: don’t mix up “climate effect” (ΔT , sea level rise, etc.) with radiative effects (ERF, RF). What you describe is the latter.

Response: We thank the reviewer for this clarification and have rephrased the manuscript to describe this more carefully.

L 26: ... for over half of aviation’s total radiative impact to date ...

- S2. L 39:** I suggest to delete “For example” and start the sentence right with “Geostationary”.

Response: We thank the reviewer for this suggestion and have updated the manuscript accordingly.

L 39: Geostationary satellite images ...

- S3. LL 60–64:** This discussion is not entirely clear. There are three possibilities for SAC fulfilled—no contrail observed: either the forecast was wrong, or the forecast is correct but no aircraft flying, or the observation is not good enough.

Response: The discussion in these lines describes the results of Schumann et al. (2017), in which contrails were observed either through in-situ measurements or through remote sensing. In every case considered there, a contrail was observed, and the result is that 93.6% of these contrails are correctly predicted by the SAC. The remaining cases, in which the SAC was not fulfilled, therefore correspond either to erroneous atmospheric data (inaccurate in-situ measurements or inaccurate NWP fields) or to a failure of the SAC to correctly predict contrail formation. The possibilities of no aircraft flying, or of the observation not being good enough, do not apply to that data. We hope that this clarifies the discussion.

- S4. LL 66–69:** The expression “lower performance of the SAC” is misleading. I think the SAC itself rests on firm thermodynamic grounds. So if forecast and observations don’t agree, the error must be somewhere either in the input data for the SAC or in the measurements, but not in the SAC.

Response: We agree that the expression “lower performance of the SAC” could be more precise. In this context, “performance of the SAC” was intended to denote the full pipeline of applying the SAC, together with aircraft performance model estimates of η and meteorological data, to predict contrail formation, and of then comparing these predictions with observations. Errors in the input data or in the observations can equally give rise to a mismatch, and are worth stating explicitly wherever this comparison is discussed. We have updated the manuscript accordingly.

LL 69–71 (Introduction): It thus remains unclear whether the lower agreement between the SAC and observations found by Low et al. (2025) and Jarry et al. (2026) arises from limitations of the observations, uncertainties in the underlying meteorological input data, or the SAC itself.

- S5. Sect. 2.1:** when I read that first, I thought that some info on the different channels (i.e. for what they are good) would be nice. This info is then given later, which is ok. Perhaps you can indicate here, that this info will follow.

Response: We added an indication that this information will follow in Table 1.

LL 101-102: The four MSI bands used in this study, along with their wavelengths, spatial resolution and what features are visible in them, are given in Table 1.

S6. L 147: What is a “first-difference standard deviation”?

Response: This is the standard deviation of the differences between successive 4 s samples. The (ordinary) standard deviation of a RH_ℓ signal is dominated by real atmospheric variability along the flight track, such that measurement noise has little impact. The first-difference method used here compares successive samples. We have updated the manuscript.

LL 148–149: ... (on average increasing the first-difference (i.e., the difference between successive samples) standard deviation by 43.3%) ...

S7. L 210: Please correct. According to Schumann (1996) T_{LM} is in °C.

Response: We have updated the manuscript to include this.

L 213: where T_{LM} is in °C and G is in Pa K^{-1} .

S8. L 220: The word “can” surprised me. Perhaps this can be made more concrete (say for the 250 hPa level).

Response: The temperature at which a given water vapour mixing ratio, for example 30 ppmv, corresponds to ice supersaturation also depends on pressure. A concrete example at 250 hPa has been added to the manuscript.

LL 224–225: ... (e.g. at 250 hPa a measurement at the upper detection limit of 30 ppmv corresponds to RH_i of 0.54 at 215 K, but to RH_i of 1.07 at 210 K).

S9. Eq. 6: Please check the units. They don’t combine to a length.

Response: Indeed, the units did not combine to a length. The mass fuel flow is expressed per unit flight distance and was missing an additional m^{-1} . With the corrected units, the numerator is in kg m^{-1} and the denominator in $\text{kg m}^{-3} \cdot \text{m}$, so that B_1 is indeed a length. We have corrected this in the manuscript, and additionally corrected ρ , which was erroneously described as the ambient air pressure rather than the density.

LL 240–241: $\dot{m}_F$ is the mass fuel flow per flight distance in kg m^{-1} ; ρ the ambient air density in kg m^{-3} ;

S10. L 235: the effective time scale is not defined.

Response: Thank you for pointing this out. The effective time scale t_0 is one of the wake vortex scales defined by Schumann (2012), given by $t_0 = 2\pi b_0^2/\Gamma_0$. We added this to the explanation of the variables in the manuscript:

LL 239–240: where $N_{dil}(t_0)$ is the dilution factor at the effective time scale t_0 , with $t_0 = 2\pi b_0^2/\Gamma_0$, corresponding to the time it takes the vortex pair to descend a distance b_0 at its initial descent velocity; ...

S11. L 244: What effect does the angle between the flight direction and the swath direction have?

Response: As shown in Fig. 4, the dominant flight direction among the contrail-forming observations is longitudinal, i.e. from east to west or vice versa. The Sentinel-2 and Landsat

satellites are in near-polar orbits with an inclination of approximately 98° , so their ground tracks run predominantly from north to south. The maximum contrail length that can be captured in a single scene is therefore set approximately by the swath width. Assuming an aircraft ground speed of 250 m s^{-1} and a Sentinel-2 swath width of 290 km, this corresponds to a maximum observable contrail age of about 1250 s, or roughly 21 minutes. Figure 9 shows that the longest-lived contrail in our dataset has an observed lifetime close to this limit, indicating that the swath geometry does constrain the upper end of the observed lifetime distribution.

S12. LL 247 ff: It remains unclear how the IAGOS data are used within CoCiP. The problem is that IAGOS has a 4 s time step, while the other met data that are used as input have a 1 hr time step.

Response: The IAGOS data are used in two ways. First, the trajectory data (time, longitude, latitude and altitude) are taken from IAGOS and resampled to a 1 s resolution. Second, the meteorological input to CoCiP is taken from ERA5, since four-dimensional fields are required to simulate the contrail at locations not sampled by the flight itself. Because ERA5 relative humidity is known to be biased for the purpose of contrail simulation (Gierens et al., 2020), it is corrected using the IAGOS measurements: a uniform scaling factor, derived from the IAGOS-measured relative humidity along the flight track, is applied to the ERA5 relative humidity field, following the approach of Teoh et al. (2022) and Teoh et al. (2024).

S13. L 353: “we make us” ???

Response: We have corrected this in the manuscript.

L 362: Specifically, we make use of the precision ...

S14. Table 5: I was surprised by the similar magnitude of the TP and TN cases. The probable reason is that this table is on contrail FORMATION, not on contrail persistence. It would be helpful to mention this in the table caption.

Response: This is correct: the table is about contrail formation, so the predicted cases follow from the SAC and the observed cases from the satellite imagery. In terms of Table 3, “predicted” refers to the SAC and “observed” to the satellite observations. We have now mentioned this in the caption of the table for extra clarity.

Caption of Table 5: Summary of performance metrics for the SAC compared against the satellite observations of contrail formation for IAGOS and ERA5 weather data. Bold row indicates best performing dataset.

S15. LL 437 ff: Is the interpretation correct that there are then 81% cloud free cases and that in the cloudy cases the increase of RH_i was not large? This could be mentioned (if true).

Response: The interpretation is correct. While going through the data, we realised that it is actually 24% cloudy cases, and 76% cloud-free cases, with most of those cases reporting cloud cover fraction below 0.10. Where the cloud cover fraction was larger, the resulting change in RH_i did not alter the entries of the contingency table, either because the case was already predicted correctly or because the change was too small to move it across the saturation threshold. The

manuscript has been revised to correct this figure and to state the proportion of cloud-free cases explicitly.

LL 450–452: In 76% of the collocations in Table 5 the cloud cover fraction was 0.0. In the remaining 24% cases with partial cloud cover, the clear-sky RH_i did not have an effect on the predictions of the SAC.

- S16. LL 445–457:** The explanation invoking the difference in spatial resolution of the sensors sounds plausible at first reading, but on further thinking, questions arise. If the resolution is 10 or 30 m, but the typical distance of the vortex centres is, say, 60–70 m, then the resolution should in both cases suffice to detect a contrail. So perhaps there is a different reason, for instance the initial optical thickness of the contrail in relation to the threshold contrast necessary for the sensors to detect anything. As you have the relative humidity, you could perhaps try to estimate the optical thickness.

Response: We thank the reviewer for raising this point. Pixel resolution is one part of the explanation, and it is indirectly related to the threshold contrast noted in the comment.

The relevant length scale is not the vortex separation but the width of the individual contrails (before they merge within the wake). In the early phases there is not yet a continuous contrail between the two vortices; rather, two separate contrails in each of the vortices are present, each considerably narrower than the 60–70 m separation. It is also worth noting that 51.8% of the contrails in our dataset never persist beyond the jet phase (defined here as 10 s (Hoshizaki et al., 1972)), which suggests that the missed detections are concentrated among the shortest-lived cases. At these small widths, distinguishing a contrail feature against a (potentially) noisy background is difficult. A sensor with 10 m resolution places several pixels across the feature and is therefore more likely to resolve a sharp gradient at its edges, whereas at 30 m resolution the initial width is covered by approximately one pixel, so that a larger fraction of the signal recorded in that pixel originates from the surrounding background, which in turn reduces the contrast.

Optical thickness could indeed explain why a contrail forms but is not observed. Because the dataset is randomly sampled, we can assume that the underlying distribution of optical thicknesses is identical for both instruments. As discussed before, the difference in resolution might then explain that contrails can be observed in Sentinel-2 while having smaller optical thicknesses than the contrails observed in Landsat, leading to the systemic difference between the two instruments. Moreover, Fig. 6b shows that the Landsat false positives occur at ambient conditions well inside the contrail-forming region, which suggests that they are more plausibly explained by the detection capability of the sensor.

We have updated the discussion in the manuscript to better clarify that the threshold contrast and resolution are related.

LL 430–432: During the early part of the jet phase (below 10 s), the widths of the vortices are in the order of 10's of meters. As a consequence, an instrument with a coarser spatial resolution of 30 m might struggle to detect these features since more background radiation is coming into a pixel, reducing the contrast with the background.

- S17. L 471:** It is unclear whether these three points are a subset of the few purple ones or of the orange ones.

Response: The figure does not distinguish which of the points correspond to the climb phase. The remark was intended as a descriptive statement rather than a reference to a feature of the figure: its purpose is to show that cases during climb and descent, for which the aircraft performance model may attribute the overall propulsion efficiency less reliably, make up only a small part of the set of incorrect SAC predictions. The text has been clarified to state that the flight phase is not indicated in the figure.

LL 484–485: ... three of which are during the climb phase (not indicated in the figure) ...

- S18. Fig. 6b: There is one of the orange bars that does not cross the zero line. Please check. In the last line of the caption there is an incomplete sentence.**

Response: We have re-examined Fig. 6b and were not able to identify an orange bar that does not cross the zero line. It is possible that we have misunderstood which marker is meant. The incomplete sentence in the caption has been corrected.

Caption of Fig. 6: ... Cross (circle) markers indicate that the contrail is observed in Landsat (Sentinel-2).

- S19. L 501: Just a comment. This points to a tremendous small-scale variability of the UT humidity field, isn't it?**

Response: We thank the reviewer for this remark. While going through the text again, we realise that this statement was poorly phrased. The original intention was to point out the small number of contrails observed with $\Delta T_{LC} > 0$. A plausible explanation is that the measured temperature and relative humidity, or the computed overall propulsion efficiency, carry sufficient uncertainty for a small number of observed contrails to fall on the wrong side of the threshold.

A further source of uncertainty concerns the pairing of the measured properties with the contrail lifetime. The relative humidity shown in Fig. 7a, for example, is the value at the estimated emission time of the oldest part of the contrail: for an aircraft observed at 12:00:00 with a contrail of 10 minutes observed lifetime, the IAGOS measurement at 11:50:00 is thus used. Over those 10 minutes the ambient conditions at the oldest end of the contrail may differ from those measured at the time of emission, both because the air mass evolves and because the contrail is advected.

We have adjusted the text in the manuscript to highlight this clarification.

LL 507–509: Another reason might be the response time of the IAGOS humidity sensor discussed in subsection 2.2. Because the humidity sensor responds more slowly in colder temperatures, sharp humidity gradients are dampened and the measured RH_i may not represent the actual RH_i .

LL 513–514: ..., similar to the three Sentinel-2 observed contrails not predicted by the SAC (see Table 5).

- S20. LL 525–526: Why this, that is why vortex separation (a horizontal distance)? Contrail spreading in the dispersion regime is mainly due to vertical wind shear. Thus the vertical extension of the contrail at the end of the vortex phase is important. As far as I know this is coded in CoCiP.**

Response: We thank the reviewer for this comment. Vertical wind shear is indeed the main driver of contrail spreading, as stated in the preceding sentence, and the vertical extent at the end of the vortex phase is what governs it. CoCiP initialises the diffusion phase with both an initial depth D_1 and an initial width B_1 . The depth D_1 indeed controls the growth due to shear and is implemented in pycontrails as expected. The sentence in question concerns the initial width, since this is what our observations are able to measure. In making that comparison we noted that the CoCiP implementation in pycontrails (Shapiro et al., 2026) uses the wake vortex separation as the initial width, whereas Schumann (2012) computes it from Eq. 6. The point of the sentence is therefore this discrepancy in the initialisation of B_1 , not a claim about the mechanism of spreading.

S21. Sect. 3.3.3: I find this section with its three examples a bit thin. What is the message of this section? What happens in case 1? What are these regions of sublimation? Or are these short paths where SAC is not fulfilled?

Response: The aim of this section is to illustrate how the observed evolution of contrail width can be compared against a model such as CoCiP, using the three longest-lived contrails in our dataset as examples. It is not intended as a quantitative validation of CoCiP, for which our sample of persistent contrails is far too small. Following a related comment from Referee 2, we have rephrased the section, together with the abstract and the conclusions, to make this purpose clearer. We hope to extend the analysis to a larger dataset in future work, which would allow conclusions to be drawn on how well CoCiP captures certain aspects of early contrail evolution.

Regarding Case 1: the SAC is satisfied along the entire contrail, so the gaps are not segments where the SAC is not fulfilled. They are regions in which RH_i was too low for the contrail to persist until the time of observation. Since the satellite records the contrail at a single instant, parts of it may already have sublimated by that time, leaving gaps between the surviving segments. A clarification has been added to the caption of Fig. 9.

Figure 9 caption: ... For Case 1, breaks in the simulated contrail indicate segments that have sublimated before the satellite scanning time.

S22. LL 575–581: Please see my comment to LL 445–457.

Response: We thank the reviewer again for pointing this out. We believe our response to the previous comment addresses this.

S23. L 634: alone (not along).

Response: We have corrected this typo.

L 650: ... by random sampling alone.

S24. References: some entries have formatting problems: namely Knight et al., Neis et al., Tompkins et al.

Response: We thank the reviewer for spotting these formatting problems. We have updated this in the manuscript.

Response to Referee #2

General comments

G1. First, I would like to sincerely congratulate the authors on a truly excellent and innovative piece of work. [...] I particularly want to highlight the rigorous attention to detail demonstrated in the data methodology. Accurately geolocating aircraft in push-broom satellite imagery is a surprisingly difficult task. The authors' corrections for parallax displacement and their intricate accounting for detector-specific scanning time offsets represent a significant and impressive technical achievement. Overall, this manuscript is fundamentally strong, well-written, and makes a substantial contribution toward bridging the gap between in-situ measurements and large-scale remote sensing. However, to ensure the conclusions are perfectly framed for the community and to prevent potential over-extrapolation of the results, there are a few structural assumptions and methodological constraints that would benefit from further contextualization. Specifically, I recommend expanding the discussion regarding fleet representativeness, particularly concerning the SAC's applicability to modern engine architectures, alongside addressing the statistical limitations in the dispersion phase analysis and acknowledging some hardware latencies.

Response: We thank the reviewer for the overall feedback and suggestions. We will address these points in M1, M2 and M5.

Specific comments (major scientific issues)

M1. [Fleet Representativeness and SAC Assumptions]

I note that the dataset currently focuses on legacy wide-body aircraft (specifically the Airbus A330 and A340 families) operating primarily on specific route structures. Furthermore, because Sentinel-2 and Landsat are sun-synchronous, the observations are naturally skewed toward standard departure schedules that cross the satellite path at fixed local times.

Your results beautifully demonstrate that the Schmidt–Appleman Criterion (SAC) performs exceptionally well for predicting contrail genesis in these older aircraft types used in the IAGOS fleet. However, it is crucial to avoid making definitive statements about the SAC's performance across the global fleet.

Recent studies using ground camera images (including findings presented in a recent CICONIA Outreach session) show that SAC prediction performance can vary significantly, dropping from an F1 score of around 0.9 for older engines (e.g. CFM56 on the A320ceo) to an F1 score of around 0.75 for modern lean-burn engines (e.g. Leap 1-A on the A320neo).

Recent modeling and empirical research supports this discrepancy. For instance, Ponsonby et al. (2025) demonstrated via microphysical models that while the SAC

critical temperature error (ΔT_{SAC}) is rather small (less than 2 K) for rich-burn (soot-rich) engines where the diameter is greater than 20 nm, the classical SAC struggles with lean-burn (soot-poor) engines. When soot emissions drop below 10^{13} to 10^{14} per kg, droplet activation relies on highly volatile particles (diameter less than 5 nm), which pushes the activation threshold to much colder temperatures and severely miscalculates the active ice crystal emission index (AEI_{ice}).

Additionally, flight campaigns like ECLIF II/ND-MAX (Voigt et al., 2021) highlight that in modern low-soot/lean-burn engine plumes, low soot concentrations fail to effectively scavenge gaseous precursors. Volatile aerosols thus grow at temperatures significantly below the standard T_{SAC} . High-fidelity 3D Large-Eddy Simulations of early-stage jet–vortex interactions paired with ground-based camera imagery validation further corroborate this.

Recommendation: It is certainly not the responsibility of this paper to improve or fix the SAC. However, I kindly request that you add a statement in the discussion or limitations section explicitly acknowledging that while these results validate the SAC’s high performance for older engine types, literature shows its predictive capability is significantly reduced for modern lean-burn, ultra-high bypass ratio, and SAF-burning engines. Consequently, operational generalizations regarding the global fleet should be carefully tempered.

Response: We thank the reviewer for this detailed and well-referenced comment. While our dataset shows that the SAC predicts contrail formation well for the aircraft in the IAGOS fleet, we agree that this should not be generalised to all aircraft–engine combinations. The SAC was originally derived and validated for turbojet and low bypass-ratio engines, for which the linear mixing line description holds well. Contrail formation by modern lean-burn and ultra-high bypass ratio engines may be captured less accurately, and requires further study.

The results presented at the CICONIA outreach session are interesting to hear about in this respect. Given that high-resolution LEO imagery resolves contrails at formation, we intend to continue this research using the automated detection approaches as mentioned in our recommendations.

The manuscript has been adjusted to make clear that the results presented here apply specifically to the IAGOS fleet and may not carry over to aircraft with modern lean-burn or ultra-high bypass ratio engines. The following changes have been made.

L 11 (Abstract): ...for the aircraft in the IAGOS fleet.

LL 399–400: The IAGOS fleet covers bypass ratios between 5.0 and 6.4. These subsequent results may therefore not apply to modern lean-burn or high-bypass engines.

M2. [Statistical Insignificance in the Dispersion Phase]

I appreciate the authors’ transparency regarding the limited sample size, correctly noting that three contrail cases survived longer than 600 seconds. While the observed horizontal growth rates align nicely with CoCiP simulations, drawing operational or definitive modeling conclusions from three isolated cases is difficult. I highly recommend not drawing any conclusions out of these results, just keep them as an example case study.

Response: We thank the reviewer for this recommendation, and agree that three cases are too few to support general conclusions. To avoid readers drawing oversimplified conclusions, we have rephrased the statements in the abstract and the conclusions so that the three contrails are presented as individual case studies illustrating a comparison, and we no longer report a level of agreement in those sections. The quantitative comparison has been kept in Sect. 3.3.3. That section already cautions the reader that the sample size does not permit statistically robust conclusions, and we consider that reporting how closely the observations and the simulations agree is necessary when a model is compared against observations.

LL 18–19 (Abstract): ...three example cases illustrate how the observed horizontal growth can be compared with contrail simulation results.

LL 604–606 (Conclusion): We presented the three oldest contrails in the dataset as individual case studies, illustrating how observations of this kind can be compared against contrail models. A larger dataset, covering a wider range of atmospheric conditions and contrail lifetimes, would be required to evaluate modelled dispersion rates quantitatively.

M3. [Manual Annotation Subjectivity]

The manual annotation of contrails using polygons and linestrings is a solid, practical approach. However, defining the edges of a dispersing contrail by the human eye in 10 m or 30 m resolution imagery is inherently subjective. A small note acknowledging the potential for non-reproducible biases in the width evolution metrics would be a great addition.

Response: We thank the reviewer for raising this point. We agree that manual annotation is a limitation of the method, which is also why automated approaches are recommended for future work. To emphasise this, two sentences have been added to Sect. 2.7:

LL 306–308: As this step is done manually, there might be non-reproducible biases in the placement of the contrail edges, and thus the derived contrail widths. This is particularly relevant for older, more diffuse contrails, where the contrail edges become less distinct from the background.

A corresponding reminder has been added where the width evolution is discussed in Sect. 3.3.2:

LL 518–519: Do note that the observed widths are subject to bias from the manual annotation process.

M4. [Wind Advection and Spatial/Temporal Mismatch]

The explicit acknowledgment that ignoring horizontal wind advection induces an error in contrail age of approximately 20% (for a 50 m s^{-1} wind speed) is much appreciated. Briefly mentioning that this spatial/temporal mismatch could introduce noise into the correlation between the modeled early evolution and the true physical state would round out this section nicely.

Response: We thank the reviewer for this suggestion. A sentence has been added at the end of Sect. 2.7 describing how this uncertainty in contrail age propagates into the later comparison against models.

LL 348–350: It is important to note that this uncertainty in the estimated contrail age introduces noise into the correlation between contrail width and age, and thus into the later comparisons against model predictions of the early evolution.

M5. [In-Situ Sensor Latency]

The choice not to apply a correction to the IAGOS capacitive hygrometer (ICH) to avoid artificially amplifying noise makes complete practical sense. However, because this leaves an uncorrected temporal lag (up to 120 seconds at 210 K), it would be beneficial to briefly discuss how this lag might affect the relative humidity values used to validate the SAC threshold within the context of your error margins.

Response: We thank the reviewer for this comment, and agree that this leaves an uncorrected source of uncertainty. For the evaluation of the SAC the effect is limited. The response time reaches its largest values at the lowest temperatures, and below approximately 215 K the SAC is generally satisfied even in completely dry air, so an inaccurate RH_ℓ does not alter the predicted outcome. More generally, the SAC is predominantly controlled by temperature (Gierens et al., 2020), which contributes more to the uncertainty in ΔT_{LC} than the humidity does, as shown in Fig. 6b.

The consequences are more relevant for the relationship between contrail lifetime and RH_i presented in Sect. 3.3.1. There, the measured value is used as a direct indicator of the ambient humidity, and at the low temperatures the lag affects the reliability of the derived RH_i . A note on this has been added where the relationship between RH_i and contrail lifetime is discussed.

LL 507–509: Another reason might be the response time of the IAGOS humidity sensor discussed in subsection 2.2. Because the humidity sensor responds more slowly in colder temperatures, sharp humidity gradients are dampened and the measured RH_i may not represent the actual RH_i .

M6. [Aircraft Performance Modelling Assumption Uncertainty]

Applying a generic 15% uncertainty buffer to bridge the gap between the BADA 4.2 and Poll–Schumann overall engine efficiency models is a reasonable compromise, though it masks the localized thrust variations that physically govern condensation. A short sentence recognizing this constraint would be helpful.

Response: We thank the reviewer for this comment. To clarify, the 15% uncertainty was not chosen to bridge the gap between BADA 4.2 and Poll–Schumann; it is the sum of the uncertainties reported by Poll and Schumann (2021a,b, 2025). We do agree that the overall propulsion efficiency varies with local flight conditions. The effect on the outcome of this study is limited, since Fig. 6b shows that the full 15% uncertainty in efficiency corresponds to a change in ΔT_{LC} of ± 0.66 K on average. We added a short sentence to section 2.4 acknowledging the lack of local thrust variation.

LL 187–188: We note that this constant 15% uncertainty does not take into account local flight conditions and can mask local thrust variations.

Minor issues

N1. [Incomplete Annotations]

There are a few examples in the dataset where only part of the contrail was annotated. Examples include:

- **Case 1:** 2019032209415402_LC82070182019081LGN00
- **Case 2:** 2020013109252202_LC80190202020031LGN00
- **Case 3:** 2014040110181403_LC81990232014091LGN01

This is completely understandable given the complexity of the imagery, but it might slightly skew the metric stating that only 3 examples lasted much longer than the small section that was labeled. Please consider adding a short disclaimer regarding partial annotations.

Response: We thank the reviewer for inspecting the dataset in this detail. The three cases listed are all Landsat collocations. For Landsat, a linestring was annotated only along the first section of the contrail, since the annotation was performed on zoomed images centred on the aircraft position (see Sect. 2.7). It is correct that several of these contrails may have persisted further and are therefore only partially annotated. The contrail evolution results presented in Sect. 3.3 are derived from Sentinel-2 observations alone, so these cases do not affect them. We believe this is already indicated at the start of Sect. 3.3, but a further clarification has been added to Sect. 2.7 stating that lifetime and evolution metrics are computed only from Sentinel-2 annotations.

LL 330–331: Because we annotate on zoomed images in Landsat, and we thus have no information on the full lifetime of these contrails, we will not use these cases for determining metrics such as contrail age and width evolution.

N2. [Unclear Imagery]

A significant amount of images in the dataset didn't allow to distinguish contrails, even though there is a label. For example: the case LC82050312013210LGN01, it is quite difficult to see anything on the cirrus band png file provided in the dataset, and only a very diffuse shape is visible on the thermal band. Could you add a brief explanatory note detailing why this specific case presents this way? Note that these are much more visible when using the uncompressed data directly from the gcp repo.

Response: We thank the reviewer for examining the dataset so closely. The case LC82050312013210LGN01 is a Landsat collocation. On re-inspection we do believe that a short contrail is visible extending towards the north-east, although we agree that it is diffuse. Several further cases are of this kind and, as noted in our response to comment M3, manual annotation always carries a degree of subjectivity.

Over the course of the labelling we identified specific features indicating the presence of a contrail. In Sentinel-2 imagery, for instance, contrails were often more clearly visible in the RGB composite while the cirrus band showed a consistently diffuse shape, so that a diffuse contrail-like feature immediately adjacent to the aircraft position could be identified with reasonable confidence. Where a contrail could not be distinguished with sufficient confidence,

the case was excluded from the final dataset; this occurred, for example, where thick cloud cover beneath the flight track made it impossible to determine whether a contrail was present. The images distributed with the dataset are produced using the visualisation functions of *pycontrails*, and the Landsat annotations were made on these images, albeit on compressed versions as noted in the comment. In this process each band is first rescaled to $[0,1]$ using the minimum and maximum values occurring in that particular scene, after which the bands are combined. The Copernicus Browser, which was used for the Sentinel-2 annotation, instead applies a fixed tone curve that is identical for every scene. A contrail may therefore be apparent in one rendering but washed out in the other as a result of this normalisation step. When working with the Copernicus Browser it is therefore advisable to adjust the settings to enhance the contrast in certain scenes.

N3. [Proposed Table Addition (Formation vs. Persistence)]

If space & time permits, it would be highly informative and interesting to see a table outlining the scores for the contrails regarding which ones were predicted to be persistent (SAC + ISSR) versus contrails that effectively lasted until the start of the diffusion phase (or those lasting more than 5 minutes).

Response: We thank the reviewer for this suggestion. We have constructed such a table, reproduced below for persistence thresholds of 2 and 5 minutes.

Table 1: Contingency table and precision and recall metrics for the prediction of contrail persistence (SAC combined with $RH_i > 1$), evaluated against observed contrail lifetimes exceeding 2 and 5 minutes, using IAGOS and ERA5 meteorological data.

Threshold	Meteorological Data	TP	FP	FN	TN	Precision	Recall
2 min	IAGOS	10	24	7	325	0.294	0.588
	ERA5	9	16	8	333	0.360	0.529
5 min	IAGOS	5	29	1	331	0.147	0.833
	ERA5	3	22	3	338	0.120	0.500

We have decided against including it in the manuscript, for the following two reasons.

There are only a limited number of cases surviving beyond the vortex phase, namely 8.7% of cases survive longer than 120 seconds. All of these are Sentinel-2 cases, because no lifetime data is available for Landsat. This naturally skews the data to be sensitive to small changes to where the persistence threshold is drawn. Combining that with the uncertainty in determining the contrail age (as also addressed in M4), it may limit what we can conclude from these metrics.

We also note that Fig. 7a already contains this information in continuous form: dividing the panel at $RH_i = 1$ and at a chosen lifetime threshold yields the four quadrants of the contingency table, with false positives in the lower right and false negatives in the upper left. Using Fig. 7a, the reader can look at the actual cases, rather than focusing on precision and recall counts. We think that the trend visible in that figure is a fairer metric. This also makes it visible that contrails can persist in slightly ice-subsaturated air, consistent with [Li et al. \(2023\)](#) and [Petzold et al. \(2025\)](#), which would normally be classified as false negatives.

We hope that this answers the comment and explains why we did not include it in the manuscript.

Technical corrections

T1. L 74: “... *either due the difficulty of performing measurements...*” Missing preposition. Please correct to “... *due to the difficulty...*”.

Response: Corrected.

T2. L 87: Typo in “(*in the case one was forrmed*)”. Please correct to “formed”.

Response: Corrected.

T3. L 322: “... *with help of other true colour and false colour images...*” Missing article. Change to “... *with the help of...*”.

Response: Corrected.

T4. L 325: “*Such annotations are flagged accordingly, to avoid potential misinterpretation of the annotation.*” The phrasing is slightly repetitive. Consider simplifying to: “*Such annotations are flagged accordingly to avoid potential misinterpretation*”.

Response: Corrected.

T5. L 353: “*we make us*” As also noted in another comment, please correct to “*we make use of*”.

Response: Corrected.

T6. L 380: “... *collocated in both climb (54 times), cruise (455 times) and descent (34 times) phases.*” The word “both” is incorrectly applied to three items. Please delete “both”.

Response: Corrected.

T7. L 384: “... *which has since been fixed, but results in validity flags of 2 for all repaired data and are therefore excluded from the analysis.*” Is the subject for the latter half of the sentence missing/mismatched? I don’t really understand the sentence...

Response: Thank you for pointing out this grammatical error. We have rephrased the sentence in the manuscript to avoid misunderstanding.

LL 392–394: The limited availability of data between 2013 and 2018 can be attributed to an error in the IAGOS data acquisition until 2018. The affected measurements have since been corrected, but all repaired data carry a validity flag of 2 and are therefore excluded from the analysis.

T8. L 407: “*Over the whole dataset, we found of a mean difference of -0.583 K ($T_{\text{ERA5}} - T_{\text{IAGOS}}$, with a standard deviation...*” Remove the extra “of” (“we found a mean difference”) and add the missing closing parenthesis after T_{IAGOS} .

Response: Corrected.

T9. L 424: “... *the probability of finding the observed differences in precision and recall are 0.005 and 0.06 respectively.*” Subject–verb agreement error. Please change “probability” to “probabilities”.

Response: Corrected.

T10. L 498: “*This is also visible in a lower R^2 value 0.23 between...*” Missing preposition. Change to “lower R^2 value of 0.23”.

Response: Corrected.

T11. L 510: “...*47.3 m for both aircraft types .*” There is an unnecessary extra space before the period.

Response: Corrected.

T12. L 589: “...*does not allow for robust statistical conclusion on the rate of contrail growth...*” Missing pluralization. Please change to “robust statistical conclusions”.

Response: Corrected.

T13. L 633: “...*that the observed differences are explainable by random sampling along.*” Typo. Please correct “along” to “alone”.

Response: Corrected.

T14. L 655–656: “...*with on PS computing engine efficiencies 0.01 to 0.02 higher than BADA 4.2.*” Typo. Please remove “on” to read “...with PS computing...”.

Response: Corrected.

T15. L 707: “...*since the azimuth angle changes discontinuously across the detector border...*” Typo. Please correct “discontinuously” to “discontinuously”.

Response: Corrected.

T16. L 733: “*TW, VM, and ZE developed the methodology and creation of the dataset.*” Typo in “methodology”. Additionally, “developed the... creation” is awkward phrasing; suggest changing to “...developed the methodology and created the dataset”.

Response: Corrected.

T17. Figure 6 caption: “...*even when accounting for uncertainty are shown in purple. with lower opacity.*” Typo with the punctuation. Please change the period before “with” to a comma, or remove it entirely.

Response: Corrected.

References

- Driver, O., Ponsonby, J., Gourgue, N., Boucher, O., Stettler, M., and Gryspeerdt, E.: Transient contrails as an opportunity for upper tropospheric humidity estimation, Tech. Rep. EGU26-20695, Copernicus Meetings, <https://doi.org/10.5194/egusphere-egu26-20695>, conference Name: EGU26, 2026.
- Ferreira, T., Figueroa, R. Q., Arnedo, M. S., and Gorle, C.: Sensitivity analysis of initial conditions for contrail formation with ground-based camera imagery, https://web.stanford.edu/group/ctr/ResBriefs/2025/22_Ferreira.pdf, 2025.
- Gierens, K., Matthes, S., and Rohs, S.: How Well Can Persistent Contrails Be Predicted?, *Aerospace*, 7, 169, <https://doi.org/10.3390/aerospace7120169>, 2020.
- Hoshizaki, H., Conti, R. J., Anderson, L. B., Redler, K. O., and Meyer, J. W.: Study of High-Altitude Aircraft Wake Dynamics. Task 1. Problem Definition, <https://apps.dtic.mil/sti/html/tr/AD0754918/index.html>, number: DOTST903, 1972.
- Kärcher, B., Busen, R., Petzold, A., Schröder, F. P., Schumann, U., and Jensen, E. J.: Physicochemistry of aircraft-generated liquid aerosols, soot, and ice particles: 2. Comparison with observations and sensitivity studies, *Journal of Geophysical Research: Atmospheres*, 103, 17 129–17 147, <https://doi.org/10.1029/98JD01045>, 1998.
- Li, Y., Mahnke, C., Rohs, S., Bundke, U., Spelten, N., Dekoutsidis, G., Groß, S., Voigt, C., Schumann, U., Petzold, A., and Krämer, M.: Upper-tropospheric slightly ice-subsaturated regions: frequency of occurrence and statistical evidence for the appearance of contrail cirrus, *Atmospheric Chemistry and Physics*, 23, 2251–2271, <https://doi.org/10.5194/acp-23-2251-2023>, 2023.
- Petzold, A., Khan, N. F., Li, Y., Spichtinger, P., Rohs, S., Crewell, S., Wahner, A., and Krämer, M.: Most long-lived contrails form within cirrus clouds with uncertain climate impact, *Nature Communications*, 16, 9695, <https://doi.org/10.1038/s41467-025-65532-2>, 2025.
- Poll, D. and Schumann, U.: An estimation method for the fuel burn and other performance characteristics of civil transport aircraft in the cruise. Part 1 fundamental quantities and governing relations for a general atmosphere, *The Aeronautical Journal*, 125, 257–295, <https://doi.org/10.1017/aer.2020.62>, 2021a.
- Poll, D. and Schumann, U.: An estimation method for the fuel burn and other performance characteristics of civil transport aircraft during cruise: part 2, determining the aircraft’s characteristic parameters, *The Aeronautical Journal*, 125, 296–340, <https://doi.org/10.1017/aer.2020.124>, 2021b.
- Poll, D. I. A. and Schumann, U.: An estimation method for the fuel burn and other performance characteristics of civil transport aircraft; part 3 full flight profile when the trajectory is specified, *The Aeronautical Journal*, 129, 825–861, <https://doi.org/10.1017/aer.2024.141>, 2025.
- Schumann, U.: On conditions for contrail formation from aircraft exhausts, *Meteorologische Zeitschrift*, 5, 4–23, <https://doi.org/10.1127/metz/5/1996/4>, 1996.
- Schumann, U.: A contrail cirrus prediction model, *Geoscientific Model Development*, 5, 543–580, <https://doi.org/10.5194/gmd-5-543-2012>, 2012.

- Schumann, U., Baumann, R., Baumgardner, D., Bedka, S. T., Duda, D. P., Freudenthaler, V., Gayet, J.-F., Heymsfield, A. J., Minnis, P., Quante, M., Raschke, E., Schlager, H., Vázquez-Navarro, M., Voigt, C., and Wang, Z.: Properties of individual contrails: a compilation of observations and some comparisons, *Atmospheric Chemistry and Physics*, 17, 403–438, <https://doi.org/10.5194/acp-17-403-2017>, 2017.
- Shapiro, M., Engberg, Z., Teoh, R., Stettler, M., Dean, T., and Abbott, T.: pycontrails: Python library for modeling aviation climate impacts, <https://doi.org/10.5281/zenodo.18260617>, 2026.
- Teoh, R., Schumann, U., Gryspeerdt, E., Shapiro, M., Molloy, J., Koudis, G., Voigt, C., and Stettler, M. E. J.: Aviation contrail climate effects in the North Atlantic from 2016 to 2021, *Atmospheric Chemistry and Physics*, 22, 10 919–10 935, <https://doi.org/10.5194/acp-22-10919-2022>, 2022.
- Teoh, R., Engberg, Z., Schumann, U., Voigt, C., Shapiro, M., Rohs, S., and Stettler, M. E. J.: Global aviation contrail climate effects from 2019 to 2021, *Atmospheric Chemistry and Physics*, 24, 6071–6093, <https://doi.org/10.5194/acp-24-6071-2024>, 2024.
- Vázquez-Navarro, M., Mannstein, H., and Mayer, B.: An automatic contrail tracking algorithm, *Atmospheric Measurement Techniques*, 3, 1089–1101, <https://doi.org/10.5194/amt-3-1089-2010>, 2010.
- Vázquez-Navarro, M., Mannstein, H., and Kox, S.: Contrail life cycle and properties from 1 year of MSG/SEVIRI rapid-scan images, *Atmospheric Chemistry and Physics*, 15, 8739–8749, <https://doi.org/10.5194/acp-15-8739-2015>, 2015.