Response to Reviewer 2's Comments

Green: Reviewer's comments

Black: Authors' response

Blue: Edits in manuscript

This manuscript investigates the use of ML approaches applied to on-board cameras for observations of contrails in their ultra early lifetimes, leveraging such observations alongside in-situ atmospheric data collected by a NASA chase aircraft during the ECLIF II/ND-MAX and ecoDemonstrator campaigns. The importance of camera properties, including mounting position and orientation, are highlighted, as well as the various forms of image processing that are required to remove obscuring factors. Comparison between the image classification and SAC/ISSR criteria based on the in-situ data demonstrates method validity, with the potential to benefit aspects of future contrail research and mitigation.

I believe that the manuscript contains informing points that will contribute to future work within the contrail, cloud-physics, and weather forecasting research communities. These largely revolve around the processing of the observations, detailing the requirements and factors of influence, and showing the path forward for high-quality observations of ultra-early contrail lifetimes from on-board camera systems (both on test aircraft such as the campaigns analysed but also potentially within commercial aviation).

However, I have several issues with the manuscript as currently written that prevent me from recommending its publication in the current form. These can be mainly categorised into (1) the proposed context for the work, and (2) underlying assumptions for the analysis, but I feel there is also a lack of interpretation in translating results into discussion and conclusions. I would seek for these to be addressed or the authors to reframe the manuscript before publishing. There is a scattering of smaller issues, such as citations and scoring metrics used, but I will leave these in favour of the major points.

We would like to thank the reviewer for their very constructive feedback. The responses to reviewer's suggestions and the associated revisions in the manuscript are provided below. We have revised the Introduction to emphasize the framing for observations of very early contrail development, the processing of onboard camera imagery and the limitations that affect its interpretation. We also acknowledge that differences between the two campaigns prevent us from attributing classification performance to camera placement alone. Following the reviewer's suggestions, we have added a comparison of SAC predictions to manually assessed contrail

occurrence for the ND-MAX imagery and take the camera observations as the reference. Our responses to the individual comments are provided below.

---

(1) Proposed Context

The framing of the work is geared around the potential for on-board cameras to function as a detection system for contrails, from which either monitoring or response mechanisms can be built to aid in efforts to mitigate the contrail climate effect (e.g., L149-151). My main issue with this relates to another point I wish to make with respect to the definition of persistence within the work. On-board camera systems of the type being analysed can only observe ultra-early contrail lifetimes. They can thus say little about contrail lifetime and consequently little about the climate impact of those contrails, even with the presence of co-located in-situ humidity measurements. A camera observing contrail formation in an ISSR is no guarantee that the contrail will live for long enough to have a meaningful climate impact, and so cannot be used to make recommendations for operational mitigation action (navigational or otherwise) targeted at strongly warming contrails.

Instead, I would be in favour of framing the work in a context similar to ground-based cameras, whereby observations of early contrail lifetimes are being used to probe model predictions of meteorology. In this way, on-board camera observations could be used for data assimilation, employing aircraft-as-a-sensor ideas and focusing on met parameters that can be inferred from contrail formation (whether such data assimilation will contribute to forecasting persistent contrails is still unclear).

We thank the reviewer for this suggestion. We have revised the Introduction to put greater emphasis on the value and the limitations of onboard camera observations for very early contrail development. In the original Introduction, we described the objective as real-time contrail detection and prediction of possible persistence. Our intention is to make the case to use low-cost cameras, potentially along with additional low cost instruments, on selected commercial flights with the longer-term motivation to support contrail prediction, the identification of potential ice supersaturated regions. We have revised the framing to emphasize the usefulness of these observations and the requirements for obtaining them. We recognize the distinction by the reviewer between observing early contrail formation and establishing its lifetime or climate impact.

We have also revised the literature discussion to support this framing. We added Brauer et al (2021) to discuss airborne measurements of young contrails and the dependence of their ice properties and extinction on atmospheric conditions. We also included and expanded on the discussion of Schumann et al. (2017) to emphasize the need for reliable observations and more complete datasets to test contrail models. We also expanded the discussion of Low et al. (2025) to describe comparisons of observed contrail lifetimes with meteorological conditions and the limitations

imposed by contrail visibility and the camera's field of view. With these changes we aim to provide context for the usefulness of additional observations and the limitations affecting their interpretation.

The revised Introduction explains how camera observations interpreted with in situ atmospheric measurements can be used to examine the relationship between visible contrail formation and thermodynamic conditions for contrail formation predicted by SAC, together with the processing requirements and the limitations that affect the observations. In "Current scope and contributions", we describe the preparation of the imagery, the classification against reference labels derived from atmospheric measurements and some suggested recommendations for future camera installations. We also now state that their deployment on commercial aircraft requires further assessment of installation, image quality and classifier performance across the aircraft and operating conditions. We agree that observing contrail formation in an ISSR does not establish subsequent contrail lifetime or climate impact. The contribution of these observations to data assimilation, improved forecasts of persistent contrails and operational mitigation remains to be demonstrated.

Edits in the manuscript – Section I. Introduction

1. Added paragraph: Cameras onboard aircraft offer an opportunity to observe the initial stages of contrail formation. When interpreted alongside in situ atmospheric measurements, these observations can be used to examine the relationship between visible contrail formation and what we expect from thermodynamics (National Academies of Sciences, Engineering, and Medicine, 2025). Their usefulness also depends on the available view of the aircraft wake, image quality and the processing to account for factors that affect contrail visibility, which in turn make camera placement and orientation important considerations for future measurement campaigns.

2. Current Scope and Contributions, fully rewritten: This study examines imagery from cameras mounted onboard aircraft from the 2018 ECLIF II/ND-MAX and 2023 ecoDemonstrator campaigns with in situ atmospheric measurements from the chase aircraft (NASA Airborne Science Program; NASA Aeronautics Mission Directorate).We investigate whether a machine learning model can distinguish thermodynamically defined categories from the imagery and also the observational limitations encountered when analyzing contrails during their very early development. The main contributions of this work are as follows:

    a. We describe the processing and filtering of onboard camera imagery from the two NASA campaigns and document visual obstructions, limitations in image quality and the available view of the aircraft wake early in the contrail development.

    b. We investigate the use of image classification to detect contrail formation and identify potential ISSRs. We evaluate classification performance against reference

labels from the DC-8's atmospheric measurements using the SAC and a threshold for RHi.

c. We discuss limitations of the camera observations and considerations for camera mounting position and orientation for future campaigns. Because the campaigns differ in aircraft, engines, atmospheric conditions and image acquisition, we cannot determine how much camera placement contributes to the differences in classification performance.

The longer-term goal would be to combine low cost onboard cameras with classification software on selected commercial flights to gather additional observations of early contrail development. The present study describes how the camera footage from the two campaigns was prepared for analysis and how the ambient conditions, such as time of day, haze, sun angle, limited the usable observations. Deployment would require further assessment of camera installation, image quality and classifier performance across aircraft and operating conditions. Whether the resulting observations could contribute to data assimilation or improved meteorological forecasts or predictions on contrail persistence remains a question for future investigation.

3. We clarified that the reference categories describe thermodynamic conditions and do not establish the observed lifetime or climate impact of a contrail.

(2) Underlying Assumptions in Methodology

The first point (smaller point) under this category relates to my suggested best use-case for the observations. The atmospheric data as measured by the case aircraft is currently used to calculate SAC pass/fail and thus characterise whether contrail formation is expected, with that calculation taken as ground truth against which the imagery is scored. I would like to see this inverted, or at least like to see the results of doing so, taking the on-board camera observations of contrail/no contrail as reference and evaluating the thermodynamic characterisation against this. The data appears to already be there and requires only a change of baseline, so not doing this feels like a missed opportunity for a fairly unique assessment on rare data.

Framed this way, the exercise contributes to an evaluation of the SAC, rather than a check on label noise. I would suggest that disagreements between observation and atmospheric data be presented with this additionally in mind and that the authors distinguish between cases where the criterion itself is at fault from cases limited by camera properties (or even the chase aircraft's sampling, see later). This also speaks to the ongoing question of whether the SAC remains a suitable assumption as aircraft technology and fuels evolve, which connects to my next (more major) point.

Following the reviewer's suggestion, we have added a comparison that takes manually assessed contrail occurrence in the ND-MAX imagery as the observational reference. The analysis uses a stratified sample of 686 images, weighted to estimate agreement for the full set of approximately 16423 images. The estimated agreement between visible contrail occurrence and SAC predictions is 90.7%, with Cohen's κ of 0.74. Most disagreements occur when SAC predicts formation but no contrail is visible. This comparison evaluates SAC predictions against camera observations separately from the evaluation of the image classifier against its thermodynamic reference labels. We discuss limited contrail visibility, differences between the atmospheric sampling and the imaged wake, and assumptions in the SAC calculation as possible contributors to disagreement. The comparison does not isolate these contributions or establish that an individual disagreement represents a failure of the SAC.

In Section III (Results):

The agreement between SAC and manual visual observation was investigated and summarized in **Table 9** for ND-MAX. Taking the camera-based manual labels as the observational reference, SAC correctly indicates contrail formation in 97.9% of frames where a contrail is visible and no formation in 71.1% of frames where none is visible. Because contrails appear in about 73% of frames, κ is the more informative. Disagreement is asymmetric since about 84% of the estimated 1525 disagreeing frames are cases where SAC predicts formation but no contrail is visible (false-alarm ratio of 9.8%), compared with 248 frames (2.1% of observed contrails) where a contrail is seen but SAC predicts none.

SAC is a condition for formation, not for visibility, so frames with marginal ΔT_SAC may produce contrails too thin or short-lived to be seen by the camera. Other possible contributors are ambient conditions such as glare or haze, timing or position differences between the chase and test aircraft, and the fixed engine and fuel assumptions in SAC. Some examples are shown in **Fig. 7.** These results limit on how well the image-based classifier trained on these SAC labels can score.

The engines mounted on the test aircraft in the two campaigns are quite different. As noted by the authors, ND-MAX used the A320's IAE V2527-A5, a conventional rich-burn combustor, while ecoDemonstrator used the 737-10's CFM LEAP-1B with a staged lean-burn design. However, what is not noted is that there are substantial differences in the nvPM emissions between these two propulsion systems at cruise-relevant power settings. The authors consider the potential differences between SAF and conventional jet fuel and find it to be minimal, which is correct for the SAC parameters in Table 2, but the combustor difference acts through an entirely different pathway and is of a different order of magnitude (see ICAO Emissions Databank for numbers). The SAC is purely thermodynamic, so nvPM should not affect the labels derived from the atmospheric

measurements. It instead will likely affect the imagery component of the analysis, as ice crystal number and the resulting optical depth of young contrails differ systematically between the two test campaigns. By not accounting for this effect, the main discussion surrounding camera properties (which builds directly from precision/recall numbers) is inhibited and conclusions are potentially dampened.

This may be expressed in the precision/recall numbers themselves (Table 5). The rich-burn ND-MAX numbers exhibit the trend that might be expected: absence is most readily classified, short-lived slightly more challenging than that (borderline ΔT_SAC cases, variation in SAC-related parameters throughout flight that fixed assumptions will not capture), and persistent being more challenging again (RHi > 95% may be sampled as small-scale structures, such as seen in Figure 6, and difficult to accurately map to precise second-level timestamps given the response characteristics of the instrumentation and the nature of chase-aircraft sampling). The ecoDemonstrator results invert this trend, with the persistent class performing the best. Tables 6 & 7 show where the difference sits, with ND-MAX confusing short-lived and persistent at 11.8 % and 14.7 %, while ecoDemonstrator does so at 5.1 % and 5.8 %.

A plausible physical explanation could be the sensitivity of the observations to the low ice crystal numbers. Low nvPM emissions (lean-burn, ecoDem) yield fewer and larger crystals, meaning that in ice-supersaturated conditions, deposition occurs readily and the contrail brightens and spreads visibly, while particle growth is more limited in subsaturated air. The difference is less pronounced in high nvPM emission, where the deposition is spread across a greater number of particles. I suggest this as a hypothesis rather than a conclusion, however, and an explicit investigation would be required accounting for modelled ice crystal numbers and optical parameters extracted from the observations. For example, the trend could also arise from the mapping of atmospheric data from the chase aircraft to the images or associated changes in thermodynamics in the test aircraft's wake.

In either case, the attribution of performance differences to camera placement is not yet supportable as stated. Across the two campaigns, the differences in combustor technology, airframe, camera geometry and orientation, and frame extraction interval (5s/2s) all mean that no single cause can be isolated, and the recommendations in Section 4 should currently be presented as informed suggestions for future campaigns, rather than as concrete findings.

Section IV (Discussion):

The results indicate the potential of airborne cameras for detecting early contrail formation and ice supersaturation. As explained in earlier sections, for the purposes of this study the classification of contrails was simplified into three classes: absent (SAC not satisfied), short-lived (SAC satisfied but not the ice-supersaturation criterion), and persistent (both satisfied). In practice, these correspond to predictions of contrail absence, contrail presence, and contrail presence under ice-supersaturated conditions, respectively. Integration with satellite imagery to determine the duration of the contrail would help establish the actual likelihood of persistence.

For the ND-MAX campaign, the absent class was predicted most accurately, followed by the short-lived class. The persistent class was the most difficult to separate, with most misclassifications assigned to the short-lived class. This is not surprising, since the determined ice-supersaturated regions can be small, as seen in Fig.~\ref{sac2} and therefore challenging to capture with indirect timestamps, the response rate of the instrumentation and the chase-aircraft sampling configuration.

The trend is inverted for the ecoDemonstrator campaign, where the persistent class performed best. One possible explanation for the inversion is the sensitivity of the observations to ice crystal number. The ecoDemonstrator's lean-burn combustor produces lower nvPM emissions, and therefore also fewer and larger ice crystals.

In turn, that makes the contrail brighter in ice supersaturated air conditions, while limiting crystal growth in subsaturated ones. At higher ice crystal numbers, the available water vapor is shared among more particles, potentially weakening this contrast. This is a hypothesis that would require testing with modeled ice crystal numbers and optical parameters extracted from the observations.

The contrasting results of the two campaigns may be attributed to several factors, none of which can be isolated with the present data. As discussed above, the differences in aircraft and engines are one significant possible factor. The amount of days and hours of data from each campaign varies, as well, with only 4 days from ecoDemonstrator footage compared to 7 days from ND-MAX. Interestingly despite the additional data, similar class counts for persistence and 2 versus 5 second image intervals, ecoDemonstrator performed better in persistence. Differences in measurement acquisition may also contribute.

Camera orientation is another possible factor. The ecoDemonstrator camera captured only a short segment of the exhaust plume close to the engine. Since the distance behind the engine at which contrails form varies, some contrails could have been outside the camera's FOV. Fig. 6 shows example images from the two campaigns. In the ND-MAX footage, the aircraft's tail obstructs part of the contrail at high roll angles because of the camera placement. The ecoDemonstrator camera is mounted lower and angled outward, which avoids this issue. In addition, a significant part of the fuselage and wing occupies the ND-MAX frame, whereas the outward orientation of the ecoDemonstrator camera minimizes intrusion from the airframe. However, the ecoDemonstrator view shows a considerably shorter segment of the contrail, whereas the ND-MAX footage extends several kilometers behind the aircraft in favorable conditions, and the length of a contrail can make it easier to distinguish from natural clouds. A higher or lower camera placement may prevent the tail from hiding the contrail at high roll angles. The loss of images to sun glare might be reduced by placing the camera high and pointing it downward, or by tinting the upper half of the FOV. An additional camera on the opposite side may help capture cases in which glare is too intense in one side's footage. Because the campaigns differ in several respects at once, these are suggestions for future campaigns rather than findings of this study.

One experiment that would help considerably (and appears feasible with the data already at-hand) would be to train on one campaign and evaluate on the other. Generalisation across the campaigns would probe a variety of experimental parameters, and any drop-off in performance would itself be informative about how transferable such a classifier is between platforms, something which matters directly for the fleet-wide deployment the manuscript proposes and that I would personally like to see happen for the context-based reasons above.

We have began the experiment but the results need further verification, thus it was left out of the paper. It is a very helpful suggestion and we will be using it in our future work.

Overall, my main requests are a reframing of the intended application, acknowledgement and discussion of the propulsion differences between campaigns, and more cautious attribution of the performance differences. I wish to say that the effort in assembling and processing these datasets, particularly the imagery, is apparent and highly commendable. I hope these comments are useful to the authors and I look forward to seeing a revised version.

We are grateful for the feedback, we have rephrased and reframed the goal of the paper throughout all sections of the paper, we have discussed the propulsive differences as explained, and elaborated on possible impactful factors to performance aside from camera orientation. Additionally, the SAC results were compared to camera observations from the ND-MAX and presented in the results.