Dear Reviewers,

Thank you for your comprehensive and thoughtful review of our manuscript. We appreciate the time you invested in providing feedback, which has been very helpful in improving the overall clarity and quality of this research.

We have carefully considered all the points raised and have incorporated your suggestions into the revised manuscript. We believe these changes have significantly strengthened the paper and addressed the areas requiring further refinement. Our point-by-point responses are provided in the following, with referee comments in blue, our responses in black and manuscript changes in grey.

Response to Anonymous Referee #1:

This paper introduces a new computer-vision method (COCOS) to automatically recognize contrails in satellite images captured by the MSG SEVIRI instrument. In contrast to many recently introduced methods for the detection of contrails in images, it does not explicitly consist of neural networks. This approach is motivated by the idea that it leads to an interpretable, adaptable detection algorithm that can be adjusted to new satellite instruments in the near-future. The improvement in performance, as evaluated using a dataset consisting of manually annotated satellite images, is considerable when compared to previous methods for contrail detection in MSG SEVIRI images

The paper is certainly within the scope of AMT, it is very well-written and it was an enjoyable read: I have only few line-by-line comments, which are given in a separate section below. I do have some more general comments. Once these are considered by the authors, I think the paper is ready to be published.

We thank the reviewer for their positive feedback. We are pleased that the reviewer found the manuscript well-written and an enjoyable read.

My first comment concerns section 2.4 “Contrail Ground Truth” which describes the dataset of annotated MSG images used for developing and evaluating the detection algorithm. After reading this section, I still have many remaining questions:

Since the submission of our manuscript to AMT, a paper on this annotated dataset has been published in ESSD (Santos Gabriel et al., 2026). In this publication the labeling process and properties of the dataset are described in detail. We have added a citation to the revised manuscript in section 2.4. We hope to highlight this other paper by marking them as companion papers.

- Broadly speaking, how were labelers instructed to identify contrails? Was this in any way similar to the instructions used by (Meijer et al., 2022) and (Ng et al., 2024)?

Labelers were all atmospheric scientists, so no special training process was done. They did, however, receive a labeling guide describing the basic properties of contrails in satellite imagery. Labelers were then asked to mark each individual pixel belonging to a contrail if it was at least 10 pixels in size. The following has been added to the manuscript in line 175:

Labelers received a labeling guide on how to identify contrails in satellite imagery and were asked to mark each individual pixel of a contrail.

- How were scenes selected for labeling, and what is the resulting spatio-temporal distribution of annotated scenes?

The scenes were created by randomly selecting times available from MSG3 ranging from 2013 through 2024. They were visually inspected with higher attention to regions more likely to contain contrails. The 140 selected scenes are randomly distributed over the whole SEVIRI disk. See Santos Gabriel et al. (2026) for a plot showing the exact spatial distribution of scenes. The following change has been made to the manuscript in line 172:

Scenes are randomly distributed geographically across the SEVIRI disk, with clusters of higher density over Europe and the North Atlantic.

- What information (other than the image itself) was presented to the labelers? For example, both previously mentioned annotated datasets provided labelers with a sequence of images leading up to the relevant image.

Images spanning one hour prior to and one hour following the timestamp of the target image were provided to account for the temporal evolution of the objects. The additional time series imagery included multiple different RGB composites and brightness temperature images. The following sentence has been added to line 176:

As additional information the labelers received a time series of different RGB composites and brightness temperature images spanning one hour prior to the target image and one hour following the target image.

For my second comment, I refer to a passage from the introduction:

By avoiding machine learning methods, the authors aim to facilitate the future adaptation of COCOS to the Meteosat Third Generation satellite. Furthermore, we gain a deeper understanding of the factors affecting contrail detection—such as why a contrail is detected or missed, which spectral channels contribute most, and which contrail characteristics are most informative, thereby strengthening the use of COCOS for assessing temporal and spatial contrail distributions, evaluating the effectiveness of mitigation strategies (such as flight rerouting), and improving the representation of aviation-induced clouds in climate models and contrail models.

I find the above motivation to avoid a machine-learning based approach sensible. However, I think that the terminology "avoiding machine learning methods" should be modified slightly, as the COCOS algorithm introduced in the paper still relies on the output of another machine learning approach: CiPS.

The terminology was modified in line 66 of the manuscript:

By prioritizing a physically based, rule-driven logic over an end-to-end machine learning architecture, the authors aim to facilitate the future adaptation of COCOS to the Meteosat Third Generation satellite.

Secondly, I think that the paper's contribution to a "Deeper understanding of the factors affecting contrail detection" could be expanded. Although the analysis of the algorithm's performance (and its dependency on background/contrail properties) is clear, I would be very interested in learning what parts of /tests within the algorithm contribute mostly to the improvement in performance compared to previous MSG SEVIRI algorithms. For example, suppose the CiPS output is a very important input to COCOS: this knowledge could be used in the design of new contrail detection approaches.

We have expanded the analysis of our algorithm's performance and have performed an ablation study excluding individual parts of the algorithm to see the contribution of each evaluation step as suggested by the reviewer. This analysis is described in a new section '4.4 Contribution of individual evaluation steps' in the revised manuscript.

Line-by-line comments

- Line 26, it may of interest to the authors to also cite (Sonabend-W et al., 2025)

We thank the reviewer for making us aware of this recently published paper. We have added a citation in the manuscript in line 27.

- Line 87, "This also ensures that contrail detection remains independent of the time of the day" Would it perhaps be better to state this as "remains possible throughout the day", since variations in scene properties may still occur throughout the day which could possibly affect contrail detection performance?

Yes. This proposed phrasing is better to avoid confusion of the reader, as scene properties might still vary throughout the day. The phrase has been changed accordingly in the manuscript.

- Line 89: there seems to be a "8" missing here

Thank you for noticing this missing number. We have added it to the revised manuscript.

- Line 116: Missing a space between "contrails" and "("

The missing space has been added.

- Line 222: Note that "gaussian weighted" is spelled differently here than for example line 220

We have checked every occurrence of "gaussian weighted" in the manuscript and have corrected all to be consistent ("Gaussian weighted")

- Line 415: Seems like there is an additional space at the start of this sentence.

The additional space at the start of this sentence has been removed.

- Figure 8: I would strongly consider replacing the pie charts with other means of visualization. For example, the distribution of "linearity" and "ice optical thickness" could be visualized with CDFs/PDFs.

We have changed figure 8. Linearity and ice optical thickness are now visualized with PDFs where each interval used for performance evaluation in figure 9 is shown by the colored area under the curve. The two categorical properties, surface types and underlying clouds are visualized as bar charts.

Response to Referee #2 Andrew Heymsfield:

I read carefully through this interesting article. I have some suggestions.

We thank the reviewer for his feedback. We are pleased that he found the manuscript interesting.

1. Contrails form at temperatures below about -40C, unless there are some extenuating circumstances. "Hole Punch Clouds". I suggest using temperature as another way of checking your results.

While using temperature as another method of validating contrail detections might seem like an intuitive method, the practical implementation is significantly hampered by vertical uncertainty. In geostationary satellite imagery like MSG, pinpointing the exact altitude of a contrail remains a complex challenge, e.g. the cirrus detection algorithm CiPS has an uncertainty of 10 % in estimating cloud top height. Because the specific height of these contrails is so uncertain, relying on temperature as a validation metric introduces more noise than clarity. Furthermore, it takes some time (probably 30-60 min or even longer) for contrails to become visible in MSG observations. Thus, we detect persistent contrails that have lived already for some time and whose age is unknown. In principle, they could have formed significant fall streaks over a range of > 1000 m below the contrail core that also contribute to the brightness temperature observations and height determination. The resulting altitude could thus correspond to temperatures higher than the Schmidt-Appleman temperature necessary for contrail formation. Consequently, we do not utilize temperature-based verification at this point.

2. It would be interesting to use ERA5 relative humidity/temperature data to see how well those fit with your analysis

While the inclusion of ERA5 humidity data might theoretically offer a physical basis for validating detected contrails, its utility is severely limited by two primary hurdles. First, the above-mentioned issue of vertical uncertainty plays a critical role. Second, the systematic underestimation of humidity in the upper troposphere within the ERA5 dataset presents a significant reliability gap (Gierens et al., 2020). Correcting for these dry biases requires complex, non-linear post-processing and calibration against observational data, such as radiosondes or aircraft sensors (see e.g. Teoh et al. 2024, Wolf et al. 2025, Wang et al. 2025). Given these two uncertainties, using ERA5 humidity as a definitive validation metric would introduce unacceptable levels of error. Consequently, we do not use this data from our current validation framework. Nevertheless, we are planning an investigation where we use detected contrails from COCOS and MSG to evaluate ice supersaturated regions in the ICON model of the German Weather Service where a 2-moment ice cloud microphysical scheme is used (see Hanst et al. 2025, Schumann and Seifert 2025) that allows for supersaturation with respect to ice.

3. EarthCare lidar data will be helpful in the future to further quantify how well your method fits the observations. You did use the CALIOP data in quantifying your results and the EarthCare lidar data can be used in the future.

We thank the reviewer for the interesting suggestion on possible future synergies with EarthCARE data. The following has been added to the manuscript in line 474:

In the future, high-resolution lidar data from the EarthCARE mission may prove helpful in addressing these gaps, providing the precise vertical profiling necessary to quantify the detection accuracy and physical characteristics of these features with far greater certainty.

Also, we would like to clarify that CALIOP data was not used to validate the results of this study. The confusion likely stems from our description of the CiPS retrieval (Section 2.2), which utilized CALIOP data during its own internal development and validation. In our study, we use this pre-existing retrieval as a tool, but our actual validation was performed against the human labeled dataset (Section 2.4).

References:

Gierens, K., Matthes, S., and Rohs, S.: How Well Can Persistent Contrails Be Predicted? Aerospace, *7*(12), 169, https://doi.org/10.3390/aerospace7120169, 2020.

Hanst, M., Köhler, C. G., Seifert, A., and Schlemmer, L.: Predicting ice supersaturation for contrail avoidance: ensemble forecasting using ICON with two-moment ice microphysics, Atmos. Chem. Phys., 25, 17253–17274, https://doi.org/10.5194/acp-25-17253-2025, 2025.

Meijer, V. R., Kulik, L., Eastham, S. D., Allroggen, F., Speth, R. L., Karaman, S., and Barret, S. R. H.: Contrail coverage over the United States before and during the COVID-19 pandemic, Environ. Res. Lett., 17, 034039, https://doi.org/ 10.1088/1748-9326/ac26f0, 2022.

Ng, J. Y. H., McCloskey, K., Cui, J., Meijer, V. R., Brand, E., Sarna, A., Goyal, N., Van Arsdale, C., Geraedts, S.: Contrail detection on GOES-16 ABI with the OpenContrails dataset. IEEE Transactions on Geoscience and Remote Sensing, *62*, 1-14, 2024.

Santos Gabriel, V., Bugliaro, L., Montag, M., Ries, S., Wang, Z., Widmaier, K., Arico, M., Unterstrasser, S., Mayer, J., Menekay, D., Marsing, A., de la Torre Castro, E., Megill, L., Scheibe, M., and Voigt, C.: A manually labeled contrail dataset from MSG/SEVIRI, Earth Syst. Sci. Data, 18, 2397–2412, https://doi.org/10.5194/essd-18-2397-2026, 2026.

Schumann, U. and Seifert, A.: On the Weather Impact of Contrails: New Insights from Coupled ICON–CoCiP Simulations, Atmos. Chem. Phys., 25, 18571–18597, https://doi.org/10.5194/acp-25-18571-2025, 2025.

Sonabend-W, A., Geraedts, S., Goyal, N., Ng, J. Y.-H., Van Arsdale, C., and McCloskey, K.: Observing long-lived longwave contrail forcing, Atmos. Meas. Tech., 19, 1951–1972, https://doi.org/10.5194/amt-19-1951-2026, 2026.

Teoh, R., Engberg, Z., Schumann, U., Voigt, C., Shapiro, M., Rohs, S., and Stettler, M. E.: Global aviation contrail climate effects from 2019 to 2021. Atmospheric Chemistry and Physics, 24(10), 6071-6093, 2024.

Wang, Z., Bugliaro, L., Gierens, K., Hegglin, M. I., Rohs, S., Petzold, A., Kaufmann, S., and Voigt, C.: Machine learning for improvement of upper-tropospheric relative humidity in ERA5 weather model data, Atmos. Chem. Phys., 25, 2845–2861, https://doi.org/10.5194/acp-25-2845-2025, 2025.

Wolf, K., Bellouin, N., Boucher, O., Rohs, S., and Li, Y.: Correction of ERA5 temperature and relative humidity biases by bivariate quantile mapping for contrail formation analysis, Atmos. Chem. Phys., 25, 157–181, https://doi.org/10.5194/acp-25-157-2025, 2025.