

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:

The authors present a manually labeled contrails dataset derived from Meteosat Second Generation SEVIRI imagery. The manuscript describes the labelers as following established best practices from the literature for their labeling campaign such as having multiple labelers annotate each scene and providing temporal context to the labelers. Additionally the authors set a nice precedent, being the first group to additionally collate and provide auxiliary data (CiPS and ProPs, Ash color scheme, reflectances/BTs, land-sea mask, surface altitude) for each scene that is useful for further scientific/ML analysis. The manuscript is well written, clear and comprehensive in its documentation of the dataset. The dataset is of high quality and fills an unmet need of contrail labels in geostationary imagery over the MSG domain, and represents a substantial amount of expert manual labor that would be nontrivial to reproduce. I had no trouble reading the article nor downloading and using the dataset based on it.

We thank the reviewer for these encouraging comments and for their positive feedback on our work. We are pleased that the reviewer found the manuscript well-written and our dataset to be of significant importance to the field.

There are however two major technical corrections needed in the data that must be resolved before acceptance:

1. CiPS fields are missing from at least 8 scenes (~6% of scenes): `scene_025.nc`, `scene_026.nc`, `scene_027.nc`, `scene_041.nc`, `scene_096.nc`, `scene_097.nc`, `scene_098.nc`, `scene_099.nc`.

We are very grateful to the reviewer for pointing out the missing data points in specific files. We have added CiPS fields for the affected files and verified the completeness of the dataset. Thank you for helping us ensure the accuracy of this work.

2. Rotation/flipping appears to be afflicting individual SEVIRI channels provided with the scenes, somewhat randomly but upon investigating a sampling it appears to be affecting what looks like possibly more than half of scenes, especially in the reflectances/NIR/WV_062. Perhaps these auxiliary data were generated by an ML dataset pipeline with rotation/flipping training augmentations accidentally left turned on?

We would like to sincerely thank the reviewer for their meticulous eye in spotting the inconsistencies in the orientation. Upon further investigation, we identified a bug in the pre-processing script that affected a subset of the images. We have since re-processed and verified the entire dataset to ensure consistency. We have uploaded a new version of this dataset to Zenodo and are now citing v2 of this dataset in the manuscript.

Minor issues to also address please:

Data:

1. Providing the data in units of 'pixels' is convenient in some regards but can make it challenging for users of the dataset to report things like contrail lengths/widths/flux in standard SI units like kilometers or Watts/meter². Consider providing an auxiliary variable for each scene that provides the area of each pixel in the scenes, or code to calculate the same. For example Appendix B of <https://egusphere.copernicus.org/preprints/2025/egusphere-2025-3739/> details a method to do so using an approximation of each pixel as a parallelogram.

We have updated the dataset to replace the previous integer pixel indices with projected meter coordinates. The x and y axes now represent the physical distance from the sub-satellite point in the native geostationary projection, rather than simple array positions. A comment on this was added in the manuscript in line 283 as well as in the ReadMe in the dataset:

The Dataset contains two spatial grids: x, y for the native 3-km SEVIRI resolution, and x_highres,y_highres for the corresponding 1-km high-resolution of the HRV channel. The coordinate units are expressed in meters, representing the scanning displacement relative to the sub-satellite point.

2. In the scenes missing CiPS fields, the variables actually contain -32767.0 despite the NetCDF metadata declaring the `_FillValue` as NaN

Thank you for making us aware of this. The fill value should normally be NaN. This was corrected. The CiPS fields for all missing files have been added and now the data should be complete and not contain any NaNs.

3. Appendix A lists variable names as for example IR108 but files use for example IR_108

We have changed variable names to be consistent on all tables in the paper as well as the dataset.

Manuscript:

1. Because the dataset spans 11 years it could theoretically be susceptible to long-term calibration inhomogeneities ('sensor drift'), which in scientific analysis might erroneously show trends over time that were in fact entirely due to sensor aging. I was not able to find any statistically significant evidence of sensor drift in the

dataset. Nevertheless, authors are advised to please see section 2.1 for details on the inter-calibration of SEVIRI (<https://essd.copernicus.org/articles/15/5153/2023/essd-15-5153-2023.html>) and affirm explicitly in the manuscript whether the shortwave auxiliary variables were derived from inter-calibrated reprocessed data ("Climate Data Record") or from archived realtime-calibrated data. I believe this applies to just the reflectances, since CiPS and ProPs are longwave only and in SEVIRI longwave channels are generally less susceptible to sensor drift.

Since the contrail observation in satellite imagery relies almost exclusively on SEVIRI longwave channels which have on-board calibration and thus are less susceptible to sensor drift, we did not use any calibration method as in Benas et al. (2023) but used the real-time calibrated data as extracted from the original HRIT MSG/SEVIRI files. This leaves to the user the possibility to change/improve calibration as if he/she was reading the original files. The following has been added to the manuscript in section 2.1 line 85:

Real-time calibrated data for all SEVIRI channels was used in this work.

2. There are a few places discussing the subjectivity of labels based on labeler (dis)agreement, where the statements appear to apply to all contrail labeling from satellite imagery; please be sure to avoid misleading readers by scoping these statements to being about the specific imager. Consider reviewing <https://amt.copernicus.org/articles/18/1115/2025/> which recently analyzed the effect of pixel size on detectability of contrails, and reports that detectable contrails contribute (70 ± 2) % of the net radiative forcing in simulated 2 km resolution images.

We reviewed the publication and added a consideration in the Introduction (Line 27):

Despite this coverage, detecting individual contrails remains challenging; Driver et al. (2025) demonstrate that only (46 ± 2) % of a modeled population of global contrail segments (in clear sky conditions) is observable using current 2 km resolution imager data. Geostationary contrail observation still remains valuable as these detectable segments account for a disproportionately high (72 ± 2) % of the total instantaneous long-wave radiative forcing.

Also, all statements in the manuscript concerning the subjectivity of labeling were scoped to the specific imager.

3. Line 20: the cited studies are simulation studies, not "Observational studies"

We corrected the sentence to say simulation studies instead of observational studies.

4. Line 34: starting the sentence with "For instance" implies that Ng et al 2024 is not a publicly available dataset, but it is actually publicly available.

The sentence was rephrased to be clearer:

A notable exception is Ng et al. (2024), who introduced the publicly available OpenContrails dataset, which ...

5. Line 88: is ECMWF really at 6 hour timesteps? I'm familiar with ERA5 that is either 1hr (nominal) or 3hr (ensemble); may help to cite the specific ECMWF product used

The skin temperature from ERA5 has been collected in the course of the last few years. For 2013 to 2018 the data was interpolated at ECMWF to a spatial grid with $0.125^\circ \times 0.125^\circ$ in latitude-longitude and output with a temporal resolution of 6 h. For 2023-2024 we used a spatial resolution of $0.25^\circ \times 0.25^\circ$ in latitude-longitude and a temporal resolution of 3 h from Copernicus. Data is first interpolated linearly in space and time to the MSG/SEVIRI grid and eventually used in the CiPS and ProPS retrievals.

6. Line 130: citation (and ideally quantification) needed for the assertion that contrails are "mostly identified as thin ice clouds" - many geostationary satellite cloud products can misclassify thin cirrus as clear sky (or water cloud if they are over a lower cloud)

The assertion that contrails are mainly identified as thin ice clouds refers to the ProPS algorithm, where we compared the contrail ground truth to the outputs of the ProPS algorithm. The point of the sentence was rather that contrail detection is more efficient than the cloud retrievals algorithms w.r.t. contrails. We have made that more clear and quantified this in the manuscript:

Of course, a part of the contrails is already considered in this statistic, as they are partly identified as thin ice clouds, namely 66.21 % of contrails in the ground truth are identified as thin ice clouds by ProPS.

7. Table 1: are these pixel-area-weighted frequencies or just pixel count frequencies?

The frequencies in Table 1 are pixel-count frequencies. This clarification has been added to the table description.

8. Line 164: "object-wise agreement" is first mentioned here but the algorithm for converting contrail pixel masks to linear objects is not described until section 4.2.1

We thank the reviewer for pointing this out. For further clarity, we have switched sections 4.1 and 4.2.

9. Section 4.1: please explicitly mention that while other labeling campaigns (such as Ng et al 2024) have provided labelers with some rasterizations of advected flight trajectories to inform their labels, they were not provided to labelers here.

The following has been added to section 3.2:

While other labeling campaigns have provided labelers with some rasterization of advected flight trajectories to inform their labels (e.g. Ng et al., 2024), they were not provided to labelers here.

And the following to section 4.2 in the revised manuscript (formerly section 4.1):

Ng et al. (2024) observed similar precision and recall values when comparing 200 examples from individual labelers with their ground truth, 64 % and 76 % respectively. Like SEVIRI, the imagery used in their study was sourced from a geostationary platform (GOES-16), which features a slightly higher spatial resolution of 2 km for infrared channels at nadir. However, they provided their labelers with some rasterization of advected flight trajectories, which we did not.

10. Conclusion: a comment would be welcome about the size of the dataset and if the authors expect it to be useful for training deep ML models (or more likely for simpler ML models or validation/qualitative insights only?)

The following comment has been added to the Conclusion:

Given the limited size of the dataset, we expect it to be useful for training and validating contrail detection algorithms using image processing techniques and simple ML models or for validating other more complex ML models.

Response to Anonymous Referee #2:

This paper describes the development and characteristics of a dataset of manually labeled contrails in images captured by the MSG SEVIRI instrument.

The paper is extremely well written: I did not find any typographical errors. It reads well and I have very few line-by-line comments, which mostly pertain to the figures.

The study contributes to our understanding of the limitations in manually labeled datasets of contrails in satellite images, which is crucial given that these datasets are typically used to create and evaluate automated contrail detection techniques.

We thank the reviewer for this positive and encouraging feedback. We are pleased that you find the study to be a useful contribution to understanding the limitations of manually labeled contrail datasets in satellite imagery.

I also have some more general questions:

- The period over which satellite data is acquired is relatively long: are there any relevant issues related to sensor degradation over this time period, that could affect the creation of this dataset? And what is the rationale behind the decision to label images in the period 2013 – 2018, but then also between 2023 and 2024?

The main channels used for contrail observation in satellite imagery are the SEVIRI longwave channels, which have on-board calibration and thus are less susceptible to sensor drift. There are methods to address this issue for the shortwave channels (e.g. Benas et al. 2023, Meirink et al. 2013), but we did not do this as the composites for the labelers that rely on the shortwave channels were provided only for additional context but are less relevant in regards to contrails.

The rationale behind these time periods for the labeled images is to avoid potential spectral inconsistencies caused by using data from different satellites. Only MSG3 (Meteosat-10) data from the operational position above 0° E is used in this study and is available for these time periods. A more detailed description of this rationale has been added to section 2.1:

To avoid potential spectral inconsistencies caused by using data from different satellites, only data from MSG3 (Meteosat-10) from the operational position above 0°E was used. This MSG3 data ranges from January 2013 through February 2018 and from March 2023 through March 2024.

- Were the images projected in some way, to address image distortion at higher satellite viewing zenith angles? Were the locations of images that were labeled in some way limited by the viewing zenith angle?

The images were not projected in any way; they are provided to the labelers in the original satellite projection. This decision was made to preserve the native characteristics of the SEVIRI observations. Reprojection would require interpolation of the pixel values, which could introduce artificial features or smoothing effects that might influence the visual identification of thin structures such as contrails.

In addition, maintaining the native satellite projection ensures that the labeled dataset is directly compatible with the original SEVIRI data. Keeping the labels in the same projection allows the dataset to be used for development and validation of detection algorithms without introducing additional uncertainties related to reprojection.

- Why was the decision made to only consider images from the hour leading up to labeled image? And was this found to be sufficient, or would a longer period have been helpful? For example, (Meijer et al., 2022) uses 2 hours.

The hour preceding the labeled image was included to provide temporal context that can help identify the evolution and movement of contrails relative to the surrounding cloud field. This time window was chosen as a compromise between providing sufficient temporal information and limiting the data volume and complexity of the dataset. In many cases, contrails can already be followed and identified within this time due to their advection and spreading. Duda et al. (2004) found a spreading rate of 2.7 km/h. While longer time windows, such as the two hours used in Meijer et al. (2022), may provide additional context in some cases, one hour was considered sufficient for the purposes of visual interpretation and dataset construction, consistent with the approach of Ng et al. (2024), who also used a one-hour window. An explanation of this chosen time window has been added to section 3.2.

- Were the labelers trained in some way? The paper mentions a labeling guide: it would be very helpful to include this with the paper, like was done by (Meijer et al., 2022; Ng et al., 2024). Additionally, I believe Ng et al. (2024) implemented a “training process” that had labelers learn on a “gold-standard dataset” before “graduating” to labeling new images.

Our labelers were all colleagues working in atmospheric sciences and they were familiar with the topic of contrails and remote sensing. In contrast to Ng et al. (2024) who had labelers with no specialized experience in atmospheric science, we did not feel the need for any specialized training process. The labeled images were subject to in-person

exchanges and discussions. Also, labelers received a labeling guide. This contains mainly an explanation of the labeling tool and the additional time-series imagery. The main contents of the guide are described in section 3.2, so we feel that the inclusion of this guide would be redundant.

- How was labeler agreement affected by the proportion of land/sea pixels in an image? How was labeler agreement affected by the number of contrail pixels present in the image?

Labeler agreement was not affected by the proportion of land/sea pixels in the images. Also, no significant correlation between labeler agreement and number of contrail pixels present in the image was found. We have added the following to the manuscript in section 4.2 describing all properties investigated:

Beyond the metrics shown in Fig. 9, which show the highest influence on labeler agreement, we evaluated a suite of additional properties to ascertain their influence on inter-labeler consistency. With ascending length of contrails, the labeler agreement increases slightly. No correlation can be found between labeler agreement and surface type (land/sea), surface altitude, solar zenith angle or satellite elevation angle. Number of contrail pixels present in the labeled image did also not have any significant effect on labeler agreement.

- How does the size of this dataset compare to previous labeled datasets (Meijer et al. 2022, Ng et al. 2024)? It would be interesting to compare the number of labeled pixels (accounting also for the lower resolution of the SEVIRI instrument) and the fraction of contrail pixels within these datasets. The distribution of contrail pixels among images might be another interesting feature to compare.

We provide 140 labeled images with a size of 256x256 pixels. Given the resolution of the SEVIRI imager this accounts for an area of roughly 82.57 million km², where 0.44 % of this area is marked as contrails in the ground truth. Meijer et al. (2022) and Ng et al. offer a total area of 2'472 million and 5'266 million km², where Ng et al. (2024) report 1.2 % of pixels as contrails. Given the limited size of our dataset, we do not expect it to be used for training deep ML algorithms, but rather training contrail detection algorithms using image processing techniques or simple ML methods as well as validation of any contrail detection algorithms.

The following has been added to the manuscript in section 4.1.2 (of revised manuscript):

The ground truth contains a total of 772 contrails, with 40812 pixels that were annotated by at least two labelers. This shows a slightly lower labeling density of 0.44 % compared to the 1.2 % observed in Ng et al. (2024).

The following was added to the manuscript in the conclusions:

The scope of this dataset is limited by its size. While Meijer et al. (2022) and Ng et al. (2024) offer a multi-billion km² coverage, we focus on a more targeted 82.5 million km² providing critical ground truth over Europe. Given the limited size of the dataset, we expect it to be useful for training and validating contrail detection algorithms using image processing techniques and simple ML models or for validating other more complex ML models.

Another general comment is that for certain words, like “sun”, the capitalization is inconsistent across the paper. It might be worthwhile to do a check throughout the manuscript.

Thank you for making us aware of this inconsistency. We have checked the manuscript and removed this inconsistency as well as any other capitalization inconsistency.

Once my questions and line-by-line comments are addressed, I consider this paper ready for publication.

Line-by-line comments

Figure 1: It might also be helpful to include the original satellite image (like figure 5a or 5e) next to this for interpretation.

Thank you for making it aware to us that this would be helpful for interpretation. We have added two panels to figure 1 showing the Ash RGB and Natural Color RGM composites of the satellite scene.

Line 101: Why does this say: (e.g. 2021)?

Thank you for making it aware that the e.g. has been misplaced here. It is now:

CiPS has been extensively validated and characterized in Strandgren et al. (2017b) and applied in e.g. Rybka et al. (2021).

Figure 2: The legend is cut-off. It might also be helpful to include the original satellite image (like figure 5a or 5e) next to this for interpretation.

The original satellite image (Ash RGB composite) has been added to figure 2 for better interpretation of the figure.

Figure 6: The resolution of this figure is a bit too low to inspect the ash RGB images. I think it may also be helpful to combine the 3 different labels in a way that highlights their agreement and disagreement, as I now have to switch back and forth between the different labels to identify differences.

For easier inspection of the images, we have decided to only show two examples from the dataset in figure 6 with each panel being bigger than before. Also, a different visualization has been chosen for the individual labeling to better highlight their differences and similarities.

Line 178: “the own image assessment” what is meant with this?

We mean that each labeler applies his/her own visual interpretation when assessing the image. For example, the perceived color or contrast of a contrail can differ slightly between individuals, and in cases where a pixel is not clearly dark blue but also too dark to be considered clear sky, labelers may differ in whether they classify it as part of a contrail or as a background pixel. The following change was made in the manuscript:

The different labeling results are thus due to the subjective decisions taken based on the individual visual interpretation of the images and on the objective challenge in identifying contrail structures of different width, length, appearance and contrast and separating them from natural clouds.

Table 3: It is not clear to me how the different columns in this table are computed. There are 6 columns, but 14 labelers in total?

There were 14 labelers in total, but each of them only received a subset of images to label. The images were distributed amongst the labelers so that each image was labeled by three different individuals. The columns represent the precision and recall values if one assumes that one of the three labelers is the truth (reference labeler) and the others are compared to this (candidate labelers). We have added two rows in the table showcasing which labeling was used as a reference and which as a candidate to produce the individual precision and recall values. This gives the reader an insight into the subjectivity on the labelers on this dataset.

Figure 7: perhaps include horizontal axis labels for plots b) and c).

Horizontal axis labels have been added to subplots b) and c) of this figure (figure 9 in revised manuscript).

Figure 8c): abbreviation “PC” not defined.

A definition of the abbreviation “PC” has been added to the figure caption.

Line 370: “as global attribute” -> “as a global attribute”

This was changed in the manuscript in line 292.

Table A1:

- “timestamp” is in UTC?

The timestamp of the labeled image is in UTC. We have added this information to the table.

- row “native_coordinates”: “boundary box” -> “bounding box”

This was changed in the table.

Table A2:

- First row: “Consensus” should this be capitalized?

The word should not be capitalized. Thank you for making us aware. This was changed in the revised version of the manuscript.

- “acquisition_time”: First letter in “Description” not capitalized

The letter was capitalized in the revised version of the manuscript.

Response to Anonymous Referee #3:

Gabriel et al. present a manually labeled dataset of aircraft-induced contrails derived from MSG/SEVIRI imagery. Given the radiative importance of contrails, particularly in the context of quantifying aviation-related anthropogenic climate forcing, this dataset represents a valuable contribution. The provision of a carefully constructed ground truth is especially relevant for the development and evaluation of machine learning–based contrail detection algorithms.

The manuscript is clearly written, and the methodology for contrail identification and labeling is well described. The inclusion of auxiliary variables improve the usefulness of the dataset. The data structure and metadata are generally well documented in both the text and the accompanying tables.

We appreciate your careful reading of the manuscript and your recognition of the value of the dataset, particularly in the context of studying aviation-induced climate effects and supporting the development and evaluation of contrail detection algorithms. We are also grateful for your comments on the clarity of the manuscript and the description of the methodology, as well as your appreciation of the inclusion of auxiliary variables and the documentation of the data structure and metadata. Your feedback is greatly appreciated and helps confirm that the dataset and its description are clear and useful for the community.

I have only a few minor comments that should be addressed before the manuscript can be considered for publication in ESSD:

Line 80: It may be worthwhile sharing why the considered dataset has a gap of 5 years between 2018 and 2023.

The reason for this gap for the labeled images is to avoid potential spectral inconsistencies caused by using data from different satellites. Only MSG3 (Meteosat-10) data from the operational position above 0° E is used in this study and is available for these time periods. Between the years 2018 and 2023 the operational satellite above 0°E was MSG4 (Meteosat-11). A more detailed description of this rationale has been added to section 2.1:

To avoid potential spectral inconsistencies caused by using data from different satellites, only data from MSG3 (Meteosat-10) from the operational position above 0°E was used. This MSG3 data ranges from January 2013 through February 2018 and from March 2023 through March 2024.

Line 122: The manuscript states that approximately 40% of scenes contain no visible contrails and 60% were selected to include contrails. Was this segregation intentional? The

wording suggests that it was. If so, this should be clarified, including the rationale for this choice.

The segregation of scenes into those containing contrails (60%) and those without (40%) was an intentional design choice aimed at enhancing the robustness of the dataset. This ratio was selected to provide a balanced training signal: the 60% majority ensures the model learns the diverse morphology of contrails, while the 40% 'null' scenes provide essential negative samples to reduce false-positive detections in clear-sky or natural-cloud conditions. We have added a more detailed explanation of this choice in the manuscript:

Around 40 % of the scenes were selected with no visible contrails, the other 60 % were selected to include contrails based on the personal judgment of one human individual. This balance enhances the robustness of the dataset and guarantees that both contrail and non-contrail cases are represented: the 60 % majority ensures that a contrail detection model learns the diverse morphology of contrails, while the 40 % 'null' scenes provide essential negative samples to reduce false-positive detections in clear-sky or natural-cloud conditions.

Line 124: The manuscript notes that the selected 140 images are not uniformly distributed in space and time. This suggests that the dataset is not intended to represent real-world contrail climatology. In addition, the relatively small sample size indicates that the primary application of the dataset is likely to be the development and validation of contrail detection algorithms (e.g., ML-based approaches), rather than for statistical analyses of contrail occurrence or detailed studies of meteorological drivers. I suggest clearly highlighting these points in the manuscript.

The expected primary application of the dataset is indeed the development of automated contrail detection algorithms and investigations of factors influencing detectability. A comment on purpose and expected use can be found in the conclusion section and has been extended in the revised version of the manuscript:

The dataset can be used for a variety of purposes. It can serve as a benchmark for developing and testing automatic contrail detection methods, as well as for studying contrail characteristics such as length, width, and orientation. The auxiliary cloud information makes it possible to look at cloud–contrail interactions and environmental factors that affect contrail detectability, which is important for understanding their radiative impact. Scene-level and pixel-level metadata make it easier to filter and analyze the data, which is useful both for machine learning approaches and more traditional observational studies. The scope of this dataset is limited by the size of the dataset. While Meijer et al. (2022) and Ng et al. (2024) offer a multi-billion km² coverage, we focus on a more targeted 82.5 million km² providing critical ground truth over Europe. Given the limited size of the dataset, we expect it to be useful for training and validating contrail detection algorithms

using image processing techniques and simple ML models or for validating other more complex ML models.

Table 3 is not discussed in the manuscript text. Also, it is not clear what the columns represent.

We discuss the content of Table 3 in section 4.2, line 227 onwards (in the revised manuscript). We added a reference to Table 3 to make it apparent that we are discussing the results portrayed in this table. The columns represent the precision and recall values if one assumes that one of the three labelers is the truth (reference labeler) and the others are compared to this (candidate labelers). We have added two rows in the table showcasing which labeling was used as a reference and which as a candidate to produce the individual precision and recall values. This gives the reader an insight into the subjectivity on the labelers on this dataset.

Figure 8 presents contrail length and width distributions in pixel units. Since pixel size varies with viewing geometry, it would be more physically meaningful to express these quantities in distance units (e.g., km).

The quantities length and width are displayed in km in the revised manuscript (figure 7 in the revised manuscript).

Line 255: It is worth noting here that the auxiliary parameters related to cloud optical properties are only available during daytime.

Outputs from the CiPS algorithm describing cloud properties, such as cloud top height or ice optical thickness, as well as outputs from the ProPS algorithm, are available both during day and night because they rely exclusively on infrared channels. Only the three SEVIRI channels HRV, VIS006, VIS008, and NIR016, which are provided in reflectance, are limited to daytime conditions.

Changes to the manuscript in line 255 in revised manuscript:

Satellite data, processed by the authors of this paper, for all 12 SEVIRI channels is provided in the NetCDF files as well. The four SEVIRI channels HRV, VIS006, VIS008, and NIR016, which are provided in reflectance, are limited to daytime conditions.

Changes to the manuscript in line 280 in revised manuscript:

Outputs from the CiPS algorithm describing cloud properties, as well as outputs from the ProPS algorithm, are available both during day and night because they rely exclusively on infrared channels.

References

- Benas, N., Solodovnik, I., Stengel, M., Hüser, I., Karlsson, K. G., Håkansson, N., ... & Meirink, J. F. (2023). CLAAS-3: the third edition of the CM SAF cloud data record based on SEVIRI observations. *Earth System Science Data*, 15(11), 5153-5170.
- Driver, O. G., Stettler, M. E., & Gryspeerd, E. (2025). Factors limiting contrail detection in satellite imagery. *Atmospheric Measurement Techniques*, 18(5), 1115-1134.
- Duda, D. P., Minnis, P., Nguyen, L., & Palikonda, R. (2004). A case study of the development of contrail clusters over the Great Lakes. *Journal of the atmospheric sciences*, 61(10), 1132-1146.
- Meijer, V. R., Kulik, L., Eastham, S. D., Allroggen, F., Speth, R. L., Karaman, S., & Barrett, S. R. (2022). Contrail coverage over the United States before and during the COVID-19 pandemic. *Environmental Research Letters*, 17(3), 034039.
- Meirink, J. F., Roebeling, R. A., & Stammes, P. (2013). Inter-calibration of polar imager solar channels using SEVIRI. *Atmospheric Measurement Techniques*, 6(9), 2495-2508.
- Ng, J. Y. H., McCloskey, K., Cui, J., Meijer, V. R., Brand, E., Sarna, A., ... & Geraedts, S. (2023). Contrail detection on GOES-16 ABI with the OpenContrails dataset. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-14.
- Rybka, H., Burkhardt, U., Köhler, M., Arka, I., Bugliaro, L., Görsdorf, U., ... & Strandgren, J. (2021). The behavior of high-CAPE (convective available potential energy) summer convection in large-domain large-eddy simulations with ICON. *Atmospheric chemistry and physics*, 21(6), 4285-4318.
- Strandgren, J., Fricker, J., & Bugliaro, L. (2017). Characterisation of the artificial neural network CiPS for cirrus cloud remote sensing with MSG/SEVIRI. *Atmospheric Measurement Techniques*, 10(11), 4317-4339.