A Manually Labeled Contrail Dataset from MSG/SEVIRI

Vanessa Santos Gabriel[1], Luca Bugliaro[2], Mara Montag[1], Sabrina Ries[1], Ziming Wang[1], Kai Widmaier[2], Matteo Arico[2], Simon Unterstrasser[2], Johanna Mayer[2,3], Deniz Menekay[2], Andreas Marsing[2], Elena de la Torre Castro[2,4], Liam Megill[2,5], Monika Scheibe[2], and Christiane Voigt[1,2]

[1]Institute of Atmospheric Physics, Johannes Gutenberg University, Mainz, Germany
[2]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[3]now at: European Space Agency, Frascati, Italy
[4]now at: GE Aerospace, Garching, Germany
[5]Delft University of Technology, Faculty of Aerospace Engineering, Section Operations & Environment, Delft, The Netherlands

Correspondence: Vanessa Santos Gabriel (vsantosg@uni-mainz.de)

Abstract. Contrails — thin ice clouds formed by aircraft — are a major contributor to aviation-induced climate forcing, yet their observational characterization remains limited. We present a manually labeled contrail dataset derived from observations of the Meteosat Second Generation (MSG) SEVIRI instrument over Europe and the North Atlantic, comprising 140 scenes of 256×256 pixels. Each scene was independently annotated by three labelers, with ground truth established via majority consensus. To provide additional context, the dataset includes outputs from two satellite retrievals: CiPS (Cirrus Properties from SEVIRI) and ProPS (Probabilistic Cloud Top Phase retrieval), offering information on cloud cover and cloud top phase, cirrus probability, ice optical thickness, and ice cloud top height. These complementary variables enable detailed investigations, such as factors influencing contrail visibility. The dataset supports analyses of contrail detection, contrail characteristics, cloud-contrail interactions, and environmental conditions affecting detection. By providing high-quality labeled data with auxiliary cloud information, this resource facilitates the development and evaluation of contrail studies, contributes to improved understanding of aviation-related cloud effects, and informs strategies for climate impact mitigation. The full dataset is available under: https://doi.org/10.5281/zenodo.17669444.

1 Introduction

Contrails are narrow ice clouds formed by aircraft when ambient temperatures are low enough. Under favorable conditions (when air is supersaturated with respect to ice), contrails can persist for hours and evolve into extensive cirrus clouds, so called contrail cirrus, contributing to aviation's impact on the climate (Lewellen, 2014; Vázquez-Navarro et al., 2015). Aviation accounts for roughly 3.5 % of anthropogenic effective radiative forcing (ERF), with contrail cirrus being the largest contributor, exceeding the radiative impact of CO_2 emissions from aircraft (Lee et al., 2021). Observational studies indicate that a small number of flights can produce disproportionately large contrail outbreaks, highlighting the importance of understanding which flights and atmospheric conditions lead to high-impact events (Burkhardt et al., 2018; Teoh et al., 2024). Despite their climatic relevance, observational knowledge of contrail properties, formation conditions, and evolution remains limited (Mannstein

et al., 1999; Meyer et al., 2007; Bedka et al., 2013; Minnis et al., 2013; Spangenberg et al., 2013; Meijer et al., 2022). Passive imaging from geostationary satellites provides a critical resource, offering continuous spatial and temporal coverage to monitor contrails and estimate their radiative impact (e.g., Mannstein and Schumann, 2005; Chevallier et al., 2023; Wang et al.,
25 2023, 2024; Vázquez-Navarro et al., 2015; Haywood et al., 2009; Atlas et al., 2006). However, the development and evaluation of contrail studies are constrained by the lack of high-quality labeled datasets that reliably indicate where contrails are present in satellite imagery. Manually labeled datasets serve as an essential foundation, enabling the assessment of contrail detection algorithm accuracy, the evaluation of contrail properties, and investigation of their visibility under different atmospheric conditions. Synthetic and model-based contrail datasets (e.g. Chevallier et al., 2023) present an alternative to human
30 labeled contrail datasets as this approach can help to build large-scale datasets with exact ground truth. However, the model-based approach uses simplified physics and therefore fails to simulate complex interactions with natural cirrus, overlapping contrails and variable background clouds. Exactly for those hard cases, human labelers are needed. Several efforts have sought to derive such human labeled data, but most of them are not publicly available (e.g. Mannstein et al., 1999, 2010; Dekoutsidis et al., 2023). For instance, Ng et al. (2024) introduced the OpenContrails dataset, which offers manually annotated contrails
35 over North and South America using geostationary satellite imagery from the GOES-16 Advanced Baseline Imager (ABI). Complementing this, McCloskey et al. (2021) developed a human-labeled Landsat-8 contrail dataset, covering a similar geographical extent but utilizing low Earth orbit (LEO) satellite imagery, and Gourgue et al. (2025) published a dataset of annotated ground-based hemispheric sky images. These open source datasets are valuable for validating automatic detection methods, for inter-comparing different detection approaches and also for studying contrail characteristics such as length, width, orientation,
40 and persistence, which are critical for understanding their formation and climatic effects. Nevertheless, current datasets remain limited in spatial and temporal coverage, particularly over regions of dense air traffic like Europe and the North Atlantic, and often lack complementary information about the surrounding cloud environment.

To address these limitations, comprehensive labeled contrail datasets with additional cloud context are required. We present a human labeled contrail dataset using the Meteosat Second Generation (MSG) satellite that includes multiple labelers' an-
45 notations to assess the inherent subjectivity in manual labeling. We incorporate auxiliary information, such as cloud phase or optical properties derived from satellite retrievals, to facilitate more detailed investigations of the factors affecting contrail detectability. High-quality labeled datasets, such as the one presented here, enable researchers to rigorously evaluate contrail detections, investigate environmental and observational influences on detectability, and support studies aimed at mitigating the climate impact of aviation through informed rerouting or operational strategies. Contrail identification in MSG is particularly
50 challenging due to the low spatial resolution of 3 km at nadir (4-6 km over Europe), compared for instance to 2 km for GOES-R/S/T or 1 km for many polar orbiting passive sensors like AVHRR and MODIS.

This paper is structured as follows: We begin by presenting the satellite data used in this study and introducing two satellite retrievals used to produce complementary information about the surrounding cloud environment in Section 2. Next, we provide a detailed description of the labeling campaign in Section 3, including the scene selection and a detailed description of the
55 labeling process. In Section 4, we discuss the results of the labeling campaign. Section 5 highlights the compiled dataset with

its contents and formats. Finally, Section 6 presents a summary of the key findings and discusses their implications for future studies on contrail detection and characterization.

2 Satellite Data and Spaceborne Cloud Retrievals

Scenes for labeling are generated from MSG satellite data. To complement the dataset, outputs from two satellite retrievals—
60 ProPS (Probabilistic Cloud Top Phase retrieval for SEVIRI) (Mayer et al., 2024a) and CiPS (Cirrus Properties from SEVIRI) (Strandgren et al., 2017a)—are also incorporated. By providing CiPS-derived cirrus properties and ProPS cloud top phase information alongside the labeled contrail dataset, the authors enable users to comprehensively examine the cloud characteristics that may enhance or obscure contrail visibility, thereby informing detection capabilities and algorithm development.

2.1 MSG/SEVIRI

65 The Meteosat Second Generation (MSG) satellites (Schmetz et al., 2002) are equipped with the Spinning Enhanced Visible and Infrared Imager (SEVIRI). The MSG program comprises four geostationary satellites, stationed at the nominal longitude of 0°E to observe Europe and Africa, though some have been operated at different longitudes or in alternative modes. SEVIRI provides imagery every 15 minutes, with a nadir spatial resolution of 3 km that degrades to about 4–6 km over Europe due to the oblique viewing geometry and Earth's curvature. The instrument observes in twelve spectral channels: four channels measure reflected
70 solar radiation and eight infrared channels capture emitted thermal radiation. One solar channel, the broadband High Resolution Visible (HRV), has a sampling distance of 1 km at nadir. Satellite instruments measure the radiance received at the sensor, which represents the intensity of radiation within a specific spectral band. However, radiance alone depends on both the illumination geometry and the spectral characteristics of the sensor, making it difficult to interpret physically. Therefore, radiances in solar (visible and near-infrared) channels are converted to apparent reflectance, a dimensionless quantity that expresses the fraction
75 of incident solar radiation reflected by the Earth–atmosphere system. For thermal infrared channels, the measured radiance originates from emitted terrestrial radiation rather than reflected sunlight. In this case, radiances are converted to brightness temperature, defined as the temperature a blackbody would have to emit the same radiance at wavelength λ. Throughout this work, we use reflectances and brightness temperatures, while SEVIRI channels are referred to by their central wavelength (e.g., IR108 denotes the 10.8 μm infrared channel). For consistency, only data from MSG3 (Meteosat-10) from the operational
80 position above 0°E was used. The data ranges from January 2013 to February 2018 and from March 2023 to March 2024.

2.2 CiPS

The CiPS (Cirrus Properties from SEVIRI) algorithm (Strandgren et al., 2017a) is designed to detect cirrus clouds and retrieve their key physical and optical properties using the SEVIRI instrument aboard the MSG satellites. CiPS employs artificial neural networks (ANNs) taking as input SEVIRI thermal infrared observations, European Centre for Medium-Range Weather
85 Forecasts (ECMWF) surface temperatures, and auxiliary data. Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) cloud products were used during training of the ANNs as a ground-truth target. The ECMWF surface temperature are spatially
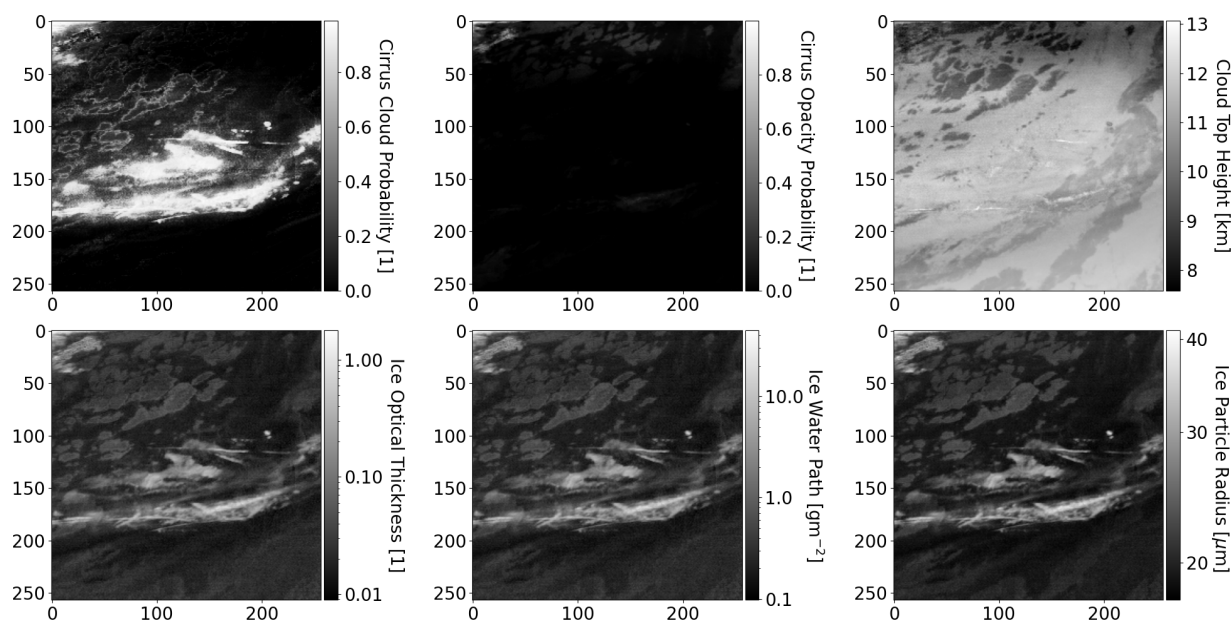
Figure 1. Example of CiPS outputs for Scene 041 of the dataset (256 x 256 pixels). The figure shows six panels in two rows, with the top row displaying (from left to right) cirrus cloud probability, cirrus opacity probability and cloud top phase, and the bottom row showing (from left to right) ice optical thickness, ice water path and ice particle radius. Contrails are located (mostly horizontal) in the central part of the image.

collocated with the satellite observations using the nearest neighbor and temporally collocated by linearly interpolating between the ECMWF 6 hour time steps. This approach allows retrievals during both day and night and provides near-continuous coverage with 15-min temporal resolution and 3 km spatial resolution at nadir. CiPS outputs include: Cirrus Cloud Probability (CCP), Cirrus Opacity Probability (COP), Cloud Top Height (CTH), Ice Optical Thickness (IOT), Ice Water Path (IWP) and Effective Ice Particle Radius (R_{eff}). In standard usage of CiPS, as recommended by Strandgren et al. (2017a), a pixel is considered a cirrus covered pixel when CCP is larger than 0.62. When COP is larger than the recommended threshold of 0.86 ice clouds are classified as thick, where CALIOP cannot provide accurate IOT/IWP retrievals. Based on the evaluation of 4.9 million CALIOP collocations, CiPS has a probability of detection for ice clouds of 71 % and 95 % of cirrus clouds with an optical thickness of 0.1 and 1, respectively. The average false alarm rate for ice clouds is 3.9 %. Identification of a thick ice cloud with COP yields a 71 % detection efficiency with a 4 % false alarm rate. CiPS has a relative error of less than 10 % regarding cloud top height and mean absolute percentage error of less than 100 % for ice optical thickness and ice water path. These outputs allow comprehensive characterization of cirrus cloud vertical structure, optical thickness and ice water path, enabling quantitative assessment of their radiative impact and life cycle evolution. CiPS outputs for one exemplary scene are shown in Fig. 1. CiPS has been extensively validated and characterized in Strandgren et al. (2017b) and applied in Rybka et al. (e.g. 2021).
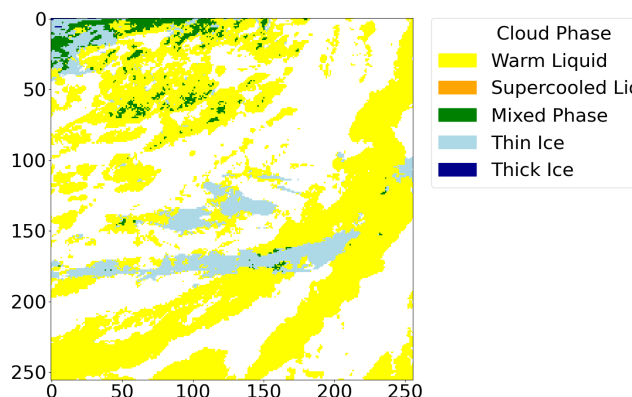
Figure 2. Example of cloud-top phases for Scene 041 of the dataset as provided by ProPS

2.3 ProPS

The ProPS (PRObabilistic cloud top Phase retrieval for SEVIRI) algorithm (Mayer et al., 2024a) is designed to detect clouds and determine their thermodynamic phase using the SEVIRI instrument aboard the Meteosat Second Generation (MSG) geosta-
tionary satellites. ProPS employs a Bayesian approach, integrating SEVIRI measurements with a large collocated dataset from the DARDAR (liDAR/raDAR) product, which provides cloud phase information from lidar–radar synergy. This methodology enables the classification of cloud states into six categories: clear sky, warm liquid (LQ), supercooled liquid (SC), mixed-phase (MP), optically thin ice (TI) and optically thick ice (IC). The algorithm operates effectively during both day and night, offering near-continuous coverage with 15 min temporal resolution and 3 km spatial resolution at nadir. ProPS outputs the most prob-
able thermodynamic phase assigned to each pixel based on the highest probability of each of the six cloud states. Validation against six months of independent DARDAR data demonstrates that ProPS achieves a daytime cloud detection rate of 93 % and a clear-sky detection rate of 86 %. For phase determination, ProPS accurately classifies 91 % of IC, 78 % of TI, 52 % of MP, 58 % of SC, and 86 % of LQ clouds. These outputs enable detailed analysis of cloud phase distributions, which are crucial for understanding cloud microphysics, radiative properties, and their role in atmospheric processes. Exemplary cloud-top phases for the same scene as in Fig. 1 are shown in Fig 2.

3 Labeling Methodology

In this section, we describe the selection of scenes for the dataset and the contrail labeling procedure.

3.1 Scene selection

The contrail scene samples were created by randomly selecting times available from MSG3 ranging from 2013 to 2024 (see Sec. 2.1 for details). These were visually inspected with higher attention to regions more likely to contain contrails like Europe

Earth System
Science
Data
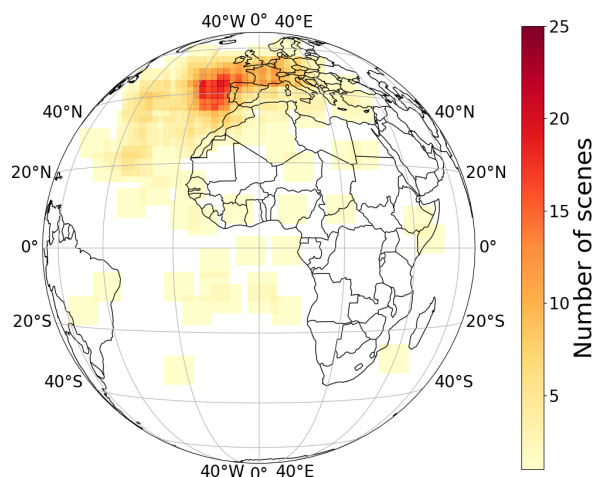
Figure 3. Spatial distribution of scenes used in labeled contrail dataset. Color values indicate how many scenes contain each individual pixel location.
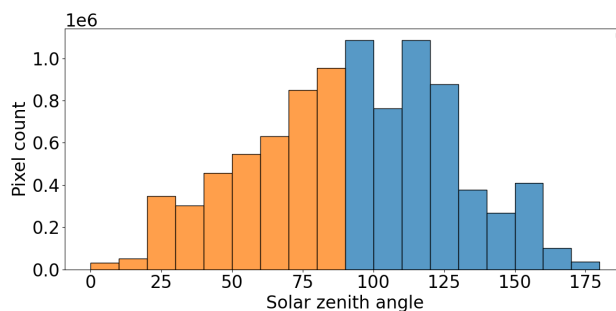
Figure 4. Distribution of solar zenith angle in the dataset. Orange and blue bars correspond to daytime and nighttime samples, respectively. The values represent the number of pixels falling into each solar zenith angle bin.

and the North-Atlantic Flight Corridor (Fig. 3). 140 images with a size of 256×256 pixels were selected (The majority of annotated contrails have a length of up to 60 pixels). 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 guarantees that both contrail and non-contrail cases are represented. The examples are not uniformly distributed in space and time as the images are sampled to include more contrail examples as described above (Fig. 3). Roughly half of the pixels are during daytime (Fig. 4), meaning that they have a solar zenith angle $< 90°$. The distribution of cloud-top phases using the ProPS algorithm (Sec. 2.3) in all pixels of our dataset (Tab. 1) shows that clear-sky conditions account for the most frequently occurring pixel type. A substantial portion of the domain is covered by warm liquid or thin ice clouds. Thick ice and mixed-phase clouds occur less frequently, while supercooled liquid clouds are observed only rarely. Of course, a part of the contrails is already considered in this statistic, as they are mostly identified as thin ice clouds.

Table 1. Relative frequency of cloud-top phases from ProPS (Mayer et al., 2024a) in the compiled contrail dataset

Cloud Top Phase	Relative Occurence [%]
Clear	38.1
Warm Liquid	27.5
Supercooled Liquid	1.4
Mixed Phase	5.2
Thin Ice	21.2
Thick Ice	6.6

Table 2. Channels and normalization ranges for Ash RGB according to EUMETSAT user manual (EUMeTrain, 2022)

Color beam	Channel / Channel difference	Normalization range [K]	
Red	IR108 - IR120	-4	2
Green	IR087 - IR108	-4	5
Blue	IR108	243	303

3.2 Contrail Labeling Process

The Open Source Darwin V7 Labeling tool (V7 Labs) was used for contrail labeling. Human labelers were presented with a subset of the 140 selected images or more precisely Ash RGB composite images in satellite projection in which contrails should be marked. The Ash RGB images are created in agreement with the EUMETSAT user manual by linearly stretching two Brightness Temperature Differences (BTDs) and one Brightness Temperature (BT) over a pre-defined range (see Tab. 2), without gamma correction (i.e. γ=1). This composite enhances thin ice clouds visibility, so contrails can be identified as dark blue linear structures in these images. Because only infrared channels are used, these composites are available during day and night.

To further aid the labeling process, additional composites and BTDs were provided as a time series. 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. Contrails are expected to suddenly appear as thin straight lines and spread out over time. As contrails grow older, they lose their linear shape and their width increases. This is different to natural clouds that usually form more gradually and often have diffuse or irregular structures from the start, that might become or include linear structures due to wind shear or waves in the upper troposphere. This temporal context is very helpful for contrail identification (see also Ng et al., 2024). The additional time series imagery included the Ash RGB, an inverted Ash RGB, two BTDs (IR108 - IR120 and IR087 - IR120) and a Natural Color RGB and HRV Cloud RGB (EUMeTrain, 2022) if the solar channels were available during daytime

Data
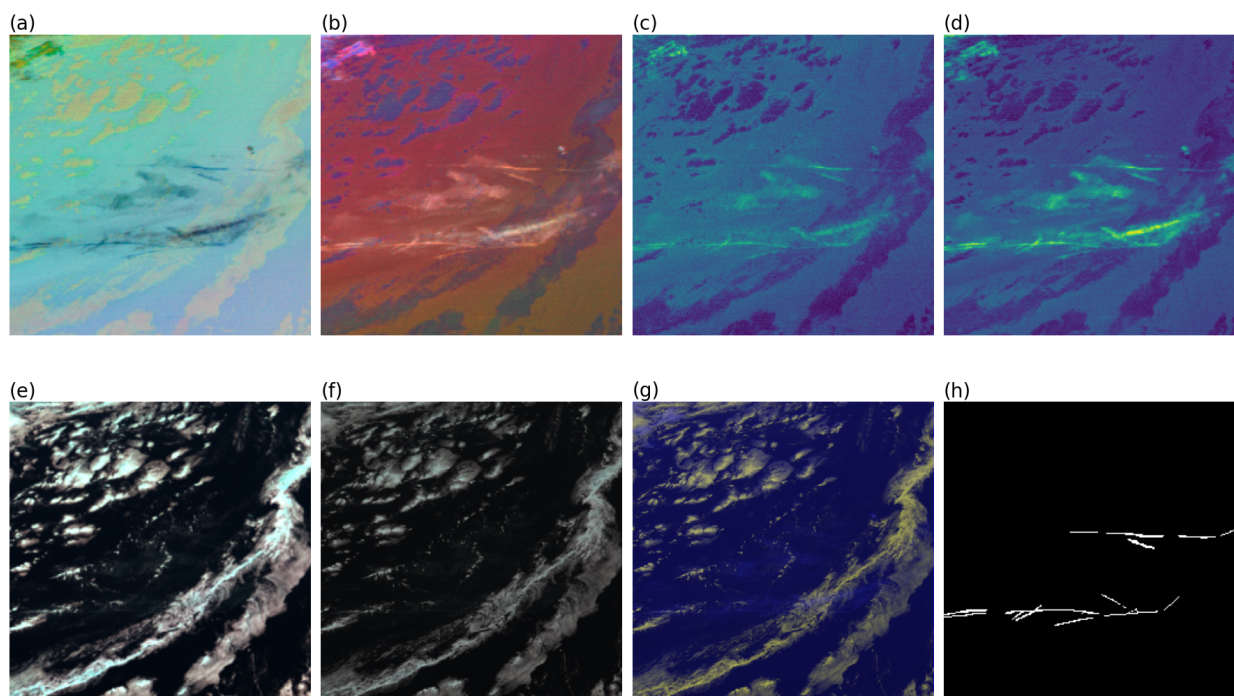
Figure 5. Example images for Scene 041 provided to labelers: a) Ash RGB; b) Inverted Ash RGB; c) BTD IR108 - IR120; d) BTD IR087 - IR120; e) Natural Color RGB; f) HRV Natural Color RGB; g) HRV Cloud RGB; and h) labeling result as the agreement between three labelers (majority mask).

(Fig. 5). Contrails appear as dark blue lines in the Ash RGB composite and bright yellow lines in the inverted Ash RGB. The two mentioned BTDs are known to enhance contrail visibility (Inoue, 1985; Mannstein et al., 1999, 2010; Mayer et al., 2024b) as contrails exhibit higher brightness compared to their surrounding in the BTDs. Additional daytime imagery (Natural Color RGB and HRV Cloud RGB) is provided to help to more clearly identify natural clouds and allow to help distinguish them from contrails.

Using a brush tool, labelers were asked to mark each individual contrail pixel in the Ash RGBs, for all contrails that contain at least 10 pixels. Labelers were provided a detailed labeling guide in a similar way as in Ng et al. (2024), explaining the labeling tool and the additional imagery provided.

To establish a validation standard that reduces subjectivity of labelers, each image was labeled independently by three different labelers. 14 total labelers were each given a subset of the 140 images, with three labelers receiving the same subset of images, making sure that each image is annotated three times by different persons. If at least two out of the three labelers marked a
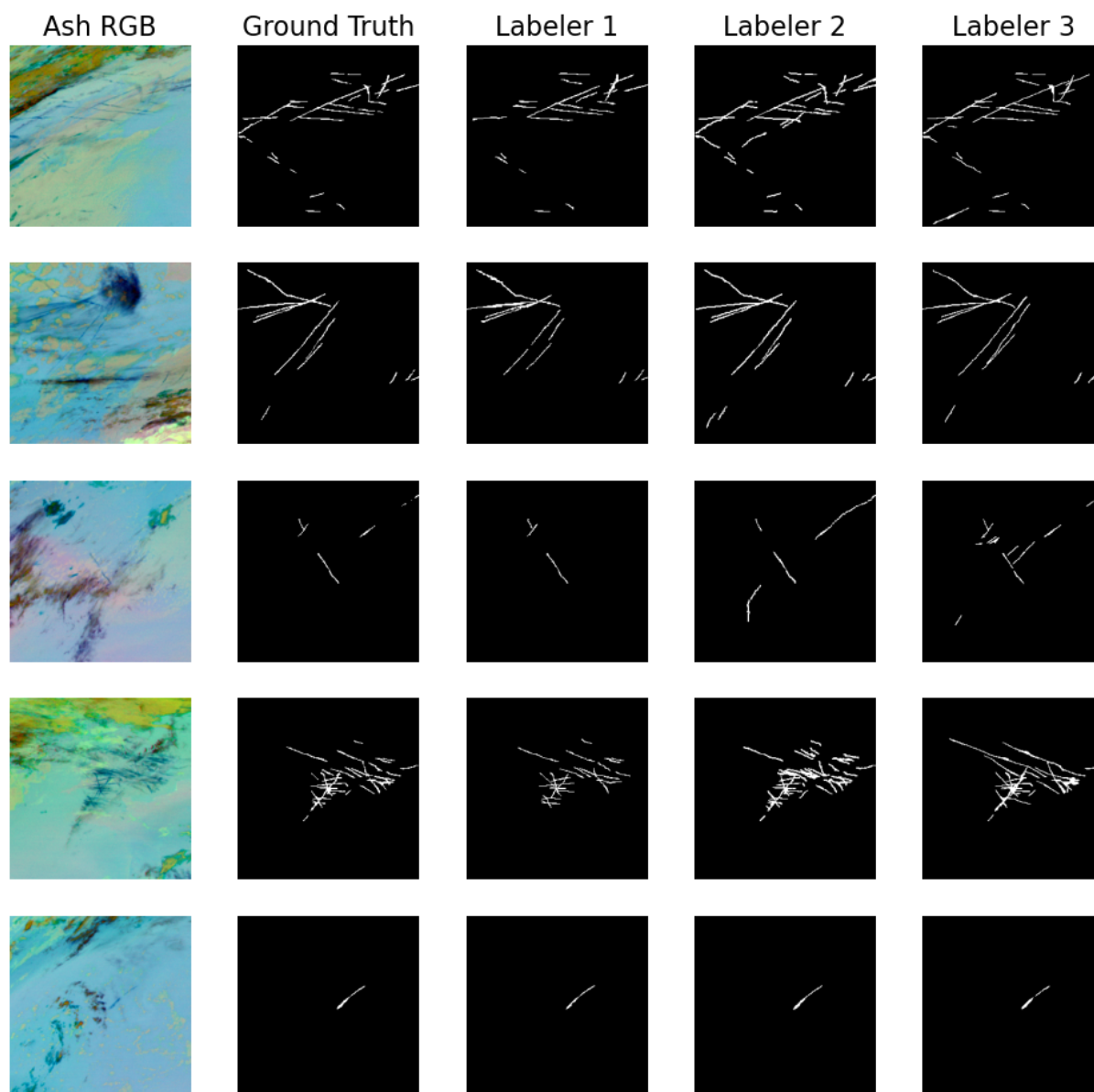
Figure 6. Examples of individual labeler annotations and the resulting consensus ground truth for selected scenes in the dataset.

pixel as a contrail pixel, it was marked as a contrail pixel in the final contrail mask (majority mask, called also ground truth). Otherwise the pixel was marked as a non-contrail pixel. Some examples are shown in Fig. 6.

4 Results

In the following, we discuss the results of the labeling campaign. The focus is on labeling dependencies, the subjectivity of human labeling and the established ground truth and its properties.

4.1 Labeling dependencies

Labeling results can be investigated pixel-wise and object-wise. An object-wise agreement was assumed if at least half of the object's pixels matched. A total of 94256 pixels were annotated by all labelers. 58 % of those pixels were marked by only one person and all three labelers agreed on only 20 % of marked pixels. Labelers marked a total of 1851 contrails in the 140 images. More than half of those contrails, namely 57 %, were only labeled by one person. All three labelers agreed on only 26 % of the labeled contrails. There was also a big difference in which pixels labelers marked in each contrail. On average, only half of the total labeled pixels for each contrail matched between the labelers. These results indicate that some labelers were stricter or more confident in identifying contrails. Also, some labelers have marked contrails more conservatively, while others included broader areas and marked contrails as longer or wider instead of labeling only the core of the contrail. The disagreement between the pixel-wise and object-wise results suggests that larger contrails may have more variation in labeling, while smaller contrails could be missed or misidentified entirely by some labelers. Labeler agreement depended heavily on the cloud phase (determined by ProPS) of clouds beneath/around the contrail. Contrails were more likely to be identified by all labelers in clear sky conditions or over warm liquid clouds. Over thick ice clouds they were more likely to be identified by only one or two persons (Fig. 7a). High linearity (see Sect.4.2) also had a big impact on labeling agreement as more linear contrails were marked by all persons more often (Fig. 7b). Narrow contrails (i.e. pixel width < 2.5) were usually marked by fewer labelers (Fig. 7c). The different labeling results are thus due to the subjective decisions taken on the basis of the own image assessment of the labelers and on the objective challenge in identifying contrail structures of different width, length, appearance and contrast and separating them from natural clouds.

As illustrated above, although detailed labeling instructions were given, the task remained subjective. Therefore, we compare the results from different labelers to assess their level of agreement. This is done by assigning/defining that one labeler is the truth, so precision and recall for the other ones are computed. The highest precision (proportion of correctly identified contrails compared to all identified contrails) observed between labelers on the full dataset was 73 %, while some comparisons showed precision values as low as 43 %. Recall (proportion of correctly identified contrails to all contrails in the truth) values similarly varied within this range. When comparing individual labeler's results to the ground truth, labelers achieve a precision of about 67.5 % and a recall of 82.3 %. 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. These results highlight that even the established ground truth (validation standard) is inherently subjective and not perfectly accurate. Defining it as objects labeled by at least two out of three labelers inevitably leads to missing an unknown amount of real contrails that were identified by only one person or even none but should also reduce false alarms.

Table 3. Pixel-wise precision and recall comparing individual labeler's results (assuming that one labeler is the truth, so precision and recall for the other ones are computed)

Precision [%]	44.04	44.21	73.22	53.80	60.10	43.99
Recall [%]	73.22	60.10	44.04	43.99	44.22	53.80
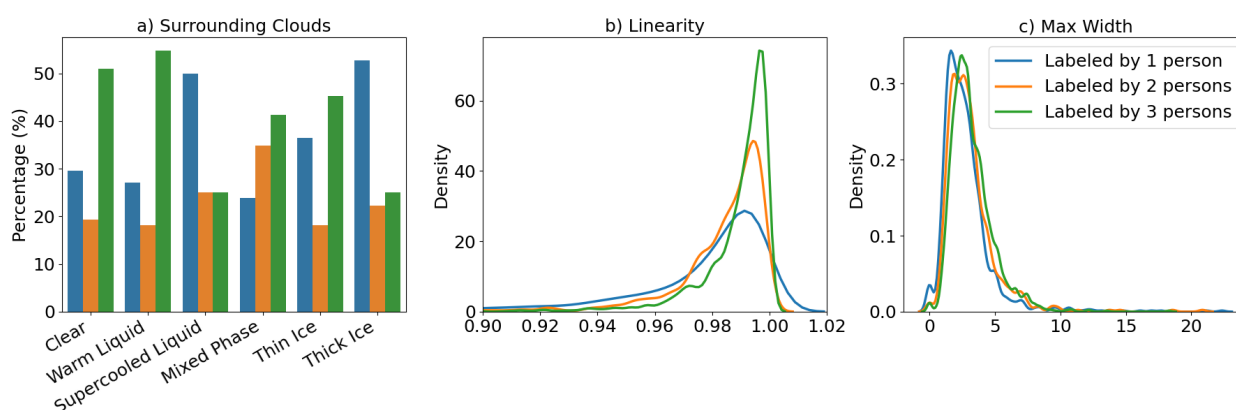
Figure 7. a) Cloud-top phase of clouds beneath/around labeled contrails; b) Linearity of labeled contrails; c) Maximum width in pixels of labeled contrails; sorted by labeler agreement (the number of labelers agreeing on a contrail). Percentage refers to the proportion of objects within each cloud type that were labeled by one, two or three people. (Probability) density indicates the likelihood of observing a given value; the area under each curve integrates to 1.

4.2 Ground Truth

4.2.1 Derivation of contrail properties

Every contrail identified in the ground truth (majority consensus of three labelers) consists of a set of pixels that have been
195 marked by the labelers. To compute properties like length, width and linearity we first fit a line to these pixels. The first and last point of the line that overlap with the pixels identified by the labelers define the two endpoints of the contrail. Contrail length is the distance between the two endpoints of the contrail.

We define the width of a contrail as twice the perpendicular distance of a pixel to the fitted line of the contrail. Mean width is the mean of those distances for all pixels that are part of the contrail, while maximum width is the maximum of those distances.
200 The average mean width of a contrail is the average of the mean width over all contrails in the ground truth.

To assess linearity, a Principal Component Analysis (Hotelling, 1933) was performed, and the proportion of variance explained from the first component is used as a linearity score for the objects. If the pixels form a line, most variance is captured
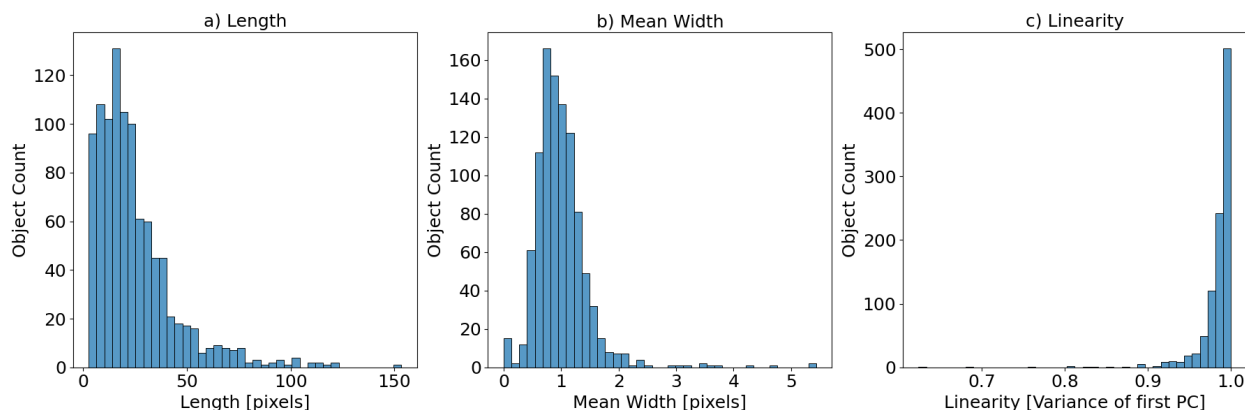
Figure 8. a) Length; b) Mean Width; c) Linearity of labeled contrails in the ground truth (annotated by at least 2 labelers).

by the first principal component; if they form a blob, variance is shared with the second. Thus, the variance fraction of the first component serves as a linearity score.

4.2.2 Contrail properties in the ground truth

The ground truth contains a total of 772 contrails, with 40812 pixels, that were annotated by at least two labelers. Most contrails in the ground truth have a length of less than 50 pixels (Fig. 8a), with an average length of 25 pixels. The average mean width is one pixel. Most labeled contrails have a mean width of up to two pixels (Fig. 8b). The maximum width of contrails in the ground truth can be up to 6 pixels. Linearity scores of contrails are mostly above 0.9 (Fig. 8c), averaging to 0.984. This indicates that contrails are easily distinguished in satellite images by human labelers if they are long, thin and linear stripes. 49 out of the 140 images contain no contrails in the ground truth (Fig. 9), thus 9 more than in the original first image selection process. This discrepancy highlights that some contrails are challenging to identify—often subtle or ambiguous—making them difficult to detect at first glance or leading to differences in classification among labelers. Across all scenes, the mean contrail number is approximately 7. When considering only scenes in which contrails are present, the mean increases to 10.9. 82.89 % of contrail pixels in the ground truth are classified as cirrus pixels by CiPS. 66.21 % of contrail pixels are correctly classified as thin ice clouds by ProPS. While pixels of a contrail might be classified correctly, those algorithms cannot distinguish between contrail pixels and other surrounding ice cloud pixels. Thus, they cannot depict the contrail in its shape. This indicates that these two satellite retrievals are not able to detect all contrails and a dedicated algorithm for contrail detection is needed. The presented ground truth provides the necessary basis for the development and evaluation of a contrail detection algorithm.

5 Dataset Description

The dataset supporting this study follows the FAIR principles to ensure it is Findable, Accessible, Interoperable, and Reusable. The published dataset (Santos Gabriel et al., 2025) contains manually annotated contrail masks derived from MSG/SEVIRI
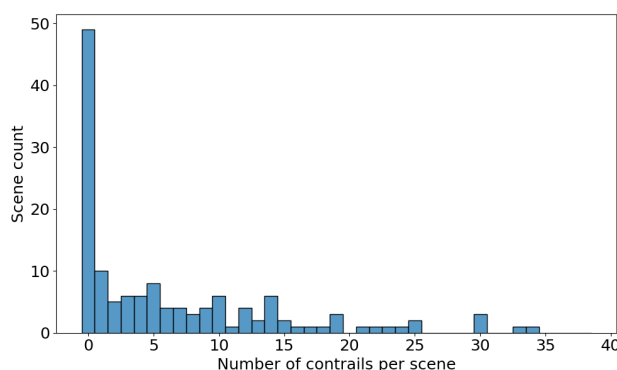
Figure 9. Distribution of contrail number per scene in the labeled dataset

imagery. Additionally, cloud properties and ancillary data are provided. Scenes are sequentially numbered from 0 to 139. The
data repository includes the following content:

225   – ash RGB images (PNG format)

– ground truth images (PNG format)

– labeler annotations, cloud properties and ancillary data (NetCDF format)

– contrails as per Ground truth as individual objects (GeoJSON format)

– file with observation times, regions, and additional info of full dataset (CSV format)

230   – documentation file (md format)

Our dataset provides the Ash RGB images (see Tab. 2) used for labeling and the ground truth (the majority consensus of
labelers) in PNG format. The main data is stored in NetCDF files, that include processed satellite data, labeler annotations,
cloud properties and ancillary information (see a list of all variables and attributes in Tab. 4). They contain the annotations of
each labeler per scene, with the additional info on labeler name (anonymized), as well as the established ground truth as the
235   majority consensus of the three labelers. Satellite data, processed by the authors of this paper, for all 12 SEVIRI channels is
provided in the NetCDF files as well. SEVIRI scans the Earth's disk line by line, requiring several minutes to complete a full
scan. Each image line is acquired at a slightly different time, so the acquisition time is provided in the dataset to identify the
timestamp at which that specific scan line was recorded. Each scene is accompanied by the corresponding latitude and longitude
grids in satellite projection, allowing the spatial collocation of detected contrails with their corresponding geophysical positions
240   on Earth. The solar zenith angle, satellite elevation angle, and satellite off-nadir angle are also provided to support spatial
collocation with respect to the satellite position. The solar zenith angle is the angle between the local zenith (straight up at
point from scene on Earth) and the direction of the sun, where a value of less than 90 degrees represents day and a value greater
than 90 degrees represents night. The satellite elevation angle is the angle between the local horizon and the line-of-sight to

the satellite. The satellite off-nadir angle is the angle between the satellite nadir vector (line from satellite straight down to
245   sub-satellite point) and the satellite-to-point-of-scene vector. The satellite viewing angle (the angle between the local vertical
(nadir direction) at the point of the scene on Earth's surface and the line of sight to the satellite) can be calculated by subtracting
the satellite elevation angle from $90°$. A land–sea mask and surface altitude map (derived from the GTOPO30 global elevation
dataset; U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center) are provided as additional auxiliary
datasets. The land–sea mask (0 = sea, 100 = land) helps distinguish between oceanic and terrestrial regions, which is essential
250   for interpreting surface-dependent features such as temperature, reflectance, or cloud characteristics. The surface altitude, given
in meters, provides topographic context and supports a more accurate interpretation of variables influenced by elevation, such
as radiance and temperature. The ice cloud properties cirrus cloud probability, cirrus opacity probability, cloud top height, ice
optical thickness, ice water path and effective ice particle radius, derived from CiPS (see Sec. 2.2), are supplemented to the
dataset. Note that all ice cloud properties are provided even if CiPS does not declare a pixel as a cirrus pixel (Calculation of
255   ice cloud properties are only verified for cirrus pixels). In addition, cloud-top phases are determined with ProPS (see Sec. 2.3)
and added to the NetCDF files. Here, a flag value is provided to indicate the cloud-top phase. A value of 0 corresponds to
a clear sky pixel. 1 are thick ice clouds with an optical thickness greater than two. A flag value of 3 represents warm liquid
and a value of 4 supercooled liquid clouds. Finally, thin ice clouds with an optical thickness smaller than two are assigned
the value of 5. The Dataset contains two spatial grids: x, y for the native 3-km SEVIRI resolution, and x_highres, y_highres
260   for the corresponding 1-km high-resolution of the HRV channel. These coordinates correspond to the standard geostationary
projection used by the SEVIRI instrument, centered on the satellite's subpoint. In addition to the pixel-wise variables provided
in two-dimensional format for each satellite scene, scene-level quantities are stored as global attributes within the NetCDF files.
Each file has a title containing the subsequent scene number. The timestamp and native SEVIRI bounding coordinates of each
scene, corresponding to a cropped region of the MSG full disk, are included to enable an exact identification of the contrail's
265   temporal and spatial context. To facilitate rapid analysis and ensure reproducibility, several secondary attributes—derived from
the primary dataset—are provided. Although these can be recomputed from the original data, their inclusion allows users to
avoid redundant calculations. Those attributes include the boreal season and daytime (day or night). Longitude, latitude, solar
zenith angle, satellite elevation angle and satellite off-nadir angle are provided for the central point in the scene. The fractions
of surface types (sea or land) in the scene, as well as the fractions of cloud-top phases are given for each labeled scene. Lastly,
270   the number of contrails in each scene is provided as global attribute in the NetCDF files. All Variables and Attributes are
presented in a detailed table in Appendix A. An overview of the whole dataset and all attributes of each scene is provided
as well to allow for a rapid overview of the entire dataset. To supplement the ground truth, provided as a binary pixel mask,
all individual contrails—including all constituent pixels—are represented in GeoJSON format. Each contrail can be accessed
either as a LineString object, containing only its endpoints, or as a MultiPoint object, containing all associated pixels. This
275   object-based ground truth is designed to facilitate analyses at the level of individual contrails. A documentation file is included
in the dataset explaining all data and its usage.

Table 4. Variables and global attributes in NetCDF files included in the annotated contrail dataset for MSG

Variables		Global Attributes
Ground Truth	Latitude	title
Labeler 1	Longitude	timestamp
Labeler 2	solar_zenith_angle	native_coordinates
Labeler 3	sat_elevation_angle	season
HRV	sat_off_nadir_angle	daytime
VIS_006	Land Sea Mask	central_longitude
VIS_008	Surface Altitude	central_latitude
NIR_016	ccp_cips	central_solar_zenith_angle
IR_039	cop_cips	central_elevation_angle
WV_062	cth_cips	central_off_nadir_angle
WV_073	iot_cips	surface_types
IR_087	iwp_cips	cloud_phase_fractions
IR_097	ireff_cips	contrail_number
IR_108	cloudphases_props	
IR_120	acquisition_time	
IR_134		

6 Conclusions

In this study, we present a manually labeled contrail dataset based on Meteosat Second Generation (MSG) SEVIRI imagery
mainly over Europe and the North Atlantic. The dataset consists of 140 scenes of 256×256 pixels, each annotated indepen-
280 dently by three labelers, with the ground truth defined through majority consensus. Beyond the binary pixel-level masks for all
contrails and the majority mask, we provide an object-based representation of individual contrails in GeoJSON format, captur-
ing both their shape and all associated pixels. To give additional context, we include outputs from two satellite retrievals—CiPS
(Cirrus Properties from SEVIRI) and ProPS (Probabilistic Cloud Top Phase retrieval)—which provide information on cloud
top phase, cirrus probability, ice optical thickness, ice water path, and effective ice particle radius. Auxiliary data, such as
285 scene-level attributes, geolocation, solar and satellite angles, land–sea masks, and surface altitude, are also included to support
detailed analyses and reproducibility.

This dataset addresses a significant gap in observational resources for contrail research. Although contrails and contrail cirrus
are known to have an important impact on climate (Lee et al., 2021), observational data remain limited, and high-quality la-
beled datasets are rare. By providing carefully labeled data along with complementary satellite-derived variables, researchers
290 can investigate contrail formation, evolution, and visibility under different atmospheric conditions. The inclusion of multiple

labelers' annotations also allows exploration of the subjectivity inherent in manual labeling, helping to quantify uncertainty and the limitations of human detection.

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.

By making this dataset publicly available, we intend to provide a resource that supports research into contrail detection, subsequently improves understanding of aviation-related cloud effects, and eventually informs strategies to reduce aviation's climate impact. We expect it to serve as a foundation for future work in atmospheric science, remote sensing, and climate modeling, supporting both algorithm development and scientific studies into the role of contrails in the climate system.

Appendix A: Variables and Attributes in Contrail Dataset

In this section, we present a table with all global attributes contained in the NetCDF files of our compiled dataset (Tab. A1). This table includes the names of the global attributes as well as a description, the units and one example of this attribute. We also provide a table with all variables in the NetCDF files (Tab. A2). This table includes the name of the variables, a description, the dimensions of this variable, the data type, units and additional attributes of the variable.

Data availability

The full dataset is available under: https://doi.org/10.5281/zenodo.17669444 (Santos Gabriel et al., 2025).

Contains modified EUMETSAT Meteosat data (2025). MSG/SEVIRI data are available from the EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) data centre (https://user.eumetsat.int/catalogue/EO:EUM:DAT: MSG:HRSEVIRI, EUMETSAT, 2025).

Author contributions. VSG and LB conceived this study. VSG prepared the data and overlooked the labeling campaign with valuable feedback from LB. CV supervised the project and provided scientific feedback. The remaining authors took part in the labeling campaign for the contrail dataset. VSG took the lead in writing the manuscript. All authors provided feedback on the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Table A1. Description of Attributes in NetCDF files in the annotated contrail dataset for MSG

Global Attribute	Description	Units	Example
scene_number	Sequential image number in dataset	-	1
title	Title of NetCDF file containing data	-	MSG Contrail Dataset - Scene 001
timestamp	Timestamp of labeled satellite image in format yyyy-mm-dd hh:MM:ss	-	2024-01-12 15:00:00
native_coordinates	Native SEVIRI coordinates of boundary box of cropped scene [x_min x_max y_min y_max]	pixels	[2600 2856 3100 3356]
season	Boreal season	-	Spring (MAM)
daytime	Day/night according to solar zenith angle	-	day
central_longitude	Longitude at central point of satellite image	°	-39.748
central_latitude	Latitude at central point of satellite image	°	45.155
central_solar_zenith_angle	Angle in degrees between the local zenith (straight up at central point of scene on Earth) and the direction to the Sun	°	66.88
central_elevation_angle	Angle in degrees between the local horizon at central point of scene on Earth and the line-of-sight to the satellite	°	82.12
central_off_nadir_angle	Angle in degrees between the satellite nadir vector (line from satellite straight down to sub-satellite point) and the satellite-to-central point of scene vector	°	7.29
surface_types	Fractions of land, ocean, and coastline pixels in image	-	{"Land": 0.016, "Ocean": 0.980, "Coast": 0.004}
cloud_phase_fractions	Fractions of cloud-top phases in image	-	{"Clear Sky": 0.0393, "Warm Liquid": 0.3161, "Supercooled Liquid": 0.0008, "Mixed Phase": 0.0063, "Thin Ice": 0.3692, "Thick Ice": 0.2681}
contrail_number	Number of contrail segments in image	-	2

Acknowledgements. This research acknowledges funding from the MEFKON project, funded by the Federal Ministry for Economics and Climate Action (Bundesministerium für Wirtschaft und Klimaschutz, BMWK) of Germany in the context of the LuFo VI-3 call with grant id 20F2202B, from the A4CLIMATE project by the European Union's Horizon Europe research and innovation program under the Grant Agreement no. 101192301, and from the German Federal Ministry of Transport and Digital Infrastructure under FE-No. 50.0391/2021. We finally thank EUMETSAT for clarifying attribution modalities and distribution conditions for the MSG data used in this study.

320

References

Atlas, D., Wang, Z., and Duda, D. P.: Contrails to cirrus—Morphology, microphysics, and radiative properties, Journal of applied meteorology and climatology, 45, 5–19, 2006.

Bedka, S. T., Minnis, P., Duda, D. P., Chee, T. L., and Palikonda, R.: Properties of linear contrails in the Northern Hemisphere derived from 2006 Aqua MODIS observations, Geophysical Research Letters, 40, 772–777, 2013.

Burkhardt, U., Bock, L., and Bier, A.: Mitigating the contrail cirrus climate impact by reducing aircraft soot number emissions, npj Climate and Atmospheric Science, 1, 37, 2018.

Chevallier, R., Shapiro, M., Engberg, Z., Soler, M., and Delahaye, D.: Linear contrails detection, tracking and matching with aircraft using geostationary satellite and air traffic data, Aerospace, 10, 578, 2023.

Dekoutsidis, G., Feidas, H., and Bugliaro, L.: Contrail detection on SEVIRI images and 1-year study of their physical properties and the atmospheric conditions favoring their formation over Europe, Theoretical and Applied Climatology, 151, 1931–1948, 2023.

EUMeTrain: Compilation of RGB Recipes, https://www.eumetrain.org/sites/default/files/2022-10/RGB_recipes.pdf, accessed: 2025-10-14, 2022.

Gourgue, N., Boucher, O., and Barthès, L.: A dataset of annotated ground-based images for the development of contrail detection algorithms, Data in Brief, 59, 111 364, 2025.

Haywood, J. M., Allan, R. P., Bornemann, J., Forster, P. M., Francis, P. N., Milton, S., Rädel, G., Rap, A., Shine, K. P., and Thorpe, R.: A case study of the radiative forcing of persistent contrails evolving into contrail-induced cirrus, Journal of Geophysical Research: Atmospheres, 114, 2009.

Hotelling, H.: Analysis of a complex of statistical variables into principal components., Journal of educational psychology, 24, 417, 1933.

Inoue, T.: On the temperature and effective emissivity determination of semi-transparent cirrus clouds by bi-spectral measurements in the 10μm window region, Journal of the Meteorological Society of Japan. Ser. II, 63, 88–99, 1985.

Lee, D. S., Fahey, D. W., Skowron, A., Allen, M. R., Burkhardt, U., Chen, Q., Doherty, S. J., Freeman, S., Forster, P. M., Fuglestvedt, J., et al.: The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018, Atmospheric environment, 244, 117 834, 2021.

Lewellen, D.: Persistent contrails and contrail cirrus. Part II: Full lifetime behavior, Journal of the Atmospheric Sciences, 71, 4420–4438, 2014.

Mannstein, H. and Schumann, U.: Aircraft induced contrail cirrus over Europe, Meteorologische Zeitschrift, 14, 549–554, 2005.

Mannstein, H., Meyer, R., and Wendling, P.: Operational detection of contrails from NOAA-AVHRR-data, International Journal of Remote Sensing, 20, 1641–1660, 1999.

Mannstein, H., Brömser, A., and Bugliaro, L.: Ground-based observations for the validation of contrails and cirrus detection in satellite imagery, Atmospheric Measurement Techniques, 3, 655–669, 2010.

Mayer, J., Bugliaro, L., Mayer, B., Piontek, D., and Voigt, C.: Bayesian cloud-top phase determination for Meteosat Second Generation, Atmospheric Measurement Techniques, 17, 4015–4039, 2024a.

Mayer, J., Mayer, B., Bugliaro, L., Meerkötter, R., and Voigt, C.: How well can brightness temperature differences of spaceborne imagers help to detect cloud phase? A sensitivity analysis regarding cloud phase and related cloud properties, Atmospheric Measurement Techniques, 17, 5161–5185, 2024b.

McCloskey, K., Geraedts, S., Van Arsdale, C., and Brand, E.: A human-labeled Landsat-8 contrails dataset, in: Proceedings of the ICML 2021 Workshop on Tackling Climate Change with Machine Learning, Virtually, vol. 23, 2021.

Meijer, V. R., Kulik, L., Eastham, S. D., Allroggen, F., Speth, R. L., Karaman, S., and Barrett, S. R.: Contrail coverage over the United States before and during the COVID-19 pandemic, Environmental Research Letters, 17, 034 039, 2022.

Meyer, R., Buell, R., Leiter, C., Mannstein, H., Pechtl, S., Oki, T., and Wendling, P.: Contrail observations over Southern and Eastern Asia in NOAA/AVHRR data and comparisons to contrail simulations in a GCM, International Journal of Remote Sensing, 28, 2049–2069, 2007.

Minnis, P., Bedka, S. T., Duda, D. P., Bedka, K. M., Chee, T., Ayers, J. K., Palikonda, R., Spangenberg, D. A., Khlopenkov, K. V., and Boeke, R.: Linear contrail and contrail cirrus properties determined from satellite data, Geophysical Research Letters, 40, 3220–3226, 2013.

Ng, J. Y.-H., McCloskey, K., Cui, J., Meijer, V. R., Brand, E., Sarna, A., Goyal, N., Van Arsdale, C., and Geraedts, S.: Contrail Detection on GOES-16 ABI With the OpenContrails Dataset, IEEE Transactions on Geoscience and Remote Sensing, 62, 1–14, https://doi.org/10.1109/TGRS.2023.3345226, 2024.

Rybka, H., Burkhardt, U., Köhler, M., Arka, I., Bugliaro, L., Görsdorf, U., Horváth, Á., Meyer, C. I., Reichardt, J., Seifert, A., et al.: The behavior of high-CAPE (convective available potential energy) summer convection in large-domain large-eddy simulations with ICON, Atmospheric chemistry and physics, 21, 4285–4318, 2021.

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.: Annotated Contrail Dataset for Meteosat Second Generation (MSG) (1.0), Data set, https://doi.org/10.5281/zenodo.17669444, 2025.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: An Introduction to Meteosat Second Generation (MSG), Bulletin of the American Meteorological Society, 83, https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, 2002.

Spangenberg, D. A., Minnis, P., Bedka, S. T., Palikonda, R., Duda, D. P., and Rose, F. G.: Contrail radiative forcing over the Northern Hemisphere from 2006 Aqua MODIS data, Geophysical Research Letters, 40, 595–600, 2013.

Strandgren, J., Bugliaro, L., Sehnke, F., and Schröder, L.: Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks, Atmospheric Measurement Techniques, 10, 3547–3573, 2017a.

Strandgren, J., Fricker, J., and Bugliaro, L.: Characterisation of the artificial neural network CiPS for cirrus cloud remote sensing with MSG/SEVIRI, Atmospheric Measurement Techniques, 10, 4317–4339, 2017b.

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, 6071–6093, 2024.

U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center: Global 30-Arc-Second Elevation (GTOPO30), https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30, https://doi.org/10.5066/F7DF6PQS, accessed: 2014-05-16.

V7 Labs: V7 Darwin: AI Data Labeling & ML Training Data Platform, https://www.v7labs.com/darwin.

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, 2015.

Wang, X., Wolf, K., Boucher, O., and Bellouin, N.: Radiative effect of two contrail cirrus outbreaks over Western Europe estimated using geostationary satellite observations and radiative transfer calculations, Geophysical Research Letters, 51, e2024GL108 452, 2024.

Wang, Z., Bugliaro, L., Jurkat-Witschas, T., Heller, R., Burkhardt, U., Ziereis, H., Dekoutsidis, G., Wirth, M., Groß, S., Kirschler, S., Kaufmann, S., and Voigt, C.: Observations of microphysical properties and radiative effects of a contrail cirrus outbreak over the North Atlantic, Atmospheric Chemistry and Physics, 23, 1941–1961, https://doi.org/10.5194/acp-23-1941-2023, 2023.

Table A2. Description of Variables in NetCDF files in the annotated contrail dataset for MSG

Variable	Description	Dimensions	Data Type	Units	Additional Attributes
Ground Truth	Ground truth as majority Consensus of labelers	(x,y)	float64	-	-
Labeler 1	Pixel-wise annotations of first labeler	(x,y)	float64	-	labeler: Name of labeler (anonymized)
Labeler 2	Pixel-wise annotations of second labeler	(x,y)	float64	-	labeler: Name of labeler (anonymized)
Labeler 3	Pixel-wise annotations of third labeler	(x,y)	float64	-	labeler: Name of labeler (anonymized)
HRV	Reflectance of High Resolution Visible channel (broadband 0.4 μm - 1.1 μm)	(x_highres, y_highres)	float64	%	-
VIS006	Reflectance of MSG channel with central wavelength 0.635 μm	(x,y)	float64	%	-
VIS008	Reflectance of MSG channel with central wavelength 0.81 μm	(x,y)	float64	%	
NIR016	Reflectance of MSG channel with central wavelength 1.64 μm	(x,y)	float64	%	-
IR039	Brightness Temperature of MSG channel with central wavelength 3.90 μm	(x,y)	float64	K	-
WV062	Brightness Temperature of MSG channel with central wavelength 6.25 μm	(x,y)	float64	K	-
WV073	Brightness Temperature of MSG channel with central wavelength 7.35 μm	(x,y)	float64	K	-
IR087	Brightness Temperature of MSG channel with central wavelength 8.70 μm	(x,y)	float64	K	-
IR097	Brightness Temperature of MSG channel with central wavelength 9.66 μm	(x,y)	float64	K	-
IR108	Brightness Temperature of MSG channel with central wavelength 10.80 μm	(x,y)	float64	K	-
IR120	Brightness Temperature of MSG channel with central wavelength 12.00 μm	(x,y)	float64	K	-
IR134	Brightness Temperature of MSG channel with central wavelength 13.40 μm	(x,y)	float64	K	-
acquisition_time	acquisition time of row in satellite image	(y)	DateTime64	-	-
Latitude	Latitude grid for the image, allowing mapping from pixel indices to geographic coordinates	(x,y)	float64	°	-
Longitude	Longitude grid for the image, allowing mapping from pixel indices to geographic coordinates	(x,y)	float64	°	-
solar_zenith_angle	Angle in degrees between the local zenith and the direction to the Sun	(x,y)	float64	°	-
sat_elevation_angle	Angle in degrees between the local horizon and the line-of-sight to the satellite	(x,y)	float64	°	-
sat_off_nadir_angle	Angle in degrees between the satellite nadir vector and the satellite-to-point of scene vector	(x,y)	float64	°	-
Land Sea Mask	Land sea mask, where 0: sea and 100: land	(x,y)	float32	-	-
Surface Altitude	Surface altitude (Above sea level)	(x,y)	float32	m	-
ccp_cips	Cirrus Cloud Probability derived from CiPS (cirrus covered pixel assumed above 0.62)	(x,y)	float32	1	-
cop_cips	Cirrus Opacity Probability derived from CiPS (thick ice cloud covered pixel assumed above 0.86)	(x,y)	float32	1	-
cth_cips	Cloud-top Height derived from CiPS	(x,y)	float32	km	-
iot_cips	Ice Optical Thickness derived from CiPS	(x,y)	float32	1	-
iwp_cips	Ice Water Path derived from CiPS	(x,y)	float32	gm^{-2}	-
ireff_cips	Effective Ice Particle Radius derived from CiPS	(x,y)	float32	μm	-
cloudphases_props	Cloud-top Phases derived from ProPS	(x,y)	float32	-	flag_description: '0: clear; 1: thick ice (OT>2); 2: mixed-phase; 3: warm liquid; 4: supercooled liquid; 5: thin ice (OT<2)'