Author Comment (AC) in response to Referee Comment 1 (RC1)

Manuscript: ContrailStruct30: a 30 m three-band thermal-infrared dataset for contrail segmentation

Discussion manuscript: essd-2026-759 | Earth System Science Data Discussions

In response to: RC1, Anonymous Referee #1 (12 November 2026)

Dear Referee,

We thank you for your encouraging assessment of the dataset and for the detailed and constructive comments. We agree that the model-assisted annotation workflow, its potential selection biases, and the interpretation of the model comparisons require particular care. We have prepared revisions that make these points explicit, strengthen the visibility of the pixel-level contribution, and clarify terminology, figures, and citations.

This is our Author Comment for the interactive public discussion. To make the responses verifiable without submitting a revised manuscript during this stage, we reproduce below the relevant wording prepared in our internal working revision. Reviewer statements are excerpted for readability. For convenience, the updated Figures 2, 3, 6 and 7 and complete Tables 1, 6 and 7 are reproduced next to the comments they address.

1. Scientific substance: iterative annotation and model-dependent bias

1.1 Potentially missed contrails during candidate screening

Referee comment (excerpt)

> *"If the computer vision model erroneously decides that a class of images doesn't contain contrails, might it never show those images to the humans, so they never get labeled? One obvious way to address this would be to run the computer vision model in a high-recall/low-precision mode, did the authors do this? If not, can they point to literature ...?"*

Author response

This is a legitimate limitation of model-guided candidate discovery. To mitigate missed detections, we combined model screening with complementary manual inspection of source scenes and applied a deliberately permissive candidate-retention rule: preliminary binary masks were generated at a probability threshold of 0.50, and any patch containing at least one positive pixel ($nz > 0$) was retained for human review, with no minimum-area, connected-component, or shape filter imposed. While this rule admits many false positives for manual examination, a patch with zero predicted foreground pixels at the initial threshold may still escape model-driven selection. We have clarified that human editing and expert checks determine labels only for the reviewed candidates, and we do not claim exhaustive recall or an unbiased inventory of all contrails in the source imagery. We have also added human-in-the-loop machine-assisted labeling references, including the related flight-informed contrail study by Dalmau et al. (2026).

Relevant wording prepared in our working revision:

> Subsequent scene searches prioritized aircraft flying above 7,000 m, with model screening and manual source-scene inspection used together to locate candidate image areas. ADS-B guided scene searches; image evidence and common annotation rules determined the labeled contrail extent.
>
> A probability threshold of 0.50 produced preliminary binary masks. Candidate retention was deliberately permissive: if the binary mask contained any nonzero contrail pixel ($nz > 0$), the associated patch was retained for manual inspection. No additional minimum-area, connected-component, or shape filter was imposed on these binary masks. Thus, even a single predicted contrail pixel led to human examination, including numerous false-positive candidates. Contrails producing no predicted foreground pixels at the initial screening threshold may therefore remain unselected. Moreover, the enhanced RGB representation used during annotation can influence the visibility and delineation of faint structures. The resulting masks represent expert interpretations of visually discernible contrails rather than an exhaustive inventory of contrail occurrence.

> This human-in-the-loop approach follows related work on machine-assisted image labeling (Budd et al., 2021; Buscombe et al., 2022; Aybar et al., 2022; Kirillov et al., 2023). Flight-informed model-assisted contrail delineation has also been explored using ground-based imagery (Dalmau et al., 2026).

1.2 Rationale and limits of the model comparisons in Section 5

Referee comment (excerpt)

"The model (MaxVit+Unet) that they used to generate the labels is also the best at classifying those labels ... I suspect that if ... SegFormer-B0 [had assisted] labeling, then that model would have performed the best."

Author response

We agree that using a MaxViT-family model in annotation assistance can influence which masks are created and how subsequent models agree with them. We have reframed Section 5 as an illustrative application benchmark, not a controlled ranking of model architectures. Table 6 evaluates six implementations on the Version 3 scene- and date-disjoint test partition. To assess sensitivity to reference-mask delineation, we additionally evaluated six earlier archived checkpoints against 60 pixel masks drawn entirely from scratch by annotators on the full 512 × 512 source images. Images and masks were subsequently centre-cropped to 256 × 256 for evaluation. MaxViT–U-Net had the highest pooled Dice (0.750) and AUPRC (0.816), while FCN–ResNet50 had the next-highest Dice (0.713).

Relevant wording prepared in our working revision:

> To illustrate the use of ContrailStruct30 for pixel-level segmentation, we compare a MaxViT–U-Net implementation with five established alternatives on the Version 3 scene- and date-disjoint test partition. This compact application example reports image-mask overlap and false detections in challenging backgrounds. The released Version 3 partitions remain the recommended basis for future independent evaluations (Sect. 3.3). A separate reference-mask check uses an earlier archived checkpoint set evaluated against hand-drawn masks (Table 7).
>
> The six-model benchmark uses the scene- and date-disjoint Version 3 test partition and quantifies agreement with its human-corrected pixel masks. The compared implementations differ in model capacity and pretraining, and the reference masks originate from a model-assisted, expert-reviewed annotation workflow. The results therefore provide a practical application reference for the released dataset rather than a controlled comparison of encoder architectures.
>
> To examine the sensitivity of the model comparison to reference-mask delineation, we conducted a complementary evaluation using 60 images independently annotated by human annotators (Table 7). The annotators manually delineated contrails directly on the full 512 × 512-pixel images, producing pixel-level reference masks entirely from scratch without using model-generated masks as starting points. After annotation, the central 256 × 256-pixel regions of both the images and their corresponding manually drawn masks were extracted for evaluation. The cropped images were then processed using the same patch-wise enhancement procedure as the main dataset. MaxViT–U-Net achieved the highest pixel-pooled Dice (0.750) and pixel-level AUPRC (0.816), followed by FCN–ResNet50 in Dice (0.713). These results provide an additional reference for assessing model agreement with independently delineated contrail structures.
>
> Table 7. Complementary evaluation of six archived segmentation checkpoints using 60 independently hand-delineated reference masks (47 contrail-positive and 13 empty-reference images).
>
> The 60 evaluated image crops correspond exactly to patches in the archived training (43) or validation (17) partitions, with none in the held-out test partition; the reference masks were independently redrawn. These results characterize in-domain agreement with independently delineated labels rather than generalization to unseen imagery and are not directly comparable with Table 6, which uses different checkpoints and metric aggregation.

Prepared revised Figure 7: qualitative examples from the Version 3 test partition

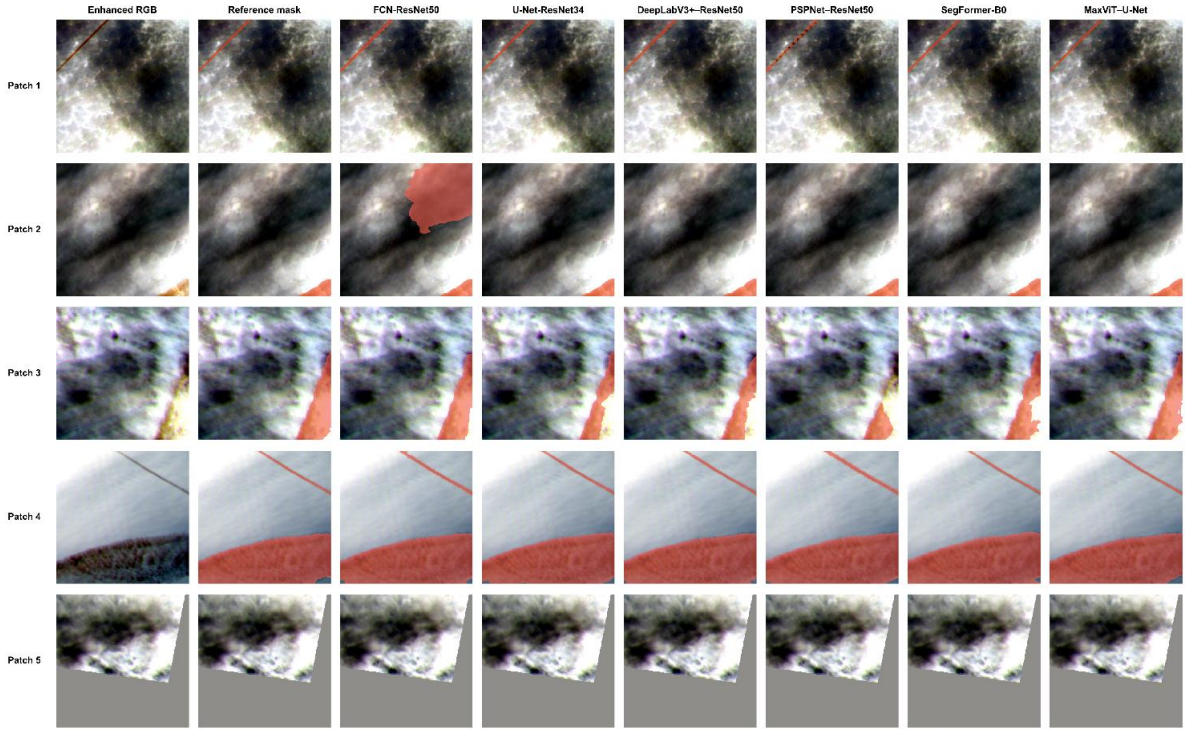

Figure 7. Example contrail segmentation predictions on five patches from the Version 3 scene- and date-disjoint test partition. Rows show four annotated patches spanning increasing contrail fractions and one difficult negative. Columns show enhanced RGB images, reference masks, and predictions from FCN–ResNet50, U-Net–ResNet34, DeepLabV3+–ResNet50, PSPNet–ResNet50, SegFormer-B0, and the latest MaxViT–U-Net checkpoint (seed 42). Coral overlays indicate predicted or reference contrail pixels. The plate is qualitative; aggregate performance is reported in Table 6.

Table 6. Six-model comparison on the Version 3 scene- and date-disjoint test partition. Values are mean positive-patch Dice and mean positive-patch IoU (calculated within each positive patch and averaged over 1,422 positive patches), pixel-pooled IoU (calculated from pooled TP, FP, and FN pixel counts), pixel precision and recall, and negative activation (percentage of the 1,700 empty-reference patches containing at least one predicted foreground pixel). All values use the original evaluation program and the seed-42 checkpoints; thresholds were fixed before test inference. The primary metric is mean positive-patch Dice; these results replace the earlier Version 2 comparison.

Model	Mean positive-patch Dice ↑	Mean positive-patch IoU ↑	Pixel IoU ↑	Precision ↑	Recall ↑	Negative activation ↓
MaxViT–U-Net	0.8202	0.7363	0.7394	0.8417	0.8588	26.59%
U-Net–ResNet34	0.8014	0.7112	0.6790	0.7460	0.8831	45.94%
SegFormer-B0	0.7851	0.6880	0.6835	0.7613	0.8699	56.41%
DeepLabV3+–ResNet50	0.7793	0.6818	0.6728	0.7583	0.8565	49.24%
FCN–ResNet50	0.7070	0.6202	0.6797	0.8332	0.7867	3.94%
PSPNet–ResNet50	0.6894	0.5868	0.6652	0.7941	0.8038	5.59%

Table 7. Complementary evaluation of six archived segmentation checkpoints using 60 independently hand-delineated reference masks (47 contrail-positive and 13 empty-reference images). Human annotators manually delineated the contrail structures directly on the full 512 × 512-pixel source images, generating corresponding 512 × 512-pixel binary masks from scratch without model-generated annotations. Following manual annotation, the central 256 × 256-pixel regions of both the images and their reference masks were extracted, and the cropped images were processed using the standard patch-wise enhancement procedure. Dice and intersection over union (IoU) were calculated from pooled pixel counts, and the area under the precision–recall curve (AUPRC) was evaluated at the pixel level. Negative activation denotes the number of empty-reference images containing at least one predicted contrail pixel. All models used RGB inputs divided by 255, without ImageNet normalization, and their previously selected validation thresholds. The 60 evaluated image crops correspond exactly to patches in the archived training (43) or validation (17) partitions, with none in the held-out test partition; the reference masks were independently redrawn. These results characterize in-domain agreement with independently

delineated labels rather than generalization to unseen imagery and are not directly comparable with Table 6, which uses different checkpoints and metric aggregation.

Model	Threshold	Dice ↑	IoU ↑	AUPRC ↑	Negative activation
MaxViT–U-Net	0.90	0.750	0.600	0.816	3/13
FCN–ResNet50	0.85	0.713	0.555	0.799	3/13
SegFormer-B0	0.75	0.701	0.540	0.754	8/13
DeepLabV3+–ResNet50	0.75	0.669	0.503	0.723	8/13
PSPNet–ResNet50	0.85	0.669	0.502	0.759	3/13
U-Net–ResNet34	0.80	0.668	0.502	0.675	8/13

1.3 The enhanced RGB representation and possible visibility bias

Referee comment (excerpt)

"The color scheme used to do the labeling is the best one to pass to the computer vision models ... Probably the contrails with high contrast in this color scheme are the ones that got labeled."

Author response

We have removed the previous quantitative RGB-versus-uint16 ranking and no longer present the enhanced RGB display as intrinsically optimal. The working revision explains that this was the consistent display used during annotation, while paired three-band uint16 values are supplied for alternative image transformations. We also explicitly acknowledge that display-dependent visibility can affect which faint structures are identified and delineated.

1.4 Within-round correction versus between-round growth

Referee comment (excerpt)

"Were the models and humans iterating on the same patches or different ones? How did the size of the dataset being labeled change between iterations?"

Author response

Within each round, human annotators corrected the provisional masks on the same candidate patches. Across rounds, accepted labels informed screening of new patches and source scenes. We have inserted the authoritative cumulative counts and explicitly documented the removal of 338 duplicate records before the Version 3 release. The cumulative values should not be added together.

Relevant wording prepared in our working revision:

> The dataset expanded through four model-assisted rounds. Within each round, the model generated a provisional mask for a candidate patch, which an annotator accepted, corrected, removed, or redrew. Accepted labels informed the next model, which screened new patches and source scenes rather than repeatedly processing a fixed annotation pool.
>
> The cumulative annotation pool grew from the initial 300 hand-labeled patches to 1,409, 2,775, 7,989, and 19,793 patch records after Rounds 1–4, respectively. The four rounds added 1,109, 1,366, 5,214, and 11,804 records in sequence. Final consolidation removed 338 duplicate records, resulting in the 19,455 patches released as Version 3.

Table R1. Cumulative annotation growth in the working revision

Stage	New patch records	Cumulative records
Pilot (hand-labeled)	300	300
Round 1	1109	1409
Round 2	1366	2775
Round 3	5214	7989
Round 4	11804	19793
Deduplication	-338	19,455 (Version 3)

Prepared revised Figure 2: the four rounds and candidate-retention workflow

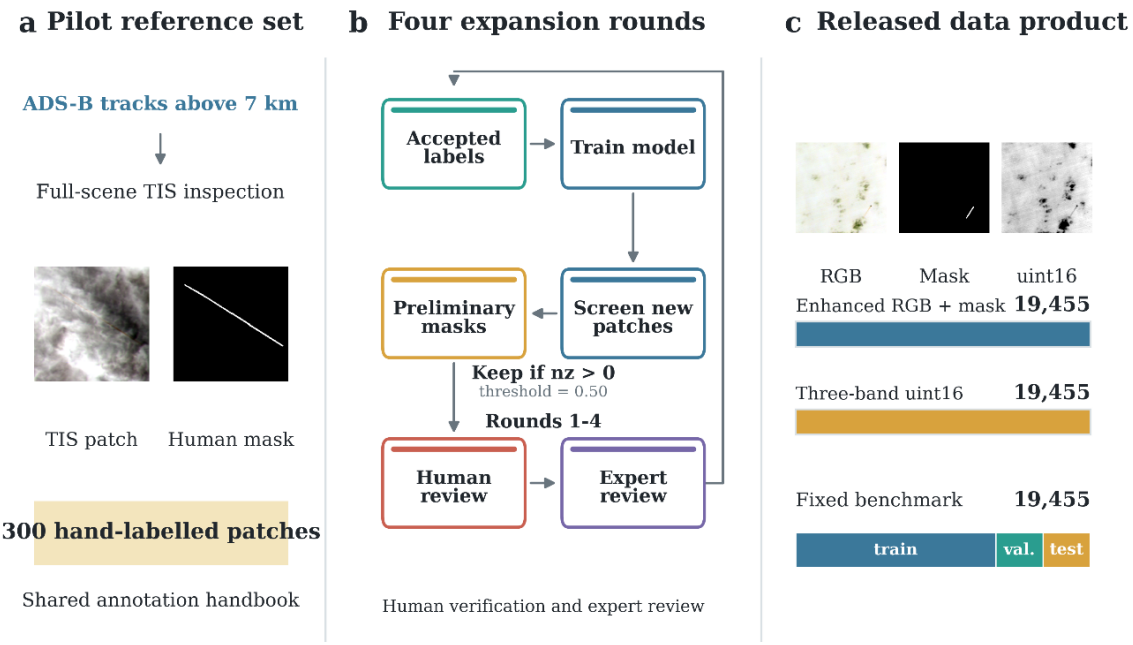

Figure 2. Construction of ContrailStruct30. (a) ADS-B-assisted full-scene interpretation and 300 initial hand-labeled patches. (b) Within each of four rounds, provisional masks were corrected on the corresponding candidate patches; accepted labels informed screening of new scenes and patches in subsequent rounds. A patch is retained when its preliminary binary mask contains at least one nonzero pixel (nz > 0) following the 0.50 probability threshold, without further area or shape filtering. Human reviewers accept, edit, redraw, or reject predictions and select recurrent false detections as difficult negatives. The cumulative pool reached 1,409, 2,775, 7,989, and 19,793 patch records after Rounds 1–4. (c) Following removal of duplicate records, released Version 3 provides 19,455 aligned RGB images, binary pixel masks, and uint16 image representations with fixed dataset partitions.

2. Accessibility of computer-vision terminology

2.1 What MaxViT–U-Net is

Referee comment (excerpt)

"Include references for what a MaxViT–U-Net is."

Author response

We have replaced the dense implementation-only explanation with a description of a feature encoder and a pixel-level decoder, and cite the original MaxViT and U-Net publications.

Relevant wording prepared in our working revision:

> MaxViT–U-Net combines a MaxViT-Base feature encoder, which integrates local image structure with wider spatial information, and a U-Net-style decoder that produces a contrail probability for each pixel (Ronneberger et al., 2015; Tu et al., 2022).

2.2 Contrail / non-contrail terminology

Referee comment (excerpt)

"Instead of 'foreground/background', how about 'positive/negative' or even 'contrail/non-contrail'?"

Author response

We now define the binary labels at the pixel level and use contrail/non-contrail where these words improve accessibility. We retain a small number of standard segmentation terms, such as foreground fraction, only where a quantitative definition requires them.

Relevant wording prepared in our working revision:

> ContrailStruct30 uses binary semantic segmentation labels: each pixel interpreted as contrail is assigned to the contrail class, and every other pixel to the non-contrail class.
>
> Pixel-level segmentation describes the target's spatial structure beyond image-level presence. This involves tracing identifiable contrail pixels, preserving non-contrail gaps between adjacent structures, and interpreting intersections, occlusions, and diffuse margins.

2.3 Definition of eight-connected components

Referee comment (excerpt)

"What does 'eight-connected' mean?"

Author response

The first substantive use now explains the concept in terms of neighboring pixels sharing an edge or a corner.

Relevant wording prepared in our working revision:

> Connectivity provides a complementary description of pixel masks. A single eight-connected contrail component—that is, a group of contrail pixels linked by shared edges or corners—occurs in 78.3% of positive masks, and the maximum count is 12 (Fig. 5b).

2.4 Explanation of timm, ImageNet-21K, and ImageNet-1K

Referee comment (excerpt)

"[Please] provide references and/or links [for] timm's maxvit_base_tf_512.in21k_ft_in1k ... ImageNet 21K ... ImageNet-1K."

Author response

We explain that timm is the PyTorch Image Models library and that ImageNet-21K and ImageNet-1K are large natural-image classification datasets used for pretraining and subsequent fine-tuning. The precise weight identifier remains in the companion code rather than being foregrounded in the Earth-science narrative.

Relevant wording prepared in our working revision:

> The encoder uses pretrained weights distributed through PyTorch Image Models (timm), an open-source model library (Wightman, 2019). Its MaxViT-Base backbone was pretrained on ImageNet-21K and subsequently fine-tuned on ImageNet-1K, two large-scale natural-image classification datasets (Deng et al., 2009).
>
> Implementation details, including the exact MaxViT weight identifier, are given in the companion code.

2.5 Definitions of segmentation metrics

Referee comment (excerpt)

"Define the metrics in Fig 7/Table 6."

Author response

Section 5.2 now defines TP, FP, FN, Dice, IoU, precision, recall, and difficult-negative activation, including their different sample denominators. Figure 7 has become a qualitative prediction plate; Table 6 gives the numerical benchmark and Table 7 defines the pixel-level AUPRC for its complementary check.

Relevant wording prepared in our working revision:

> We report mean positive-patch Dice as the primary metric. Dice is calculated within each positive patch as 2TP/(2TP + FP + FN) and then averaged over the 1,422 positive test patches. Mean positive-patch IoU is calculated analogously. Pixel-pooled IoU, precision, and recall use TP, FP, and FN counts pooled over all test pixels. Negative activation is the percentage of the 1,700 empty-reference patches containing at least one predicted foreground pixel. The table therefore reports both typical positive-patch performance and pooled pixel-level behaviour; mean positive-patch Dice remains the primary metric.
>
> Here, TP denotes correctly identified contrail pixels, FP denotes non-contrail pixels incorrectly predicted as contrails, and FN denotes reference contrail pixels missed by the prediction. The corresponding pixel-level definitions are IoU = TP/(TP + FP + FN), precision = TP/(TP + FP), and recall = TP/(TP + FN). IoU and Dice measure overlap with the reference contrail mask, precision characterizes false detections, and recall characterizes missed contrail pixels.

> Dice and intersection over union (IoU) were calculated from pooled pixel counts, and the area under the precision–recall curve (AUPRC) was evaluated at the pixel level.

3. Precision and consistency of language

3.1 Meaning of region, pixel, and difficult-negative image area

Referee comment (excerpt)

"What does 'region' mean ... in the first sentence ... [and] in the second sentence?"

Author response

The revised class definition refers to individual pixels. The selection of difficult negatives is described at the patch level, with image areas as the local features from which candidate patches were chosen.

Relevant wording prepared in our working revision:

> ContrailStruct30 uses binary semantic segmentation labels: each pixel interpreted as contrail is assigned to the contrail class, and every other pixel to the non-contrail class.
>
> The two experts jointly selected difficult-negative patches from image areas prone to human misinterpretation or repeated model false detections. These include natural cirrus, elongated cloud edges, linear surface boundaries, and image artifacts. Using the same interpretation rules as for positive samples, such patches were assigned empty contrail masks. They provide targeted examples for distinguishing contrails from visually similar backgrounds.

3.2 Definition of a source scene and its relation to image patches

Referee comment (excerpt)

"What is a 'scene' ... 'SDGSAT-1 takes images in X pixel by X pixel scenes, [which are] divided into 256 × 256 pixel patches'."

Author response

We now define a source scene as a complete TIS source observation before patch extraction. Source-image dimensions vary among products, so providing a single nominal source-scene pixel dimension would be misleading. We define the 256 × 256 image patches, their nominal footprint and crop procedure.

Relevant wording prepared in our working revision:

> Here, a source scene denotes a full TIS source observation before cropping into smaller image patches; source-image dimensions vary among products.
>
> All samples were cropped into 256 × 256-pixel patches, equivalent to approximately 7.68 km × 7.68 km at nominal sampling. Extraction combined overlapping grid crops with a 128-pixel stride and selection within approximately 1,024-pixel neighborhoods around candidate targets.

4. Highlighting the pixel-level contribution

4.1 Contrast with Yu et al. (2024) and the morphological value of pixel masks

Referee comment (excerpt)

"The big difference ... [is] per-pixel annotations ... Mention it more ... [link] detailed morphology ... [to] per-pixel level labels."

Author response

We have made per-pixel reference masks a central contribution from the opening abstract onward. The Introduction explicitly distinguishes our pixel-masked segmentation resource from the image-level SDGCONTR classification resource (Yu et al., 2024). The Discussion now connects widths, gaps, intersections, continuity, and local morphology specifically to the pixel masks, not merely to imagery.

Relevant wording prepared in our working revision:

> Aircraft contrails are an important source of aviation's non-CO2 climate effects, and their fine spatial structure requires observations and reference labels at the pixel level.

> The earlier SDGCONTR dataset addressed image-level recognition and classification in SDGSAT-1 imagery (Yu et al., 2024). ContrailStruct30 extends these observations with binary pixel-level masks that preserve each discernible target's width, continuity, boundaries, and spatial relationships.
>
> This observational capability is particularly relevant to narrow contrails. The median contrail-pixel fraction among positive patches is only 3.54%, indicating that many targets form sparse linear structures in a much larger background. Pixel-level masks make such targets explicit for both training and morphological analysis, rather than allowing their structure to disappear within a single image-level class label.
>
> Narrow, compact contrail shapes provide candidate observations for studies of newly formed or early-stage contrails. Spreading and deformation depend on atmospheric conditions and time since formation (Schumann and Heymsfield, 2017; Kärcher, 2018). Visible widths, adjacent lines, and local broadening can be related to flight tracks and meteorological evidence, but acquisition time, aircraft position, and transport conditions are needed to establish formation stage. The value of the pixel-level masks is their explicit spatial morphology, not an age assignment based on shape alone.

4.2 Table 1: explicit annotation type for every dataset

Referee comment (excerpt)

"I think every row should say either pixel masks, polygons, or image-level labels."

Author response

We have revised the annotation-type column so every dataset explicitly identifies its annotation representation. The complete revised Table 1 is reproduced below, including the distinction between polygon delineation and rasterized pixel masks.

Relevant wording prepared in our working revision:

> Table 1. Complementary contrail reference datasets. The annotation column explicitly distinguishes pixel-level masks, polygons, and image-level labels. Spatial sampling refers to the source sensor; the detail represented by the labels also depends on annotation and preprocessing. Dataset sizes follow the cited publications.

Table 1. Full revised dataset comparison, with explicit annotation types

Dataset	Observations	Annotated data	Annotation type and procedure	Context
OpenContrails (Ng et al., 2024)	GOES-16 ABI; 2 km infrared sampling at nadir	20,544 training and 1,866 validation samples	Pixel-level masks; four-annotator consensus	Eight-frame sequences and advected flight density
Landsat-8 (McCloskey et al., 2021)	30 m cirrus band; 100 m native thermal-infrared bands	4,289 scenes	Hand-drawn polygons, rasterized to pixel masks	Multispectral imagery and flight-density information
MSG/SEVIRI (Santos Gabriel et al., 2026)	SEVIRI; 3 km nominal sampling	140 scenes	Pixel-level masks; three independent annotations and consensus	Multitemporal imagery, satellite channels, and cloud properties
GVCCS (Jarry et al., 2026)	Visible whole-sky camera	122 video sequences	Instance polygons; expert review	Image sequences and flight-trajectory overlays
SDGCONTR (Yu et al., 2024)	SDGSAT-1 TIS	Image patches for classification	Image-level classification labels	Enhanced three-band imagery
ContrailStruct30 (this study)	SDGSAT-1 TIS; 30 m nominal sampling	19,455 patches	Binary pixel-level masks; model-assisted delineation, manual correction and expert review	Source-scene interpretation and corresponding uint16 images

5. Minor comments and presentation

5.1 Comparison of GEO and LEO contrail observations

Referee comment (excerpt)

"[The] introduction ... should probably include ... differences and relative merits of contrail detection on LEO vs GEO satellites."

Author response

We have added a concise observing-system comparison that describes the temporal advantage of GEO and the finer spatial sampling of LEO, presenting the two as complementary.

Relevant wording prepared in our working revision:

> Geostationary (GEO) instruments, including GOES-16 and MSG/SEVIRI, provide frequent views useful for following contrail evolution, but their kilometer-scale infrared sampling may not resolve narrow structures. Low Earth orbit (LEO) imagers provide finer spatial detail at the cost of less frequent observations of a given location. The two observing modes are therefore complementary (Meijer et al., 2022; Ng et al., 2024; Santos Gabriel et al., 2026; Euchenhofer et al., 2025).

5.2 Related machine-assisted contrail labeling literature

Referee comment (excerpt)

"Should probably cite https://doi.org/10.5194/egusphere-2026-4608 which also does some machine-assisted contrail labeling."

Author response

We have cited Dalmau et al. (2026) in the iterative-labeling discussion and included its EGUsphere preprint in the reference list. The wording identifies its ground-based, flight-informed scope without presenting it as identical to the SDGSAT-1 workflow.

Relevant wording prepared in our working revision:

> Flight-informed model-assisted contrail delineation has also been explored using ground-based imagery (Dalmau et al., 2026).

5.3 Clarification of ADSB.lol

Referee comment (excerpt)

"I didn't know what ADSB.lol was ... provide a citation or even just a hyperlink."

Author response

We define Automatic Dependent Surveillance–Broadcast (ADS-B) at first use, describe its use for source-scene searches, and give the service hyperlink.

Relevant wording prepared in our working revision:

> Historical Automatic Dependent Surveillance–Broadcast (ADS-B) aircraft position broadcasts obtained from ADSB.lol (https://adsb.lol/) helped locate candidate structures in full-scene imagery, producing an initial set of 300 manually annotated patches.

5.4 Figure 3c legend: Records to Patches

Referee comment (excerpt)

"Fig3c legend: records -> patches."

Author response

The Figure 3c legend has been changed to "Patches" and the caption now distinguishes source-scene entries from image patches. The underlying seasonal totals and interpretation are unchanged.

Relevant wording prepared in our working revision:

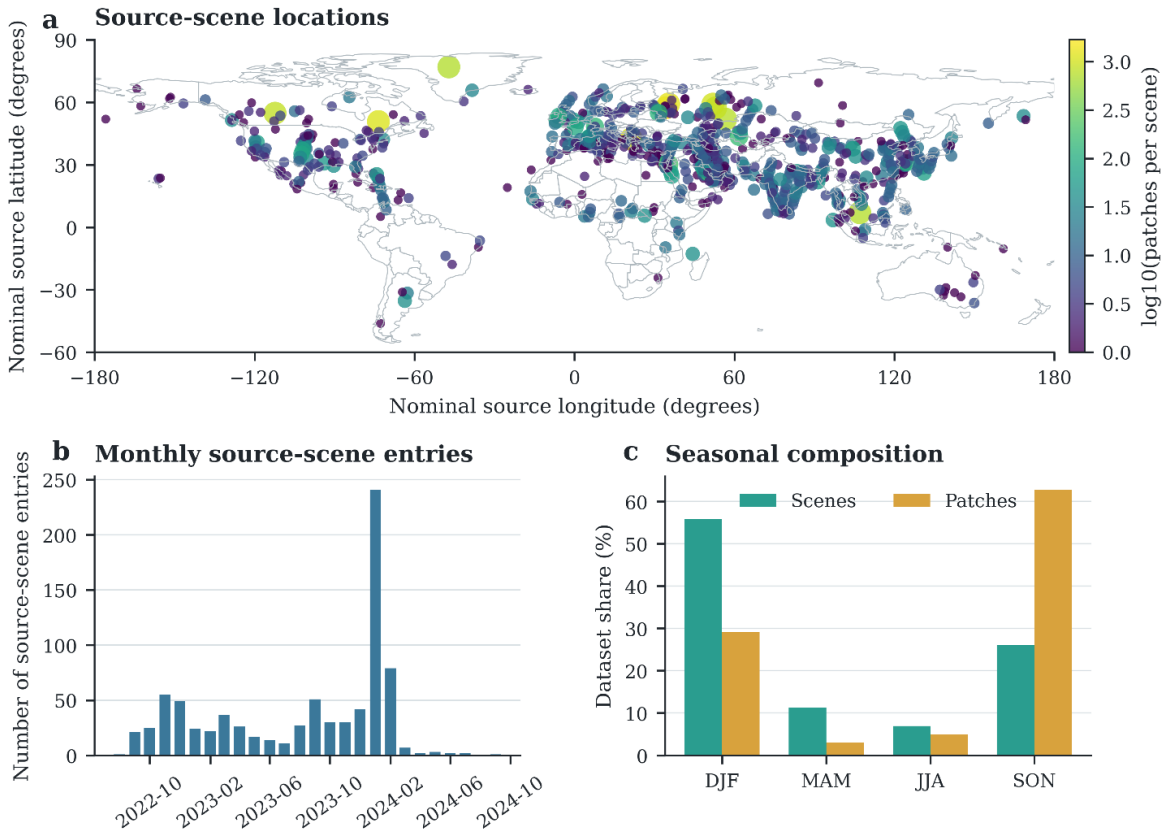

Figure 3. Spatial and temporal composition. (a) Nominal source-scene positions; symbol area scales with the square root of contributed patch count and color indicates its base-10 logarithm. Country boundaries are from Natural Earth. (b) Monthly source-scene entries. (c) Seasonal proportions of scenes and patches. These distributions describe a purposefully assembled dataset rather than climatological contrail occurrence.

5.5 Figure 6: visibility of empty and nonempty masks

Referee comment (excerpt)

"Fig 6: I don't see much difference between b/c and e/f. Is there supposed to be a contrail mask in the blue box?"

Author response

We have revised the figure to show explicitly that panel (c) has zero contrail pixels and to enlarge the small labeled segment in panel (f). We also clarified the meaning of the cyan location boxes and coral reference-mask overlays.

Relevant wording prepared in our working revision:

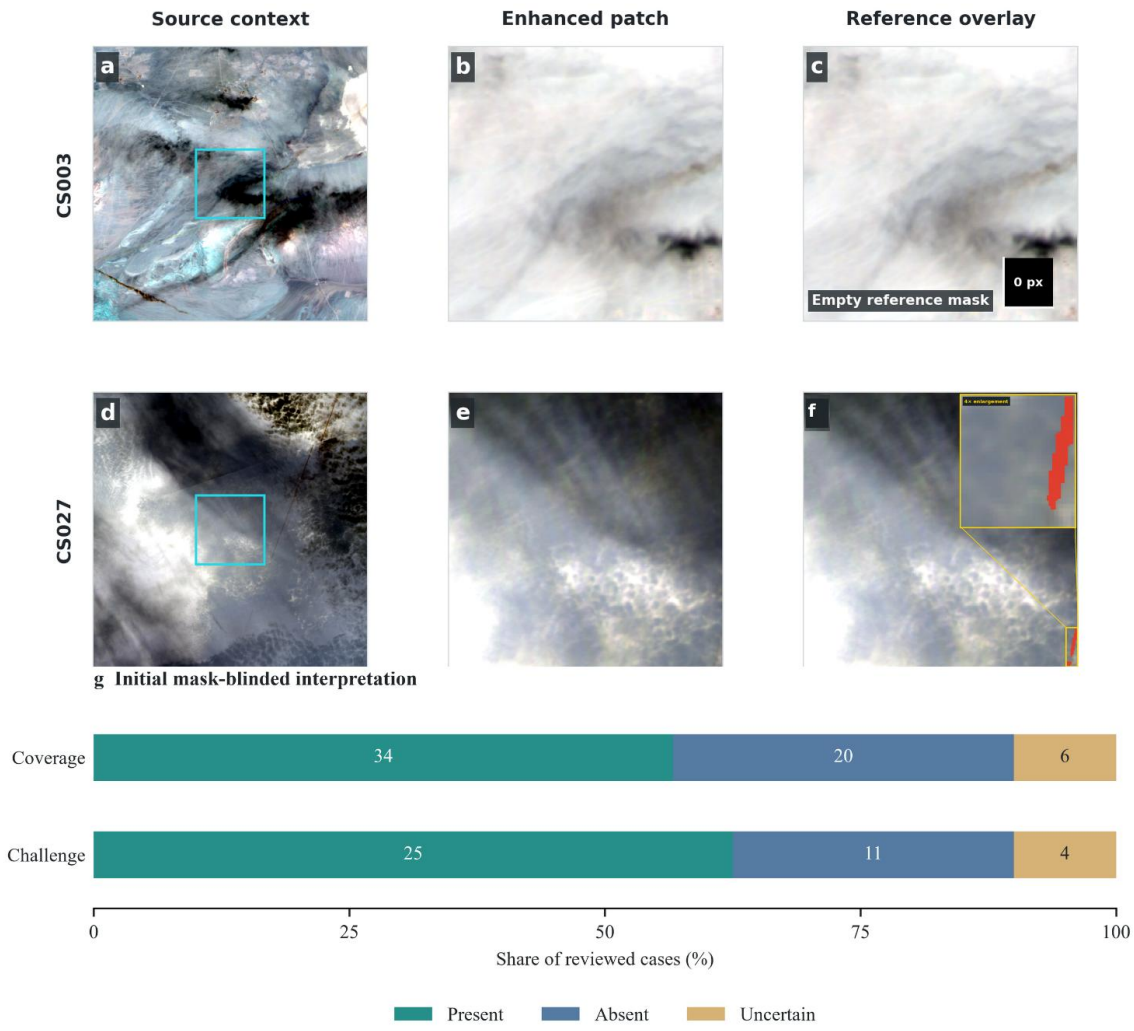

Figure 6. Source-scene context and expert interpretation. (a–c) A patch initially judged to contain a contrail but retained with an empty reference mask after contextual review; panel (c) explicitly indicates zero contrail pixels. (d–f) A case initially judged empty but confirmed to contain a small contrail segment, with a 4× enlargement of the labeled location in (f). Columns show source-scene context, enhanced patches, and reference overlays. Cyan boxes mark patch locations, and coral denotes reference contrail pixels. Context images use per-band 2nd–98th percentile stretching. (g) Initial mask-blinded judgments in the coverage and challenge groups.

Closing comment

We thank the referee again for these constructive suggestions. The wording reproduced above is part of our internal working revision and is supplied here to make the planned changes transparent during the public discussion. We will finalize and submit the revised manuscript and its marked changes at the appropriate post-discussion revision stage, taking into account any further referee, community, and editor comments.

On behalf of the co-authors,
The authors of ContrailStruct30