S1 Transfer learning for contrail detection

The pretrained CoaT segmentation model Ortiz et al. (2025) was adapted to MSG imagery using the labelled dataset by Santos Gabriel et al. (2026). All scenes acquired on the same UTC day were assigned to the same partition, yielding 39 training, 14 validation and 14 test days. Training augmentation comprised joint rotations in multiples of 90° and horizontal flips of the images and corresponding masks. Transfer learning followed two stages. First, the encoder was frozen and the segmentation decoder was trained for five epochs with a learning rate of 10^{-4}. This stage adapted the decoder to MSG imagery while preserving the pretrained encoder features. The encoder was then unfrozen, and the complete model was fine-tuned with encoder and decoder learning rates of 10^{-6} and 10^{-5}, respectively. The smaller encoder learning rate limited changes to the pretrained representation while allowing adaptation to the target imagery. Optimization was performed using AdamW with a weight decay of 0.01 and an equally weighted combination of binary cross-entropy (BCE) and Dice losses:

$$\mathcal{L} = \frac{1}{2}\mathcal{L}_{\mathrm{BCE}} + \frac{1}{2}\mathcal{L}_{\mathrm{Dice}}, \tag{S1}$$

where

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right], \tag{S2}$$

and

$$\mathcal{L}_{\mathrm{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i y_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \epsilon}. \tag{S3}$$

Here, N is the number of pixels, p_i is the predicted contrail probability, $y_i \in \{0,1\}$ is the corresponding ground-truth label, and ϵ is a small constant introduced for numerical stability.

Binary cross-entropy penalizes incorrect pixel predictions, including false detections in negative scenes, while Dice encourages overlap with the sparse contrail labels. Fine-tuning allowed up to 30 epochs, with early stopping after five epochs without improvement in validation loss. Global probability thresholds were selected by maximizing validation Dice. The resulting thresholds were 0.20 before transfer learning and 0.55 after transfer learning.

Figure S1 compares the models on the same 23 held-out test scenes from 14 days. At their respective validation-selected thresholds, transfer learning increased precision from 0.416 to 0.664, recall from 0.644 to 0.679, and Dice from 0.506 to 0.671. Boundary sof dice with a one-pixel matching tolerance increased from 0.728 to 0.826, indicating closer agreement between predicted and labelled boundaries. False-positive pixels decreased from 46.9 to 17.8 per 10,000 valid pixels, a reduction of approximately 62%. Thus, the improvement mainly reflects fewer false detections, accompanied by a smaller increase in the fraction of labelled contrail pixels recovered.

The lower panels compare predicted and labelled contrail coverage for each test scene. Coverage is defined here as the percentage of pixels labelled as contrail. Before transfer learning, predicted coverage frequently exceeded labelled coverage, consistent with the lower precision. After transfer learning, the points lie closer to the one-to-one line. Mean coverage bias decreased from $+0.284$ to $+0.011$ percentage points, while mean absolute error and root-mean-square error decreased from 0.311 to 0.086 and from 0.549 to 0.142 percentage points, respectively. These results indicate improved agreement in scene-level contrail coverage as well as pixel-level segmentation. Coverage agreement alone can conceal compensating false positives and missed detections, so it should be interpreted together with precision, recall and Dice values from the panels in the first row.

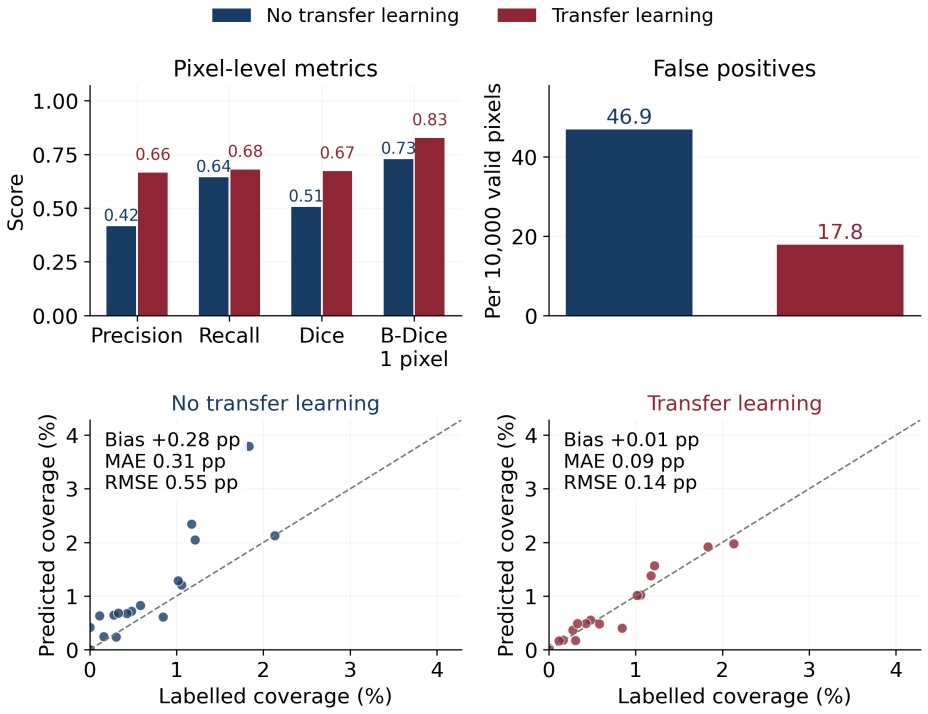


Figure S1. Comparison before transfer learning (dark blue) and after transfer learning (dark red) on 23 held-out MSG dataset test scenes. Global probability thresholds of 0.20 and 0.55, respectively, were selected using validation data only. Top left: pixel-level precision, recall and Dice, together with boundary soft dice at a one-pixel tolerance (labelled B-Dice in the panel). Top right: false-positive pixels per 10,000 valid pixels. Bottom left and right: predicted versus labelled contrail coverage before and after transfer learning, respectively. Each dot represents one scene; dashed lines indicate equal predicted and labelled coverage. Coverage is expressed as a percentage of valid pixels. Bias, mean absolute error (MAE) and root-mean-square error (RMSE) are calculated across scenes and expressed in percentage points (pp).

S2 Detection threshold selection for central, lower, and upper contrail coverage estimates

Throughout the main document, contrail probabilities are binarized at a threshold of 0.55, which maximized the validation-set Dice score (0.671). To assess sensitivity to this choice, coverage was also evaluated at thresholds of 0.05 and 0.95, which maximized recall (0.747) and precision (0.767), respectively, among the tested thresholds $(0.05, 0.10, \ldots, 0.95)$, while maintaining Dice scores near 0.65 (Figure S2). Compared with the central threshold, the lower threshold increased recall from 0.679 to 0.747 but reduced precision from 0.664 to 0.566, yielding greater detected coverage. Conversely, the higher threshold increased precision to 0.767 but reduced recall to 0.561, yielding lower coverage. Both settings maintained Dice scores within 0.028 of the maximum. These alternative coverage estimates characterize sensitivity to the precision–recall trade-off rather than statistical confidence intervals or bounds on true contrail coverage.

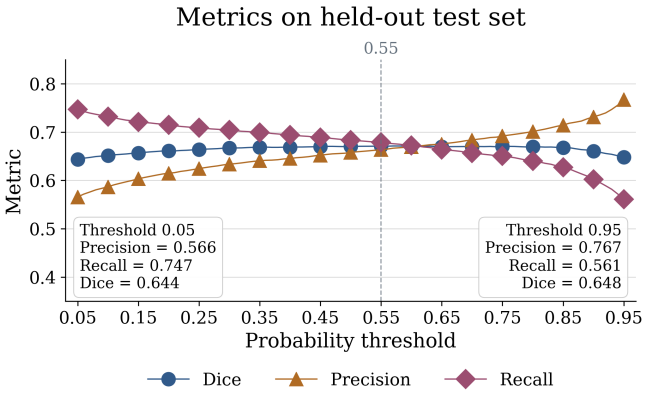


Figure S2. Threshold dependence of recall, precision, and Dice score for the transfer-learned contrail detector on the test set. The vertical dashed line marks the threshold maximizing Dice.

S3 Cloud retrievals

This section evaluates the discrepancies between OCA cloud-property retrievals and the CiPS (reference retrievals for high-altitude cirrus clouds observed by SEVIRI). It also examines how these discrepancies propagate into radiative forcing estimates, highlighting the impact of inaccurate cloud characterization and motivating the use of the regression-based calibration adopted throughout the main manuscript.

S3.1 Comparison of OCA with CIPS

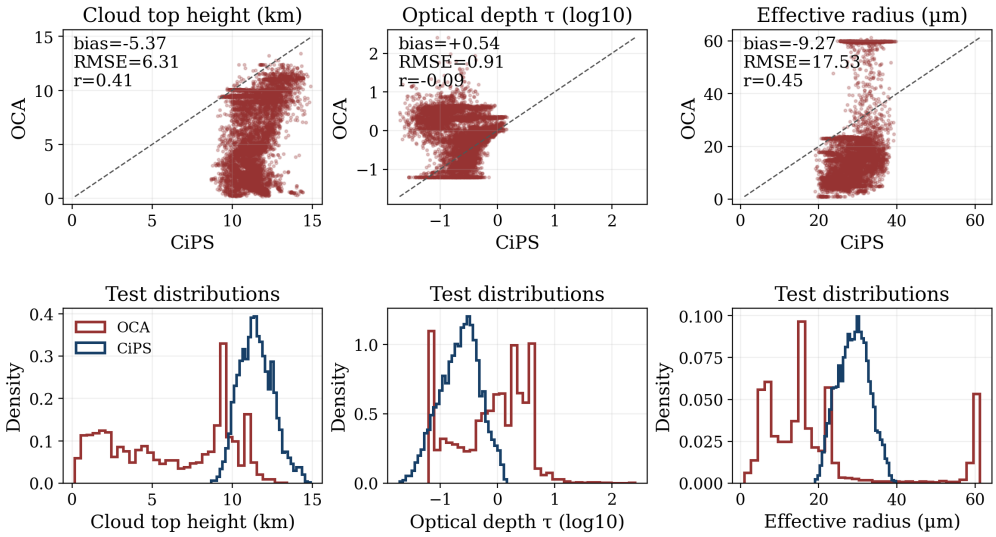


Figure S3. Comparison of original OCA retrievals with CiPS reference retrievals for cloud-top height, logarithmic optical depth, and effective radius on the held-out test set of human-labeled SEVIRI images. Upper panels show pixel-wise scatterplots with the 1:1 reference line, with annotations reporting bias, RMSE, and Pearson correlation. Lower panels show the corresponding OCA and CiPS distributions. Optical depth is expressed as $\log_{10}(\tau)$.

Figure S3 reveals substantial discrepancies between OCA and CiPS cloud-property retrievals for the labelled contrail pixels in the test set of dataset by Santos Gabriel et al. (2026). The retrievals show weak correlations for cloud-top height ($r = 0.41$) and effective radius ($r = 0.45$), and no correlation for optical depth ($r = -0.09$). OCA systematically underestimates cloud-top height (bias of -5.37 km) and overestimates optical depth (bias of $+0.54$ in $\log_{10}$ units), frequently producing values inconsistent with the expected properties of high-altitude, optically thin contrail cirrus (Schumann et al., 2017). The effective radius estimates also exhibit substantial discrepancies, with OCA values clustering near the smallest possible sizes.

S3.2 Effect of the cloud retrieval on the radiative forcing estimates

Here, we show the impact of inaccuracies in the OCA-retrieved contrail properties on the resulting radiative forcing estimates, motivating the use of the linear regression model presented in the main text to obtain cloud properties consistent with those retrieved by the established CiPS cirrus retrieval algorithm.

S3.2.1 Comparison between OCA and CiPS-calibrated retrieval on the benchmark dataset

Using the test set of Santos Gabriel et al. (2026), we assess the impact of calibrating OCA retrievals on radiative forcing estimates. The original OCA retrievals show substantial deviations from the CiPS reference across UTC, viewing zenith angle, and solar zenith angle, although no clear trends emerge, as radiative forcing depends on the combined influence of cloud properties and local radiative conditions. In contrast, the calibrated retrievals consistently yield mean differences close to zero, demonstrating improved agreement also in the radiative forcing estimates across the different examined conditions.

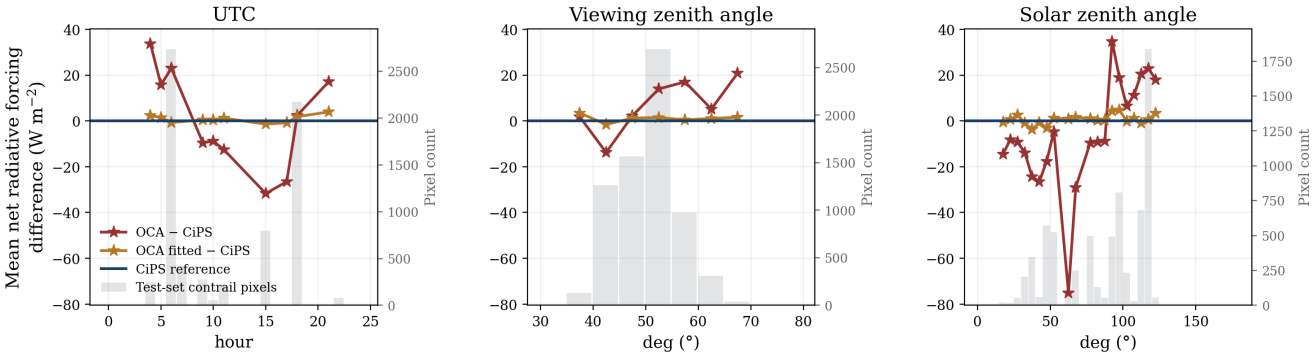


Figure S4. Deviation in net radiative forcing estimates obtained using raw OCA retrievals (red) and fitted OCA values (orange) relative to the CiPS reference (blue), as a function of UTC hour (first panel), satellite viewing zenith angle (second panel), and solar zenith angle (third panel). Light-gray background bars indicate the number of pixels in the test set corresponding to each interval. The fitted OCA values show deviations close to zero across all regimes, whereas the raw OCA retrievals exhibit substantially larger discrepancies relative to CiPS.

S3.2.2 Comparison between OCA and CiPS-calibrated retrieval on the 2023 dataset

Figure S5 compares monthly radiative forcing (RF) estimates obtained using the original and CiPS-calibrated OCA retrievals. The first row presents mean RF within detected contrail pixels, while the second row shows the corresponding regional means over the full Euro–Atlantic domain, as reported in the main manuscript. At the pixel level, the original OCA retrievals slightly overestimate longwave warming but substantially overestimate shortwave cooling, particularly during spring and summer. Consequently, net RF is strongly underestimated, with negative values during several months that turn positive after calibration. This discrepancy also affects the regional estimates, where the stronger shortwave cooling from the original OCA retrievals produces negative net RF from April through August, in contrast to the consistently positive values obtained with the calibrated retrievals. The largest negative regional estimate occurs in May, the month with the highest detected coverage.

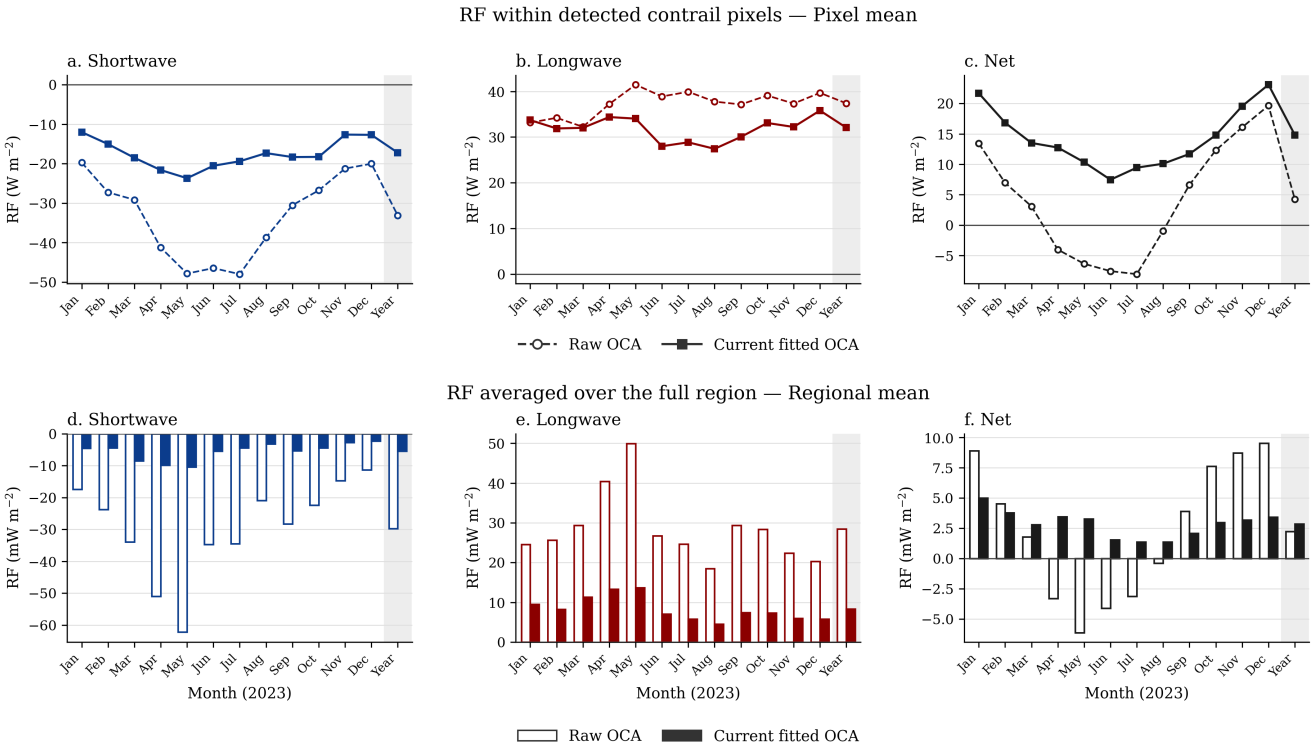


Figure S5. Monthly and annual shortwave, longwave, and net radiative forcing (RF) estimates derived from raw and fitted OCA retrievals for 2023. (a–c) Mean RF within detected contrail pixels (W m^{-2}), reflecting the effects of contrail optical properties and background atmospheric conditions alone. (d–f) Regional RF estimates (mW m^{-2}), accounting for the combined effects of contrail coverage, optical properties, and background atmospheric conditions, as described in Equation 7 of the main text.

S4 Regional contrail coverage and distance flown

Table S1. Regional contrail coverage, cruise distance, the fraction of cruise distance within persistent contrail-forming regions (PCR), and accumulated contrail area per cruise distance for 2023, corresponding to panels (d)–(f) of the six-panel figure. The regional box limits in panel (a) are listed below. All boxes lie within 20–60°N. Detections use CoaT transfer-learning masks binarized at $p \geq 0.55$.

Region	Mean coverage (%)	Contrail-pixel share (%)	Cruise distance (km)	Distance share (%)	Cruise distance within PCR (%)	Area / total cruise distance ($m^2\,m^{-1}$)	Area / distance inside PCR ($m^2\,m^{-1}$)
N. Atlantic (40–60°N)	0.068	29.024	1 161 971 925.109	14.770	6.369	152.801	2399.275
N. Atlantic (20–40°N)	0.072	33.376	254 695 614.839	3.238	5.442	992.404	18234.445
Western Europe	0.103	23.917	2 247 263 796.113	28.566	6.752	69.693	1032.179
North Africa	0.005	1.339	27 812 635.294	0.354	1.526	393.229	25768.123
Mediterranean / S. Europe	0.026	6.796	2 283 098 323.172	29.021	3.392	20.936	617.211
Central–Eastern Europe	0.023	5.549	1 892 106 149.808	24.051	5.394	15.502	287.378
Total	–	100.000	7 866 948 444.335	100.000	5.333	85.802	1608.920

Region	$\lambda_{\min}$ (°)	$\lambda_{\max}$ (°)	$\phi_{\min}$ (°N)	$\phi_{\max}$ (°N)
N. Atlantic (40–60°N)	-60.000	-12.000	40.000	60.000
N. Atlantic (20–40°N)	-60.000	-12.000	20.000	40.000
Western Europe	-12.000	5.000	30.000	60.000
North Africa	-12.000	40.000	20.000	30.000
Mediterranean / S. Europe	5.000	40.000	30.000	46.000
Central–Eastern Europe	5.000	40.000	46.000	60.000

S4.1 Distance flown inside persistent contrail forming regions

Table S2. Fraction of cruise distance within persistent contrail-forming regions (PCR) in 2023, calculated as $100\,D_{\mathrm{PCR},r}/D_r$. The total uses summed PCR distance divided by summed cruise distance.

Subregion	Within PCR (%)
N. Atlantic (40–60°N)	6.369
N. Atlantic (20–40°N)	5.442
Western Europe	6.752
North Africa	1.526
Mediterranean / S. Europe	3.392
Central–Eastern Europe	5.394
Total	**5.333**

S4.2 Land–ocean partition of regional contrail-pixel shares

Table S3. IGBP land–ocean partition of regional grid pixels and detected contrail pixels in 2023. The two within-region pairs each sum to 100%. The final land and ocean columns instead share a domain-wide denominator; their sum reproduces the regional contrail-pixel share.

	Regional grid pixels (%)		Regional detections (%)		Share of all domain detections (%)		
Subregion	Land	Ocean	Land	Ocean	Land	Ocean	Total
N. Atlantic, 40–60°N	2.78	97.22	0.15	99.85	0.043	28.981	29.024
N. Atlantic, 20–40°N	2.91	97.09	0.23	99.77	0.077	33.299	33.376
Western Europe	46.58	53.42	37.70	62.30	9.016	14.901	23.917
North Africa	95.58	4.42	82.91	17.09	1.110	0.229	1.339
Mediterranean / S. Europe	52.39	47.61	32.71	67.29	2.223	4.573	6.796
Central–Eastern Europe	84.92	15.08	73.16	26.84	4.059	1.490	5.549
Total	37.47	62.53	16.53	83.47	16.528	83.472	100.000

Table S4. Partition of detected contrail pixels among land, Atlantic waters, and other water bodies within western Europe in 2023. The two rows use different denominators, as indicated.

Percentage denominator	Land	Ocean / Atlantic	Ocean / Non-Atlantic	Total
Western-European detections (%)	37.696	56.387	5.917	100.000
All domain detections (%)	9.016	13.486	1.415	23.917

S4.2.1 Radiative forcing in the western Europe region

Table S5. Contributions of land, Atlantic waters, and other water bodies within the western European region to the annual signed net radiative forcing in 2023. IGBP class 17 defines water. Radiative forcing (RF) is weighted by physical pixel area and averaged equally over the all the yearly scenes. The two rows use different percentage denominators.

Percentage denominator	Land	Ocean / Atlantic	Ocean / Non-Atlantic	Total
Western-European net RF (%)	43.56	52.68	3.76	100.00
All domain net RF (%)	10.89	13.18	0.94	25.01

S5 Comparison with Low-Earth-Orbiting satellite data

The marked geographical differences in coverage may partly reflect satellite resolution and viewing-geometry limitations. We therefore sampled higher-resolution overpasses to compare the geographical distribution directly with the 2023 geostationary results. The data used for this comparison comes from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard Suomi NPP. VIIRS has a nominal spatial resolution of approximately 750 m at nadir, about four times finer than SEVIRI. Following Euchenhofer et al. (2025), we use the NASA resampled VIIRS product on a regular latitude–longitude grid with an approximate spacing of 500 m. To capture seasonal variability, we analyze four one-week samples from 2023: 25–31 January, April, July, and October, and compare the resulting patterns with those from SEVIRI.

To quantify contrail coverage, we first verified that the detection algorithm performed sufficiently well on higher-resolution VIIRS data using the dataset of Euchenhofer et al. (2025). Figure S6 shows a mean Dice score of 48.2 %, increasing to 61.7%

with a 2-pixel tolerance. Although the model tends to overestimate contrail width and performs less well than on geostationary data, its performance is sufficient for comparing the geographical distribution of detected contrails across sensors.

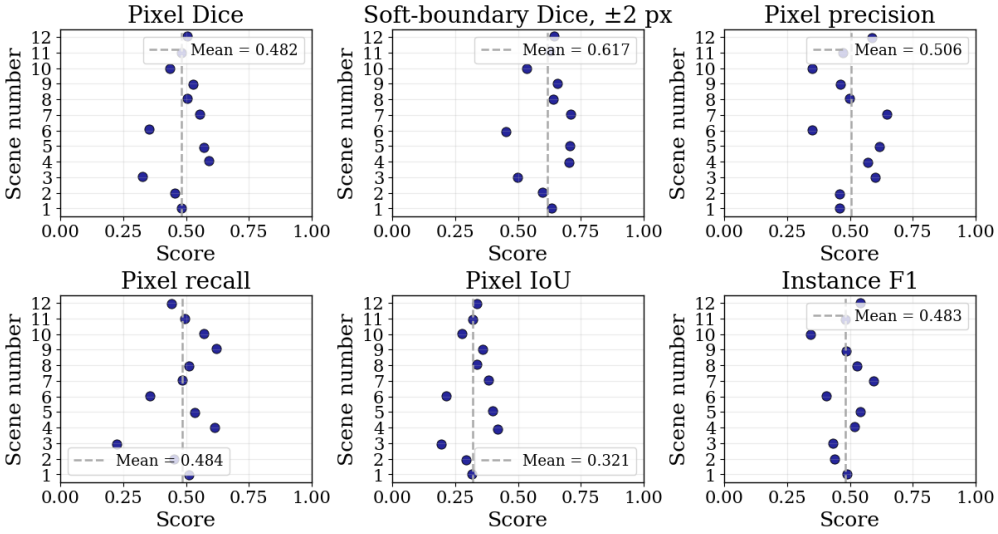


Figure S6. Performance of the contrail-detection algorithm on 12 VIIRS validation scenes. Points show scene-level pixel Dice, soft-boundary Dice with a ± 2-pixel tolerance, precision, recall, intersection over union (IoU), and instance F1 scores. Dashed lines indicate the mean value across all scenes.

The VIIRS distributions shown in Figure S7 are more broadly spread in longitude than the strong SEVIRI concentration over the eastern North Atlantic and the longitudes around Portugal, Spain, and France. VIIRS also shows relatively more detections above approximately $55°$N, mainly over oceanic regions, supporting an association with transatlantic traffic. This stronger high-latitude signal suggests that SEVIRI may underestimate contrail occurrence there because of pixel enlargement and geometric distortion toward the edge of its field of view.

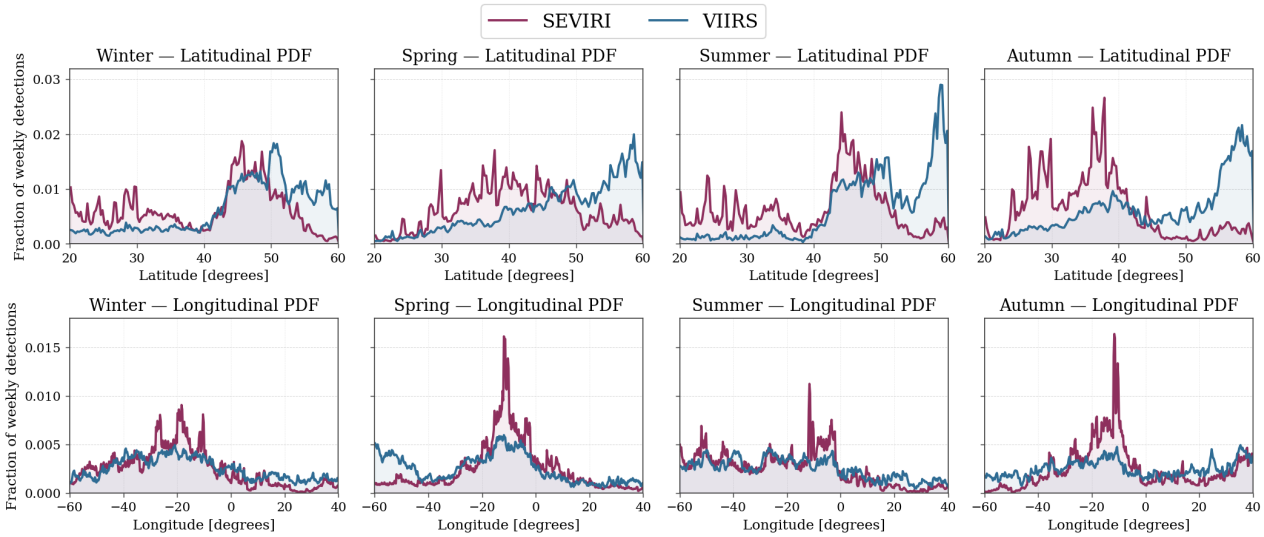


Figure S7. Seasonal latitudinal and longitudinal probability density distributions of contrail detections derived from SEVIRI and VIIRS for the selected representative weeks. VIIRS shows a weaker longitudinal contrast but retains higher detection probabilities over the eastern North Atlantic and the longitudes around Portugal, Spain, and France than over eastern Europe. It also shows relatively more detections above approximately 50°N, suggesting a possible high-latitude detection limitation in SEVIRI.

S6 Contrail coverage location with respect to jet axis

Figure S8 presents the contrail detection probability maps shown in Fig. 4 of the main manuscript, but averaged over one-week periods to reveal spatial patterns that may be hidden by longer-term averaging. The weekly mean jet-stream axis and its variability are also included to explore potential relationships with contrail detection hotspots. Although the spatial correspondence is not straightforward, regions of enhanced detection probability tend to occur near or south of the jet-stream axis, consistent with the preferential occurrence of ice-supersaturated regions on its southern side (Irvine et al., 2012). Over eastern Europe, increased contrail detections also appear to coincide with a northward displacement of the jet stream. The association of jet streams and frontal systems with extensive natural cirrus cloud cover may partly explain the lower detection probabilities inside the jet itself, as these clouds can reduce the contrast between contrails and their background. Consequently, the observed patterns may reflect both atmospheric conditions favourable for contrail persistence and variations in satellite detectability. A more detailed analysis accounting for the instantaneous jet-stream position, cloud conditions, and air traffic would be needed to clarify these relationships in future studies.

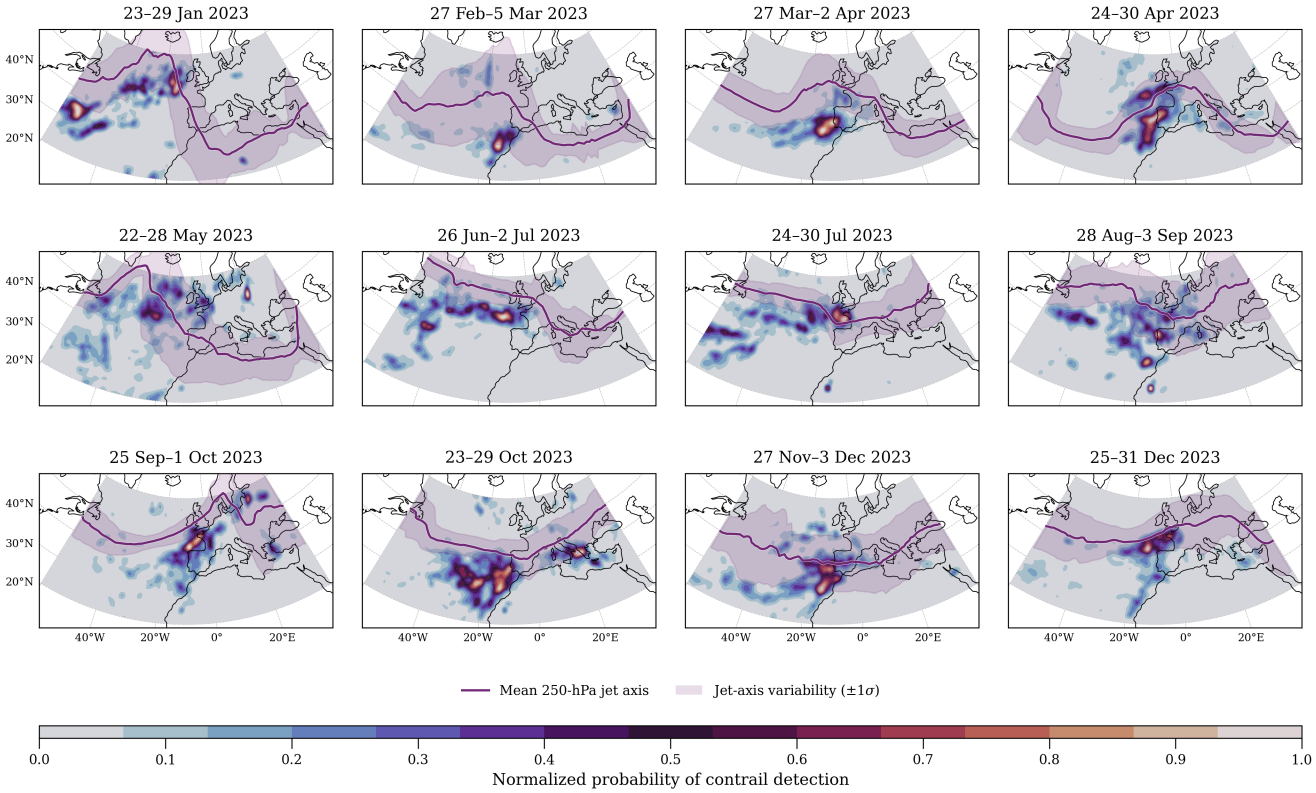


Figure S8. Normalized contrail detection frequency and jet-stream position during the first Monday–Sunday week starting in each month of 2023. Contrail cover is calculated from detections at a probability threshold of 0.55. Each weekly field is independently scaled from 0 to 1, showing relative spatial patterns within that week rather than absolute differences between weeks. The purple line indicates the weekly mean 250-hPa jet-stream axis, while the shaded band represents its ± 1 standard deviation in position. Panel titles give the corresponding date ranges.

S7 Monthly coverage

Table S6. Monthly values for all curves in the contrail-cover and cruise-distance figure for 2023. All cover values are percentages; cruise distance is the monthly mean daily distance in million km day^{-1}. COAT detections use $p \geq 0.55$. Values are rounded to three decimals.

	Panel (a)		Panel (b): regional mean cover (%)					
Month	Domain cover (%)	Cruise distance (million km day^{-1})	N. Atlantic 40–60°N	N. Atlantic 20–40°N	W. Europe	N. Africa	Mediterranean/ S. Europe	Central–E. Europe
Jan	0.047	22.665	0.078	0.072	0.064	0.004	0.016	0.005
Feb	0.048	23.390	0.078	0.073	0.047	0.007	0.029	0.014
Mar	0.065	24.525	0.036	0.105	0.178	0.003	0.039	0.013
Apr	0.081	25.668	0.076	0.114	0.205	0.003	0.039	0.036
May	0.084	27.246	0.117	0.132	0.123	0.008	0.024	0.043
Jun	0.057	29.241	0.084	0.068	0.108	0.003	0.030	0.036
Jul	0.046	30.729	0.106	0.059	0.051	0.005	0.002	0.012
Aug	0.041	30.179	0.086	0.035	0.091	0.006	0.006	0.017
Sep	0.053	28.646	0.059	0.057	0.094	0.008	0.036	0.076
Oct	0.043	27.379	0.031	0.057	0.108	0.004	0.036	0.021
Nov	0.038	27.646	0.028	0.051	0.104	0.007	0.032	0.006
Dec	0.031	28.051	0.034	0.040	0.068	0.007	0.023	0.004

S8 Regional radiative forcing figure

Table S7. Monthly and annual regional shortwave, longwave, and net contrail radiative forcing for 2023. All forcing values are in $\mathrm{mW\,m^{-2}}$. The central estimate uses $P > 0.55$; the confidence limits describe its 95% scene-bootstrap confidence interval. Means at $P > 0.05$ and $P > 0.95$ describe detection-probability sensitivity.

Statistic	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Shortwave													
Mean $P > 0.55$	-4.52	-4.50	-8.52	-9.90	-10.42	-5.51	-4.40	-3.21	-5.38	-4.42	-2.77	-2.35	-5.48
95% CI lower	-4.88	-4.76	-8.98	-10.40	-10.86	-5.76	-4.63	-3.37	-5.67	-4.66	-2.95	-2.52	-5.58
95% CI upper	-4.19	-4.23	-8.07	-9.40	-9.99	-5.26	-4.19	-3.05	-5.11	-4.18	-2.59	-2.19	-5.39
Mean $P > 0.05$	-8.96	-9.19	-18.01	-21.03	-22.18	-11.46	-9.07	-6.57	-11.67	-10.07	-6.35	-5.34	-11.64
Mean $P > 0.95$	-1.77	-1.82	-2.87	-3.48	-3.82	-2.10	-1.71	-1.25	-1.96	-1.52	-0.94	-0.80	-2.00
Longwave													
Mean $P > 0.55$	9.53	8.28	11.33	13.35	13.69	7.05	5.77	4.56	7.47	7.38	5.95	5.78	8.33
95% CI lower	9.14	7.96	10.99	13.02	13.35	6.83	5.59	4.41	7.24	7.16	5.74	5.58	8.24
95% CI upper	9.92	8.60	11.69	13.71	14.04	7.26	5.96	4.71	7.72	7.62	6.17	5.98	8.42
Mean $P > 0.05$	19.62	17.38	23.93	28.58	29.10	14.71	11.85	9.31	16.27	16.84	13.68	13.31	17.85
Mean $P > 0.95$	3.72	3.28	4.02	4.80	5.06	2.71	2.27	1.80	2.78	2.61	2.08	2.05	3.09
Net													
Mean $P > 0.55$	5.01	3.78	2.80	3.45	3.27	1.53	1.37	1.35	2.09	2.97	3.18	3.42	2.85
95% CI lower	4.77	3.52	2.54	3.15	2.96	1.37	1.23	1.24	1.92	2.78	3.02	3.26	2.79
95% CI upper	5.24	4.04	3.08	3.75	3.58	1.70	1.51	1.46	2.26	3.16	3.35	3.60	2.91
Mean $P > 0.05$	10.66	8.19	5.92	7.55	6.91	3.25	2.78	2.74	4.61	6.77	7.33	7.97	6.21
Mean $P > 0.95$	1.95	1.46	1.15	1.32	1.24	0.61	0.56	0.55	0.82	1.09	1.14	1.25	1.09

References

Euchenhofer, M. V., Prashanth, P., Parke, S. A., Eastham, S. D., and Waitz, I. A.: Contrail Observation Limitations Using Geostationary Satellites, Geophysical Research Letters, 52, e2025GL118 386, https://doi.org/10.1029/2025GL118386, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2025GL118386, 2025.

Irvine, E. A., Hoskins, B. J., and Shine, K. P.: The dependence of contrail formation on the weather pattern and altitude in the North Atlantic, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL051909, 2012.

Ortiz, I., García-Heras, J., Jafarimoghaddam, A., and Soler, M.: Robust Evaluation of OpenContrails Neural Networks with Temporal Corrections via Optical Flow, IEEE Transactions on Geoscience and Remote Sensing, pp. 1–1, https://doi.org/10.1109/TGRS.2025.3629628, 2025.

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

Schumann, U., Baumann, R., Baumgardner, D., Bedka, S. T., Duda, D. P., Freudenthaler, V., Gayet, J.-F., Heymsfield, A. J., Minnis, P., Quante, M., Raschke, E., Schlager, H., Vázquez-Navarro, M., Voigt, C., and Wang, Z.: Properties of individual contrails: a compilation of observations and some comparisons, Atmospheric Chemistry and Physics, 17, 403–438, https://doi.org/10.5194/acp-17-403-2017, 2017.