Supplement of Atmos. Meas. Tech., 17, 6145–6162, 2024
https://doi.org/10.5194/amt-17-6145-2024-supplement
© Author(s) 2024. CC BY 4.0 License.

Atmospheric Measurement Techniques
Open Access

Supplement of

Contrail altitude estimation using GOES-16 ABI data and deep learning

Vincent R. Meijer et al.

Correspondence to: Vincent R. Meijer (vrmeijer@tudelft.nl)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Locating contrails in CALIOP data

The process of locating contrails detected in GOES-16 ABI imagery in CALIOP LIDAR data involved two steps. The first "coarse" step considers whether contrails are detected within 50 km of the CALIPSO ground track as it passes over the domain, which is defined by the bounds of the orthographic projection used for the contrail detections. This is done by using the contrails detected at the time corresponding to approximately halfway the CALIPSO overpass. Given that such an overpass lasts at most 30 minutes, the resulting maximum difference between a local CALIOP measurement time and the detection time is 15 minutes. Combined with a maximum wind speed of about 60 m/s, this leads to a distance threshold of about 50 km. An example of two coarse collocations, one with and one without collocated contrails, is shown in Figure S1.

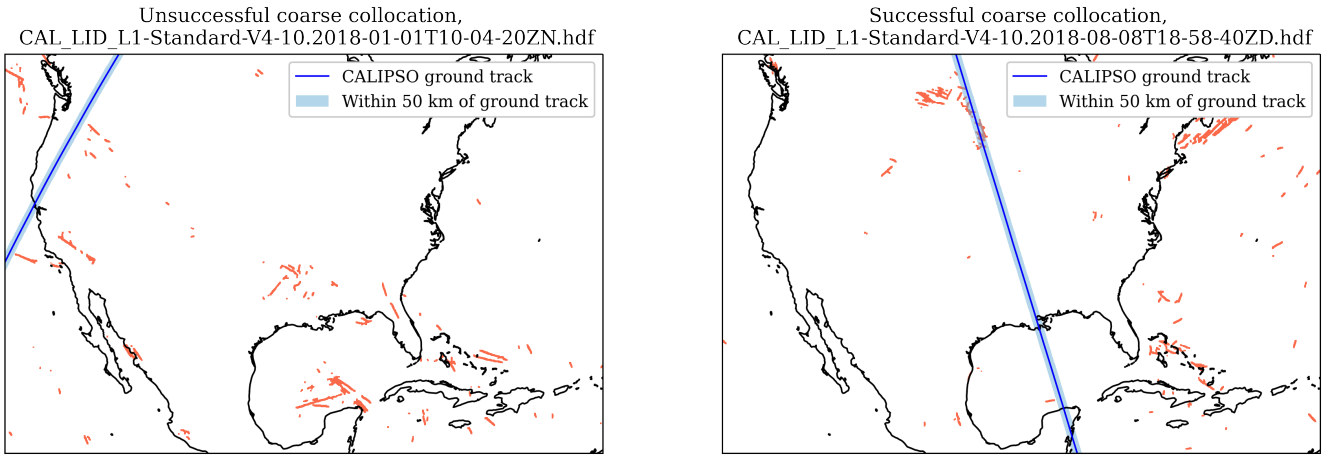


Figure S1. Examples of the coarse collocation process, which ascertains whether contrails are found within 50 km of the CALIPSO ground track corresponding to a particular overpass. The contrail detections that are used for this process are shown in red, and have been dilated using a 10 by 10 pixel square as a structuring element to enhance visibility.

If contrails are found in the coarse collocation step, the "fine" collocation step is performed. The first stage of the fine collocation process is to apply a high cloud filter to the CALIOP L1B data between 8 and 15 km. This filter involves thresholding the sum of the attenuated backscatter β' in the 532 and 1064 nm channels as suggested by Iwabuchi et al. (2012)

$$\beta'_{532} + \beta'_{1064} > 0.007 \text{ km}^{-1}\text{sr}^{-1}. \tag{1}$$

For all remaining features (regions of non-zero backscatter) those with a width smaller than 1000 m and a thickness smaller than 240 m are filtered out. These filtering parameters are different than those used by Iwabuchi et al. (2012), who use a threshold of 0.003 $\text{km}^{-1}\text{sr}^{-1}$ for the backscatter and a thickness threshold of 60 m (the width threshold is the same). We found that the backscatter and thickness thresholds used by Iwabuchi et al. (2012) do not filter out all noise, based on inspection of the resulting CALIOP profiles and comparison with GOES-16 ABI and IIR imagery. Iwabuchi et al. (2012) analyze CALIOP data for the years 2007 and 2009 in a larger geographical domain than the analysis here. Gradual deterioration of the CALIOP

instrument with time as well as different background conditions may have caused these changes in noise level. These higher backscatter and thickness thresholds used here may lead to filtering out contrails. However, we did not find any situations during the manual inspection process (discussed at the end of this section) where a contrail was visible in GOES-16 ABI and/or IIR imagery but no backscatter was present in the filtered CALIOP L1 product. Instead, contrails missed by the detection algorithm but visible in the CALIOP data were more common. Furthermore, we only consider CALIOP data for the 8 km to 15 km altitude interval.

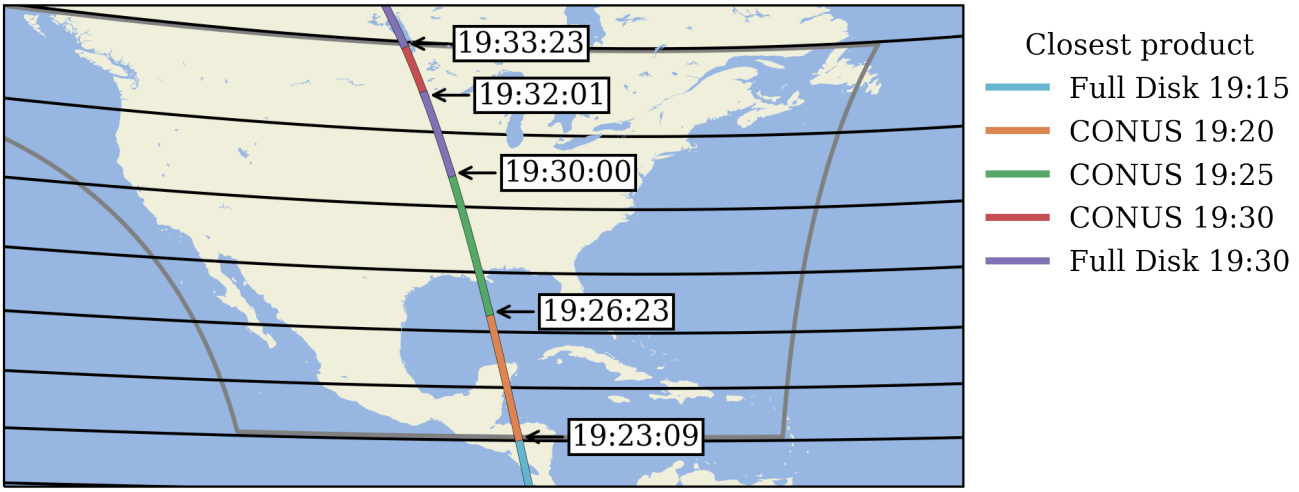


Figure S2. Example of the process of finding the GOES-16 ABI product where the pixel time is closest to the measurement time of the CALIOP profile. The CALIPSO overpass ground track is segmented (different colors) according to the closest GOES-16 ABI product, with the transition locations labeled with the corresponding time. Shown also are the GOES-16 ABI swath boundaries as black lines, and the CONUS product boundary in gray. The overpass shown here is the same as the one in the right plot of Figure S1. All times shown are UTC.

After these filtering steps are applied, for the remaining CALIOP profiles which feature a "cloud" we find the GOES-16 ABI product where the pixel capture times in the vicinity of the CALIOP profile are closest in time to that of the profile. To estimate the pixel capture time, we use the models introduced by Carr et al. (2020). An example of the output of this process for a CALIPSO overpass is given in Figure S2, showing how this process leads to a "segmentation" of the CALIPSO ground track where each segment has a different "closest" GOES-16 ABI product. For the first portion of the overpass in Figure S2, the closest product is the Full Disk (i.e. ABI-L2-MCMIPF) scan that started around 19:15Z, despite the local overpass time being close to 19:22Z. This is because this portion of the domain is not captured by the GOES-16 ABI CONUS (i.e. ABI-L2-MCMIPC) products. When the CALIPSO ground track enters the CONUS domain (gray boundary in Figure S2), the closest product switches to the CONUS scan starting around 19:20Z. The next switch occurs at 19:26:23Z, when the closest product is found to be the CONUS scan starting around 19:25Z. The reason that this "switching" time of 19:26:23Z does not correspond directly to the CONUS scan start time is that the different swaths (with boundaries shown as black curves in Figure S2) are

scanned sequentially, from North to South. As the CALIPSO satellite moves northward, the closest GOES product changes to "later" instances. Before the overpass ends, at 19:33:23Z, the closest product switches from the CONUS scan started around 19:30Z to the Full Disk scan started around the same time. This is because the domain of the CONUS scans does not capture this northern-most part of the overpass, similar to what happened at the start of overpass near the southern boundary of the domain.

Once the local pixel capture time is found, we advect the cloud top in the CALIOP profile from the measurement time to this pixel capture time using ERA5 wind data. Finally, we estimate the parallax displacement of the advected cloud top that would occur if viewed from the perspective of the GOES-16 ABI by use of the method from Bieliński (2020). The resulting position of the parallax-displaced cloud top is then mapped to the closest GOES-16 ABI pixel. If that pixel was also part of a detected contrail, we consider the CALIOP profile collocated to that contrail pixel.

The collocations resulting from the fine collocation process are manually inspected and corrected by analyzing figures such as Figure S4 and Figure S5 within an image annotation tool. If a collocation is considered incorrect, it is marked using an annotation and removed. Incorrect collocations occurred because of incorrect contrail detections or misalignment. Multiple

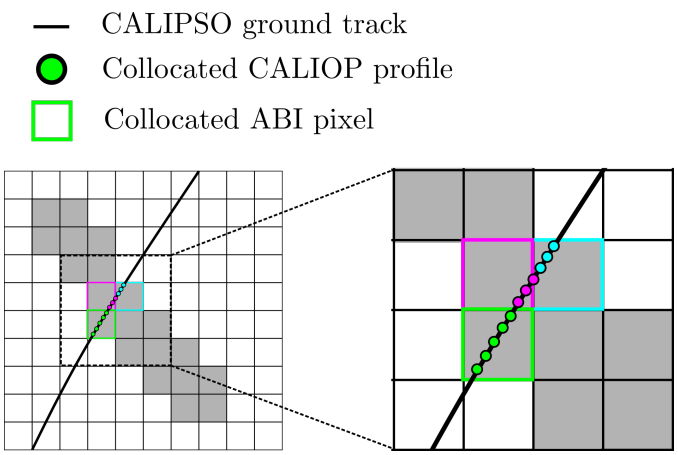


Figure S3. Ontology of the collocation dataset, showing the relationship between collocated CALIOP profiles, pixels and contrails. The gray pixels are part of a detected contrail.

CALIOP profiles can be collocated with the same GOES-16 ABI pixel, as is shown in Figure S3. Additionally, each contrail can have have multiple collocated GOES-16 pixels. To obtain the contrail top altitude associated with each GOES-16 ABI pixel, we average those associated with each collocated CALIOP profile. For the statistical analysis of the dataset of collocated contrails, the top altitude averaged over the collocated pixels in a contrail is used as the top altitude of the contrail.

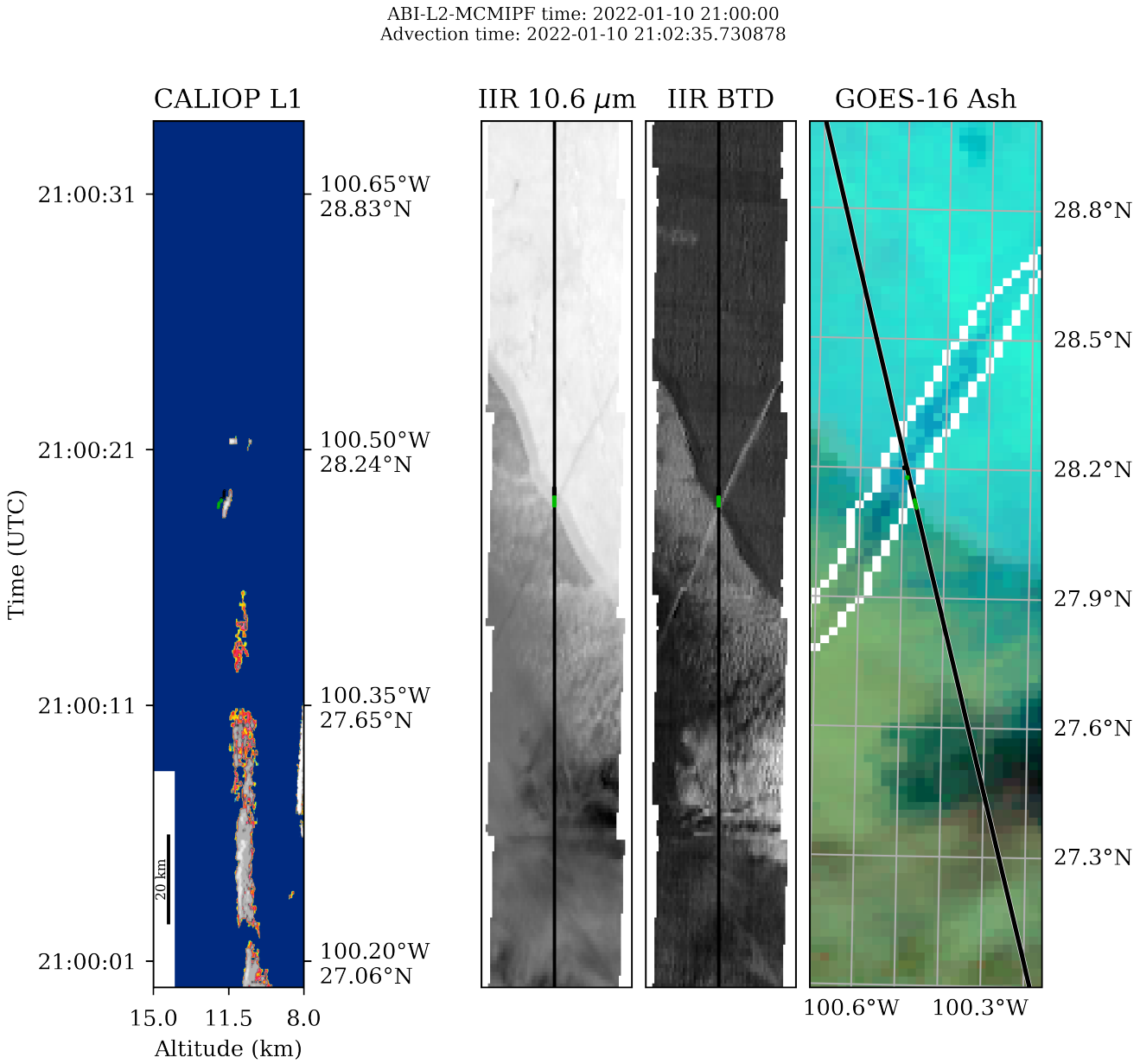


Figure S4. An example of the figures used for manual correction of the collocation between CALIOP L1B and GOES-16 ABI data. From left to right: cloud-filtered CALIOP L1B data, IIR brightness temperature, IIR brightness temperature difference, GOES-16 ABI ash. The colored dots in each plot show the location of the collocated CALIOP L1 profiles. In the CALIOP data, the dots are raised slightly above the cloud altitude, for clarity. In the GOES-16 image, diamonds are also shown which indicate the location of the CALIOP L1 profiles that are collocated. The circles indicate the collocated pixels, which are offset from the diamonds due to the parallax effect. This example did not require any correction.

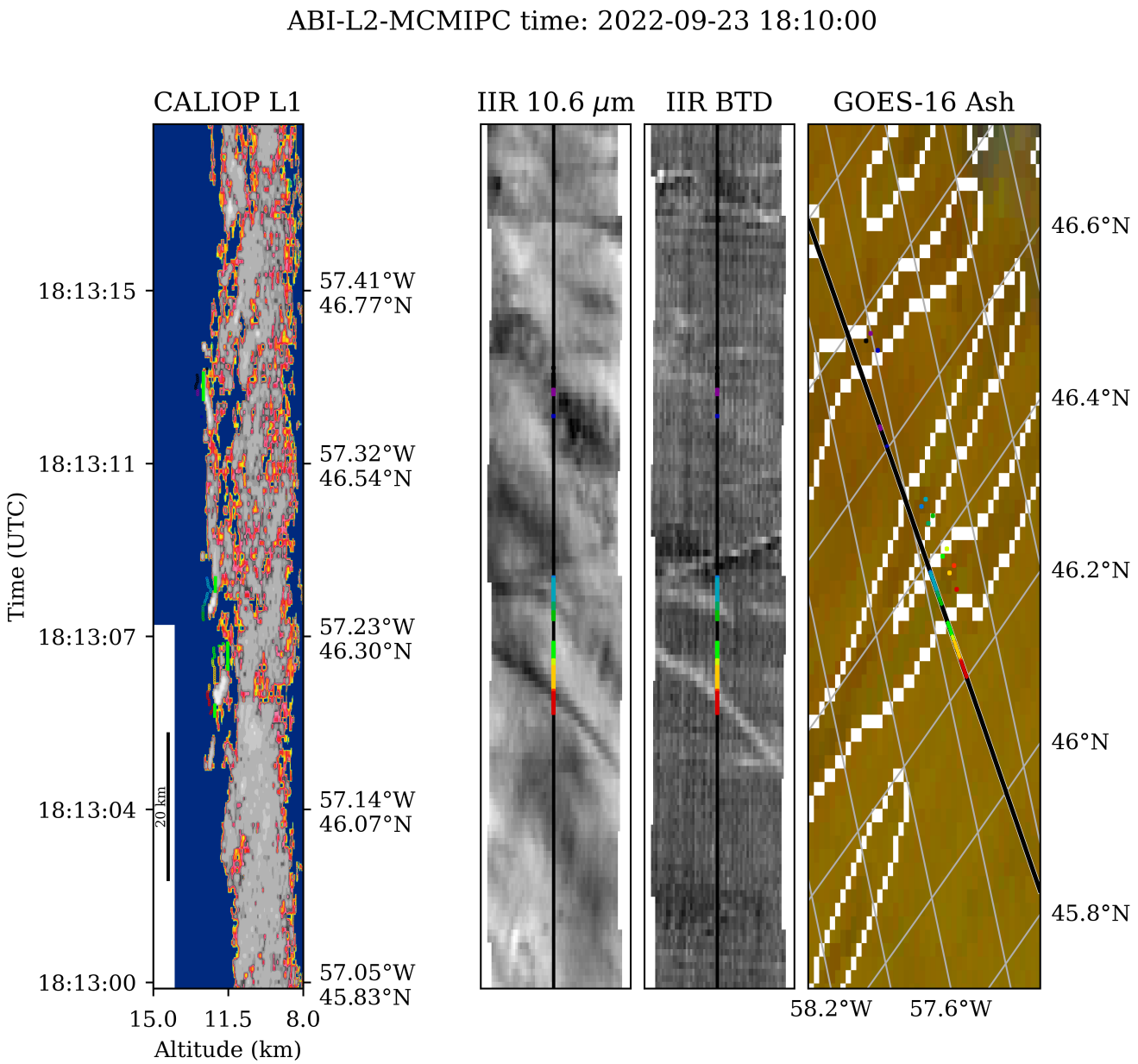


Figure S5. Same as Figure S4, but for an example requiring correction. The collocations that are removed are underlined with green in the CALIOP figure.

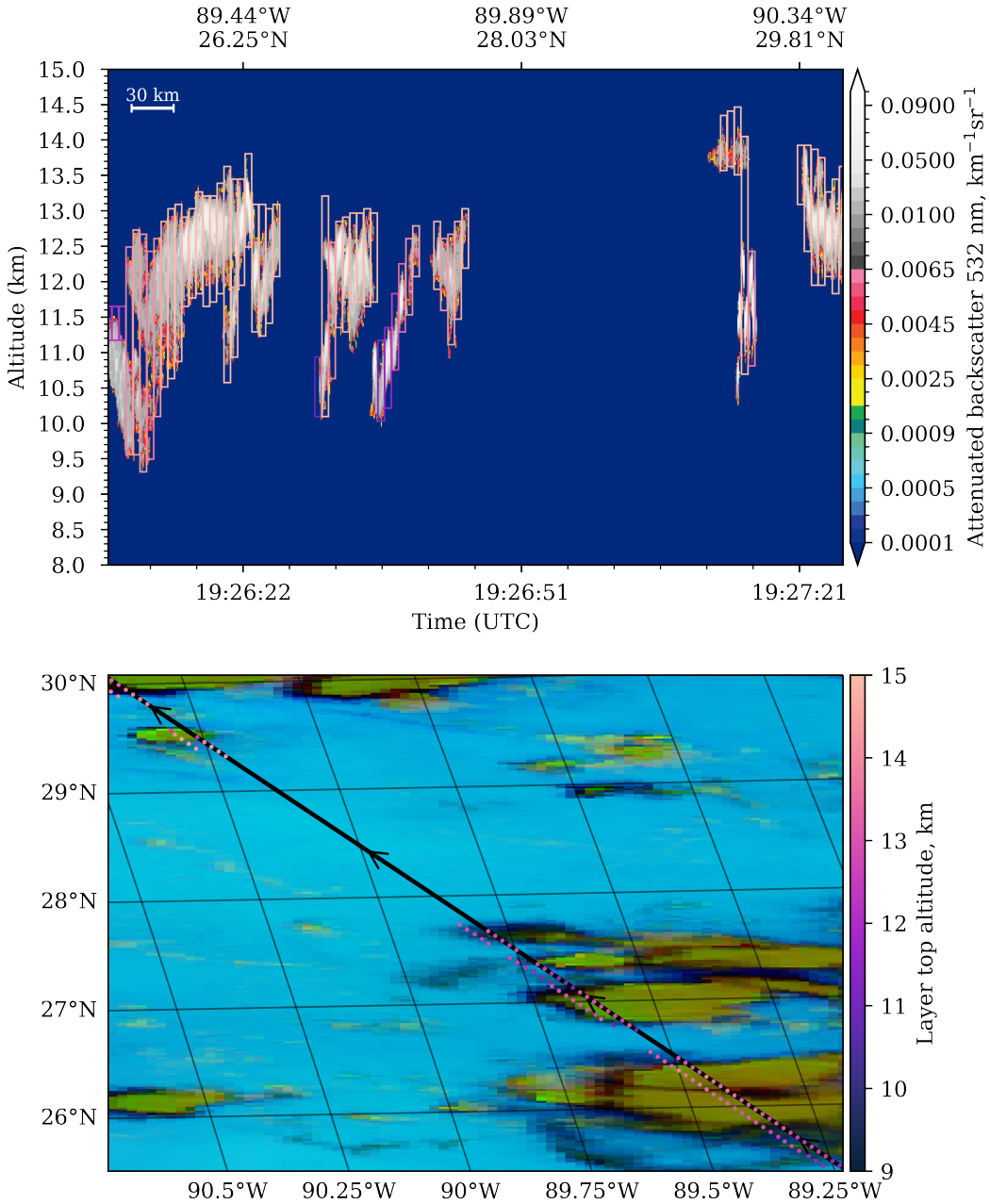


Figure S6. An example collocation of GOES-16 ABI imagery (bottom) and cirrus layers in CALIOP data (top). The GOES-16 ABI image is a false color ash product, created using the ABI-L2-MCMIPC data for August 8 2018 19:25Z. The ground track of the CALIPSO satellite is shown in the GOES-16 ABI image, with the arrows indicating the direction of movement. Collocated cirrus layers are shown in the CALIOP L1B data using boxes, colored according to the cirrus top altitude. The middle of each collocated cirrus layer is shown in the GOES image using a circle, using the same colors. The pixel collocated for each layer is also indicated on the GOES image.

S2 Collocation of CALIOP L2 and GOES-16 ABI data

Similar to Strandgren et al. (2017), we use the CALIOP L2 5 km cloud-layer product to create a dataset for developing a cirrus altitude estimation algorithm. For each "high" cloud layer in this product that was found to be consisting of "ice" or "oriented ice particles" and of the "cirrus (transparent)" or "deep convective (opaque)" cloud subtype, we perform the same fine collocation process as for finding contrails in CALIOP L1 data. We only consider CALIOP L2 layers for which 5 km horizontal averaging was used and for which the quality assurance flag is "high" or "confident" for the feature type, cloud phase and cloud subtype. Additionally, during the fine collocation process we no longer require the closest GOES-16 ABI pixel for a layer to be part of a detected contrail, as we consider all cirrus cloudiness. Only cloud layers for which the CALIPSO overpass time and the local GOES-16 ABI pixel capture time are within 2 minutes are collocated. An example collocation is shown in Figure S6.

S3 Dataset statistics

The statistics on the vertical location of contrail tops, measured in pressure, pressure altitude and geometric altitude are shown in Table S1. The geometric altitudes for the contrails and persistent contrail forming conditions (ISS & SAC) are shown in Figure S7.

Table S1. Collocated contrail dataset statistics

	Pressure, hPa			Pressure altitude, km			Geometric altitude, km		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Season									
DJF	223.67	217.66	35.72	11.14	11.25	0.99	11.27	11.38	1.19
MAM	217.58	213.56	32.84	11.31	11.37	0.93	11.49	11.50	1.12
JJA	202.20	196.88	32.12	11.76	11.87	0.90	12.21	12.31	1.02
SON	210.55	205.71	30.34	11.50	11.60	0.87	11.79	11.89	1.02

S4 GRUAN data processing

To evaluate the Schmidt-Appleman criterion, we find the slope of the mixing line G as (Schumann, 1996)

$$G = \frac{c_p p}{\epsilon} \frac{EI_{\mathrm{H_2O}}}{LHV(1-\eta)}$$

where the parameters in Table S2 are used. The pressure p is obtained from the GRUAN radiosonde data. To obtain the temperature corresponding to the point at which the slope of the liquid saturation curve is equal to G, the approximation by Schumann (1996) is used.

The seasonal distributions of the altitude of SAC & ISS regions within Lamont (SGP) GRUAN radiosonde launches, as a function of year, are shown in Figure S8. The seasonal pattern where regions of SAC & ISS occur at higher altitudes during

Table S2. Parameters used to evaluate the Schmidt-Appleman criterion.

Parameter	Symbol	Value
Engine efficiency	η	0.4
Emissions index of water vapor	$EI_{\mathrm{H_2O}}$	1.23
Ratio of molar mass of water to air	ϵ	0.622
Lower heating value of fuel	LHV	42.3 MJ/kg
Specific heating coefficient	c_p	1005 J/kg/K

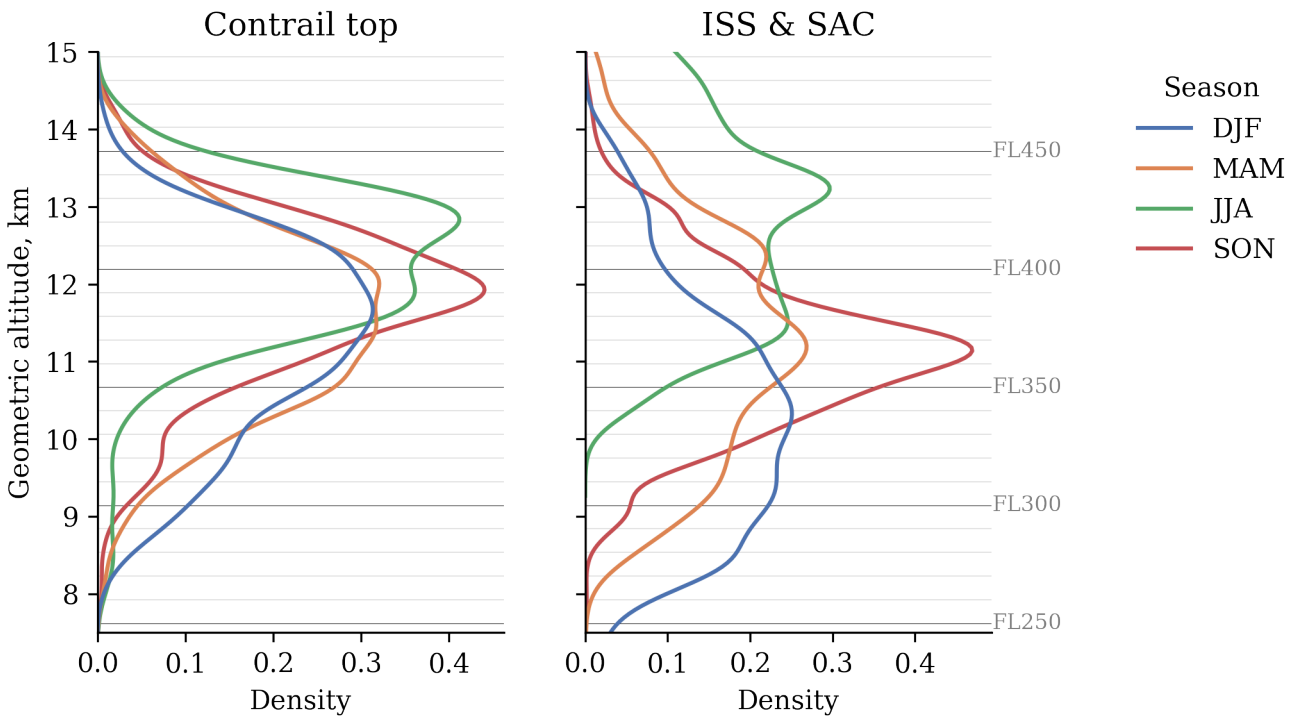


Figure S7. Kernel density estimates for the vertical distribution of contrail top altitude, persistent contrail forming regions (ISS & SAC) and flight traffic as a function of season. Each curve integrates to 1. The flight distribution is weighted by the distance flown at each altitude, and is based on OpenSky ADS-B data. The ISS & SAC data is obtained from GRUAN profiles.

Summer than in Winter is consistent throughout the years. Additionally, no general trends, such as increasing or decreasing altitudes over the years are visible.

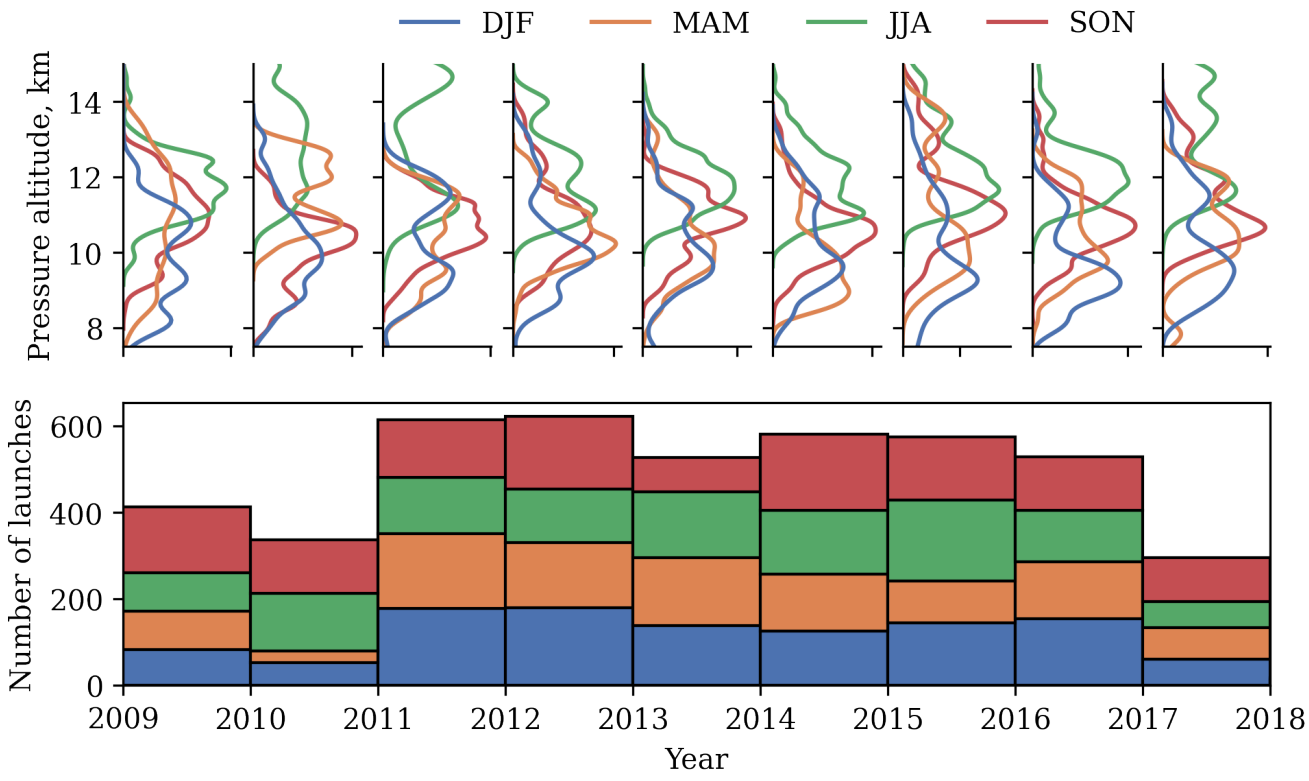


Figure S8. Seasonal distribution of the altitudes of regions satisfying the Schmidt-Appleman criterion and being ice super-saturated as measured by radiosonde launches at the Lamont (SGP) GRUAN location, as well as the number of launches by year and season. The small multiples are kernel density estimates for each year of data, and are aligned with the horizontal axis of the bottom plot.

S5 Contrail locations relative to the tropopause

The CALIOP L1B data comes with an estimate of the local tropopause altitude (above mean sea level) and corresponding temperature. This tropopause altitude is derived from MERRA-2 reanalysis data (Gelaro et al., 2017) and is defined as the minimum of the dynamical (based on potential vorticity) and thermal (based on temperature) tropopause altitudes. This tropopause altitude estimate is used to assess the relative location of the contrails found in CALIOP L1B data to the tropopause. The contrail top temperatures are estimated by use of the temperature profile data within the CALIOP L1B product, which also comes from the MERRA-2 reanalysis.

The resulting relative locations and temperature differences (between the tropopause and the contrail top) are shown in Figure S9. 86.6% of collocated contrails are found to be below the local tropopause altitude. Furthermore, 98.6% of contrail top temperatures are found to be warmer than the local tropopause temperature. Figure S9 shows that most contrail top altitudes (67%) are within a 2 km region below the local tropopause, and that the majority (77%) of contrail top temperatures is between

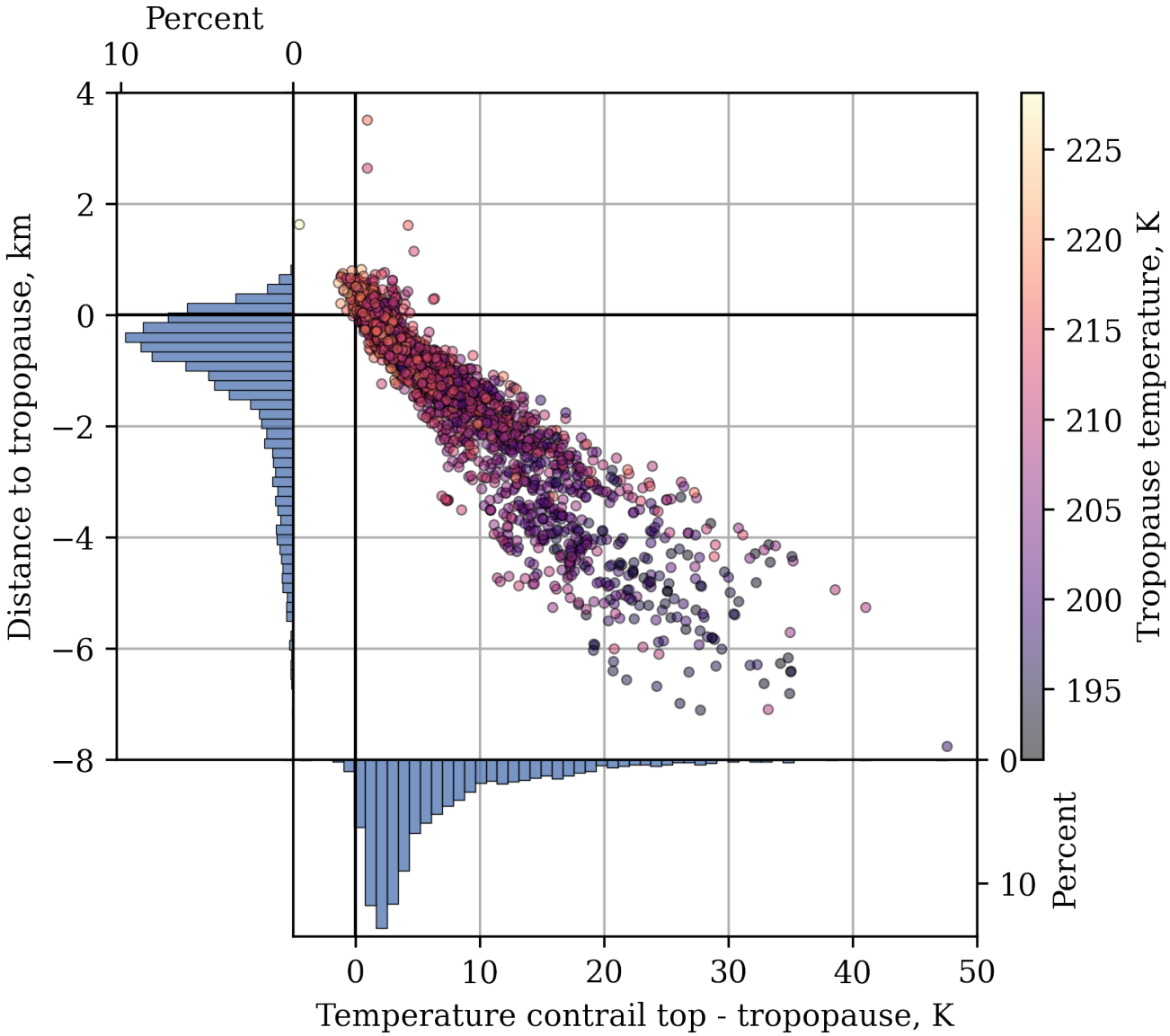


Figure S9. Relative location to the tropopause of contrails found in CALIOP L1B data, as well as the difference between tropopause and contrail top temperature.

0 and 10 K warmer than the tropopause temperature. The correlation between the distance to the tropopause and the associated difference in temperature shown in Figure S9 is most likely due to the tendency of temperature to decrease with increasing altitude in the troposphere. Lastly, the colormap in Figure S9 also indicates that contrails tend to be located closer to the tropopause for warmer tropopause temperatures which generally corresponds to lower tropopause locations. These contrails

that are closer to the tropopause have been found mostly at higher latitudes (not shown). This is likely due to aircraft flying at constant pressure levels, and the typical tropopause pressure levels decreasing in altitude at higher latitudes.

S6 Neural network development

The convolutional neural network used is based on that presented by Amell et al. (2022). It uses an encoder-decoder architecture as shown in Figure S10, featuring Xception blocks (Chollet, 2017). The activation functions used are Gaussian Error Linear Units (GELU) (Hendrycks and Gimpel, 2023). During training of the image-based model, the loss function is computed only

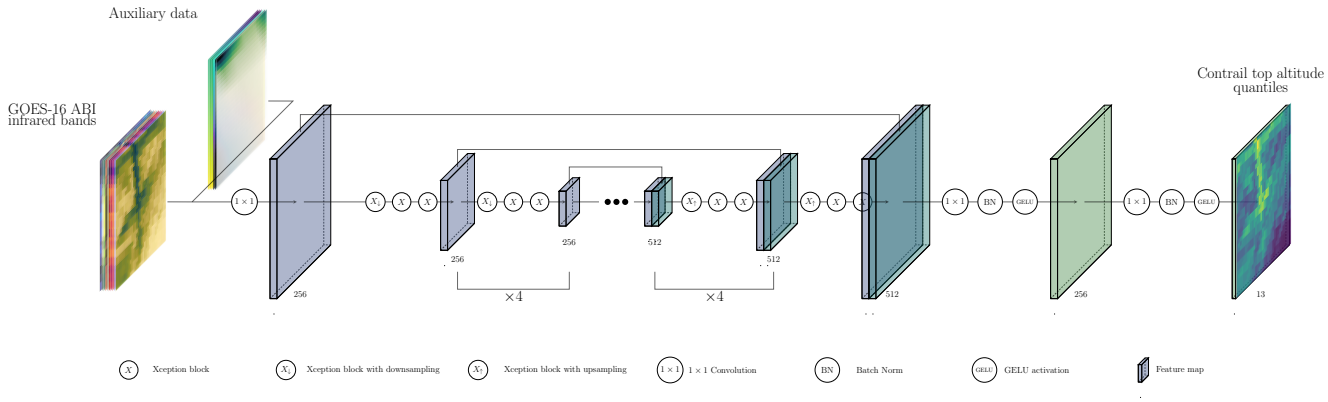


Figure S10. Architecture of the convolutional neural network used for contrail and cirrus altitude estimation. The various Xception block architectures are shown in Figure S12.

for pixels for which CALIOP data is collocated, as is shown in Figure S11. All networks are trained using the Adam (Kingma

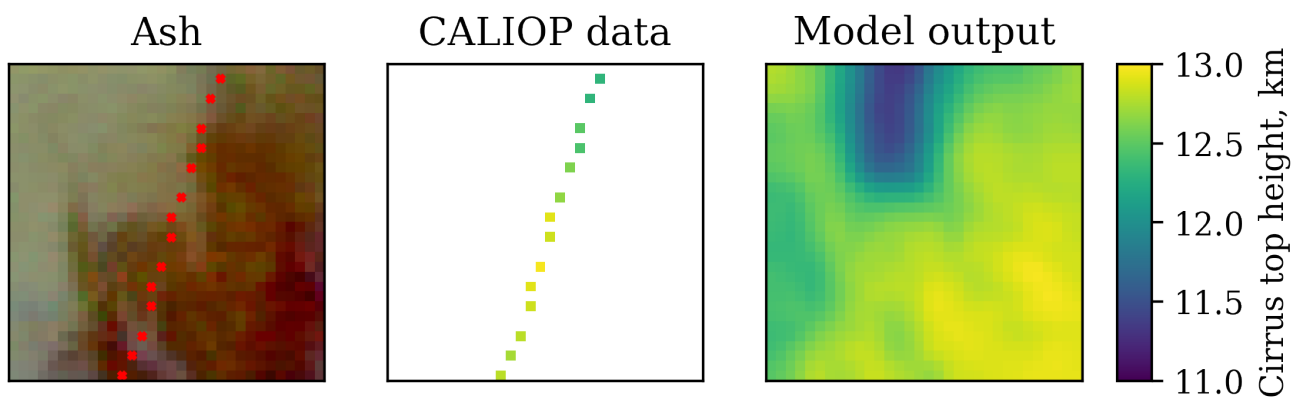
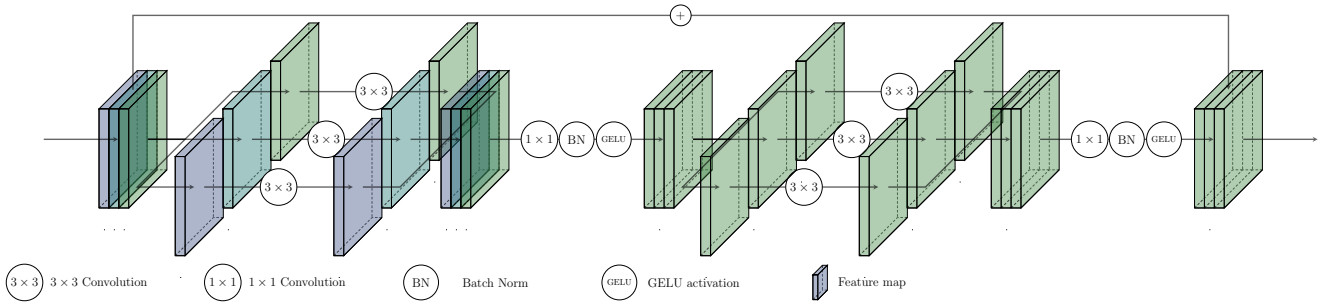
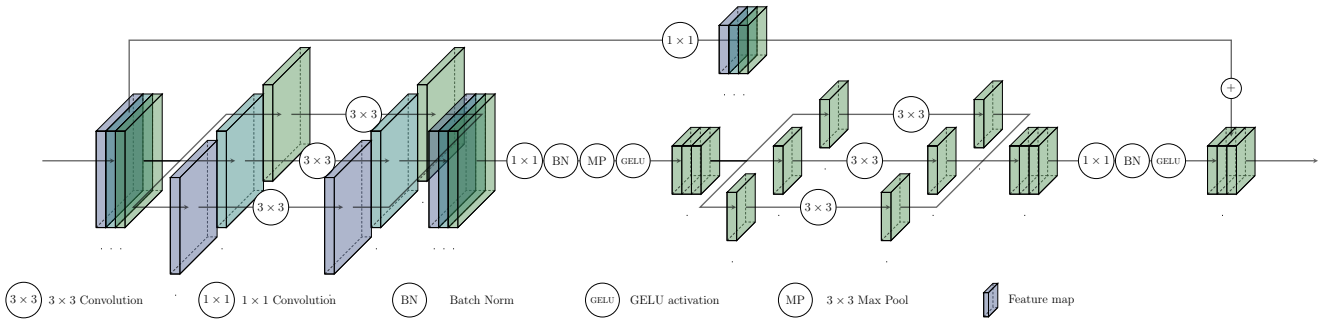


Figure S11. Example of the training setup for the image-based model. (Left) The ash product for the 32 by 32 pixel crop that is presented to the model during training. (Center) The cirrus top altitude data from CALIOP, for the collocated pixels indicated by red crosses on the ash image. (Right) The output of the CNN for the image crop. Only the collocated pixels are used for training.

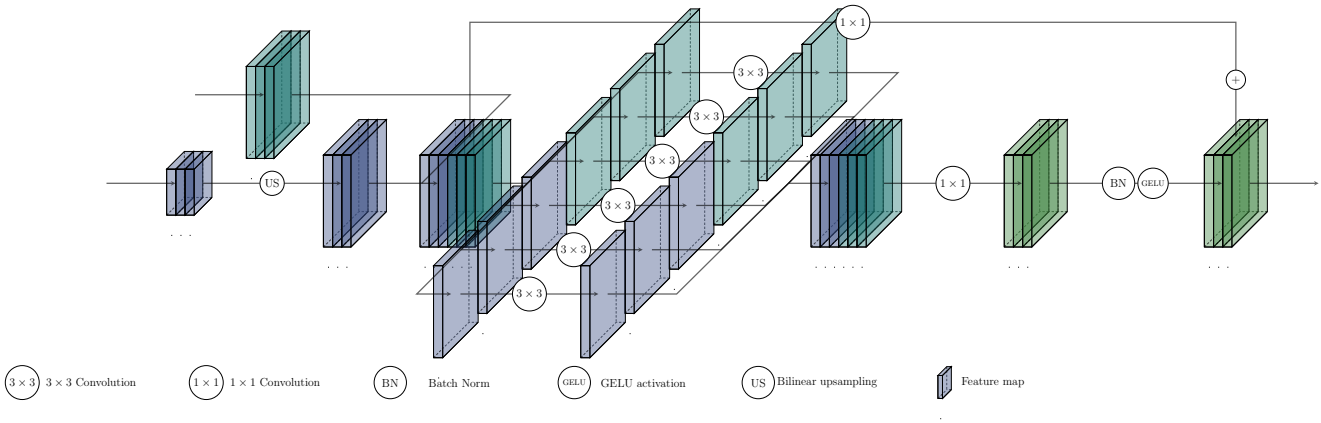
and Ba, 2014) and AdamW (Loshchilov and Hutter, 2019) when zero weight decay is used. The learning rate, weight decay, batch size and number of training epochs for all models can be found in Table S3. The "Cirrus MLP" model has 2 hidden layers



(a) Xception block.



(b) Xception block with downsampling.



(c) Xception block with upsampling.

Figure S12. Xception blocks used as part of the convolutional neural network in Figure S10.

with 128 neurons each, and the "Contrail MLP" model has 2 hidden layers with 16 neurons each. Both MLPs use the ReLU activation function (Nair and Hinton, 2010).

Table S3. Training details for the neural network models

Model	Learning rate	Weight decay	Batch size	Epochs
Cirrus MLP	10^{-4}	0	1000	500
Cirrus CNN	10^{-3}	10^{-3}	128	50
Contrail MLP	10^{-4}	1	100	600
Contrail CNN	10^{-3}	0	200	40

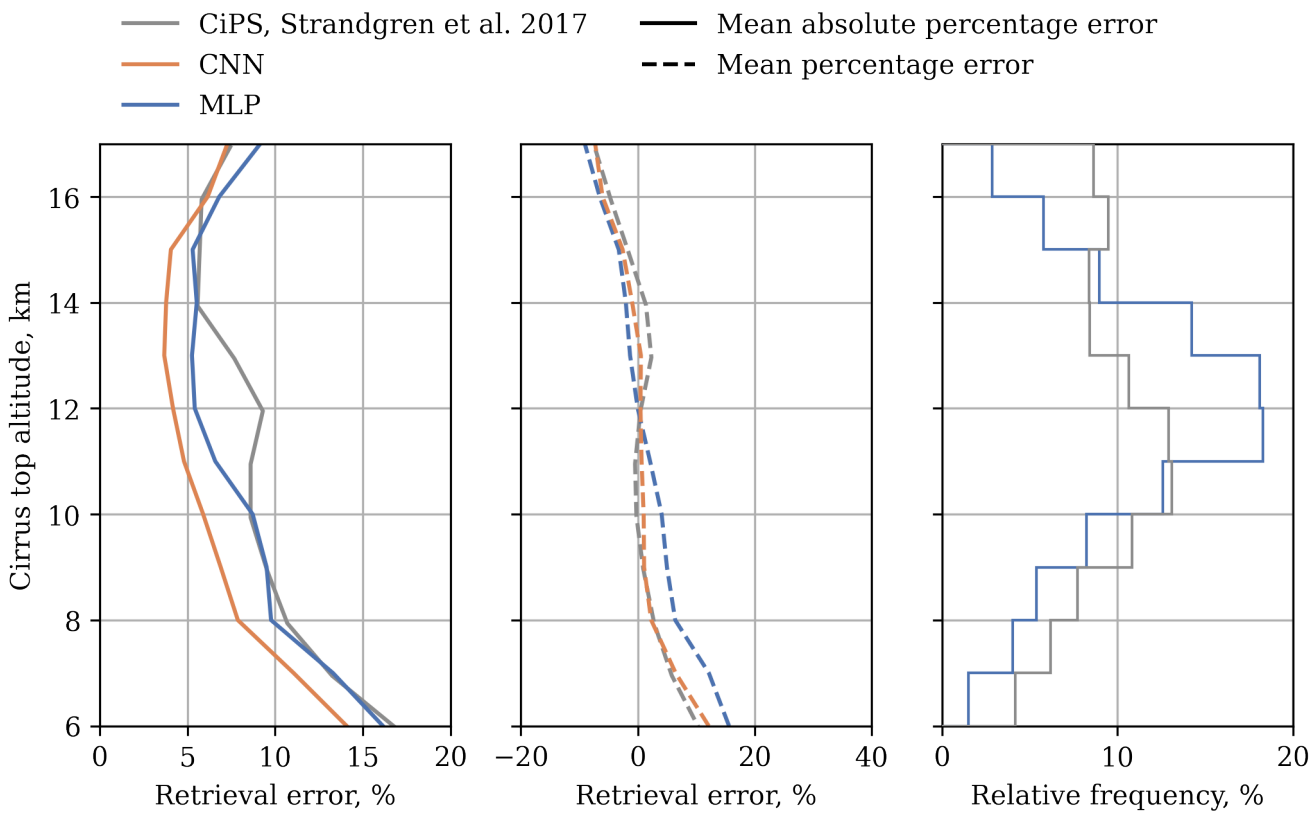


Figure S13. Performance of the cirrus altitude estimation algorithms on the cirrus test dataset as a function of cirrus top altitude (left), as well as the entire dataset distribution (right). The performance of CiPS (Strandgren et al., 2017) for retrieving cirrus altitude is also shown, as well as the distribution of the dataset used for training the algorithm. Data on CiPS performance has been extracted from Figures 2 and 9 from Strandgren et al. (2017).

S7 Comparison of cirrus altitude algorithms with previous work

Here we analyze the performance of the pixel-by-pixel and image-level models for cirrus altitude estimation, as evaluated using the cirrus test dataset. The mean percentage error (MPE) and mean absolute percentage error (MAPE) as a function of cirrus top altitude are shown in Figure S13. Performance data for CiPS (Strandgren et al., 2017) is also shown. The image-based cirrus altitude estimation algorithm (CNN) has a lower MAPE than CiPS for all cirrus top altitudes below 16 km. The pixel-by-pixel model (MLP), which is more similar in its structure to CiPS, has a lower MAPE for altitudes between 10 and 14 km. However, CiPS has lower MAPE values for cirrus above 16 km. The MPE, and thus bias, of the CNN is comparable to that of CiPS, but the MLP shows higher bias than CiPS. Over the entire test dataset, the root mean square error of the CNN is 0.81 km, with a correlation coefficient of 0.93. Based on the data for CiPS presented in Figure S13, we estimate that CiPS has a RMSE of

0.95 km. Strandgren et al. (2017) report a correlation coefficient of 0.90 for the CiPS altitude retrieval, whereas the CNN has a correlation coefficient of 0.93. The MLP algorithm achieves a RMSE of 1.04 km on the test set and a correlation coefficient of 0.88, both of which are worse than CiPS.

We also show the distribution of cirrus top altitude in the datasets used to develop both algorithms in Figure S13. The distribution of our dataset shows that more data is present for cirrus with top altitudes between 11 and 14 km, and fewer data for very low and high clouds when compared to the Strandgren et al. (2017) dataset. This difference in distributions is likely a consequence of our dataset consisting of cirrus clouds only in the domain of the contrail detection algorithm (see Figure 2), whereas the CiPS algorithm uses data from cirrus clouds found within the entire viewing domain of the SEVIRI instrument.

S8 Analysis of predictive uncertainty

Examples of the predictive probability density functions (PDFs) are shown in Figure S14 for the CNN, CALIOP dataset distribution, and error statistics model. For the CNN, these PDFs are obtained by first linearly interpolating the quantile values to obtain the cumulative density function (CDF) and then evaluating its derivative at the midpoints between the quantile locations. Thus, "spikes' or "peaks" in the CNN PDFs in Figure S14 correspond to short segments in the CDF where the slope is slightly larger or lower. This does not necessarily imply a "peak" in probability, as the PDF would need to be integrated with respect to altitude and the altitude interval covered by these peaks tends to be small. Smooth(er) PDFs could be obtained by either using a different type of interpolation of the CDF, curve fitting or modifying the CNN's output (e.g. parameters of a Gaussian).

Figure S14 also shows that the 95% CIs by the three models have varying widths, or "thicknesses". The 95% CI width is constant for the CALIOP dataset distribution and the error model, but varies for the CNN. In order to investigate whether the varying 95% CI width from the CNN model improves upon the two baseline models, we show its variation as a function of error of the mean in Figure S15. The Pearson correlation coefficient between the absolute error (not shown) and the 95% CI width is 0.21, with its own 95% CI of [0.165, 0.273] based on 10,000 bootstrap samples. The mode of the 95% CI width distribution is slightly larger than 1.5 km, and the mean is given by 2.05 km. For 67.7% of the data points the 95% CI by the CNN is smaller than that of the error statistic distribution CI width of 2.19 km. The number of data points within the 95% CIs that are smaller than the error statistics 95% CI is 90%, compared to 92.6% for the entire dataset. Despite this local reduction in calibration, for 58.8% of the test set data points does the CNN give better (in terms of width and correctness) confidence intervals than the error statistics model. This 58.8% is the fraction of the 95% CIs containing the data points where the CI width is smaller than that of the error statistics model, minus the fraction of the data where the 95% CI misses the data point but the CI by the error statistics model does not. Additionally, the error statistics model here has been constructed using the test data and evaluated using that same data. The performance of this model on "unseen" data could possibly be worse than quantified here.

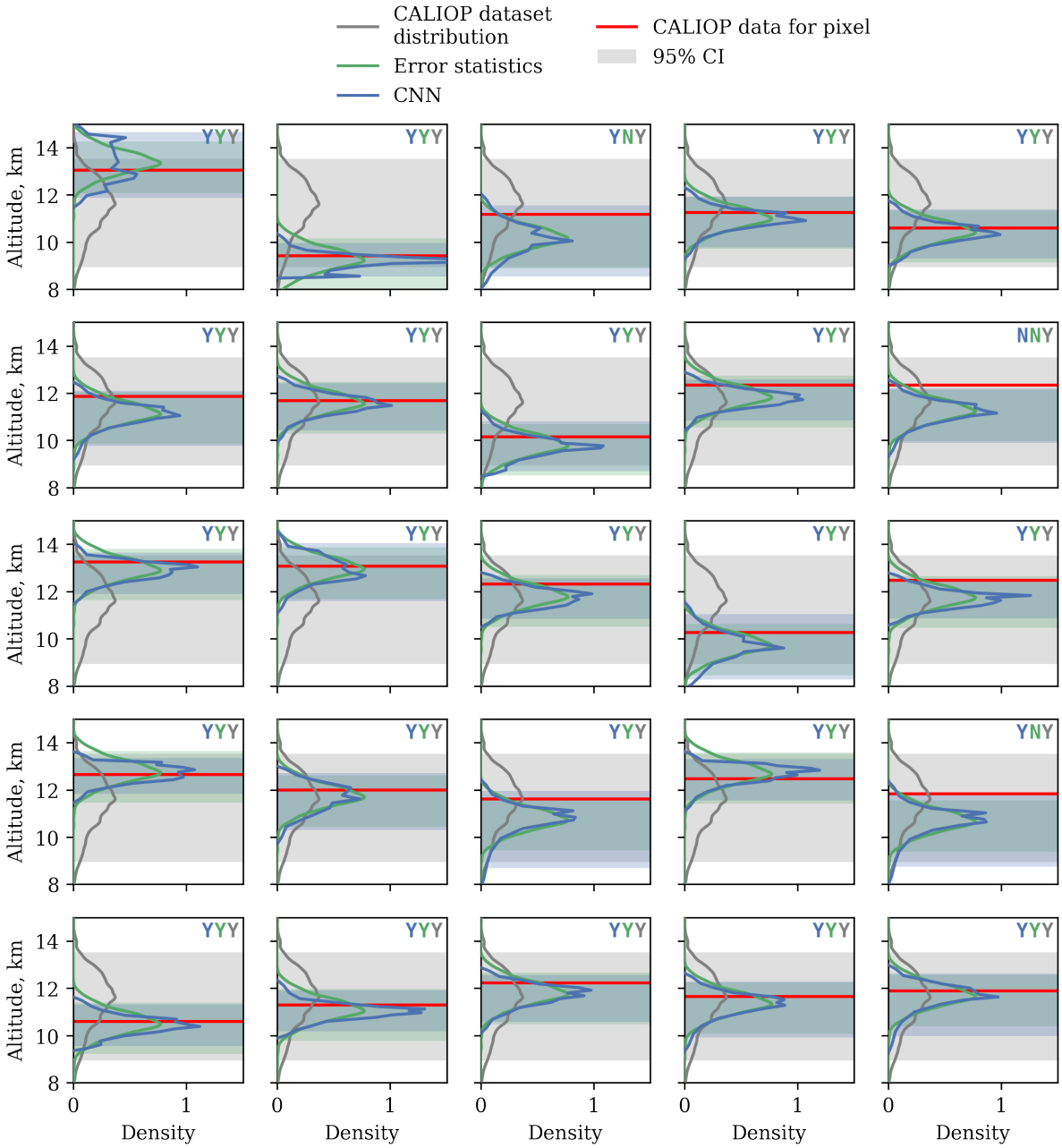


Figure S14. Examples from the contrail altitude test set and the predicted density by the 3 models, as well as the 95% confidence intervals (CI) they imply. The letters in the top right of each plot indicate whether the CALIOP data point is within the 95% CI of each of the three models with "Y" indicating a hit and "N" a miss. These "Y/N" are colored according to the respective model, and are always in the order (left to right) CNN, Error statistics, CALIOP dataset distribution.

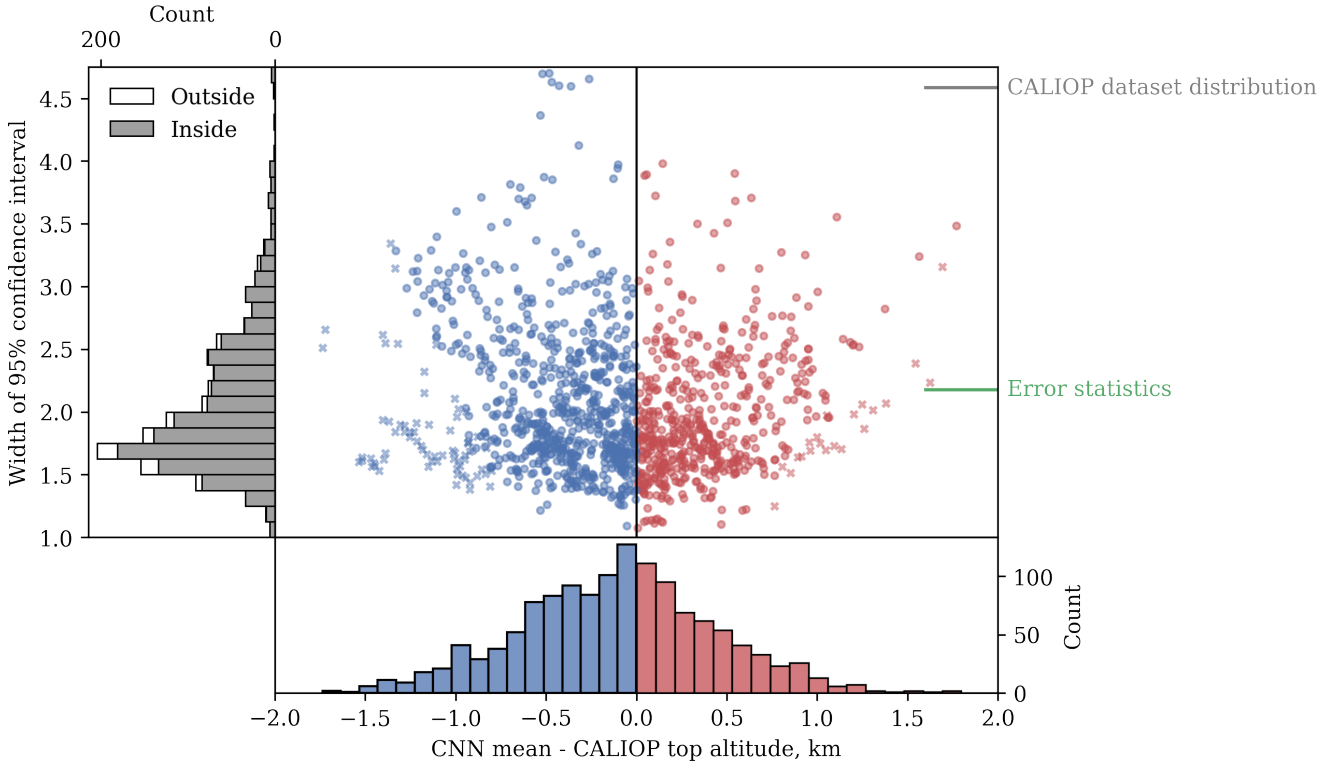


Figure S15. Scatter plot showing the 95% confidence interval width predicted by the CNN model and the error of the predicted mean, for the test dataset. Points that are within the confidence interval are marked as dots, those that are outside are marked with crosses. Furthermore, the horizontal insets show the 95% confidence interval width for the Error statistics and CALIOP dataset distribution model. The histograms show the marginal distributions of the two variables.

S9 Distributions of distance flown per flight level

For the comparison of contrail top altitudes and the probability distributions predicted by the algorithm with flight data, statistics of distance flown per flight level are computed. The following two subsections describe the approach taken for the comparison with the test set data and the 24-hour case study for August 21 2023.

S9.1 Comparison with test set estimates

Flight track data from the 2 hours leading up to the observation of the contrail is advected to the observation time. The choice for advecting only the past 2 hours of flight data is motivated by the results from Chevallier et al. (2023) who show that lifetimes of contrails observed on GOES-16 imagery tend to be shorter than 2 hours. In order to identify flight tracks suitable for this advection process, we advect the contrail location backward in time at the ERA5 pressure levels between 450 and

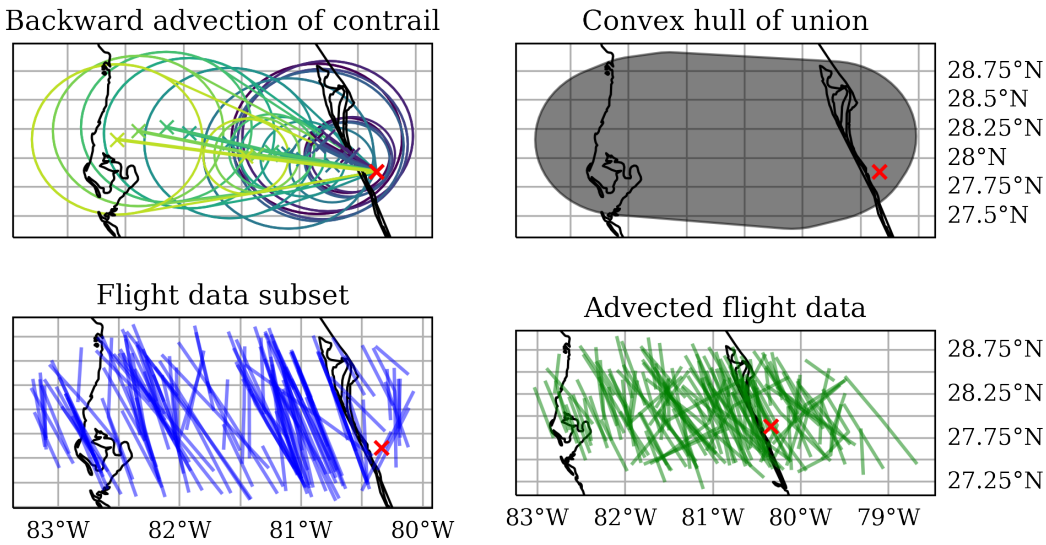


Figure S16. Example of the approach to flight subsetting and advection. The red marker is the position of the observed contrail. (Top left) The contrail location advected backward in time for 2 hours, with markers indicating the positions after 0, 1 and 2 hours. Different colors are used for each pressure level (brighter is lower pressure, thus higher altitude) and the associated ellipses. (Top right) The convex hull of the union of the ellipses in the top left plot, which is used to subset the flight data. (Bottom left) The flight data for the 2 hours leading up to the contrail observation time, within the region found in the previous step. (Bottom right) The flight data from the previous step, advected up to the observation time. The observation time is 19:25Z on November 10, 2019.

100 hPa. The objective of this "subsetting" process is to include only flight tracks in the analysis that could plausibly have formed the relevant contrail. An exact backward trajectory analysis however - involving a single trajectory that starts at the contrail location - might miss the flight that has formed the contrail due to errors in the wind field. In order to account for this possibility, we assume a maximum ERA5 wind error of at most 10 m/s, which is 3-5 times larger than the long-term average of first-guess upper wind departures from observations (Hersbach et al., 2020). At each point along the simulated trajectory, we then compute the “ellipse” that covers all possible horizontal trajectory locations for a wind error between 0 and 10 m/s. The convex hull of the union of these ellipses, for all pressure levels, is the region for which all flight data will be advected. An example of this approach is shown in Figure S16. There is the possibility that some waypoints corresponding to a part of a flight that is flown close to the observation time, but further away from the contrail observation location, are advected as well. This causes some of the advected flight tracks in Figure S16 to be located far away from the contrail observation. To filter out such flight data, a final filtering step is performed as shown in Figure S17. This filtering step ascertains whether a particular advected waypoint is sufficiently close to the observation point, based on the assumed wind error.

The advected flight tracks are used to find the distribution of distance flown. Each flight segment (defined by two successive waypoints) is placed in the bin corresponding to flight level FL if the mean flight level of the two waypoints is in the interval $[\mathrm{FL}-5, \mathrm{FL}+5)$. The distance of the segment, computed using the haversine formula, is then added to the total distance flown in

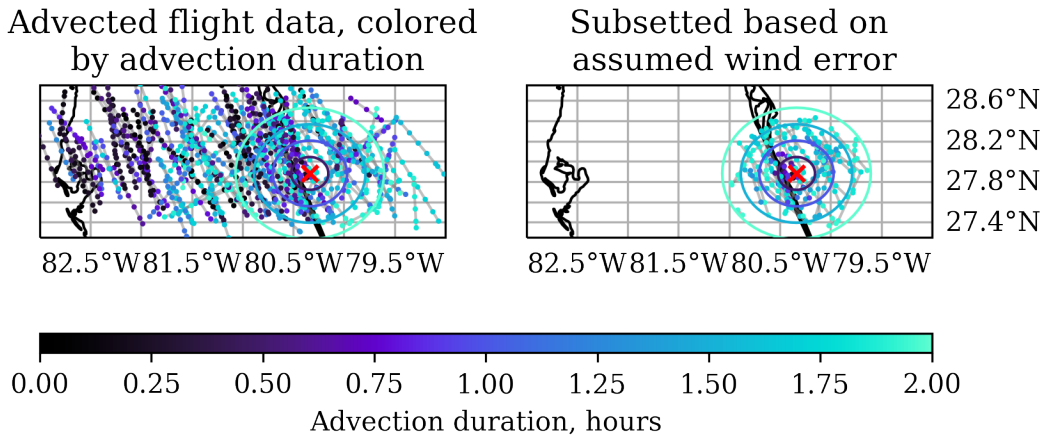


Figure S17. Example of the final step of the approach to to flight subsetting. The red marker is the position of the observed contrail. The data in the left plot corresponds to the advected flight tracks (green in Figure S16), with waypoints colored according to the time over which they have been advected. Also shown are the ellipses (centered on the observation) that contain all waypoints which could have ended up at the contrail location (red cross) for a given advection time (color of ellipse) and the assumed wind error of 10 m/s. These ellipses are used to subset the advected flight data, resulting in the plot in the right. This subsetted, advected flight data is used to construct the flight altitude distribution. The observation time is 19:25Z on November 10, 2019.

that bin. The distance flown in each bin is normalized by the overall distance flown (at all flight levels), such that a percentage remains. The distribution corresponding to the advected flight track data in Figure S16 is shown in Figure S18.

S9.2 Analysis of full day of data

The approach taken for the analysis of a full day of contrail altitude estimates is less involved than that for the test set. No selection of flight data for advection, besides requiring the flight to be above 8000 meters in pressure altitude, is performed. All waypoints flown in the 2 hours leading up to the ABI-L2-MCMIPF product time (every 10 minutes) are advected. Then, given the location of a contrail pixel, we compute the great circle distance between the pixel and all advected waypoints. The maximum allowed distance depends on the time over which the waypoint has been advected, and is found by multiplying this time by the maximum assumed wind error of 10 m/s. If the distance between waypoint and contrail pixel is lower than this distance, the waypoint is included in the flight altitude distribution. This process is identical to that shown in Figure S17. The result of this process are distributions similar to the one shown in Figure S18.

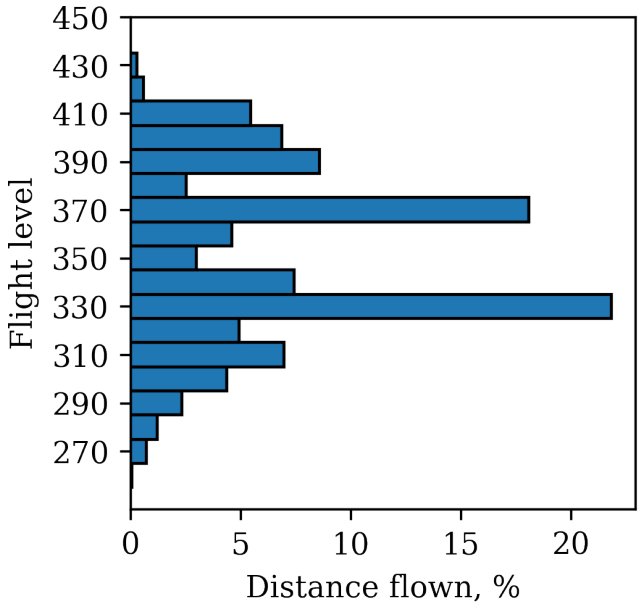


Figure S18. Distribution of distance flown by flight level, resulting from the approach shown in Figure S16.

References

Amell, A., Eriksson, P., and Pfreundschuh, S.: Ice water path retrievals from Meteosat-9 using quantile regression neural networks, Atmospheric Measurement Techniques, 15, 5701–5717, publisher: Copernicus GmbH, 2022.

Bieliński, T.: A parallax shift effect correction based on cloud height for geostationary satellites and radar observations, Remote Sensing, 12, 365, publisher: MDPI, 2020.

Carr, J. L., Wu, D. L., Daniels, J., Friberg, M. D., Bresky, W., and Madani, H.: GEO–GEO stereo-tracking of atmospheric motion vectors (AMVs) from the geostationary ring, Remote Sensing, 12, 3779, publisher: MDPI, 2020.

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, https://doi.org/10.3390/aerospace10070578, number: 7 Publisher: MDPI, 2023.

Chollet, F.: Xception: Deep Learning With Depthwise Separable Convolutions, pp. 1251–1258, https://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html, 2017.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. d., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, publisher: American Meteorological Society Section: Journal of Climate, 2017.

Hendrycks, D. and Gimpel, K.: Gaussian Error Linear Units (GELUs), https://doi.org/10.48550/arXiv.1606.08415, arXiv:1606.08415 [cs], 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Iwabuchi, H., Yang, P., Liou, K., and Minnis, P.: Physical and optical properties of persistent contrails: Climatology and interpretation, Journal of Geophysical Research: Atmospheres, 117, publisher: Wiley Online Library, 2012.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, https://doi.org/10.48550/arXiv.1711.05101, arXiv:1711.05101 [cs, math], 2019.

Nair, V. and Hinton, G. E.: Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814, https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf, 2010.

Schumann, U.: On conditions for contrail formation from aircraft exhausts, Meteorologische Zeitschrift, pp. 4–23, publisher: Schweizerbart'sche Verlagsbuchhandlung, 1996.

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, publisher: Copernicus GmbH, 2017.