# A human-labeled Landsat-8 contrails dataset

**Kevin McCloskey** [*1] **Scott Geraedts** [*1] **Brendan Jackman** [1] **Vincent R. Meijer** [2] **Erica Brand** [1] **Dave Fork** [1]
**John C. Platt** [1] **Carl Elkin** [1] **Christopher Van Arsdale** [1]

## Abstract

Contrails (condensation trails) are the ice clouds that trail behind aircraft as they fly through cold and moist regions of the atmosphere. Avoiding these regions could potentially be an inexpensive way to reduce over half of aviation's impact on global warming. Development and evaluation of these avoidance strategies greatly benefits from the ability to detect contrails on satellite imagery. Since little to no public data is available to develop such contrail detectors, we construct and release a dataset of several thousand Landsat-8 scenes with pixel-level annotations of contrails. The dataset will continue to grow, but currently contains 4289 scenes (of which 47% have at least one contrail) representing 950+ person-hours of labeling time.

## 1. Introduction

Contrails are cirrus clouds caused by jet aircraft which, by trapping outgoing infrared radiation, are likely aviation's largest contribution to global warming (Burkhardt & Karcher, 2011; Chen & Gettelman, 2013; Schumann et al., 2015; Boucher et al., 2014; Bock & Burkhardt, 2016; Bickel et al., 2020; Lee et al., 2021). Contrails are started by jet engine exhaust, which contains both water vapor and soot. If the surrounding air is cold and humid enough (Appleman, 1953), the water vapor condenses and freezes into ice clouds, using the soot particles as nucleation sites. The lifespan of these clouds depends strongly on the ambient humidity level. With low humidity, the ice crystals rapidly (in seconds or minutes) sublimate away or never even form, but with higher relative humidity contrails can persist for hours. While relevant jet engine soot regulation (Jacob & Rindlisbacher, 2019) as well as hydrogen or electric aircraft are on the horizon, these mitigations are likely decades away from comprehensive impact. Fortunately, recent simulation studies indicate a small percentage of contrails are responsible for the vast majority of contrail warming. This provides an opportunity to significantly reduce global warming by diverting that subset of flights a short distance vertically or laterally. Although diversions may sometimes increase fuel burn and associated emissions, these negative effects are estimated to be more than offset by the reduction in contrail radiative forcing (RF) (Avila et al., 2019; Teoh et al., 2020).

[*]Equal contribution [1]Google, Mountain View, California [2]Laboratory for Aviation and the Environment, Massachusetts Institute of Technology. Correspondence to: Kevin McCloskey <mccloskey@google.com>.

*Tackling Climate Change with Machine Learning* workshop at ICML 2021. 

## 2. Context and related work

In order to divert aircraft to avoid creating warming contrails, one needs to be able to predict which flights will create them. Models for this, such as CoCIP and APCEMM (Schumann, 2012; Fritz et al., 2020) explicitly model the microphysics and radiative transfer effects of individual contrails from formation to dissipation, depending on input data such as the estimated aircraft fuel burn rate (i.e. the jet exhaust quantity and qualities) and ambient weather conditions. The physics are fairly well established, but the inputs are subject to uncertainties from multiple sources. These uncertainties - including the spatial resolution of numerical weather models and limitations of high-altitude humidity measurements (Gierens et al., 2020) - have resulted in fairly large error bars on contrail RF estimates from simulations (Lee et al., 2021; Sanz-Morère et al., 2020).

Empirical analysis of contrails in satellite imagery also comes with uncertainty, stemming from the fundamental physical and visual similarity between contrails and naturally occurring cirrus clouds - and the absence of satellite orbits that can yield both high spatial and high temporal resolution imagery simultaneously. Contrail detection in satellite imagery for the last two decades has largely been accomplished with the algorithm of Mannstein et al (Mannstein et al., 1999; Meyer et al., 2002; Palikonda et al., 2005; Minnis et al., 2005; Meyer et al., 2007; Vazquez-Navarro et al., 2010; Duda et al., 2013). This algorithm is a series of hand-engineered convolution and thresholding operations that operate on brightness temperature imagery, followed by identifying linear connected components of appropriate size. It has been tuned to have either high Precision or high Recall in contrail detections but a single model has not achieved both simultaneously, necessitating wide lower-bound/upper-bound approaches in empirical measurements of contrail

coverage (Duda et al., 2013). One notable exception to the use of the Mannstein et al algorithm is (Kulik, 2019; Meijer et al., 2021) which trained a deep learning model for pixel-level contrail detection on GOES-16 satellite imagery.

Highly accurate contrail detection in satellite imagery is not just useful for validating contrail simulations. Once a contrail is detected, its lifetime impact can be estimated (Vazquez-Navarro et al., 2010; 2013; 2015), or attributed to possible flights that may have caused it. Cumulatively, these techniques allow measuring the effectiveness of flight diversions in avoiding contrail formations. Making the climate case for flight diversions to policy makers, airlines and the general public will be much easier with satellite verification capability in hand.

Geostationary satellites (e.g. GOES-16) and satellites in low-earth orbits (e.g. Landsat-8) have complementary strengths and weaknesses when applied to the contrail detection task. The main advantage of Landsat-8 is its high spatial resolution of 30m and 100m per pixel (cirrus and thermal infrared bands respectively), allowing for more confident labeling of young contrails. A geostationary satellite's strength is its temporal resolution, with as low as 5-minute refresh rates, allowing tracking of contrails (once detected) and enabling the use of the relatively sudden onset of contrails as a feature to distinguish them from natural cirrus. However, GOES-16 cirrus and infrared bands have a nadir spatial resolution of 1km and 2km respectively; by the time a contrail has persisted long enough (20-60min) to be visible across multiple GOES pixels, it has often diffused and become non-linear enough to be difficult even for experts to distinguish from natural cirrus clouds.

We do accept trade-offs to gain Landsat-8's high spatial resolution. Firstly it is temporally sparse, imaging the same area only once every 16 days. It also has a strong local time-of-day covariance with the imagery location, due to Landsat-8's sun-synchronous polar orbit. That is, each latitude is imaged at almost the same time of day (or night) on every orbit, with the daytime equator crossing occurring between 10:00-10:25am local time. Landsat-8 never captures low-to-mid-latitude imagery at dusk or dawn. In isolation then, this dataset is not ideal for analysis that requires disentangling physical contrail phenomena that might vary with time-of-day and machine learned model errors that might vary with time-of-day.

We anticipate this dataset will be useful for a number of purposes, including: as a benchmark for evaluating contrail detection methods; for validating high-confidence subsets of GOES-16 contrail labels/predictions; as training data for super-resolution models focused on GOES-16 contrail detection/tracking; for analyzing how frequently contrails exist but are invisible in colocated GOES imagery; and for contrail prediction via analysis of the weather conditions of contrails that persist long enough to become detectable in aligned GOES imagery.

## 3. The dataset

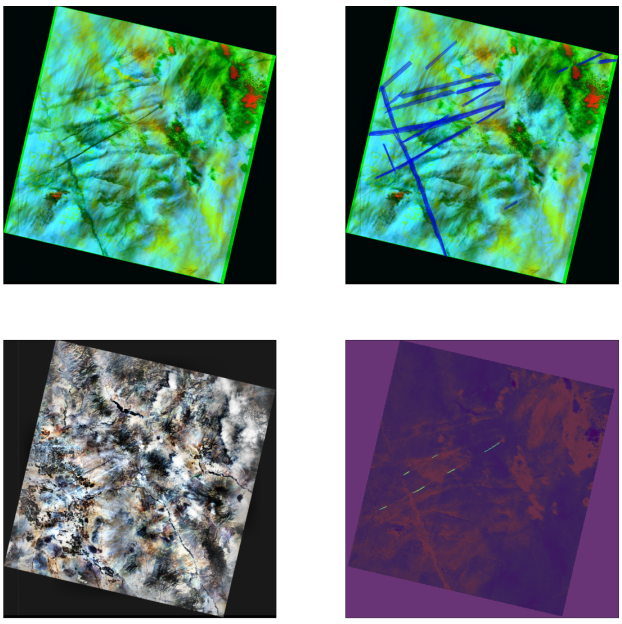

*Figure 1.* An example of a false color Landsat-8 scene as shown to labelers (top left) and its labeled contrails overlaid in blue (top right), along with the true color RGB (bottom left) and highlighted contrails identified by the algorithm of Mannstein et al (bottom right).

The dataset includes Landsat-8 scenes (primarily from 2018 and inside the viewable extent of the GOES-16 satellite) which have been reviewed by human labelers taught to identify and mark the bounding polygon of each contrail in the scene. The labelers were instructed in contrail identification initially by training slides containing general instructions and examples of different types of edge cases and their ideal labels (the instructions given to the labelers are released here alongside the data). Labelers then proceeded to iterative rounds of labeling Landsat-8 scenes and reviewing corrections to their labels provided by the paper authors. Once corrections were no longer necessary, labelers graduated to labeling imagery that appears in the dataset. Each Landsat-8 scene was labeled by two to four labelers. In our own analyses we have tended to retain only the contrails which were agreed upon by a majority of labelers who reviewed the scene, but the dataset includes all labels to allow other options.

An initial randomly selected sample of 1000 Landsat-8 scenes (from 2018 and in viewable extent of GOES-16) reviewed by one of the authors resulted in 16.2% of scenes

containing at least one contrail (data not shown). Therefore, when selecting Landsat-8 scenes for labeling, scenes which were likely to contain contrails were oversampled to improve the class balance in the dataset. Scenes identified by a high-Recall tuning of the Mannstein et al algorithm as containing a contrail have been labeled. Next priority for labeling was given to scenes which contained non-zero advected flight density and weather conditions satisfying the Schmidt-Appleman criteria (Appleman, 1953). To help mitigate dataset bias due to covariance with these factors, approximately 20% of Landsat-8 scenes introduced to the labeling pipeline are selected by random sampling.

Multiple types of imagery were shown to the labelers to aid in their decision making. The primary image was a false color image with the red channel being the brightness temperature difference between the 12 $\mu$m and 11 $\mu$m bands, the green channel being the 1.37 $\mu$m "cirrus" reflectance band (omitted for nighttime images), and the blue channel being the 12 $\mu$m brightness temperature. In this false color scheme, contrails appear as black linear clouds (see Figure 1). Two supplementary images were included: the true-color RGB, and an image representing advected flight density. The advected flight density was generated starting from waypoint data and advecting waypoints forward in time – using ECMWF ERA5 wind data (Hersbach et al., 2020) and the Runge-Kutta method (Bogacki & Shampine, 1989) – until they reach the imagery timestamp. Then the position uncertainty (estimated as a linear function of advection time length) was used as the standard deviation in a Gaussian distribution that was plotted orthogonally to the advected flight trajectory. See the labeler instructions document available alongside the dataset for examples.

Initial rounds of labeling (done by the authors when developing the labeling instructions and the imagery shown to labelers) made use of full resolution Landsat-8 imagery. However loading full-resolution imagery caused noticeable latency in the labeling web app. After reviewing contrail labels identified in full resolution imagery compared with contrail labels identified in imagery at 300m per pixel (still approximately 8x higher resolution than is available in GOES-16 infrared bands), accuracy did not appear to suffer and subsequent imagery has been labeled at 300m per pixel.

The dataset is provided in JSON string format and includes: the Landsat-8 filename, polygon bounds of contrails in the scene, and de-identified advected flight waypoints for each labeled Landsat-8 scene. It is available for download at `https://console.cloud.google.com/storage/browser/landsat_contrails_dataset` and at `gs://landsat_contrails_dataset` via the command line. Please see the open source code accompanying this paper alongside the dataset for an example of parsing it, as well as an implementation of the Mannstein et al algorithm (the first to be open-sourced that we are aware of). The Landsat-8 scenes themselves are available from multiple sources including as described at `https://cloud.google.com/storage/docs/public-datasets/landsat`. Landsat-8 imagery is courtesy of the U.S. Geological Survey. The de-identified advected flight waypoints are derived from terrestrial ADS-B data licensed for publication from FlightAware, LLC (`https://flightaware.com`).

## 4. Mannstein baseline results

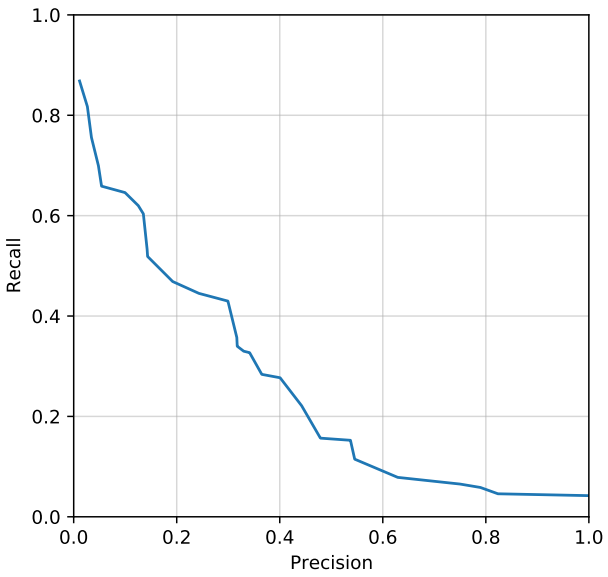


*Figure 2.* Precision/Recall curve for the Mannstein et al algorithm on this dataset.

The Mannstein et al algorithm (Mannstein et al., 1999) was developed for use with the AVHRR instrument aboard the NOAA-12 satellite. The algorithm has tunable parameters allowing for adapting the algorithm to new types of satellite imagery. Figure 2 demonstrates the performance of our open-sourced implementation on a randomly selected 20% of the scenes of our Landsat-8 dataset. With 10 parameters to tune, we treated it as a black box optimization and selected the Batched Gaussian Process Bandits method to maximize Precision and Recall (Golovin et al., 2017). Examples from the 10 tuned parameters include: thresholds on brightness temperatures that indicate cirrus clouds, the widths of the difference-of-gaussian line kernel used to highlight linear cirrus clouds, and the size and threshold of the "large scale gradient" operator used to filter out coastline false positives. See the accompanying open source code for more details.

There are multiple choices of metrics for evaluating satellite contrail detection models, and the ideal metric depends on which downstream analysis will be performed. For evaluating radiative forcing impact of contrails in aggregate, a per-pixel metric may be simplest. For analysis connecting flights to contrails (or further investigation of weather conditions under which contrails were formed) a per-contrail metric may be appropriate.

We observe that almost all of the identified contrails in this dataset are relatively young (likely less than 4 hours old - analysis not shown) and thus still fairly linear. We suggest for per-contrail metrics that post-processing both model predictions and polygon labels into linear objects can provide a common format to evaluate per-contrail metrics that is fairly easy to implement and which does not introduce too many additional hyperparameters that would require tuning.

Therefore the details on what constituted a true positive detection in this work are as follows: labeled polygons from individual labelers were converted to pixel masks and then linearized using a least squares regression of the contrail pixel coordinates in the image to generate linear contrail endpoints. These linear contrails were then combined across labelers to create the ground truth labels, keeping only those which were agreed upon by at least two labelers. A linear contrail was considered a match to another linear contrail if it had the same slope within 10 degrees and a mean distance of points on the contrail line within 10km to the other contrail. The Mannstein et al algorithm pixel masks were similarly converted to linear contrail endpoints and thresholded with the same angle and distance criteria to determine if they matched a ground truth label.

The line slope and distance thresholds used to match one linear contrail to another were selected by review of approximately 100 Landsat-8 scenes from the dataset. Each scene was rendered multiple times with the labels that passed the "two labelers agree" criterion for a few different options of line slope and distance thresholds. The authors then ranked which label set for the scene was most preferable, and the label sets resulting from "10 degrees slope" and "10 km mean distance" had the best (or tied for best) rank by a wide margin.

## 5. Conclusion

To our knowledge we are publicly releasing here the first large dataset of pixel-level contrail locations labeled by humans in Landsat-8 satellite imagery. We expect the labelers have been able to differentiate contrails from naturally occurring cirrus with good accuracy due to Landsat-8's high spatial resolution and the advected flight history information provided to labelers. This dataset will be useful in benchmarking contrail detection models and validating contrail research in colocated geostationary satellite imagery.

The results from the Mannstein et al algorithm above demonstrate there is room for improvement on this challenging first task of contrail detection, and we invite others to develop machine learning models and perform other analyses with this dataset. We also look forward to joining datasets of detected and tracked contrails with their ambient meteorological conditions to use as training data for machine learned models which are able to predict the formation and severity of contrails.

We intend to occasionally release new versions of the data available for download from the same location, with additional and refined labels. We recommend making a local copy of the dataset and accompanying license and documentation, since we may remove older versions of the dataset without advanced warning. We are hopeful this dataset will contribute to progress in contrail detection accuracy as well as provide a foundation for further empirical studies into contrail prediction, tracking, attribution, climate impact analysis and mitigation.

## Acknowledgements

The authors gratefully acknowledge: Rachel Soh, Jessica Ferguson, Jeanie Pearson, Ladislav Honsa and Rob von Behren for assistance acquiring and processing flight path data; Nathan Kiner for assistance setting up the contrail annotation pipeline; and the Google Research data operations team for labeling the Landsat-8 imagery.

## References

Appleman, H. The formation of exhaust condensation trails by jet aircraft. *Bulletin of the American Meteorological Society*, 34(1):14–20, 1953.

Avila, D., Sherry, L., and Thompson, T. Reducing global warming by airline contrail avoidance: A case study of annual benefits for the contiguous united states. *Transportation Research Interdisciplinary Perspectives*, 2:100033, 2019.

Bickel, M., Ponater, M., Bock, L., Burkhardt, U., and Reineke, S. Estimating the effective radiative forcing of contrail cirrus. *Journal of Climate*, 33:1991–2005, 2020.

Bock, L. and Burkhardt, U. Reassessing properties and radiative forcing of contrail cirrus using a climate model. *Journal of Geophysical Research: Atmospheres*, 121(16):9717–9736, 2016. doi: https://doi.org/10.1002/2016JD025112. URL

https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025112.

Bogacki, P. and Shampine, L. F. A 3 (2) pair of runge-kutta formulas. *Applied Mathematics Letters*, 2(4):321–325, 1989.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S., Sherwood, S., B., S., and X.Y., Z. *Clouds and Aerosols*, pp. 571–658. Cambridge University Press, 2014. doi: 10.1017/CBO9781107415324.016.

Burkhardt, U. and Karcher, B. Global radiative forcing from contrail cirrus. *Nature Climate Change*, 1:54–58, 2011.

Chen, C.-C. and Gettelman, A. Simulated radiative forcing from contrails and contrail cirrus. *Atmospheric Chemistry and Physics*, 13(24):12525–12536, 2013. doi: 10.5194/acp-13-12525-2013. URL https://acp.copernicus.org/articles/13/12525/2013/.

Duda, D. P., Minnis, P., Khlopenkov, K., Chee, T. L., and Boeke, R. Estimation of 2006 northern hemisphere contrail coverage using modis data. *Geophysical Research Letters*, 40(3):612–617, 2013.

Fritz, T. M., Eastham, S. D., Speth, R. L., and Barrett, S. R. The role of plume-scale processes in long-term impacts of aircraft emissions. *Atmospheric Chemistry and Physics*, 20(9):5697–5727, 2020.

Gierens, K., Matthes, S., and Rohs, S. How well can persistent contrails be predicted? *Aerospace*, 7(12):169, 2020.

Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J., and Sculley, D. Google vizier: A service for black-box optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1487–1495. ACM, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Jacob, S. D. and Rindlisbacher, T. The landing and take-off particulate matter standards for aircraft gas turbine engines. 2019. URL https://www.icao.int/environmental-protection/Documents/EnvironmentalReports/2019/ENVReport2019_pg100-105.pdf.

Kulik, L. Satellite-based detection of contrails using deep learning. Master's thesis, Massachusetts Institute of Technology, 2019.

Lee, D., Fahey, D., Skowron, A., Allen, M., Burkhardt, U., Chen, Q., Doherty, S., Freeman, S., Forster, P., Fuglestvedt, J., Gettelman, A., De León, R., Lim, L., Lund, M., Millar, R., Owen, B., Penner, J., Pitari, G., Prather, M., Sausen, R., and Wilcox, L. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. *Atmospheric Environment*, 244:117834, 2021. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2020.117834. URL https://www.sciencedirect.com/science/article/pii/S1352231020305689.

Mannstein, H., Meyer, R., and Wendling, P. Operational detection of contrails from noaa-avhrr-data. *International Journal of Remote Sensing*, 20(8), 1999. doi: https://doi.org/10.1080/014311699212650. URL https://www.tandfonline.com/doi/abs/10.1080/014311699212650?journalCode=tres20.

Meijer, V. R., Kulik, L., Eastham, S. D., Allroggen, F., Speth, R. L., Karaman, S., and Barrett, S. R. Contrail coverage over the united states before and during the covid-19 pandemic. Submitted to Environmental Research Letters, 2021.

Meyer, R., Mannstein, H., Meerkötter, R., Schumann, U., and Wendling, P. Regional radiative forcing by line-shaped contrails derived from satellite data. *Journal of Geophysical Research: Atmospheres*, 107(D10):ACL–17, 2002.

Meyer, R., Buell, R., Leiter, C., Mannstein, H., Pechtl, S., Oki, T., and Wendling, P. Contrail observations over southern and eastern asia in noaa/avhrr data and comparisons to contrail simulations in a gcm. *International Journal of Remote Sensing*, 28(9):2049–2069, 2007.

Minnis, P., Palikonda, R., Walter, B. J., Ayers, J. K., and Mannstein, H. Contrail coverage over the north pacific from avhrr data. *Meteor. Z*, 2005.

Palikonda, R., Minnis, P., Duda, D. P., and Mannstein, H. Contrail coverage derived from 2001 avhrr data over the continental united states of america and surrounding areas. *Meteorologische Zeitschrift*, 14(4):525–536, 2005.

Sanz-Morère, I., Eastham, S. D., Speth, R. L., and Barrett, S. R. Reducing uncertainty in contrail radiative forcing resulting from uncertainty in ice crystal properties. *Environmental Science & Technology Letters*, 7(6):371–375, 2020.

Schumann, U. A contrail cirrus prediction model. *Geoscientific Model Development*, 5(3):543–580, 2012.

Schumann, U., Penner, J. E., Chen, Y., Zhou, C., and Graf, K. Dehydration effects from contrails in a coupled contrail–climate model. *Atmospheric Chemistry and Physics*, 15(19):11179–11199, 2015. doi: 10.5194/acp-15-11179-2015. URL https://acp.copernicus.org/articles/15/11179/2015/.

Teoh, R., Schumann, U., Majumdar, A., and Stettler, M. E. Mitigating the climate forcing of aircraft contrails by small-scale diversions and technology adoption. *Environmental science & technology*, 54(5):2941–2950, 2020.

Vazquez-Navarro, M., Mannstein, H., and Mayer, B. An automatic contrail tracking algorithm. *Atmospheric Measurement Techniques*, 3(4):1089–1101, 2010. doi: 10.5194/amt-3-1089-2010. URL https://amt.copernicus.org/articles/3/1089/2010/.

Vazquez-Navarro, M., Mayer, B., and Mannstein, H. A fast method for the retrieval of integrated longwave and shortwave top-of-atmosphere upwelling irradiances from msg/seviri (rrums). *Atmospheric Measurement Techniques*, 6(10):2627–2640, 2013.

Vazquez-Navarro, M., Mannstein, H., and Kox, S. Contrail life cycle and properties from 1 year of msg/seviri rapid-scan images. *Atmospheric Chemistry and Physics*, 15 (15):8739–8749, 2015.