Review 2

Summary:

The authors use two deep learning techniques to estimate contrail cloud top height in GOES infrared imagery trained with CALIOP lidar data. Over 3000 contrails over five-year period are collocated with the lidar data to allow for the building and testing of the deep learning methods. The more successful contrail height estimate method was developed from a convolutional neural network (CNN), estimating contrail height with a root mean square error of 570 m in test data. An analysis of the CNN method results show that the predictive probability of the CNN method is generally well calibrated and has a smaller 95% confidence interval than the confidence intervals derived from flight altitude data alone. The authors also processed a 24-hour period of GOES data to show the spatial and temporal distribution of the contrail height estimates.

General comments:

The overall quality of the manuscript is good. The authors explain the methodology and results of the research concisely, and reach logical and consistent conclusions.

We thank the author for the time taken to review the manuscript and the comments provided.

Although the topic of the manuscript is contrail height estimation, the authors include discussion of thin cirrus height estimation that is unnecessary (considering the title of the manuscript) and confusing (especially subsection 3.2). Unless the authors can show why the cirrus altitude estimation is integral to the research presented in the main manuscript, I suggest that discussion about cirrus altitude estimation be removed from the paper.

We think it is essential to include this discussion for the purposes of this manuscript. One of the core objectives of the paper is to develop a contrail altitude estimation algorithm, but we consider it crucial to motivate this objective by showing that existing approaches (similar to what we refer to as “Cirrus MLP”) for cirrus height estimation – which have been applied to contrails as well – do not perform as well as expected. Moreover, the development of the “Cirrus CNN” algorithm is essential for the development of the “Contrail CNN” algorithm. This is because the Contrail CNN is constructed by taking the Cirrus CNN and fine-tuning it on the dataset of collocated contrails. In turn, the Cirrus CNN has been developed using collocations of cirrus clouds in CALIOP L2 data with the corresponding GOES-16 ABI infrared radiances. Hence, omission of a discussion of thin cirrus height estimation would not make sense from this perspective either.

With this in mind, we realize that the manuscript as previously written was not making this point adequately. We have added a statement in lines 52-53 of the introduction which highlights the reason for investigating the use of cirrus altitude estimation models for contrails. We also reiterate this point when introducing our cirrus-trained algorithm, and when comparing its performance to that of the contrail-trained algorithms (lines 290-292). Finally, in the conclusions we use our evaluations of the improved performance of the contrail-trained algorithm to argue that future research should not rely on approaches which were trained only on cirrus data, while also recommending that multi-pixel (CNN) methods can outperform single-pixel methods (lines 483-484).

The authors make multiple references to the Supplementary Materials, so much so that it is nearly impossible to understand the manuscript without also reading those pages. As a result, the manuscript is incomplete and might not stand alone. The reader should not have to rely on the Supplementary Materials to read the principal paper. Finally, the paper lacks references to multiple concepts that should be explained in the paper (not just the Supplementary Materials). I can find no references for the various height conversions (between geometric, geopotential, and pressure altitudes), the advection of contrails, or the parallax correction used in the main paper. Add these references to the manuscript.

The comments by the reviewer on the large number of references to the Supplementary Materials are appreciated and have been addressed. We have expanded the discussion in the manuscript on the following topics by use of material found in the Supplementary Materials:

- Construction of flight altitude distributions (section 2.9)

Furthermore, we have added new material on the following topics to the main paper:

- Conversion between geometric, geopotential and pressure altitudes (section 2.7)
- Advection (section 2.6)
- Parallax correction (section 2.5)

Specific comments:

Lines 100-104: Could not the width of contrail 2 also be the result of the geometry of the contrail relative to the CALIPSO ground track? Most of the other contrails are nearly perpendicular to the ground track, while the angle between contrail 2 and the satellite track is much more acute?

Part of the increased width (as viewed in the CALIOP data) is due to the angle this particular contrail makes w.r.t. the CALIPSO ground track. However, we think that this alone is not enough to explain the relatively large width of the contrail, compared to contrails 3 and 4. We have added the following sentences to the manuscript:

"The increased width of this contrail in the CALIOP data is partially caused by the angle between the CALIPSO ground track and the contrail. When correcting for this angle however, contrail 2 is still found to be wider than contrails 3 and 4."

Lines 109-117: This paragraph is unclear.. The authors state "The collocation process is nearly identical at that for contrails, except that the contrail detection masks are no longer involved." I can't find any mention of cloud masks up to this point.

We have modified the sentence to:

"The collocation process is nearly identical as that for contrails, except that the contrail detections described in subsection 2.2 are no longer involved."

Line 156: What are "normal" operations? Even after reading the paper, it is not clear to me what that means.

We have modified the sentence to:

"When using the CNN for contrail altitude predictions after training is finished, the resulting output (for the whole image) is only considered for pixels where contrails are detected."

Lines 172-173: '...all inputs to the neural networks are "observational"...' What is "observational" in this context? Why not say instead that the inputs are derived from the satellite radiances alone with no additional (NWP data) used?

Indeed, the wording suggested by the reviewer here is more exact. We have modified the sentence to:

"With the exception of the ERA5 land/sea mask and some of the inputs to the ACTP processing algorithm (Heidinger et al., 2020), all inputs to the neural networks used for altitude estimation are derived from observed satellite radiances."

Figure 2 and elsewhere: It is not always clear which altitude the authors are using. In Figure 2, for example, what type of altitudes are plotted here? Geometric? Geopotential? Pressure altitudes?

We have modified the manuscript in the following places to clarify which altitude is referred to:

- Figure 2
- Figure 4
- Figure 5

- Figure 8
- Figure 9

Lines 194-195: What is the "thickness" of the altitude distribution? It appears in Figure 2 that the variance of the altitude distribution increases as the latitude increases, contrary to the text.

We sincerely thank the reviewer for noticing this error. Indeed, the reviewer is correct in that the variance of the altitude distribution increases as the latitude increases. We have modified the text to read:

"The mean contrail top altitude decreases with increasing latitude, as was found by Iwabuchi et al. (2012). The variance of this contrail top altitude distribution is found to increase with latitude."

Figure 3: Why are the contrail tops generally so much higher than the ISS & SAC regions (except for summer)?

We think this is mostly due to sampling effects, given the limited spatial extent of the GRUAN data at the SGP site and the amount of temporal overlap between the ISS & SAC measurements and the contrail altitude data points. We have added a sentence on this to the manuscript:

"The contrail top altitudes in Figure 3 are generally found to be higher than the regions that are ice supersaturated and satisfy the Schmidt-Appleman criterion. Given the limited spatial coverage of the GRUAN data as compared to the contrail top altitudes (whose spatial distribution was shown in Figure 2) as well as the different times at which this data was captured, this discrepancy is likely due to sampling effects."

Earlier in the discussion of figure 3, we have also added a sentence that emphasizes the utility of the data in figure 3 to be mostly to compare seasonal trends:

"Finally, the flight altitude distributions in Figure 3 result from 1000 randomly sampled hours of ADS-B data in the years 2018 and 2019. Given the different locations and times at which the data in the three different plots is collected, Figure 3 serves mostly to compare seasonal trends in the altitude of flights, regions conducive to persistent contrail formation, and observed contrail top altitudes."

It is apparent from the paper that the three profiles (Flight, ISS & SAC, Contrail top) represent entirely different times, locations, and number of observations. It would be better to make this distinction much more clear to the reader, otherwise they may be confused by this figure.

We agree with the reviewer that this might be confusing. We have revised Figure 3 to include an inset that shows the location of the GRUAN site from which data is used, and have modified the caption to point out that the SAC & ISS data is from a single GRUAN site. The caption now also repeats the point about the flight data being from randomly sampled times in the years 2018 and 2019. In the response to the previous comment, we have also added another sentence to the manuscript that re-iterates this point.

Section 3.2: The authors refer much more to the Supplementary Materials here than the manuscript itself. Many of the values stated in the text don't match any of the values presented in Figure 4. This is very confusing! As stated earlier, I suggest the authors remove any discussion of cirrus altitude from the paper. It is superfluous and not presented well.

We agree with the reviewer that the discussion of Figure 4 in the original manuscript could be improved significantly. We have therefore made the following changes:

- Figure 4 itself has been revised to more clearly indicate the correspondence between the scatter plots and the models they are for, as well as more descriptive axis labels.
- We have revised subsection 3.2 completely. We have added an introductory paragraph that we hope clarifies the context of this subsection and its relation to the rest of the manuscript. We have also repeated some of the details of the four different altitude estimation models, and have expanded on the discussion of Figure 4.

Figure 5: Green line, blue line, black line. Which models do they represent? A legend would make this figure much easier to understand.

Our intention was to avoid visual clutter by using the same color scheme between the two plots in this figure (and hence the legend in the plot on the right also applied to the one in the left). To avoid any such future confusion, we have added the same legend to the plot on the left. We thank the reviewer for pointing out this source of confusion.

Line 311: Why is "simulate" in quotation marks? It appears to be a simulation (*i.e,* it imitates the appearance of) in the true sense of the word. The parallax correction is actually making the alignment of the flight tracks match better with the detected contrails.

This is a good point: we have removed the quotation marks.

Figure 7: Unless they looked that the Supplementary Materials, the reader would not know what "% of distance flown in 2 hours before" would mean. Some description of

how this quantity was obtained must be included in the manuscript, not just the Supplementary Materials.

We completely agree with the reviewer here and we apologize for the confusion. We have modified the manuscript as follows:

- Under methods, the section on “ADS-B data” has been extended with a discussion of the methodology used to obtain these quantities.
- In the paragraph starting with *“The comparison of flights to contrails is more complicated in areas of higher traffic density, and makes analyses such as that presented in Figure 6 infeasible given currently available tools. For these cases, we compare estimated contrail top altitudes to the distribution of distance flown, rather than individual flight tracks and their altitudes, as shown in Figure 7.”,* we have removed the reference to the supplementary material.