Review of "Improving discrimination between clouds and optically thick aerosol plumes in geostationary satellite data" by Robbins et al.

Clouds are one of the prime remote sensing targets from space-based instruments which can provide critical information on the global distribution of clouds horizontally and vertically along with their phases. Clouds also present a serious problem for retrieval of aerosol and gas species from such instruments. Classifying a layer accurately as cloud is thus very important. In this paper, the authors describe the development and assessment of a cloud mask for Advanced Himawari Imager using a neural network algorithm and collocated CALIPSO data for training and validation. The paper is well within the ambit of AMT and all the sections are well organized and written clearly. In particular I enjoyed reading the section on case studies showing the strengths and caveats of the network. However I have some comments primarily dealing with the authors' use of CALIPSO data. I recommend publication of the paper after revision.

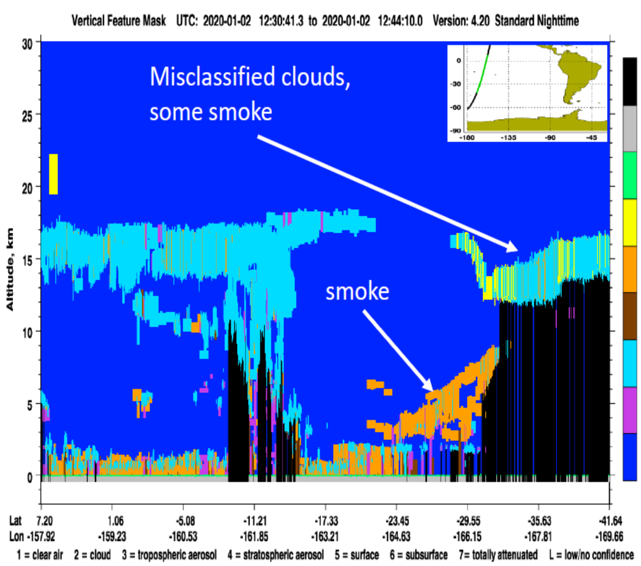
1. Since much of this work depends crucially on the Cloud Aerosol Discrimination (CAD) by CALIPSO, it is essential to be aware of some aspects of the latter. The authors state that they have analyzed both the 5km and 1km CALIPSO cloud products and then decided to go for the 1km product. Although the CAD algorithm in version 4 CALIPSO data has been extended to classify both the single shot (333m) and 1 km layers, the training sets used to generate the 5 dimensional pdfs that are used for the CAD algorithm never really used the optical properties of the 333m and 1km layers. In other words, the pdfs were generated using the measured variables of 5km layer products only and then applied to the 333m and 1km layers, so the quality of the CAD for 1km layers is largely unknown. Clearly this has ramifications for using the collocated CALIPSO dataset for training and validation. I would urge the authors to carry out a comparison of CAD at the two resolutions (5km vs 1km) to check on this. On line 102, the authors state that optimum results were found using the 1km layers. It is not clear how this was achieved, since no extinction retrievals are done for 1km cloud layers and an evaluation of the cloud optical depth bias being ingested in the training set is difficult to estimate. Perhaps a figure showing an example of the performance using both these resolutions and discussing the trade-off will be good.

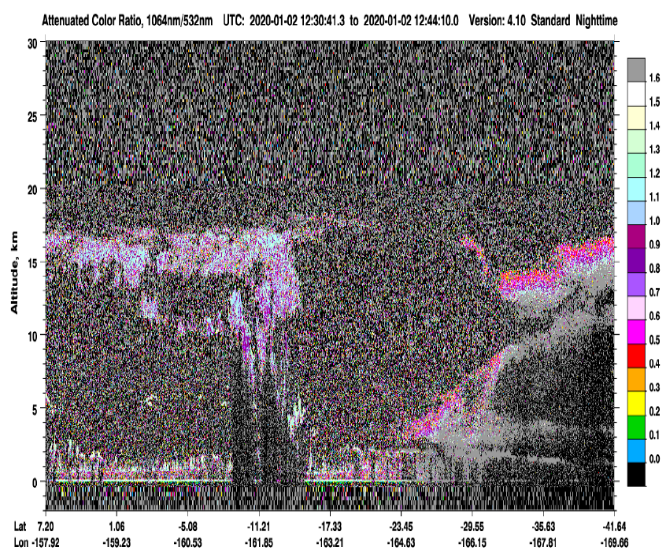
2. I think that more stringent filtering of CALIPSO data will be needed for using in the training and validation of the neural networks. For instance, since late 2016, CALIOP has been having issues with low energy laser shots---these primarily affect the data quality above the South Atlantic Anomaly (SAA) region but are increasingly affecting other parts of the globe. This can lead to artifacts in the data including false layer detections at all altitudes, particularly in the dayside (see data advisory: https://www-calipso.larc.nasa.gov/resources/calipso_users_guide/advisory/advisory_2018-06-12/CALIPSO_Laser_Energy_Technical_Advisory.pdf).

   In fact one can identify some of these false layers as vertical streaks in Figure 4, say near 42ºS. These false layer streaks probably have low CAD scores for the most part, which would take them out below 7.9 km by the CAD score filtering. It's not clear if the authors use the CAD score criterion above 7.9 km as well—if not, these false layer detections will contaminate the training sets at higher altitudes. These effects can be alleviated by using a threshold of minimum laser energy field (say >0.08 joules) provided in 1 km/5km layer data files or vfm files.

3. Section 3.2 on CALIOP should be expanded giving some details of CALIPSO version 4 CAD algorithm and also more substantial references (see the special issue on CALIPSO version 4 algorithm in AMT).

4. Lines 165-170---what are the implications of this altitude threshold? The Labonne et al. (2009) paper that the authors use is rather old and more recent data from CALIPSO and other instruments indicate that biomass burning plumes can be injected at altitudes higher than 7.9 km. In particular the pyroCb events can transport smoke plumes to very high altitudes into the stratosphere. These high altitude smoke plumes often have high depolarization ratio similar to clouds and CALIPSO CAD algorithm classifies a good number of them as clouds which appear in CALIPSO browse images side by side with smoke layers as can be seen in the browse image below. This scene is from January 2, 2020 and the arrows mark smoke plumes downwind from the Black Summer bushfires similar to the one shown in Figure 10.





(full CALIPSO imagery at:

https://www-calipso.larc.nasa.gov/products/lidar/browse_images/show_v4_detail.php?s=production&v=V4-10&browse_date=2020-01-02&orbit_time=12-03-45&page=3&granule_name=CAL_LID_L1-Standard-V4-10.2020-01-02T12-03-45ZN.hdf)

Note that smoke layers at low altitudes are classified correctly as such by CALIPSO but at high altitudes near 12-15 km, there are a large number of layers which are likely misclassified as clouds. As the authors point out in section 5.3.3, there is possibly some ice mixed with smoke coming from the pyroCb event. However note the very high attenuated color ratios (right panel) towards the base of the layers. This is a tell-tale signature of smoke in CALIOP products and suggest that most of the layers should have been classified as smoke. This might relate to the cloud misclassification by NN in Figure 10 that the authors discuss in lines 376-384 and indicates the caveats in using CALIPSO as “truth label”.

5. Lines 227-233 and Figure 4. Please provide the CALIOP granule information and also state the month and day rather than in abbreviated form (for this Figure as well as others). Why is the CAD score scale going from -100 to 100—for clouds it should be only from 0-100. What are the blue layers poleward of 50^{o}S and why are they appearing below the surface level?

6. The case study dealing with a strong dust storm (section 5.3.2) is somewhat intriguing. Asian dust and pollution events occur every year in spring (Huang et al., 2015, doi:10.1088/1748-9326/10/11/114018 , Di Pierro et al., 2011, doi:10.5194/acp-11-2225-2011) and plumes from these events often travel at high altitudes ~ 3-8 km towards the arctic. One major problem with CALIPSO CAD in version 3 was the misclassification of these thick aerosol plumes as clouds, but these should have been corrected in version 4 through improved CAD. Therefore I am somewhat surprised that the NN didn’t classify the 2021 dust storm correctly and once again wondering about the CALIOP CAD performance at 1km as possibly contributing to this.

7. Typo in line 413: “bask”→ “mask”

8. Line 94: CALIPSO→ CALIOP and Calipso→ CALIPSO

9. Line 339 and Figure 7: I had a hard time finding pale blue color in the sun glint area in the mask comparison plot!

10. Line 90: it would be better to specify the input variables for which the mean and standard deviation are being used.

11. In Figures 7,8 and 10-12, in the cloud mask comparison panels, dark blue is labelled as “all clear”—should it be “non-cloud” instead---as stated on line 328.