The response to the reviewers' comments is in italic.

Abhay Devasthale, SMHI, Norrköping, Sweden 2017-02-06

Review of Liu et al. doi:10.5194/acp-2016-1132, 2017

While I am a core satellite believer, I do understand and appreciate the importance of in-situ measurements, especially in anchoring space based observations. And there is perhaps no other region in the world where we desperately need more in-situ observations than in the Arctic. Combining these two (space based and in-situ) observing systems is even better. So I really appreciate the work done by the authors in this regard. I have few issues mentioned below that I regard minor in nature, but need to be explained/elaborated. I also had an opportunity to go through the comments posted by the other reviewer and I broadly agree with her/him and I hope the authors will address them as well.

We appreciate Dr. Devasthale's valuable comments. The manuscript becomes better with revisions in response to reviewer's comments and suggestions. We have responded to other reviewer's comments point by point, and made correspondent revisions in the revised manuscript.

1) The authors discuss a great deal about how they compute vertical cloud fraction, but very little (or almost nothing if I haven't missed anything obvious) about the spatial (and temporal) collocation of space based and in-situ measurements. The impact of uncertainties arising from these issues is not be underestimated, especially when you compare and combined products with different spatial resolutions (even at monthly mean scale). Let's say that you (or CALIPSO team) use 15 CALIOP single shots (1/3 km each, 5x3) to generate 5 km product. What happens when this 5 km product is not centered over Barrow or Eureka and you are inconsistently selecting single shots? Have the authors evaluated few individual cases manually to check what to expect when they merge 1/3, 1 and 5 km data with reference to the station in question?

We totally agree with the reviewer's comments, and thank for his insight. These issues, e.g. cloud frequency from surface observations v.s. spatial coverage from space-based observations, different spatial resolutions, viewing angles, vertical resolution among satellite products, all contribute to the shown differences in this manuscript. By using long-term observations, e.g. over 4 years at Eureka and over 2 years at Barrow (all data we have right now), we believe the temporal and spatial average would mitigate these issues. When longer term data from both surface-based and space-based are available, it is worth to revisit this, and see how the differences would change.

Inspired by the reviewer's comments, we add a paragraph in the "Conclusion" as the following, "Cloud frequency from surface is calculated in the temporal domain, while the cloud fraction from space-based observations is calculated in the spatial domain although near the surface sites. Differences in spatial resolution, viewing angles, vertical resolution, instrument sensitivity to clouds and retrieval algorithms may all contribute to the differences in the cloud vertical distributions from different instruments. Long-term averages of products may mitigate the impacts of some of these factors. Causes of the remaining differences are worth further investigation.
"

2) It would be helpful if the authors also provide some physical explanation of the seasonal highs and lows in cloud fractions seen in the results. For example, in the case of Barrow, why is cloud

fraction peaking in Feb, Apr and Oct months? Why is there a minimum in Jun and Jul? This is different from Eureka. Why? Perhaps Shupe et al (2011; 2015) already discuss this, but I think the reader still needs at least a brief description of it to make full sense of the differences you observe from these two observing systems.

A short description of the difference between Barrow and Eureka has been added to the end of Section 3.1.2. This explanation also links to a more detailed discussion of the matter in Shupe (2011). The discussion is "In additions, both satellite and surface observations reveal a key difference to the annual cycles of clouds at Eureka versus Barrow. While both sites have a similar annual cycle of ice cloud occurrence with a relative decrease in summer (Figure 8a, and 8d), there are less frequent liquid-containing clouds at Eureka with the annual maximum of these generally shifted to the autumn. These relative annual cycles explain the key differences in total cloud occurrence fraction over the annual cycle and are explained by generally colder and drier conditions in Eureka relative to Barrow (e.g., Shupe 2011)".

3) In the case of Barrow station, I am bit surprised at the differences in CF between 2B-GEORPFO and 2B-GEORPOF-Lidar in Aug (Fig. 2). When you add CALIOP there seems to be increase in clouds in the free troposphere from 1 to 5 km. Instinctively, I would have thought that, in the free troposphere, CALIOP would add those subvisual or super thin clouds that are missed by CPR, located in the upper troposphere lower stratosphere. Nearly 30-40% more clouds are added by GEOPROF-Lidar compared to GEOPROF in the lower and middle troposphere and it seems that even surface measurements missed these clouds. Even more confusing is the fact that CALIPSO 5 km doesn't show these clouds in Aug. So what is happening here? Part of this discrepancy can be due to the attenuation of CALIOP signal and part of it due to high amount thin clouds in the middle and lower troposphere (Devasthale et al. 2011). But it is difficult to say without further investigations.

I agree with the reviewer's comment. The GEOPROF-Lidar has higher values than the sum of those from 2B-GEOPROF and CALIPSO 5 km in August at Barrow. The reviewer gave some possible causes, and we appreciated that and have included such discussion in the revised manuscript. However, it is still unclear why the 2B-GEOPROF-lidar has higher values than the sum of those from 2B-GEOPROF and CALIPSO 5 km. Though finding the causes is beyond the scope of this study, it is worth further investigation in future work. The following discussion has been added in the revised manuscript.
"It is worth pointing out that the 2B-GEOPROF-LIDAR shows higher cloud amount values from 1 km to 5 km in the troposphere than the sum of cloud amounts from 2B-GEOPROF and CALIPSO 5 km. The differences can be partially attributed to the attenuation of CALIOP signal and high amount thin clouds in the middle and lower troposphere (Devasthale et al. 2011). Though attribution investigation is beyond the scope of this study, it is worth further investigation in future studies".

4) The authors say that the blended cloud vertical distribution provides a complete picture. But how do we quantitatively know this? After all, we need a third independent reference to make that conclusion.

We totally agree. A 3-D cloud distribution product would be ideal with known uncertainties. However, such a product does not exist, and probably will not be available in the near future. So, in my humble opinion, we need to work hard on getting the uncertainties of the existing products, and hopefully merging them for better quality. That is the motivation of this study.