Response to

Review of "On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO-CALIOP data" by Karlsson and Johansson, (amt-2013-3)

by Karl-Göran Karlsson, 15 April 2013

The response is given as inserted blue text below:

This paper uses CALIPSO/CALIOP cloud data to evaluate cloud occurrence and cloud top height in the CM SAF CLARA-A1 dataset, which is based on AVHRR observations. They have identified 99 orbits of the NOAA-18 satellite matched in space and time to the CALIPSO orbit. This provides a matched AVHRR-CALIOP dataset which can be used to characterize retrievals from not just the NOAA-18 AVHRR but from all AVHRR instruments. The authors take some care in preparing an appropriate version of the CALIOP data for the comparison, develop a number of different statistics for evaluating the PPS cloud data and provide a good discussion and interpretation of the results. The paper provides new and interesting results and merits publication, but would benefit from some revisions.

General comments on description of the CALIOP cloud data products:

The structure of the CALIOP cloud products is rather complicated and not directly suited to the comparison the authors wish to perform in this paper. This is the motivation for their development of an "optimal" method. The CALIOP cloud products are described in sections 2.2and 3.1, but not clearly or in detail, and several key papers describing the retrieval algorithms are not cited. Because of an apparent misunderstanding of the algorithm used to produce the cloud data products, they raise concerns in section 3 which are unwarranted. In its description of the CALIOP dataset the paper repeats some common misconceptions about the nature of the CALIOP data. In the end, I don't think these have a significant impact on the results but it is better to have a clear understanding of the data products being used in any validation study.

Karl-Göran: We have understood that there are indeed some misinterpretations of both the content of the CALIOP dataset and the method used to produce them (more about this further down) in our descriptions. The whole motivation for this work was the discovery that sometimes the total cloud fraction for a full global orbit was higher for CALIPSO 1 km data compared to CALIPSO 5 km data. This was found frustrating when given the task to validate cloud results from 5 km AVHRR GAC datasets which apparently would fit best in resolution to the CALIPSO 5 km datasets. At that point we did not realize that the two datasets should not be seen as completely independent from each other and we sought explanations that could have to do with the different averaging approaches. We will provide a better motivation for the study and give a more appropriate description of the CALIOP products in the final revised manuscript.

Apart from this, we still maintain the view that we need to use finer resolution CALIOP datasets together with the 5 km resolution dataset to get the optimal information for evaluating global cloud retrievals from passive imagery such as from AVHRR and MODIS. This comes from the fact that sub-resolution (i.e., below 5 km scale) cloud elements – in this case dense but isolated convective boundary layer clouds – will appear very bright in passive imagery and are often detected as cloud-contaminated pixels or even as if being covering the pixel completely even on the 5-km scale. The detection of these sub-pixel clouds are important, e.g. for SST applications. Thus, to not have information about their presence in the reference CALIOP dataset (here the 5 km dataset, which obviously do not include information about such clouds at a finer resolution) would not be fully acceptable. For this reason it was found important to add this information about sub-pixel cloudiness to the 5 km CALIOP datasets in order to utilize the CALIOP datasets in an optimal way for evaluating cloud retrievals from passive imagery. We are glad that the reviewer still gives us green light for this concept and we hope that the revised manuscript gives a better description on how to achieve the optimal use for our purposes.

To clarify my concerns: The CALIOP retrieval algorithms report clouds detected at 5 different spatial resolutions. The paragraph beginning on line 8, page 1100, and the discussion in section 3.1, describe these as if they are equivalent in some sense. Clouds detected in single shot profiles and in profiles averaged over three shots are reported in the 1/3 km and 1 km cloud layer products, respectively. Clouds detected after averaging to 5 km, 20 km, and 80 km are all reported in the 5-km Layer Product. When the authors discuss using the "5 km dataset" I assume they are using clouds detected at all three resolutions, but this is not clear. The 1-km dataset is independent of the other four and was intended to provide a product with spatial resolution matched to passive sensors having 1 km IFOV. As described in the CALIPSO ATBDs (Vaughan et al. 2005) and several papers (Vaughan et al 2009, Winker et al. 2009 – see reference list below) the clouds detected at 1/3, 5, 20, and 80 km are complementary and, together, provide the complete description of clouds observed by CALIOP. (Clouds detected at 1 km can be substituted for 1/3 km clouds, somewhat less optimally, as described below.) Clouds detected at higher resolutions are subtracted from Level 1 profiles before averaging, to avoid smearing dense clouds together with more tenuous clouds (and aerosols).

Karl-Göran: We admit that we should have better distinguished between the 1/3 km, 1 km and 5 km datasets. A clear mistake was the belief that products were also given at 20 km and 80 km resolutions which is not true. Results after averaging also at scales 20 km and 80 km are indeed included in the 5 km dataset. Thus, equation 1 in the manuscript is formally wrong already from this point of view (i.e., there are no 20 km or 80 km products).

And you are right, we did originally think that the different datasets were comparable, just differing by the fact that more and more thin clouds would be detected for coarser resolutions. We seriously thought that, even if (as you state in the last sentence above) higher resolution clouds were removed during the averaging process at a coarser resolution, the information about those clouds would still be retained and transferred for inclusion in the coarse resolution datasets. If that was true inequality (1) would still have been valid. But we realize now that this is not the case. This equation will be reformulated or even removed in the final manuscript.

The authors point out that Eqn (1) "is not fulfilled in all cases". This is correct, because Eqn (1) does not follow from the way the products were designed and implies an algorithm different from the one that is used to produce the CALIOP data products. To explain: In trade cumulus regions, for example, where clouds tend to be small and bright, most cloud is detected in 1/3 km, single-shot profiles. These detected clouds are subtracted from the Level 1 profiles before averaging to 5 km and often the 5-km average profiles are found to be cloud-free. Thus, there is no general expectation that CFC_0.3 km < CFC_5km. Similarly, clouds detected at 5 km are subtracted from the Level 1 profiles before averaging to 20 km, and clouds detected at 20 km are subtracted before averaging to 80 km (Vaughan et al. 2005). We may find that CFC_20 km < CFC_80 km, but that depends on the nature of clouds and is not an inherent behavior produced by the algorithm.

Karl-Göran: I completely understand this now. Once again, I can only confirm that we misinterpreted the content of the different datasets. To warn other users I would strongly recommend them to pay full attention to the meaning of the following statement in the CALIPSO Quality Statements Lidar Level 2 Cloud and Aerosol Layer Products Version Releases: 3.01, 3.02 (stated after the description of the three different spatial resolutions for the layered products):

"Users should be aware that while the 5 km layer products are reported on a uniform 5 km grid, the amount of horizontal averaging required to detect a layer may exceed 5 km. For example, detection of subvisible cirrus during daylight operations may require averaging to 20 km or even 80 km horizontally. In these cases, the layer properties of the feature detected are replicated as necessary to span the full extent of the averaging interval required for detection. For example, the layer properties for an aerosol layer that could only be detected after averaging over 20 km horizontally will be repeated over four consecutive 5 km columns."

Thus, there should be no doubt that information on cloud layers from coarser scales than 5 km is indeed included in the 5 km dataset. I can only regret that we did not draw the right conclusions from the beginning here.

Line 23, page 1101, notes that the 1 km dataset sometimes has larger cloud fractions than the 5-km dataset. This is not an inconsistency and does not indicate a problem in the dataset. This has been reported by other researchers because it results from the design of the algorithms. As described above, cloud layers detected at 0.3 km are subtracted from Level 1 profiles before averaging to 5 km. They are not subtracted before averaging to 1 km, however. Therefore, CFC_0.3km is less than or equal to CFC_1km in all cases, because 1 km cloud detection is based on simple averaging of three 1/3 km profiles. In trade cumulus regions, where cloud horizontal extents are often less than 1 km, this leads to an overestimate of cloud fraction in the 1-km product. As described above, if all cloud in a scene is dense and detected at 1/3km, then no cloud will be reported at 5 km and coarser resolutions.

Karl-Göran: Perfectly clear now but there are still ambiguities that sometimes make the interpretation difficult. For example, many of the clouds detected at single-shots at 1/3 km and in the 1 km dataset (treating them now as being more or less the same) extend over scales that are larger than 5 km. So what are you really removing when averaging at coarser scales (e.g. 5 km)? How do you guarantee that you only remove features at the finer scale and retaining the larger scale clouds? This is a critical issue. I understand that this is taken care of by the so-called "Boundary Layer Cloud-Clearing" method as described in the ATBD document. This identifies in particular small-scale clouds that are much smaller than the analysed 5 km scale but which produces quite strong signals in the individual single-shots about to become averaged. So, you identify these clouds and remove them from the averaging. However, it turns out that it is exactly those clouds that give us the problem. As these clouds are intensively reflecting they will strongly affect AVHRR-measured FOVs even at the 5 km GAC scale. Thus, they are very likely to be detected (even if they from a theoretical point of view are sub-pixel clouds). So, they (as you state it) do explain why the 1/3 km and 1 km datasets sometimes can have more clouds than the 5 km dataset. But we are definitely also interested in knowing about the additional thin cloud layers detected after averaging of the 5 km resolution (and higher resolutions). So to have a complete picture of how clouds appear we concluded that the 1 km and 5 km datasets should be merged in some sense.

Finally the author's conjecture, at the beginning of Section 3.1, that "lost clouds" in the 5km dataset are "most likely included" in the 1 km dataset is correct, by the design of the algorithm. These clouds are not "lost", they are detected and reported at 1/3 km and then show up in the 1 km dataset, after averaging. Clouds detected at all 5 resolutions are reported together in the VFM, although only the cloud occurrence and not the optical depths.

Karl-Göran: Perfectly clear.

In spite of the problems in Section 3, the CALIOP data seem to have been used in a reasonable manner and the results in Section 4 appear to be reasonable and meaningful. Figure 4 indicates the detection threshold of CLARA-A1/PPS cloud mask algorithm is about 0.3, in optical depth, which is similar to the ISCCP threshold (also largely based on AVHRR).

Karl-Göran: Well, it seems sometime you do things correctly but for the wrong reasons (☹). Nevertheless, we are glad that misinterpretations did not in any way remove the motivation to find an improved methodology compared to using exclusively data from the 5 km CALIOP datasets.

Detailed comments:

1) There is no discussion in the paper on how the method adopted is "optimal." That is, how this method is better than all other methods. In any case, an optimal method is only optimal for certain applications. While the method described here is reasonable and defensible, I would encourage removal of the word "optimal" from the title.

Karl-Göran: Optimal means that we need also to take into account the strongly reflecting boundary layer clouds that are not included in the CALIPSO 5 km datasets. If not restoring this information from the 1 km datasets, we would risk concluding that many cloudy AVHRR-GAC pixels are misclassified cloud-free pixels. But they are not – they do contain sub-pixel clouds – and this must be acknowledged. Such knowledge is extremely important for some applications (e.g. SST-retrievals).

2) There are quite a few references to a CALIOP "5 km FOV." The CALIOP FOV is 100 m. Averages over 15 laser shots are usually referred to as "5-km averages" or "5-km segments".

Karl-Göran: This will be corrected in the revised manuscript.

3) The citation for Stubenrauch, et al. (2012) can be updated. The final report (WCRP Report No. 23/2012) is available at:

http://www.gewex.org/gdap/gdap_assessment_wgs.html

I'm not sure what document is referenced on the NASA NTRS server. When I attempted to locate it, access to the NASA NTRS data server was blocked.

Karl-Göran: Yes, we will update the reference.

4) A slight comment on page 1098, lines 11-12, which imply CALIOP data products contain only geometric cloud top height: They also contain cloud top temperature. The justification for only looking at the CALIOP cloud height product is that cloud height is directly measured, whereas temperature is derived from a meteorological analysis product and the observed cloud height.

Karl-Göran: Sure, agree. Will make this clearer in the text.

5) Page 1099, line 24: the lidar signal becomes fully attenuated after penetrating an optical depth of 3 to 5, not 6-10. Larger optical depths are sometimes reported in the data products, but these are highly uncertain and should be ignored.

Karl-Göran: OK, I will modify this statement accordingly.

6) Page 1099, line 27: a better URL for the CALIPSO data archive would be: http://eosweb.larc.nasa.gov/PRODOCS/calipso/table_calipso.html

Karl-Göran: Good, I will update (even if this link temporarily (?) seem to not work right now).

7) I'm afraid my communication quoted on lines 1-2 of page 1102 will be misunderstood, as it is taken out of context. My statement that the CALIOP dataset is not "ideally suited for this task" was not that cloud information was lacking, but that the information was split between the 5-km and 1/3-km (or 1-km) Cloud Layer products and so required the combination of multiple products. When constructing a Level 3 monthly gridded cloud product we used the 1/3-km and 5-km Cloud Layer Products (Chepfer et al. 2013), avoiding the 1-km product which can overestimate low cloud cover.

Karl-Göran: OK, I will remove this statement in the text.

8) On page 1102, lines 11-15, the results of Chan and Comiso (2011) are referenced. Unfortunately, some misconceptions in that paper are repeated. Chan and Comiso report on MODIS and CALIOP observations at high latitudes. They show several scenes which CALIOP retrieves as "clear" but MODIS retrieves as cloudy: marine stratus with cloud top heights of 1.5 km or more. CALIOP Level 1 data shows these 'clouds' are fog layers confined to within about

300 m of the surface. MODIS retrieves a reasonable optical depth, but places the layers much too high above the surface. Because the fog layer extends to the surface it cannot be distinguished from the ocean surface by the CALIOP Version 3 algorithm, and is classified as part of the surface return. The limitation of the CALIOP retrieval has nothing to do with the optical depth of the layers (as stated by Chan and Comiso) but that they are immediately adjacent to the ocean surface. Technically, this is not a problem of missed detection, but a problem of being unable discriminate the fog layer from the ocean surface.

Karl-Göran: OK, I will remove the reference to the Chan and Comiso paper.

9) Page 1103, line 10: I don't understand why a 5-km segment is set to 'clear' if one or two of the 1-km segments within the 5-km is cloudy. The 5-km segment is obviously partly-cloudy if at least one and less than 5 of the 1-km observations is cloudy. Why draw a threshold at 50%? Will the AVHRR threshold test be unable to detect cloud if the 4-km AVHRR pixel is less than half-filled? In any case, the CALIOP swath can only characterize the cloud cover of partly cloudy AVHRR pixels in a statistical sense.

Karl-Göran: At some point we need to judge whether we think that a cloud may be detected in a 5 km GAC FOV or not. This is really difficult since it depends on how strong signal these sub-pixel resolution clouds is really giving in the measurement. We thought it was reasonable to set the limit to 50 %. In the revised manuscript we have also studied the impact of setting the limit to 25 % or 75 % (i.e., a sensitivity test was requested by another referee). In any case, it is judged as too strong to label the pixel as cloudy if just one of five 1 km CALIPSO segments is cloudy. But we admit that this is a weak point in our method/study, showing that in this situation we rely more on the 1 km dataset than on the 5 km dataset. However, it is difficult to judge how weak it is in comparison to other uncertainties. For example, how certain can we be that thin cloud layers detected at 5 km resolution are really covering the entire 5 km segment? I could imagine that this uncertainty is at least equally important.

10) Page 1104: I'm not sure what the concern is in lines 8-16. It is not clear the authors are aware that, at altitudes below 4 km, clouds detected on single shots are subtracted from 1/3 km profiles before being averaged to 5 km. At low altitudes, the cloud layers reported at 5 km tend to be either the bases of dense overcast clouds or the averages of thin, broken cloud which is missed by 1/3 km detection. Because 1/3 km cloud detection is already more sensitive than passive cloud detection, we tend to ignore low-altitude 5-km cloud layers when computing cloud cover. At altitudes above 4 km, the cloud fraction within 5 km layers could be tested by comparison with 1/3 km or 1 km cloud detections. The final paragraph is correct, however, that results based on both 1 and 5 km products will be better than if 1 km or 5 km products are used alone.

Karl-Göran: Well, maybe this part can be omitted (at least, we have reduced the size of the discussion here) since it is kind of biased and influenced by the misinterpretations discussed previously. On the other hand, one could wonder what happens to fractional cloudiness detected at a few single shots and being above 4 km altitude (thus not being removed by the Boundary

Layer Cloud-Clearing)? Will this end up as a thin single layer cloud in 5 km data? Maybe this case could correspond to the thin broken clouds you talk about. But this case shows that the 1/3 km and 5 km datasets are not only complementary. There is also some overlap that makes it a bit difficult to define exactly what is the global cloud coverage (estimated over a full orbit).

Additional references:

Chepfer, H., G. Cesana, D. Winker, B. Getzewich, M. Vaughan, and Z. Liu, 2012: "Comparison of two different cloud climatologies derived from CALIOP attenuated backscattered measurements (Level 1): the CALIPSO-ST and the CALIPSO-GOCCP" , *J. Atmos. Oceanic Technol.*, doi:10.1175/JTECH-D-12-00057.1, in press

Vaughan, M., K. Powell, R. Kuehn, S. Young, D. Winker, C. Hostetler, W. Hunt, Z. Liu, M. McGill, and B. Getzewich, 2009: "Fully Automated Detection of Cloud and Aerosol Layers in the CALIPSO Lidar Measurements", *J. Atmos. Oceanic Technol.*, **26**, 2034–2050, doi: 10.1175/2009JTECHA1228.1

Vaughan, M., D. M. Winker, and K. Powell, 2005: CALIOP Algorithm Theoretical Basis Document Part 2: Feature Detection and Layer Properties Algorithms. PC-SCI-202.02. available at http://www-calipso.larc.nasa.gov/resources/project_documentation.php

Winker, D. M., M. A. Vaughan, A. H. Omar, Y. Hu, K. A. Powell, Z. Liu, W. H. Hunt, and S. A. Young, 2009: "Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms", *J. Atmos. Oceanic Technol.*, **26**, 2310–2323, doi:10.1175/2009JTECHA1281.1.

Karl-Göran: Will be added.