

Review of ‘Evaluation of the Aqua MODIS Collection 6.1 multilayer cloud detection algorithm through comparisons with CloudSat CPR and CALIPSO CALIOP products’ by Marchant et al.

This manuscript by Marchant et al. shows an interesting evaluation of the MODIS/Aqua multi-layer product through comparisons to a lidar-radar cloud detection product from CALIPSO-CloudSat. These comparisons are thoroughly done for different multi-layer conditions, i.e. depending on the thermodynamic phase, optical depth and distance between the two layers. The performance of two versions of the MODIS algorithm (including the operational C6) are discussed, in terms of absolute accuracy against active products but also in terms of significance to avoid and/or flag biases in other MODIS cloud products.

The manuscript is clear, well written, of good scientific significance and absolutely fits the scope of AMT. I therefore advise for publication of this work, once the authors will have addressed the following (relatively minor) comments.

---

General comments:

1. Are the findings shown in this manuscript really limited to Aqua? I realize that the evaluation only is possible for the instrument onboard Aqua but is there a reason to think the conclusions are not just as valid for MODIS/Terra? If not, I wouldn't emphasize the Aqua dependence in the title and abstract.
2. The quality of the 2B-CLDCLASS-lidar product used in this study should at least be briefly discussed. For instance, I assume that the identification by lidar-radar of the thermodynamic phase becomes increasingly less accurate for the lowermost detected layers – is that of significance for the results presented here? Also, please explicit what is meant by “mixed phase”.
3. It would be useful for readers and MODIS users if the authors further relate their results to the actual Cloud Multi Layer Flag SDS. In section 3 the authors describe the 4 methods / tests for multi-layer detection and explain that they are merged into a single confidence-level metric that ranges from 2 to 10 in case of multi-layer. I think that a couple more sentences explaining how the cumulative weight is obtained would be helpful. I realize that this paper does not aim to be too technical or replace the ATBD but it will likely become a reference paper for those interested in the multi-layer detection product. Also, it is unclear how the MODIS multi-layer cases that are shown in the manuscript actually relate to the SDS value, do they correspond to all cases with a value greater or equal to 2?
4. Related to the previous comment, and because this paper is likely to become a reference for the C6 multi-layer product, it would be very helpful if the authors included a brief bullet list of the practical implications of their findings, which users could easily refer to. For instance reminding that i) the MODIS multi-layer detection is to primarily be used as a retrieval quality indicator, ii) the flag should mainly be used when interested in ice cloud retrievals (as liquid cloud retrievals are by construction not too impacted?), iii) perhaps a word on the SDS values to be used (2 or higher?) for different cases, etc.

---

Specific comments:

1. p.6 l.145: Do I understand correctly that the L2 product in C6 includes the PH04 but the corresponding L3 product does not? If so, it would be worth emphasizing this by repeating it somewhere that be more visible to the readers (introduction or conclusion).

2. Fig. 3 and its analysis: It is interesting that the proportion of true/false detection of multi-layer cases in MODIS remains the same with or without using PH04. In both cases there is a 50% agreement with 2B-CLDCLASS and only the overall proportion of multi-layer detection changes. Would you then consider that PH04 does not significantly improve the quality of multi-layer detections or does the 8 vs 12% detection rate still make a difference to avoid biases on cloud properties? Fig. 11 indicates that PH04 does improve a bit the agreement ice cloud CER retrievals obtained in single- and multi-layer conditions, but I wonder if it is significant enough to risk higher false rejection rates.

3. Fig. 8 and its analysis: Why not also use the  $OD > 4$  threshold here, for a better consistency with the following results related to Fig. 9-11?

4. p. 13 l. 303–305: It is typically considered that effective radii retrievals associated with optical depth below 3 or 4 are not accurate, then is it really worth showing and discussing the results of Fig. 11?

---

Technical comments:

1. p3 l53: “a two-layer cloud overlapping model” sounds like the layers are not vertically separated, which would be surprising. Perhaps “a two-layer model” is sufficient?

2. p5 l105: might be worth precisising “thermal IR”.