Anonymous Referee #2

**General comments:**

Summary of paper: The authors describe a data set of global cloud properties based on AVHRR observations, available since 1982. The data set is an update of version 2, with the main changes the use of artificial neural networks for cloud mask and cloud phase detection, and additional cloud radiative properties. Both versions 2 and 3 are evaluated against the best available retrievals from other satellite and ground-based products. Standard verification metrics indicate overall improvement in most cloud properties, with some deterioration in ice cloud top height. The cloud radiative properties compare well against CERES observations.

Review: This paper is generally very well written. It is mostly complete and useful information for anyone wishing to use this data set. Comments mostly concern some further clarification sought and perhaps slightly different presentation of the differences between version 2 and 3. The DOI links to a suitably presented web page describing the data. Overall, the recommendation is to accept this paper with minor corrections.

**Minor comment:**

*Referee comment:*

Statistical significance. It is not immediately clear whether any of the differences in skill between v2 and v3 reported in the tables are statistically significant, although the large sample suggests these are. However, it should be possible and it would help the reader if the maps in figures 1, 2, 5, 6, 7, and 8 could include (i) difference/bias maps and (ii) stippling/hashing/shading for statistically significant differences. Most of these maps are visually similar and might hide key differences due to the colour scale used. A different way of presenting the various data sets, including additional maps of bias and statistically significant differences, would help inform the reader how the new data set compares against existing data sets.

*Author's response:*

Thank you for this suggestion. We updated Figures 1,2,3,5,6,7 and 8 to include difference maps as suggested. Unfortunately, we have not always all information available to calculate and provide the statistical significance of the differences, which are for this reason not included.

*Author's changes to the manuscript:*

Updated Figures 1,2,3,5,6,7 and 8, now including difference maps.

**Line-by-line comments:**

*Referee comment:*

p2. l28-30. This sentence is difficult to read, especially the first part.

*Author's response:*

Thanks. We will rephrase that sentence.

*Author's changes to the manuscript:*

The rephrased sentence will read "For the MODIS cloud record however there is the potential to be combined with high quality TOA radiation measurements made by the Clouds and the Earth's Radiant Energy System (CERES) sensors mounted on board the same platforms (Terra and Aqua)."

*Referee comment:*
p2. l33. "limitations". Limitations to do what?
*Author's response:*
We mean the limitation to resolve cloud and their radiative effect at smaller scales that the CERES footprint size. We will rephrase that sentence.
*Author's changes to the manuscript:*
The rephrased sentence will read "Limitations to resolve small scale clouds and their radiative effect might arise from the coarse spatial resolution of CERES (footprint size of approximately 30 km) and from the fact that the clear-sky fluxes are exclusively based on clear-sky pixels (and interpolation of clear-sky fluxes for gap filling on monthly scales), for which the spatio-temporal sampling is reduced and the meteorological conditions are likely to be biased."

*Referee comment:*
p3. l3-7. Please provide references for the WCRP GEWEX data and the ISCCP DX data.
*Author's response:*
Thank you. We will include the reference Stackhouse (2011) and Rossow and Schiffer (1990).
*Author's changes to the manuscript:*
Included reference Stackhouse (2011) and Rossow and Schiffer (1990).

*Referee comment:*
p3. l10-14. "based on the rationale above". It is clear why these data are required, compared to the MODIS/CERES and GEWEX data sets. However, what are the other data sets based on AVHRR lacking (PATMOS-x, CLARA-A2, Cloud-cci) that this paper will address with Cloud cci 3? A sentence on p2, line 19-21 would help clarify the shortcomings of those existing data sets.
*Author's response:*
Thanks for this suggestion. We will add the suggested sentence; however we find that sentence better placed on p3 l 13, as we need the provision of a full suite of cloud and radiation properties to make the point.
*Author's changes to the manuscript:*
We will add the following sentence near line 13 on page 3: "The availability of the full suites of cloud and radiative flux properties will also make these data superior to the already existing AVHRR-based datasets mentioned above."

*Referee comment:*
p3. l20-31. This paragraph pre-empts the findings ("superior") and methodology. The relevant information is better placed in section 2.1.
*Author's response:*
Thank you for this suggestion. We kind of see the point. However, algorithm developments (content of section 2.1) are only one reason for this finding. The longer time period covered and in particular the inclusion of the radiative broadband fluxes are strong contributors to this finding/conclusion as well. As these characteristics are motivated in the introduction we would like to leave that paragraph where it is right now.
*Author's changes to the manuscript:*
None.

*Referee comment:*
p4. l2. Please add that table 1 contains all abbreviations used throughout the text. CER had not been introduced in the main text prior to p6 and it took a moment to figure out its meaning.
*Author's response:*
Thank you. We'll do that.
*Author's changes to the manuscript:*
We will modify that sentence: "The set of cloud properties included in v3 is identical to v2 and is outlined in the upper part of Table 1, which also gives all cloud property abbreviations used throughout the paper."

*Referee comment:*
p4. l16. "much larger set". How do the two sets of training data compare? Did both v2 and v3 use CALIOP, but v2 just used fewer overpasses?
*Author's response:*
Yes. Both used CALIOP and for v2 we used much fewer overpasses compared to v3 (about a factor of 10). We will clarify in the text.
*Author's changes to the manuscript:*
The corresponding sentence will be changed to "The ANN for cloud detection (ANNmask) has been retrained using a much larger set of training data (approx. 10 times more collocation data used for v3 than for v2), which is composed of…."

*Referee comment:*
p6. l16. A "lower" CTP mean is not explained by more very low-level clouds, which have higher CTP. It appears that over the West Pacific and Maritime Continent, mean CTP has generally increased, which could be due to detection of more low-level clouds. Please re-consider this statement.
*Author's response:*
Thank you. This is indeed a type. A "higher" CTP it is. We will revise that sentence.
*Author's changes to the manuscript:*
That sentence will be changed to "Mean CTP is higher in v3 than in v2 in the Tropics…."

*Referee comment:*
p7. l1. Regarding the validation, did the authors consider performing the validation separately for daytime and nighttime observations? The algorithms use different channels and the authors consider nighttime COT and CER "experimental". It would be useful to understand the algorithm performance for different times of the day.
*Author's response:*
It is indeed useful. For this manuscript however we would prefer presenting just the general figures as the discussion would get to extended otherwise. Important to consider in this respect that there is a Project Validation Report soon being issued in which we plan to stratify the validation results wrt. illumination conditions. We will add this link to the text.
*Author's changes to the manuscript:*
At the end of Section 2.3 we will add the following sentence "An even broader assessment of the quality of the presented dataset can be found in PVIR (2019), in which the results are also stratified by illumination conditions among others."
Along with adding the reference:
PVIR – Product Validation and Intercomparison Report (PVIR) ESA Cloud_cci, 2019, Issue 6, Rev. 0; DD/MM/2019, available from http://www.esa-cloud-cci.org/?q=documentation_v3, 2019.

*Referee comment:*

p7. l3. Please, briefly explain how the collocation is carried out. In particular, what is the impact of the temporal mismatch between CALIOP and AVHRR? And what is the impact of the mismatch in footprint?

*Author's response:*

The collocations are done identically to Stengel et al. (2017). Most important facts are that all collocations are based on searching for the nearest neighbour in the Cloud_cci Level-3U data to each CALIOP observation. Given that a L3U grid box is usually of smaller size than 5 km, 5 km is the maximal allowed spatial mismatch. Most collocations have much lower spatial mismatches. Due to the similar orbital characteristics of NOAA-18 and NOAA-19 compared with CALIPSO, a very large set of collocations can be retrieved, which in turn also allows for very strict criteria and still infer a sound collocation database. In this context, the temporal mismatch criteria was set to a time window of±3 min. While the systematic deviations to CALIOP do not significantly depend on the match-up criteria, the random deviations do. We will summarize all this information in the text.

*Author's changes to the manuscript:*

We will add the following sentences: The collocations between CALIOP and the AVHRR-PM data were done as reported in Stengel et al. (2017) with the most important fact being that only those collocations were included for which the spatial and temporal mismatch was below 5 km and 3 minutes, respectively. These criteria were chosen as compromise between using best spatial and temporal matches, and allowing for a compositions of a sound data basis to be used in the validation. Important to note that the random deviations of AVHRR-PM to CALIOP depend on the defined criteria, while the systematic do most likely not.

*Referee comment:*

p7. l22-23. Why would improved identification of liquid clouds lead to reduced POD for ice clouds? This suggests that some ice clouds are now erroneously identified as liquid. Does that mean there are more "false alarms" in terms of liquid cloud detection?

*Author's response:*

That sentence is indeed misleading. For the v3 algorithm the probability of detection (POD) of liquid clouds is improved while the POD for ice clouds is slightly reduced, still leading to an overall increased (improved) HSS values.

*Author's changes to the manuscript:*

We will modify that sentence to: "Comparing the HSS score as overall measure for the correct cloud phase detection, v3 performs better than v2. The POD of liquid clouds is significantly improved in v3, while a small degradation in POD of ice clouds is found in v3 compared to v2."

*Referee comment:*

p8. l5-15. It would be helpful to consider the results from Tables 4, 5, and 6 through a visual comparison, as done in Figure 4. A scatter plot (or 2D histogram) of CTH, LWP, and IWP comparing the data set with the "truth" could help identify where biases are most likely to occur. For instance, the CTH bias of ice cloud could be mostly due to the highest clouds, even at high COT, as these might have a region of low extinction coefficient near cloud top, that would lead to higher CTH in CALIOP. A scatter plot could show this clearly. Similarly, LWP and IWP are highly skewed variables and the metrics presented could be affected by a few outliers. A scatter plot or 2D histogram (perhaps shown on a logarithmic scale) could indicate whether LWP and IWP estimates are typically good, or whether there is a consistent bias across cloud types of all LWP and IWP values.

*Author's response:*

Thank you for this suggestion. We have included 2d histograms for CTH, LWP and IWP.

*Author's changes to the manuscript:*

We have included 2d histograms for CTH, LWP and IWP.

*Referee comment:*

p9. p10. p11. Please rename standard deviation to "root mean squared error", which is presumably what is reported.

*Author's response:*

We actually mean the standard deviation of the error (which is basically identical to the bias-corrected root mean squared error). We will add to the captions of tables 4,5 and 6 that we mean the standard deviation of the error and the mean error.

*Author's changes to the manuscript:*

We will modify the captions of tables 4,5 and 6 to "….standard deviation of the error (Std), the mean error (bias) …."