

Earth Syst. Sci. Data Discuss., referee comment RC3
<https://doi.org/10.5194/essd-2022-282-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-282**

Anonymous Referee #3

Referee comment on "Updated observations of clouds by MODIS for global model assessment" by Robert Pincus et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2022-282-RC3>, 2022

Summary

The authors describe a new cloud dataset derived from MODIS satellite observations intended for direct comparison to the output from the MODIS simulator that is part of the broader COSP satellite simulator package run in many modern climate models. The paper details the salient portions of the MODIS cloud retrieval and processing stream, the manner in which the present dataset is constructed (and how it is different from other MODIS level 3 products), and provides some visualization of the types of fields that are readily available from the dataset and their potential scientific uses. This will certainly be a useful dataset for the community working on clouds, in particular those evaluating clouds in climate models.

Major Comments

I am a bit surprised that the dataset is not produced in such a way as to be applicable to models "right out of the box" as is the case for datasets like GOCCP. The motivation for leaving some processing as an exercise for the end-user, which seems ripe for accidental misuse, was not clear to me (though the provided python code is of course welcome). Is it an attempt to leave some flexibility for end-users' diverse needs? If the COSP run in climate models produces monthly mean cloud property fields (cloud fraction, joint histograms, $\log(\tau)$, etc), why can't this dataset provide the same right out of the box? Lines 257-260 indicate that even the model output will have to be further processed in order to match what is provided here, another place where user error can creep in (which fields require weighting by cloud fraction? Does cloud fraction have to be weighted by cloud fraction? Which cloud fraction -- mask or retrieval -- should be used as the weight?). Is there a plan to provide python code for processing the MODIS simulator

output in such a way as to be directly comparable to what is produced by the provided python code? Could all of this post-processing be avoided from the outset by just providing an idiot-proofed dataset that is as close as possible to MODIS simulator output?

I am also a bit surprised (as I am currently downloading the dataset via the github instructions) that the filenames have such cryptic names, particularly the timestamp which seems to be reporting a Julian day at the start of the month rather than a format like YYYYMM. This results in further reliance on the python script rather than being able to quickly assess what a file contains.

For several fields, it is not clear to me that there is a COSP counterpart; what is the reasoning for including these fields in the "MODIS-COSP" dataset if COSP does not provide them? These fields include the two versions of cloud fraction (from the mask and from the retrieval); the partly cloudy pixel fields; and the additional statistics like standard deviations, sum-of-squares, etc.

Minor Comments

- Throughout, the low cloud criteria is listed as $p > 440$ hPa (line 129, Figure 4 caption, Table 1 caption) where as I believe it should be something else like $p > 680$ hPa. Also, in the caption of Figure 4 and Table 1, mid-level clouds are listed as those with CTP greater than 680 hPa and smaller than 440 hPa
- L164: "also able compute"
- L178: I believe "optical" should be capitalized.
- L187-190: Suggest reassuring the reader here that a script is provided to do this
- L200: "for much the"
- Figs 4 and 5: The sum of these maps equals 1 (not the retrieved total cloud fraction) everywhere, right? In this case, should the colorbar label instead be just "fraction"?
- Figs 5,8,9: Suggest putting ice clouds above liquid clouds, which seems a more natural orientation.
- Figure 5 caption: should be "phase" singular, I believe
- Figs 7-9: please provide units on the axes. Also, the y-axis in Figure 7 is labeled as "cloudtoppressure" – suggest breaking it up into 3 words.
- L261: "variables" should be singular
- Tables 1 and 2: Is it possible to report the name of the equivalent field from the MODIS

simulator in a 4th column, or state that it is not available?

- Figures: The authors clearly had a lot of fun trying out various matplotlib color schemes, including my favorite, the Eye of Sauron scheme in Figure 4. I wonder if this may be distracting and unintentionally conveying differences that are not meant to be conveyed, considering many figures show the same field (cloud fraction). I am not sure whether any of these are unfriendly for color-blindness, but that should also be considered. The scheme in Figure 9 seems to artificially distinguish cloud fractions larger than about 0.007 from those below, but I'm not sure why that would be useful. In some figures lighter colors = larger values, but the opposite is true for Figures 3, 7-9.