

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2022-282-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-282**

Anonymous Referee #1

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-RC1>, 2022

Review for "Updated observations of clouds by MODIS for global model assessment" by Pincus et al.

This manuscript documents a new dataset that allows for the comparison of clouds observed by the National Aeronautics and Space Administration (NASA)'s two Moderate Resolution Imaging Spectroradiometer (MODIS) instruments onboard the Terra and Aqua satellites and clouds that are simulated by climate models. The manuscript describes the idiosyncrasies of the product, its technical implementation and caveats regarding bugs in the processing of the dataset.

Overall, the documentation is easy to read and well-written. The background is presented quite well, except for a few ambiguities that are detailed below. My main suggestions would be to further clarify the purpose and scope of the product and manuscript about the product, as well as the variables that are described in the manuscript. I am also wondering if it would make more sense for the dataset to be completely finished before writing the documentation in a separate publication from than the User's Guide since it may lead to confusion or disorganization if the same product is documented in different places. The unfinished products were mentioned twice in the manuscript. Specific comments follow.

- The authors state that the product is made for the "convenience" of end users, but also

mention that on line 39: "The system was also quite fragile and ceased production when NASA updated the production of MODIS datasets" in the Introduction. It's not transparent to me what is meant by that the system was quite fragile and why production ceased. Also the statement on line 46: "The dataset, produced using a system designed to be more robust to changes in the upstream data, provides a set of custom cloud-related parameters using specific dataset definitions more closely aligned with the MODIS simulator than are the standard datasets" is also not transparent to me. What upstream data are the authors referring to and what changes were made to them? Also, importantly, does this mean that using the standard MODIS product to compare against climate models is incorrect or is this product really only designed for the convenience of end users? Please clarify.

- A number of variables are described but I would recommend that in each case the variables themselves (as listed in Table 1) are spelled out to be clear and equations written out if applicable, which would be relevant for a publication in ESSD, e.g. line 257: when weighting by the "cloud fraction" in the MODIS simulator, Section 3.3 variables.
- Why is there no height-resolved cloud retrieval fraction saved as a separate product? Only the height-resolved cloud fraction from the cloud mask seems to be reported.
- Lines 260-263: "To facilitate comparisons with the MODIS simulator we have provided Python code, described below, that transforms a set of monthly files containing all variables to datasets with time series of each variables, which may be written as netCDF files and/or Zarr stores." Does this sentence mean that the Python code takes individually saved netCDF files that each contain one month of data and simply concatenates the files into a time series for any individual variable and time period of choice? Please clarify.
- Is there a reason why this manuscript was submitted when the complete dataset is still under development? It seems that it would make more sense for the product to be finished first and then a final paper to be published on it for completeness and to avoid confusion in the future where the same product might be separately documented. It seems that the current User's Guide that is already available on the Internet is sufficient in the interim?
- Figure 8: So if a user wanted the total joint histogram for liquid and ice, could they simply sum up the liquid and ice separately?
- Lines 136-137: "...and the condensed water path estimated from the product of optical thickness and particle size". For ice clouds, what equation was used?

Typos and other minor edits:

- Line 211: "Liquid clouds are substantially more common than ice clouds (see Figure 5)". This is a general statement but does not seem to hold in the middle right part of the plot. Also, although the outline of the continents is somewhat visible from the cloud fields themselves, it would be nice to have the outline of the continents drawn in.
- Typo on line 109: "of the of the"
- Typo on line 127: "is is ratio..."
- Why are low clouds consistently labelled as p_c greater than or equal to 440 hPa? Is this a consistent typo? I found this in a couple of places. I think you mean 680 hPa? But then line 229 uses 800 hPa as the threshold which is inconsistent with the

previously mentioned definitions.

- Although it is probably clear to most readers, please spell out all acronyms to avoid any possible ambiguity, including MODIS and ISCCP