Response to Reviewer # 1 (Manuscript ID: amt-2019-223)

First of all, we would like to thank the reviewers for their valuable comments. In the revised manuscript, we have accommodated all the suggested changes into consideration and revised the manuscript accordingly. The reviewers' comments are copied here as texts in BLACK. The authors' responses are followed in BLUE, and our changes in the manuscript are in *italics*.

Reviewer # 1

This paper by Yao et al., evaluates qualities of cloud properties in three reanalysis datasets, namely, China Meteorological Administration Reanalysis data (CRA), ECMWF's Fifth-generation Reanalysis (ERA5), and Modern-Era Retrospective Analysis for Applications version 2 (MERRA-2). A radiance-based evaluation approach is utilized with reflectance and brightness temperature observations from the Advanced Himawari Imager (AHI) onboard the Himawari-8 satellite. A radiative transfer model (CRTM) is used to link cloud related variables from reanalysis to satellite observations.

Overall, I believe this work is very valuable, which enhances our understanding of cloud representation in those reanalysis products. However, I have some concerns about the structure and some details of this paper.

Several major concerns I have about this paper include:

1. This paper uses observations from AHI/Himawari-8 to evaluate reanalysis. It is very important to mention that which satellite products (in particular cloud related datasets) are used as input in the three reanalysis products.

Response: Thanks for the suggestion. Yes, it is necessary to introduce satellite observations assimilated for the reanalysis, because the differences on satellite datasets assimilated may be a potential reason for different performances of the reanalysis. Thus, we added the related contents in Section 2. Both ERA5 and CRA consider Himawari-8 observations, whereas MERRA-2 does not. This may be one of the reasons that MERRA-2 has relatively poor performance in the Asian region. To address the reviewer's concern, we included the following discussion in the revision (Lines 401-403):

"*It should be noticed that both ERA5 and CRA reanalysis consider Himawari-8 observations for assimilation (see Section 2), whereas MERRA-2 dose not. This may be one of the reasons that MERRA-2 has relatively poor performance on cloud representation in the Asian region.*"

2. The advantages of a radiance-based evaluation approach are discussed in the abstract and introduction. I don't understand why the authors still use a lot of space describing AHI cloud products in Section 4?

Response: In the original submission, we try to demonstrate more clearly that direct retrieval-based evaluation may be problematic, so Figures 1 and 2 as well as the corresponding discussions give comparisons based on the cloud products retrieved based on different bands (i.e., the solar channels and thermal infrared channels). We agree with the reviewer that the purpose of the study is to evaluate different reanalysis datasets based on the radiance-based approach. Considering that the Introduction Section is clear enough to demonstrate the disadvantage and uncertainties related to the retrieval-based evaluation (as noticed by the reviewer), we have removed the details related to the retrieval-based evaluation (i.e., Figs. 1 and 2 as well as the corresponding discussions), and the part related to AHI cloud products has also been removed.

3. This paper uses almost 4-pages to describe a case (a snapshot on a particular day) assessment, which I think is not necessary. In my point of view, the authors should pay more attention on long-term cloud representation (e.g., cloud monthly mean, seasonal/annual variability).

Response: Actually, the "case study" mentioned in this study is not a snapshot for a particular day, and we consider results over eight days with over 30 realizations. To avoid such misunderstanding, we have the added the following sentence in the revision:

"Noted that even for this case study, we consider a period over eight days covering 32 time steps."

We think the case assessment is meaningful as well for the following reasons:

(1). The results in Figures 11 and 13 indicate that the evaluations are generally stable over time. The results of the case study are universalistic and representative, and the corresponding conclusions are actually consistent with those from the long-term evaluation. However, because the forward radiative transfer simulation is computationally expensive, this study considers results from a typical case with eight days and a generally evaluation with 144 realizations over one year.

(2). In fact, we use the case study results to present more details of the three reanalysis, whereas use the long-term results for the general evaluation. As a result, we think both parts are necessary.

(3). Both the case study and the 144 realizations spanning over one year indicate that our methodology, i.e., the radiance-based evaluation, is feasible, and the results are reliable.

Meanwhile, we agree with the reviewer that more attentions should also be paid to cloud monthly mean, seasonal/annual variability, and we have extended these discussions. Furthermore, we would like to investigate the long-term cloud representation in details in our future studies.

Some minor suggestions include:

1. Page 2, large advantages of spatial distributions –> large advantages of spatial coverages.

Response: Thanks for your suggestion, and the phrase is corrected.

2. Page 6, CTT from two satellite retrieved cloud datasets (i.e., from solar and thermal infrared) How to use AHI solar bands to get CTT, can you give more details on this?

Response: Sorry for the confusion because of my incorrect description. The cloud top in the product from Letu et al. (2018) is retrieved based on the observations in the infrared window channel (11.2 μm), and the cloud product of Iwabuchi et al. (2018) is based on observations in the 10.4 μm channel. However, the atmospheric profiles used in the cloud retrieval are different, and Letu et al. (2018) and Iwabuchi et al. (2018) cloud products use profiles from the GPV (the Grid Point Values of atmospheric) and MERRA reanalysis, respectively. As mentioned above, we think this study should focus on the radiance-based evaluation, so we have removed the section on cloud retrieval products.

3. Figures 3, 5, and 7. The plots in Figures 5 and 7 use all pixels (i.e., clear + cloudy) in Figure 3? If yes, I suggest remove clear pixels or only focus on the regions of interest. I noticed that a large number of pixels in Australia are clear and reflectances from models are much higher (brighter) than AHI observations. This can significantly bias your plots in Figs. 5 and 7, and statistics.

Response: Yes, both clear and cloudy pixels are considered in Figs 5 and 7. Because we consider different clouds by using different BTs or BTDs, even with all pixels considered, the problems related to the reanalysis over cloudy regions can be illustrated by the figures. We think the reviewer gives an excellent comment to consider only cloudy pixels, so we added a new Figure 5 in the revision with clear and cloudy pixels considered separately. We found that the cloud property representation contributes more to the differences than the atmospheric profiles.

Meanwhile, as there is no "truth" for the classification of clear/cloudy pixels (again, we do not want to use the retrieval results due to their own uncertainties), we can only use reanalysis data for the classification. This is also a reason that we mostly consider all pixels in the discussions.

4. Figures 11 and 12 and corresponding text: The authors use BT 11um as a proxy to differentiate clouds on low, mid, and high levels. This is problematic since high and thin cirrus may be attributed to low clouds.

Response: Thanks for the suggestion. In the revision, the widely-used thresholds based on BTDs between the 6.2- and 11.2-μm channels are used to differentiate clouds over different layers (Mecikalski and Bedka, 2006; Yao et al., 2018). Because of strong water vapor absorption in the 6.2-μm channel and the temperature lapse rate within the troposphere, the BTDs between 6.2- and 11.2-μm are usually negative. The BTDs increase as the cloud top height increases and larger negative BTDs often corresponds to clear-sky pixels. We use the thresholds of -45 to -30 K to infer pixels with low cloud tops, and those with low- to mid-layer cloud are represented by BTDs between -30 and -10 K following Mecikalski and Bedka (2006). The BTDs less than -45 K normally correspond to clear pixels and those larger than -10 K are from high cloud pixels. With the improved classification, most results and conclusion are similar, and slight differences are noticed for mid-layer clouds (The mid-layer cloud in CRA is closest to the observation.) Thanks for your suggestions, and we have updated the corresponding classification, figures, and the corresponding discussion in the revision.

References:

Letu, H., Nagao, T. M., Nakajima, T. Y., Riedi, J., Ishimoto, H., et al.: Ice cloud properties from Himawari-8/AHI next-generation geostationary satellite: capability of the AHI to monitor the DC cloud generation process, IEEE Trans. Geosci. Remote Sens. 12, 1-11, 2018.

Iwabuchi, H., Putri, N. S., Saito, M., Toloro, Y., Sekiguchi, M., et al.: Cloud property retrieval from multiband infrared measurements by Himawari-8, J. Meteor. Soc. Jpn, 96, 27-42, 2018.

Mecikalski, J. R. and Bedka, K. M.: Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery, Mon. Wea. Rev., 134, 49-78, 2006.

Yao, B., Liu, C., Yin, Y., Zhang, P., Min, M., and Han, W.： Radiance-based evaluationo WRF cloud properties over East Asia：Direct comparison with FY-2E observations, J. Geophys. Res., 123, 4613-4629, 2018.