Response to Reviewer #2 of gmd-2019-104

Dear Reviewer #2,

Thank you very much for taking your time to review our paper. We think that your comments greatly help improve the manuscript. We have revised the manuscript according to your comments as explained below with point-by-point responses to your comments. We hope that the revision is enough to address your comments to make the manuscript now acceptable for publication in *GMD*.

[RC]: ***Referee comment***
[AC]: Author comment

Reviewer #2:

[RC] *Michibata et al. propose new inline warm rain diagnostics for a community satellite simulator project (COSP) to enable process-oriented model evaluation. The diagnostics are well described and motivated. The diagnostics are similar to those developed by Kay et al. (2018) for precipitation frequency, but include important process-inspired criteria beyond Kay et al. (2018) for warm rain. The implementation of these new diagnostics in MIROC6 reveals interesting results. I have only minor comments. I recommend publication.*
[AC] We would like to thank the referee #2 for his/her positive comments. We have revised the manuscript according to the referee comments as listed below. Note here that, page and line numbers denoted in the authors' responses below correspond to the track-changes file, not original manuscript.

Specific points:

[RC1] *Abstract. Can you more specifically describe the diagnostics beyond "two diagnostics for warm rain processes"? What are the new outputs in COSP?*
[AC1] We have added more specific description about the two diagnostics as follows: "Here, we incorporate two diagnostics for warm rain microphysical processes into the latest version of the simulator (COSP2). The first one is the occurrence frequency of warm rain regimes (i.e., non-precipitating, drizzling, and precipitating) classified according to CloudSat radar reflectivity, putting the warm rain process diagnostics into the context of geographical distributions of precipitation. The second diagnostic is the probability density function of radar reflectivity profiles normalized by the in-cloud optical depth, the so-called contoured frequency by optical depth diagram (CFODD), which illustrates how the warm rain processes occur in vertical dimension using statistics constructed from CloudSat and MODIS simulators."
In accordance with the rearranged above, the sentence was modified as follows: "… produce statistics online along with the subcolumn information during the COSP execution, …".

[RC2] *Line 5-7, page 2. I recommend removing "apple-to-orange" and "apple-to-apple". Instead - can you describe what makes the comparisons more credible when they are done with COSP? For example, the reader should be aware that "definition aware" and "scale aware" comparisons are made possible by the use of satellite simulators and a sub-column generator respectively.*
[AC2] Thank you for suggestion. We removed "apple-to-orange" and "apple-to-apple". Instead, we have modified the sentence that explains the merit of COSP as follows:
Page 2 Line 15: "… same algorithms applied to each satellite sensor for consistent ("definition-aware") comparison. Furthermore, the process evaluation among models and observations should be done under the same spatiotemporal scale for consistent ("scale-aware") comparison."

Page 2 Line 33: "… from multiple instrument simulators in a "definition-aware and scale-aware" framework (Kay et al., 2018)."

[RC3] *Line 12, page 2. "such as CMIP6" to "including CMIP6" or "e.g., CMIP6". Much was done with CMIP5 as well…*
[AC3] We have modified the sentence in the revised manuscript.

[RC4] *Line 30-31, page 4. Please explain how the number of sub-columns (140) was selected. The rationale behind the selection of the number of sub-columns is important to describe, especially for those new to COSP.*
[AC4] For CFMIP2 experiments, COSP users are recommended to assume ~100 subcolumns per 1 degree of model grid spacing to enable comparison to satellite sampling at the kilometer scale, as described in Page 3 Line 7. This statement is based on the README file in CFMIP2 experiments for recommended configuration (cfmip2/cosp_input_cfmip2_long_inline.txt). The model horizontal resolution is T85 (~1.4 degree in longitude and latitude) in this study, and hence we prepared 140 subcolumns in COSP execution. We have added this note in the revised manuscript.

[RC5] *Line 12-14, page 5. "The spatial resolution of the reference A-train data …". This sentence is incorrect. The A-train native spatial resolution is much higher than 1.5 degrees – For CloudSat it is ~1 km. While the statistics of both the models and the observations are compared at 1.5 degrees – this study has taken a lot more care to make "scale aware" comparison not at 1.5 degrees. Specifically, the climate model data were "down-scaled" to the A-train data native resolution using a sub-column generator in COSP. Please describe in detail so that the reader does not confuse the resolution of the grid at which the statistics are reported (1.5 degrees) and the resolution at which the comparisons are being made («« < 1.5 degrees).*
[AC5] Thank you for such an important comment, and we agree with the reviewer. The following notes have been added in the revised manuscript: "Note that although the reference A-Train statistics is shown at 1.5º x 1.5º resolution, which is close to that of MIROC6-SPRINTARS, the statistics are constructed from the native CloudSat resolution (1.4 x 2.5 km) and subcolumns in the host model prepared by COSP (kilometer scale) to achieve the "scale-aware" model-satellite comparison.".

Thank you very much again for reviewing our paper.

Sincerely yours,

Takuro Michibata