

Comment on [essd-2021-298](https://doi.org/10.5194/essd-2021-298)

Anonymous Referee #2

Referee comment on "Resilient dataset of rain clusters with life cycle evolution based on observations from the GPM DPR and Himawari-8 AHI" by Aoqi Zhang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-298-RC2>, 2022

The authors present an interesting and useful data set of precipitation tracks for MCSs in Eastern Asia. The manuscript is very well written, but lacks some necessary details. I have more concerns about the data set, which is not properly described.

Response: Thanks a lot for your encouragement and helpful suggestions! The following is our replies point-by-point to your issues (presented in blue color in the “**track change file**”).

Major comments:

1. The dataset: Neither the metadata or the readme file explains what the "past", "present", and "future" suffixes actually mean. Why can, e.g., `life_past` be zero all the time? I would assume it would mark the hours the track has persisted. I am obviously wrong, but this is an example of where documentation is needed to make sure the data is correctly used and interpreted. Also, the time stamp is missing in all files. I encourage the authors to use the CF-convention format of the time vector.

Response: Thanks a lot for your constructive suggestions! **We have carefully revised the readme file and dataset. The latest version of the dataset is available at <https://doi.org/10.5281/zenodo.6198716> (Zhang et al., 2022).**

1) Sorry for our carelessness. We have explained the “past”, “present”, and “future” suffixes in the Readme file and metadata. The life-cycle information of MCS was stored in three portions with different suffixes including “past”, “present” and “future”. We tracked both backward and forward the MCS with a time interval of 1 hour. The “present” suffix indicates the unique time of MCS within ± 30 min of the corresponding RC. The “past” suffix indicates backward track times before the “present” time, and the “future” suffix indicates forward track times after the “present” time.

2) “`life_past`” indicates backward life time of the corresponding MCS track, and

“life_future” indicates forward life time of the corresponding MCS track. The total life time of MCS track (units: hour) equals $life_past + 1 + life_future$. Note that numerous MCS tracks are short-term tracks with backward (or forward) life time less than 1 hour, the “life_past” (or “life_future”) of many MCS tracks are stored as 0.

3) We have provided the time stamp for each file and used the CF-convention format of the time vector in the latest dataset.

2. The manuscript provides the spatial and temporal limitations of the data set at first on L103. This should be clear from the title, abstract and introduction. Space: eastern Asia; time: April to June 2016-2020.

Response: Thanks for your nice suggestion. We have clarified the spatial and temporal limitations of the data set in the title, abstract and introduction [**Line 1-2, 21-23, and 80-81**].

3. Some details of the processing are not explain in sufficient detail. See comments below.

Response: Thanks for your helpful suggestions! We have added these details in the manuscript.

Detailed comments:

L17-20: Please rewrite this sentence. It is too long, not clear about how the different data sets are employed, and has some grammar issues ("tracking algorithm" should be "tracks").

Response: Sorry for our unclear descriptions. We have reorganized these words [**Line 18-21**].

L103: The data set ends in 2020, which is hardly present day. Please explain why the geographical domain is not the complete disk of the geostationary satellite. Would it be technically possible to do the full domain (not considering the work of course).

Response: Thanks. We have corrected the unsuitable words [**Line 108**]. In the present

version, the derived dataset only covers the eastern Asia. This is because that the researches are familiar with the precipitating systems in eastern Asia. In our future work, the spatiotemporal coverage of the dataset will be further expanded to the full disk of Himawari-8.

L117: Was the contours made again on the new grid? Or were the contours remapped? Figure 4b indicates that the MCSs were remapped, rather than the DPR. Please clarify.

Response: Sorry for our unclear descriptions. We have corrected them in the manuscript.

- 1) We compared the remapped contours of the initial RCs and the contours of MCSs to determine the mapping relationships between them. Specifically, we remapped the DPR pixels of initial RCs to $(0.05^\circ \times 0.05^\circ)$ grids (consistent with the MCSs) and determined the overlapping grids between the initial RCs and the MCSs at the nearest time (± 30 min) [**Line 123-125**].
- 2) Figure 4b indicates the pixels of RCs after the two steps of RC segmentation algorithms [**Fig. 4**].

L120: Please specify from which point you calculate the 100 km. From the nearest pixel of the contour, or the centroid of the contour? Please also explain how the centroid of the contour is defined (here and in the data set). E.g., does the centroid need to lie within the contour (comparing to a banana shape area where it might end up outside the banana).

Response: Thanks. 1) The distance was calculated from the nearest pixel of the contour [**Line 127**].

2) The dataset provides the central latitude and longitude of RC and MCS. The central latitude and longitude are defined as the average of all RC pixels (or MCS pixels), so the centroid may lie outside the contour [**Line 240-241 & readme file**].

L138: Please explain what is meant by "restore the overlapping grids".

Response: Sorry for our unclear words. We have corrected them [**Line 144-146**]. In the

previous algorithms, we gridded the DPR pixels to $(0.05^\circ \times 0.05^\circ)$ grids to obtain the overlapping grids between the initial RCs and the MCSs. The first step was to remapped the overlapping grids to the DPR pixels using just the reverse method.

L142: What is meant with "gain weight", and "round by round"? Do you mean "we iteratively increase the RC cores by adding pixels around the area until..."

Response: Yes, thanks a lot for your nice suggestion [Line 150-151]!

L144: How is the "smallest rain rate" defined? From which relative point in space? Is the gradient determined using two points, or more?

Response: Sorry for our unclear words. We have reorganized these sentences [Line 151-155].

If, at a certain round of the collision process, one certain pixel was allocated to multiple RC cores, its nearest 8 DPR pixels, including non-precipitation pixels, must contain allocated precipitation pixels from different RC cores. We would then calculate the rain rate gradients between the certain pixel and allocated precipitation pixels using rain rate difference divided by distance, and the certain pixel would be reallocated the RC core with minimum rain rate gradient.

L225: Please add some information about the final data set. How many tracks, which information is in the files etc.

Response: Thanks very much for your nice suggestion. We have added some information of the final dataset in the draft [Line 236-240].