Comment on essd-2021-298
Anonymous Referee #2

Referee comment on "Resilient dataset of rain clusters with life cycle evolution during April to June 2016–2020 over eastern Asia 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.

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.

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.

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

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").

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).

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.

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).

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

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..."

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

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