General comments:

The paper uses a symmetric rain rate to define the radar reflectivity error in the assimilation algorithm based on the symmetric rain rate referring to the symmetric error model in satellite all-sky assimilation. The paper is well-structured but still, there are many ambiguous sentences in the paper which need to be rewritten/clarified.

major revisions:

- The very important point that is missing in the paper is that the reflectivity error in an assimilation algorithm is a representative error based on each assimilation system/algorithm meaning that if anything changes in the assimilation algorithm either the NWP model or the number/type of observations, the reflectivity error need to be recalculated /modified. But in this paper, it seems there is not any assimilation algorithm/system. The equivalent reflectivity is based on the 6-hour model forecast which does not represent the equivalent reflectivity after assimilation. The very large standard deviation (up to 35 dbz) clearly shows this inconsistency.
- Besides, defining a better reflectivity error is supposed to improve the assimilation results. However, the paper did not show any plots related to applying the newly defined reflectivity error in an assimilation algorithm and the comparison with the constant error (which was claimed in the paper is not suitable for radar assimilation).
- Line 92: What is the purpose of defining the radar composite based on the vertical maximum reflectivity? Do you use this composite in your assimilation algorithm? Why not the radar reflectivity composites of a specific level (which will be used later in the assimilation algorithm)?
- It was mentioned that to match with the rain rate resolution of 4 or 5 km, the reflectivities with the resolution of 1 km were interpolated! Actually, this procedure is extrapolation. However, the normal trend in the radar assimilation is defining/using the super observation which defines the reflectivity over a larger grid box (compared to the original grid box) to match with other observations in the assimilation algorithm which have lower resolution.
- Fig 2: It clearly can be seen that the simulated reflectivity and the observed reflectivity are far from each other. This means the model has a very poor ability to capture the convective events. As I mentioned, this could not be the representative reflectivity of an assimilation algorithm. It can be easily seen that defining the equivalent reflectivity based on this kind of plot can cause a high value for the reflectivity standard deviation which is not realistic.
- Fig3: The two plots look identical. It couldn’t/shouldn’t be like this. Please check the plots.
- What is the purpose of excluding the false and missed events? and defining the 'both-reflectivity‘? At the end, they all need to be included in defining the standard deviation.
- As it was written the CMPAS rain rate is the most complete one (or maybe the reference ones) what is the purpose of using another product (FY-4A)?

- In Fig 6 (or later in Fig 8), the departure or the standard deviation needs to be defined clearly. It is not clear that the departure means 'obs-model' or 'model-obs'.
- Fig 8 (Line 258): It was mentioned that the standard deviation increases after 9 mmh^-1 because of the error in the WRF model or the initial data. Actually, a big part of the error could be due to the deficiency in the forward model (the process of converting the model state to the radar reflectivity).
- Fig8: When there is (kind of) reference data set of rain rate, why the standard deviation should be plotted based on the derived rain?
- Line 279 or Line 283: What was here exactly normalized? Fig 8 and Fig 9 are the standard deviations based on the rain rate. Did you normalize any variables here?
- Fig 8: The black dashed line shows the log of sample numbers that reach less than 2 after 15 mmh^-1 meaning that the number of samples is less than 100! If this is the case, means that the number of samples in these bins is not enough to do the standard deviation. As it was shown the number of samples in some bins can reach 10^6 so there is a big inconsistency in defining the standard deviation in different bins. There should be a limit based on the number of samples to do the standard deviation. I would suggest a limit of 10^3 or 10^4.

minor correction:

- Line 76: "stage IV precipitation" → What is the defintion of stage IV precipitation?
- Line 99: "on May 3th 2021" → on 3rd of May 2021
- Line 218: "…, chosen to be larger than 100 samples": What does this mean? If it means that you calculate the reflectivity departure for the grid box which has more than 100 samples, why the colorbar starts from 1?
- Line 219: "excessive reflectivities" → overestimation in simulating the reflectivities
- Line 224: "It could be argued" → It shows that
- Line 235: "exists" → defines the relation between …
- Line 257: What is the "geophysical boundary"?
- Line 262: The constant reflectivity error is usually between 5 and 10 dbz
- Line 323: "… from the large mislocation errors in the main" → What is main?