Comment on essd-2022-133
Anonymous Referee #1


The manuscript presents the 48-member Noah-MP simulations in CONUS and the evaluation results. Common terrestrial water budget variables are provided. Comprehensive evaluation is performed based on multi-source reference dataset. The manuscript is well written. The dataset will be useful for diverse applications. I think the manuscript is suitable for publication on ESSD. Below are some comments which could be useful to the authors.

The dataset is not developed by this paper according to the description in the manuscript. For example, Fei et al. (2021), which is a publication of the same authors, already evaluated part of the 48-member Noah-MP model outputs. However, the description in the manuscript is kind of misleading (e.g., the abstract and line 79), making the readers have an impression that this manuscript runs the ensemble simulation. I recommend that the authors re-organize relevant contents, clearly stating the development and evaluation history of the 48-member simulations, and the role of this manuscript (e.g., evaluation and data release?) in the introduction part.

Line 104: Can you talk more about the “pitfalls”?

Section 2.2: Why these parameterizations are chosen? Can they represent the full range of uncertainty? Besides, I think the introduction to parameterizations can be moved the appendix. As a dataset description paper, these technical details could weaken the readability of the paper for most readers.

Line 279: Is this recursive spin up in a single year?
I have some doubts about Sections 3.1 and 3.2. I think there is a mistake. In Eq 34, you should not subtract \( r_{clim} \) (see Eq 8 in Dirmeyer et al., 2006). Otherwise, \( r_{clim} \) is subtracted twice in Eq 37. For the subscript \( t \) in Section 3.2, I did not find any explanation (please correct me if I made a mistake). To be honest, the two sections use more equations and symbols than Dirmeyer et al., (2006) but make the same concept much less straightforward and harder to understand. Probably the authors want to use more symbols to make the definition clearer, but it turned out making things worse from my opinion. I suggest that the authors reorganize these sections.

Section 3.5: I understand that the reference datasets are important. But this section is too long for a dataset description dataset on ESSD. This can be a distraction from your core dataset. I am wondering whether you can remove some contents or move some contents to the appendix.

Line 492: Can you explain it more clearly?

Figures 5 and 6: According to the second column, Noah-MP EM not is not notably better than NLDAS EM. Can you explain how this affects the results in the third column? Besides, the statement “four estimates’ arithmetic average outperforms the three-model NLDAS ensemble mean at almost every NASMD site” is not always true (e.g., Fig. 6c and 6f). I suggest adding some quantitative statistics in the figures (e.g., the median value, or the ratio of positive values). This will make comparison more straightforward.

Figure 7: The figure caption is unclear. Besides, I think you mean “difference” (Line 524) instead of “relative bias” in the figure caption.

Line 530-532: Any explanation?