

esd-2021-53 version 2 Review

The authors have incorporated or responded satisfactorily to all my comments on the initial submission. There remain two essential issues to rectify before the paper is ready to publish (corresponding comments are marked with an asterisk):

1) The authors conclusions regarding the number of ensembles needed to detect variance changes are too strong with regard to precipitation.

2) The method for estimating the empirical CDF, from which the blue dots in figure 5&6 are obtained, should be discussed. (I missed this the first round, apologies).

In the interest of time, I think the editor can determine whether the authors sufficiently address the points above, so I do not need to see the second revision, though I am happy to review it should the editor request.

Detailed comments:

*Line 15: This implies variance change can be detected with 5-10 members. The results do not support this for the case of Rx5Day (see comments below re lines 278, 343.)

Line 127: I would change 'responsible for' to 'related to' since the location parameter is not exactly the mean of the tail and variability and tail behavior and not well defined.

Line 133: The URL runs off the page.

Line 232 consistently → consistent

Line 276-278 While this is true, the actual size of the confidence intervals will not be known in your scenario of 5 initial runs, as a larger ensemble would be needed.

*Line 278 The 'counting' approach should be inaccurate for long return periods using a small number of ensembles. For N-yr events where N is larger than $n \cdot 11$, the CDF values need to be extrapolated. Consider removing the blue dots (and associated discussion in the text) or only using a single estimate from the full ensemble. If left in, provide a description of how the CDF was estimated.

Line 299 Consider rephrasing "bound to be an upper bound".

Line 341 Correct 'a spatially noisier pictures'.

*Line 343 The results do not support this conclusion, in my opinion. There seem to be quite a number of locations where n is larger (often much larger) than 5. To quantify this, show the histograms corresponding to the panels in Figures 9 & 10 in the supplementary material.

*Line 398 - 400. Related to the above comment, this statement is far too strong. There are a lot of point where $n > 5$ in the top right panel of Figure 9, corresponding a change in Rx5Day between 1950 and 2100 (the largest time difference). My interpretation is that for rain metrics, you need a minimum of ten ensemble members to be reasonably confident you can detect variance changes in most locations where they occur (around 90%) or so. But this is hard to estimate from the maps, which is why I suggest showing the histograms in the supplement.

Figure 3, 4 9 & 10 are sideways.