Comment on hess-2021-345
Anonymous Referee #1

The paper of Brunner and Slater demonstrates the utility of a technique based on reforecast ensemble pooling to assess the frequency of extreme floods. Starting from selected catchments available in the EFAS database, they apply this technique in central-northern Europe, deriving several outcomes concerning also the possibility of regional flooding.

The paper is generally well written and provides useful insights for the application of the reforecast pooling (or UNSEEN) approach to flood frequency analysis (the authors claim that this is the first application with such a variable). In the Discussion section, the limits of the methodology are also clearly outlined (the main being, as usual, the availability of observed streamflow data). However, the paper could be made clearer and more straightforward. I have two main comments.

My first comment concerns the question: why should one prefer reforecast pooling to other methods such as “classical” stochastic simulations? This question is fundamental to highlight the utility of the proposed method. It should be considered both in the Introduction (which could be enlarged) and Discussion sections.

My second main comment concerns methodology, especially Sections 2.2 and 2.3. Applying the reforecast pooling technique does not look very straightforward, given that several preliminary steps are needed. Another aspect that made it difficult for me understanding these sections, which I had to reread several times, is the sudden description of operations that had not been introduced previously. E.g., the need for bias correction (L124) comes abruptly, such as the use of linear regression models (L166). I suggest introducing better the different steps, linking them to specific objectives, possibly aided by a flowchart.
Please find below some specific comments. I hope my review can help to improve the quality of the paper.

LL12-13: “... specific flood return levels are highest in ...”: not very clear. I would better tell that, given a return period, specific floods are higher in steep and wet regions etc.

L49: are they mean elevations?

L54: what does “acceptable” mean in this context? Maybe some details could be disclosed here

Section 2.2: as I wrote before, a flowchart would help a lot

Fig. 2: it looks like not always quantile mapping produces better results (e.g. fig. 2b). Could the authors provide more details about the overall analysis? However, please change colours. Shades of red are too similar.

Fig. 3 is a bit minimalistic. I’m confident it can be improved

L180: the the

L188: r%?

Figg. 9(c) and (d) are not very clear. Maybe the x-axis could have a log scale.

Fig. 10: please correct the typo in the caption ”)“). Furthermore, the five maps are very similar. Maybe some of them could be removed.

L280: what about the results concerning latitude? In Fig. 11 it is dark green for all return periods. Please check. Anyway, results concerning the other variables (i.e., slopes, mean precipitation, dams and snowmelt) are quite obvious and make the analysis less interesting.
L299: the the

L314: “...any such...” please check