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.
Reply: *Thank you for pointing out the need to better discuss the relationship of reforecast pooling with other methods available for increasing sample size. We did not intend to suggest that reforecast pooling is to be preferred over other methods. Rather, we wanted to present it as an alternative to existing methods such as stochastic simulation or large climate ensembles. To clarify this, we substantially reworked and extended the introduction and added a short paragraph to the discussion section, where we discuss how the reforecast ensemble pooling method relates to other methods: 'Therefore, streamflow reforecast ensemble pooling represents a suitable alternative to stochastic or climate model large ensemble approaches for studying the frequency and magnitude of rare extreme events. Similar to large ensemble approaches but in contrast to stochastic approaches, reforecast-based simulation approaches rely on physical representations of the hydrological cycle. Such physical representation may be especially valuable if relationships between different variables are of interest and if one wishes to study the physical drivers of flood events. In contrast, stochastic models have the advantage of being relatively straightforward to implement and are potentially less computationally intense.'*

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.
Reply: *Thank you for pointing out that the methods section needed greater clarity. We adopted your suggestion and designed a flow chart to visualize links between the different working steps. We also readjusted the order in which the different working steps are presented by reorganizing the methods section in a chronological way.*

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.
Reply: *Thank you very much for this rephrasing suggestion, which we adopted.*

L49: are they mean elevations?
Reply: *Yes, we specified that we are talking about mean elevations.*

L54: what does "acceptable" mean in this context? Maybe some details could be disclosed here
Reply: *We added the model evaluation criteria introduced in Section 2.2 to the introductory paragraph of the Methods and Materials section.*

Section 2.2: as I wrote before, a flowchart would help a lot
Reply: *Thank you very much for this great suggestion, which we adopted by adding a new Figure 1, a flow chart indicating the most important analysis steps and links between them.*

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.
Reply: *It is correct that median correction may perform as well or even slightly better than quantile mapping for certain parts of the distribution as illustrated in Figure 2b (now 3b). However, we still found that overall, quantile mapping resulted in more satisfactory results. The main reason for not just correcting by the median was because our study focuses on extreme flows, which can be corrected by applying quantile mapping. We changed the colors to make the different lines easier to distinguish.*

Fig.3 is a bit minimalistic. I'm confident it can be improved.
Reply: *The main idea behind the figure was to illustrate which part of the dataset was used for the frequency analysis and which part was excluded. As part of the information previously included in Figure 3 is now included in the new Figure 1 (10 perturbed members, 24 lead times, overall sample size), we removed the previous Figure 3.*

L180: the the
Reply: *We removed the redundant 'the'*

L188: r%?
Reply: *We corrected this to 'p%'*

Fig. 9(c) and (d) are not very clear. Maybe the x-axis could have a log scale.
Reply: *Thank you very much for this great suggestion, which we adopted and helped to improve the figure.*

Fig. 10: please correct the typo in the caption "))". Furthermore, the five maps are very similar. Maybe some of them could be removed.
Reply: *We replaced the inner pair of brackets by '[]'. It is true that the spatial patterns shown on the five maps look quite similar but we would prefer to retain all of them because the different return periods are again picked up in Figure 11.*

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.
Reply: *We agree that it is difficult to argue why latitude should physically be an important predictor of flood magnitudes. We therefore excluded longitude and latitude as potential predictors from the analysis and redid figure 11. We now see a relatively strong negative relationship between temperature and flood magnitude. That is, higher flood magnitudes for catchments with colder climates (e.g. those in the Alps).*

L299: the the
Reply: *We removed the redundant 'the'*

L314: "…any such…" please check
Reply: *We removed 'and'*