The authors would like to thank the Anonymous Referee #1 for the detailed review of our paper and the many constructive comments. We believe that they will enhance the document significantly. In the following, we provide our answers. The reviewer comments are printed in black and our replies, in blue.

RC1: Anonymous Referee #1

1. **RC1:** *The research is a comparison between different forecasting systems. Quantitative results are needed to justify the conclusions. However, the authors tend to use sentences like 'Line 467-468: We also observe that post-processing precipitation forecasts have a* ***much higher*** *impact on the quality of precipitation forecasts (Fig. 9) than on the quality of streamflow forecasts, as evaluated by the BIAS score', or 'Line 491-491: It is interesting to note that, for systems B, C, and D, streamflow forecasts based on raw precipitation forecasts are always* ***much better*** *than streamflow forecasts based on post-processed precipitation forecasts in system'. They don't give the readers an objective description for the research. The qualitative results cannot help the scientific choice. Please provide more numerical results, especially in conclusion.*

   Authors Replay (AR): We will revise the description of the results to provide a more in depth numerical analysis.

2. **RC1:** *Concerns about the multimodel approaches. From Figure 8, the multimodel approach seems to bring additional uncertainty to streamflow forecasts, as system C always has the worst results in term of BIAS and IQR. From Figure 5, the models are with an average KGEm of 0.64 in validation without EnKF. This is a quite low value. When the hydrological uncertainty is dominant, it is difficult to analyze the effect from precipitation post-processor. So, the boxplot for system C with or without post-processor for lead time 1 day in Figure 11 is similar. It is not indicated that the precipitation post-processor brings in no improvements. The improvements probably are too minor to offset the hydrological errors.*

   AR: This is a very interesting point, and we thank the reviewer for drawing attention to it.

   The 0.64 is the mean of all models in all catchments considered separately, not as a multimodel. That is, the performance of each of the models was determined individually and then averaged. In the case of the multimodel, its average yield is 0.73 (see figure attached to question 12 of reviewer #2). The simulations of each model are considered as an ensemble and then their performance is determined.
   It is also good to remember that extreme values affect the mean. As shown in Figure 5, models 1, 4, 6, and 7 experience difficulties in simulating some basins, which is to be expected since no single model excels in all situations. In the case of systems using only one model, the one used is the median during calibration. As we responded to reviewer 2, it is almost impossible to predict which model will be the best predictor on any given day and basin, and that is when a multimodel has value.

In the revised version, we will add the performance of the multimodel in Figure 5 to avoid confusion, and we will also soften our wording to recognize the value of the post-treatment. This point raised by the reviewer is worth mentioning, and we will include a comment on it.

3. **RC1:** *The author's language usage was difficult to read at times. Too many adverbial clauses and attributive clauses make the sentence too long to understand.*

   AR: Fixing this will be a priority when producing a revised version.

4. **RC1:** *In Section 2.1.1, the authors reduced the ECMWF database to 0.1° and then spatially averaged forecasts to the catchment scale. I am confused by the resolution reduction as the catchment areal forecasts were used. It is more simple to use the archived 0.25° to calculate the catchment average. The resolution reduction might bring in additional uncertainty to precipitation forecasts.*

   AR: We fully agree with the reviewer that resolution reduction might bring additional uncertainty to precipitation forecasts. However, the original resolution of the ECMWF is too coarse for this application, especially for the smaller group of catchments (11 in total). The surface of this group is less than 800 km2, and in many cases, only one meteorological forecast grid point falls within catchment boundaries and in others none. Therefore, downscaling ensures that several points fall within the catchment boundaries.

   In addition, Thiboult et al. 2016 suggested that when we reduce the spatial resolution, we consider the contribution of the points close to the catchment boundaries, which allows us to have a better description of the meteorological conditions of the catchments. This is also corroborated by Scheuerer and Hamill 2015a who demonstrated that it is beneficial to add forecasts from grid points within a certain neighborhood of the location of interest as potential predictors to account for position uncertainty.

   We will clarify the purpose behind the downscaling in the revised version.

5. **RC1:** *Line 170-171: The authors mentioned they used an adapted CSGD to post-process ECMWF precipitation. Whether the only difference is that the original CSGD used neighboring grid points while they used all grids in a catchment?*

   AR: The reviewer understood this issue correctly. The original CSGD uses neighborhood information from grid-based forecasts to compute the ensemble statistics, thus accounting for displacement errors. In our case, we computed the ensemble statistics directly from the mean areal precipitation over catchments. As mentioned in the previous answer, the downscaling implicitly allows considering the contribution from neighboring grids.

6. **RC1:** *Line 221-222: the same….as….*

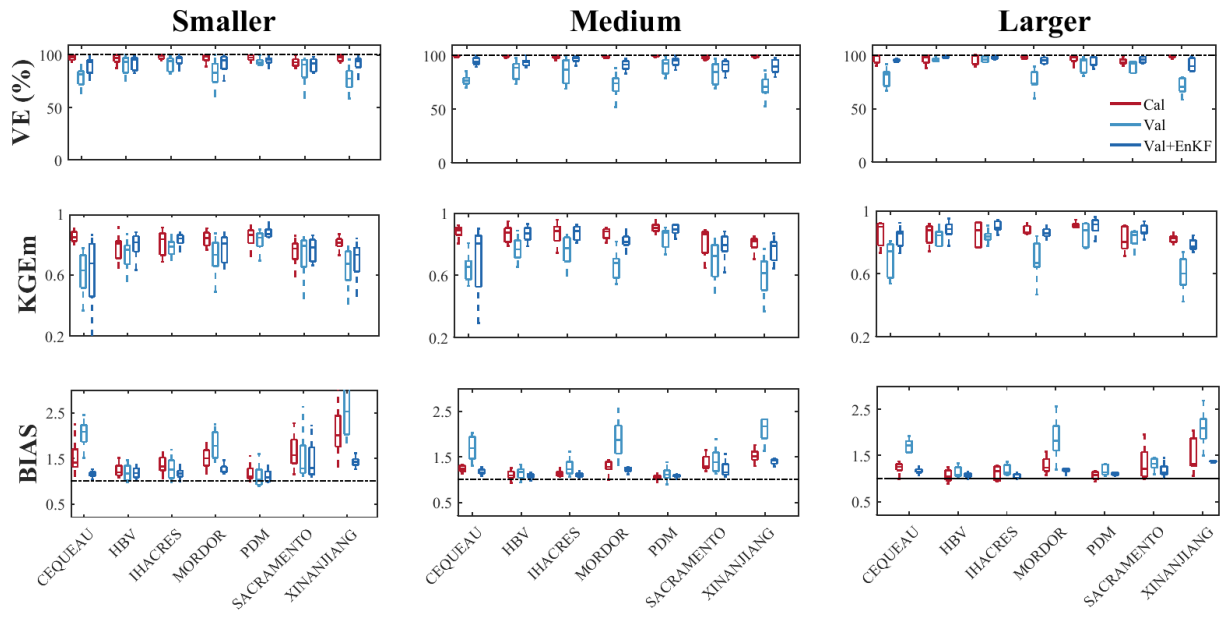
   AR: We will correct in the revised version

7. **RC1:** *Line 506-509: NWP products often fail to capture precipitation forecasts in small domain, yet behave better in large catchments. Lumped hydrological models are more likely to better model streamflow at small basins, where the hydrological process is simpler and easier to be simulated by those simple lumped models.*

   AR: Thank you for pointing this out. As suggested by Reviewer # 2, we will shift the emphasis to the properties of the catchments rather than the modeling approach (see our answer to question 13 of Reviewer #2 for details).

8. **RC1:** *Since the authors elaborately analyze the performance for different catchment size in Section 3.4, they should provide detailed validation results of the hydrological models in different catchments in Section 3.1. It would help the readers to understand the forecast skill over catchment sizes.*

   AR: Thank you very much for this suggestion. We propose to include the figure below and move Section 3.1 to the supplementary material as recommended by Reviewer #2 to decrease the length of the paper.



References

Thiboult, A., Anctil, F., and Boucher, M.-A.: Accounting for three sources of uncertainty in ensemble hydrological forecasting, Hydrol. Earth Syst. Sci., 20, 1809–1825, https://doi.org/10.5194/hess-20-1809-2016, 2016.

Scheuerer, M. and Hamill, T. M.: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions, Monthly Weather Review, 143, 4578–4596, https://doi.org/https://doi.org/10.1175/MWR-D-15-0061.1, 2015.