

## ***Interactive comment on “Exploring hydrologic post-processing of ensemble stream flow forecasts based on Affine kernel dressing and Nondominated sorting genetic algorithm II” by Jing Xu et al.***

**Anonymous Referee #2**

Received and published: 14 October 2020

Dear authors,

Thank you for presenting this interesting work. With application of both EnKF for hydrologic initial condition uncertainty, and hydrologic post-processing, this is a potentially valuable case study on hydrologic ensemble prediction. The paper is overall clearly written, particularly the Data and Methodology sections.

My main questions and concerns are the following:

The objective(s) of the research is(are) in my view not clearly stated, nor the intended

C1

contribution to the literature. Could the authors describe these?

This perhaps also makes the literature review rather general, not zooming-in to identify a gap or under-represented aspects/applications of ensemble prediction, or a particular forecast challenge in the case study catchment.

As I understand, post-processing of the meteorological ensemble forecasts was not done. Could the authors comment in the paper on the performance of the meteorological ensemble forecasts and state the reason for not also applying meteorological post-processing?

Based on Figure 4, presenting one forecast, I do not understand how it can be concluded that the members generated from the meteorological eps as forcing are not fully interchangeable, which is the basis of applying weights with NSGA-II. Perhaps that the video (I am sorry I could not find it, this is probably my omission) shows this, but this is not explicitly stated in the paper.

I do not understand why the authors choose to calibrate the post-processors on only one forecast horizon (day-4) and validate on the other horizons (1-3, 5-7). Because of a generally present decrease of skill with increasing forecast horizon, usually a post-processor is calibrated for each forecast horizon separately. It also seems that the analysis period for which re-forecasts have been prepared has not been split in a calibration and validation period (or a leave-one-out approach). There may be good reasons for choosing this approach, e.g. stemming from catchment or application characteristics versus limited data availability, or as a research objective, but I missed the explanation in the paper. Could the authors perhaps explain the chosen calibration/validation approach?

Lastly, I would kindly encourage the authors to expand the presentation and interpretation/discussion of the results. For example, why present the 5 sub-catchments, what should we learn from the results? What about the inflows to the reservoirs? Why present the 5 single h-eps, what should we learn from the results? Why not assess

C2

the performance of the combined grand multi-model ensemble? How does the performance of the raw and post-processed forecasts compare with the performance of a reference forecast such as climatology or persistence (forecast skill)?

Detailed comments:

Introduction: Could you add explanation why AKD and NSGA-II have been chosen for this research? (line 51)

Introduction, data description, and/or Results section: Could you comment on observational uncertainty?

Line 92: Does the analysis of forecast performance take into account these different flood generating processes, and related seasonality? Would be interesting.

Line 99: In this section, kindly add some information on catchment response time to rainfall/snow melt, and travel time (routing), to inform us about potential forecast lead times without meteorological forecasts as forcing.

Line 104: Could you mention why inflow to the reservoirs is not measured (for some reservoirs), and how the inflow time series have been constructed?

Line 105/106: Could you briefly describe the observational network, and methods used to create sub-basin average precipitation? (to inform observational uncertainty, and perhaps a reason for not going for meteorological forecast post-processing)

Line 168: I think we are missing here, what the parameters are optimised on. From the later paragraph on Experimental set-up it seems that parameterisation of AKD was done by minimising MCRPS.

Lines 224-226: Kindly explain why these parameter values were chosen, and if a sensitivity analysis was done? Was the maximum evolution runs a result of a stopping criterion? If so, please mention this.

Lines 272-273: Please introduce the use of a moving window in Section 3.2, and ex-

C3

pand explanation. Also the mentioning here of operational requirements is interesting and further explanation and discussion would be welcome.

Figure 5: The differences in bias and NSE over the range of the Pareto front are small. Please discuss. What weights are in the weight matrices of these solutions?

Line 338: This is interesting. Could you discuss what could be the reason? Something specific about Model M05?

Editorials

Lines 25-27. Deterministic systems do not assess/quantify uncertainty, so the superiority question, I think, did not concern uncertainty quantification, that difference is simply a given. The superiority question concerned more the value when using the forecasts in decision making, and ensemble mean versus deterministic forecast performance.

Figure 1: Please indicate in the map more clearly the main river reach and flow direction.

Lines 82-84: Not clear from this sentence if in Section 4 the results are analysed for each model individually first (not taking into account model structure uncertainty), and then are considered and processed as a grand multi-model ensemble, which does take into account model structure uncertainty.

Lines 137-139: Consider to move up to Introduction for literature review, or down in the sections below. In these few introductory sentences to the methodology I would focus on announcing what was the general approach followed to reach the research objectives. After having introduced the overall methodology, going into the details of the two post-processing methods as of section 3.1 makes sense.

Figure 3: Qobs is not output, so can be left out on the right. It is also not indicated that the output or final results concerns post-processed (interpreted) Qfcsts. The flowchart ending in only one set of post-processed forecasts is confusing, because up to now I was under the impression that AKD and NSGA-II would be used independently to

C4

post-process and hence each method to result in a set of post-processed forecasts, after which the performance of each method will be analysed and compared.

Line 256: Spread Skill plots are announced, but later not presented.

Figure 8: Presenting results in spider plots is a nice idea, but with the scores selected this does not work well, because some scores indicate a better performance with lower value (RMSE) while others the other way around (NSE), and some have a scale only to 1 (NSE) while others are not limited. This makes interpretation of the plots rather difficult.

Lines 357-362: General/Literature - I suggest to delete or move to Introduction

Lines 366-369: Consider to move to Introduction or Methodology

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-238>, 2020.