**Revision notes for "Potential application of hydrological ensemble prediction in forecasting flood and its components over the Yarlung Zangbo River Basin, China" (hess-2018-179)**

Dear Editor and referees,

Thanks a lot for your great efforts to read through this manuscript and give very valuable comments. We agree with your suggestions which will be of great help to improve the quality of our manuscript. Here we have addressed the comments from you and the detailed description is attached in this document.

Best regards,

Li Liu, Suli Pan, Zhixu Bai, Yue-Ping Xu

## Response to main comments:

**1**. The application of flow separation in forecasting and analysis of forecast skills of different flow components, including flood volumes, base flow, first flood in a year and annual maximum flood are the main novelty of the study. The second aim of the paper is to study "the impact of an ensemble of Pareto optimal solutions on model simulations". I feel that the authors have failed in combining those aims together and it is the main drawback of the paper.

**Response:** Thank you very much for your comments. We agree that the original manuscript fails to combine the two aims properly. As the streamflow components are unknown, a plausible total runoff doesn't mean accurate streamflow components. Under this circumstance, we attempt to capture the most possible flow components by applying an ensemble of parameters to take account of more scenarios. The accuracy of multiple parameters in total runoff is the precondition for application in further flow components forecasts. Thus, the evaluation of N-simulations (simulation from ensemble of multiple parameters) in Subsection 4.1 and 4.2 is to demonstrate the efficacy of N-simulations and to prove the feasibility for flow components simulation. However, we didn't mention it in the original manuscript

and make the paper somewhat confused and fragmented. Our solution is to clarify the purpose to adopt an ensemble of Pareto optimal parameter for flow components forecasting to make the paper more logical and integral.

**2.** The available observations were divided into calibration, verification and evaluation periods. The authors refer to their previous paper published in Journal of Hydrology (Liu et al., 2017) when listing the additional datasets used in the present paper. I advise them to repeat that information in the present paper to help the reader. Also, the authors refer to that paper while describing the calibration stage and as a result the description has become not very transparent. It is not clear if the snowmelt component was previously used and the number and names of parameters optimised in the present paper are missing. The authors should state clearly which parameters they optimize and how the separation into snow-induced runoff component is calibrated. The authors mention some validation but the description is not clear. In summary, the Data section should be extended and the Methodology section re-organised.

**Response:** Thanks for your comments. We are so sorry to omit the relevant data description and methodology introduction in the original manuscript and we agree that this part should be rephrased and reorganized. Different from the previously published paper, in current study the snow model related parameters are calibrated with the other normally calibrated parameter and detailed calibrated parameter will be clarified in Subsection 3.1 in the revised manuscript. The separation into snow-induced runoff components is calibrated based on two model parameters: the maximum temperature at which snow can fall and the minimum temperature at which rain can fall. These two parameters can separate the input precipitation into two parts: the liquid and the iced portion, and furtherly the total runoff is separated following the hydrological separation in Subsection 3.2.

**3.** The authors compare an ensemble of multiple objective Pareto simulations and single best in their ability to forecast different flood components. The authors apply Preference Ordering Routine (POR) to choose the N-Pareto-optimal sets. There is no explanation of why that particular method was used nor what is the physical meaning of the applied numerical procedure. The authors set the number of

ensemble members to ten, but it is not explained why that number was chosen in statistical terms.

**Response:** The explanation for why the POR method was adopted and why the number of ten was set will be added in the revised manuscript.

In practical sense, users of automatic calibration routines have to face the task of selecting a set of suitable model parameters (preferred solution set) from the numerous Pareto-optimal sets. This is also the challenge for our study. The POR is proposed exactly to solve this kind problem by sorting out small number of preferred solutions. This method has been demonstrated for calibration of MIKE11/NAM rainfall-runoff model and is able to provide the best estimated parameter sets with good overall model performance (Khu, 2005). These are the reasons why POR was adopted in this study.

The number of 10 is set mainly for the consideration of computational capacity. The number of parameter sets would be more than 10 when superadded the extreme value and the compromise value in any two-objective trade-off. Given the large number of ensemble meteorological forecasts and different lead times, the parameter sets less than 20 would be manageable and achievable. In the work of Khu (2005), two solutions with efficiency of order 3 with degree 4 are able to provide still good performance, so in this study, the number of 10 is enough for representative of model parameter. As a matter of fact, when applying POR for all possible combinations of the four objective functions in this study, the final number of preferred parameter set is less than 10. Though we presupposed the upper limit, the final results didn't reach it.


**4.** In the hydrograph separation subsection 3.2, the authors do not explain which data were used for the calibration/validation of the separation parameters.

**Response:** There is no parameters needed to be calibrated in Subsection 3.2. All the variables in the formulas are either known (model outputs) or unknown but able to be obtained by an iteration process like $f_{W,snow,t}$. The only related parameters are the Maximum Temperature at which snow can fall and the Minimum Temperature at which rain can fall. These two parameters determine the amount of rainfall $Rain_t$ and snowmelt $M_t$ in Eq. (5) ($R_{snow,t} = R_t f_{R,snow,t} = R_t f_{i,snow,t} = R_t \frac{M_t}{M_t + Rain_t}$) and are calibrated with other soil parameters in VIC based on streamflow related objective functions (Eq. (1)-(4)).

**5.** The post-processing of ECMWF forecasts is performed in an arbitrary way, without checking if it is necessary and provides better forecasts. Bias correction does not always give positive results regarding forecasting (Kiczko et al., 2015, Benninga et al., 2018).

**Response**: As a matter of fact, we conducted a preliminary analysis of the performance of parameterized QM on ECMWF forecasts and we found that the CRPS from bias-corrected forecasts is much smaller than that from the raw forecasts especially for temperature forecasts. Given the already redundant figures in the original manuscript, we left them out in the submission. If the referee thinks it is necessary to include the evaluation of NWP forecast in the paper, we would like to add them in the revised manuscript or just add some comments about the post-processing for simplicity.


**6.** The Results section includes hydrological model performance and an assessment of flood volume and flood component forecasts. This section I find very confusing. It does not help that the authors use abbreviations that the reader needs to be acquainted with.

**Response**: Thanks for your comments. Originally, we thought the abbreviations would be beneficial for the readers to read and understand the manuscript. However, according to the suggestions from the referee, we will reduce the abbreviations in the revised manuscript. The hydrological station names and the flood events can be indicated by full name which would be helpful for reading and make less abbreviations that the reader needs to be acquainted with. At the same time, we will improve our expression and wording to make the paper more readable.


**7.** The authors conclude that 7-day accumulated flood volumes are easier to forecast than the peak flows. The snow-induced flood component is not well captured whilst the rainfall-induced floods are forecast well. Taking into account the fact that the snow and glacier melt forecasts were not available that conclusion is not surprising.

**Response**: We agree with you that since snow and glacier melt forecasts are not available, the performance in these streamflow components is unsurprisingly inferior due to various uncertainties. We

know that the snow/glacier melting is influenced by not only the input rainfall and temperature but also the ability of model to capture melting process. However, considering the rainfall-induced components are also unavailable and the portion of rainfall used to generate this component is determined by the same procedure used to determine the snowfall to generate snowmelting, the comparison between these components can tell some relatively valid conclusion. We will add the reason you suggested in the revised manuscript and it would make the conclusion more strongly justified.

**8.** The authors find that the base flow component forecast is insensitive to the forecast lead. As the base flow dynamics is slower, the forecast lead may not cover the base flow variability. However, on page 12, the authors state that for NX, the base flow forecasts show a deterioration with a lead time. A synthesis of the overall results is missing.

**Response**: We totally agree with the referee that the insensitivity of base flow to lead time is also caused by the slower flow dynamics and the lead time doesn't cover the flow variability. We will add this possible reason in the revised manuscript to make the study strongly justified.

Thanks for the reminding. Due to our negligence, the unique behavior of base flow at NX is missing in the conclusion. We will add the contents related to NX into the final section and make the conclusion more accurate and thorough.

**9**. The language requires correcting by a native English speaker. The text is rather difficult to follow.

**Response**: We are so sorry and the manuscript will be carefully checked and polished by the native English speaker.

## Response to Editorial comments:

**1**. Page 1: The Abstract conclusions are not transparent. It is not mentioned that the forecasting performance varies within the catchments studied.

**Response**: The forecasting performance does vary for different sub-catchments and the detailed difference in three sub-area will be added in the Abstract. The main difference for three sub-catchments is that baseflow components at NX tends to change with the lead time.

**2**. Page 2, lines 5-9 Style should be corrected.

**Response**: We will correct the style for lines 5-9 in Page 2.


**3**. Page 2., lines 25-27: Is this snow tracking model used in the present paper?

**Response**: Yes, this snow tracking model is used in this study.


**4**. Page 3, lines 111-13: style should be corrected.

**Response**: We will correct the style for lines 111-113 in Page 3.


**5**. Page 5, lines 17 and 20: instead of the word "theorem" I would use "attribute".

**Response**: We will replace the word "theorem" with "attribute" in the revised manuscript.


**6**. Pages 7- to the end: there are language problems in nearly all pages and language editing by a native English speaker is needed.

**Response**: Native English speaker will be asked to check and polish the entire manuscript.


**7**. Figure 5 – there should be some quantitative assessment of the differences between simulated and observed snow cover. The comparison does not look well!

**Response**: We will add the quantitative assessment for the simulated and observed snow cover in the revised manuscript. The special correlation coefficient and the overall bias in entire study area will be calculated to show more direct evaluation.


**8**. Figures 10-12 - it would help if the columns were named (snow-melt -induced and rainfall-induced components).

**Response**: We will name the columns in Figures 10-12 to make the figure more accessible.


**9**. Benninga et al. (2018) is not referred to in the text.

**Response**: The reference for Benninga et al. (2018) will be added in the revised manuscript.