Comment on hess-2021-562

Anonymous Referee #2

Referee comment on "A comparison of hydrological models with different level of complexity in Alpine regions in the context of climate change" by Francesca Carletti et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-562-RC2, 2022

This paper addresses an important and interesting topic- comparison of the hydrological models with different level of complexity in the high Alpine catchments. The authors compared two degree-day models and one full energy-balance models in the context of climate change. Overall the paper is well organized and the presentation is good. I suggest a major revision, and there are several issues to be further improved. The comments are as follows:

On Section 1:

Are there any other similar models that could also reach the goals? Why do you decide to select these two models for comparison? I suggest some literature review and explanation could be given in section 1.

This section listed many references that are mainly related to the comparisons of the Alpine3D model and the degree-day model. However, there is a lack of the summary of the relation and innovation of this research which differs from the previous studies. Some discussion in more detail on the relevance of the references to the present research are needed. The innovation of this study should be highlighted.

One of the aims of the study is "getting a better understanding of the conditions under which one kind of melt scheme and/or hydrological model outperforms the other". The study only considered two catchments, thus I regard it as a case study. We don't know the how do the models perform in other cases. I'm concerned that the cases in the

research are not strong enough to support the generalization.

On Section 2:

"68 model chain outputs are provided under three Representative Concentration Pathways: RCP8.5, RCP4.5 and RCP2.6. In this paper, we considered a selected subset of 17 out of the original ensemble". Do the models you selected in this study outperformed others? Is there any assessment of the historical performance of the GCMs and RCMs before they are selected for the study area? Please explicit the reason why you choose the subset.

On Section 3:

The the model description part, two models are introduced separately. Since the title is compare the models with different levels of complexity. I think more focus could be paid on the summarizing the overall differences in terms of, for instance, the models structure and modules, hypothesis, parameters and etc. And how the complexity differences are embodied in the models. I think it would be easier for readers to obtain the most important information about the differences of the models.

Please give the equation for the calculation of the statistical scores RMSE, NSE and KGE in this section.

On section 4:

Did you do calibration for the A3D? If not, please clarify the reasons. If so, please list the parameters and their ranges for the calibration of the A3D model, and the calibration results for the A3D model.

The calibration scores for the PH model listed in table 8 is not ideal, especially in Dischma catchment with only 0.36 measured in NSE. I just wondered how much credit could we give to the models? Though there are analysis for the performance of the model simulation. Could you add the comments on the major contribution for such errors? I strongly recommend adding some references to support the results and it is necessary to make an explanation for the errors. It would be helpful for the readers to interpret the results if the explanation is given.

"PHR delays the spring snowmelt-induced discharge by one month compared to

observations" Why does the PH reproduce a delayed melt season? It's noticed that the PH also has a lower snow melt volume. How do could it be explained in terms of model structures, mechanisms and hypothesis differences?

In the figure 7 and 8, It seems that the performance of the A3D and A3Ddd is very close to each other, although a simpler melt-factor energy balance mode is applied in the A3Ddd. Could it be interpreted as the differences of energy balance modules for the A3D model does not have a significant effect in simulating the runoff?

Please add some interpretation of the α, β and components o the KGE scores in table 9.

The discussion part is suggested to be in a new section after all the results are listed.

On section 4:

The errors here are attributed to the dams, " The explanation is twofold. First, the Mera catchment is highly regulated by dams, which is not accounted for in the models." However, in section 2.3.2 you mentioned "Discharge modeling here may be slightly disturbed by hydropower regulation... However, at the daily scale and at longer time scales, streamflows are not largely disturbed overall, and hydrological modeling exercise provides acceptable results."I think the arguments are controversial. Besides in the conclusion part, you also emphasized the effect of reservoir regulation on the discharge simulation. As far as I see, the impact of hydropower regulation could not be easily neglected for this study.

It's interesting to notice that on average the peak of snow melt and discharges in RCP2.6 is higher than those in RCP8.5. With higher temperature increase in RCP 8.5, what makes the peaks of the discharge and snow melt being less?

On section 5:

"Our interpretation is that the calibration process for strongly regulated catchments as Mera overshadows the benefits of a full energy balance scheme showing good performances in reproducing snow melt." Maybe it's true in this case that the calibration offset the errors from regulation to some extent. But I think as the conclusion it is more important to know implication from the study. In which case the calibration could overshadow the benefit of the physical scheme? Could benefit from the calibration also be applicable under climate change scenarios, and what is the limitation of the models through the comparison?