Point-by-point Response to the Reviewers for the Manuscript:
"A comparison of hydrological models with different level of complexity in Alpine regions in the context of climate change"

Francesca Carletti, on behalf of the author team

April 2022

1 Introduction

This document contains a point-by-point reply to the comments to the manuscript made by Reviewer 1 (Section 2) and Reviewer 2. We thank both the Reviewers for all their valuable comments to our manuscript, knowing that this review has been a significant time investment and therefore especially appreciating their feedback and commitment. All the comments have been addressed, and the vast majority of them has led to additions or clarifications in the text. All comments were relevant and well placed and did certainly contribute to increasing the clarity and overall quality of our work.

This document is divided in three sections: Response to Reviewer 1 (Section 2), Response to Reviewer 2 (Section 3) and a Summary of the relevant changes made to the manuscript (Section 4). The issues raised by the Reviewers are subdivided according to the sections of the first version of the manuscript they refer to.

Comments from the reviewers are reported in **bold font**. Replies from the authors are reported in regular font.

2 Response to Reviewer 1

2.1 Section 2

Spatial resolution of 500m seems coarse for this region. Why didn't authors select a higher resolution to run their models?

Our choice is based upon the work of [Schlögl et al., 2016]. An extensive explanation and justification to this choice was added in Section 2.6.1 of the reviewed manuscript, P11L242-255.

2.2 Section 3

Both hydrologic models are semi-distributed models even though one is grid based. In such complex topography a fully distributed hydrologic model (even based on energy balance) with multi parameter regionalization approach would be more suitable than those two semi distributed models. mHM is one these models with day degree method. However, it is open source model in which other complex formulations could be introduced in its Fortran code.

Distributed models have the advantage to explicitly take into account spatial variability, which is particularly useful if snow dynamics are important. This is why we used a fully distributed model, Alpine3D, to describe surface processes and (part of) sub-surface soil processes. In order to separate effects that come from lumping surface processes from effects of Energy-Balance versus Degree-Day modelling, we introduced a variant of Alpine3D, in which we replaced the Energy-Balance with a Degree-Day approach but keep other model features unchanged. Since this comparison shows that the Alpine3D in its Degree-Day configuration runs don't lead to much new insight, we don't execute them for the future climate. Concerning sub-surface hydrology, in our opinion a more complex (3D or distributed) model makes only sense if you have sufficient sub-surface soil and hydro-geology information, which is not the case.

Our work compares the fully distributed model Alpine3D, in its Energy-Balance and Degree-Day configuration, with the semi-distributed Degree-Day model Poli-Hydro. The use of a semi-distributed modelling approach rather than a fully distributed one can be justified by the following reasons. In the first place, the increase in spatial resolution introduces two well-known problems: overparametrization and equifinality [Beven, 1989, Beven, 1993, Beven, 2006]. More complexity in spatial detail results in more parameters involved and, therefore, potential issues in calibration, such as equifinality. Secondly, computational times are consistently higher for fully distributed models. For these reasons, semi-distributed or purely conceptual models are preferred in most studies, justifying our choice to rather assess those kinds of models. In the revised version of the manuscript, we added these clarifications in P12L283-288.

On the other hand, the use of a Degree-Day approach rather than an Energy-Balance can be motivated as follows. In the first place, the Energy-Balance approach usually requires a large amount of measured weather data (i.e. wind speed and direction, soil moisture, radiation), rarely, if ever available in scarcely monitored and/or inaccessible areas. One of the scopes of the paper was to underline the strengths and drawbacks of a relatively simple model against a fully distributed, Energy-Balance one for climate change impact studies. One of the strengths of the Degree-Day approach is the possibility of obtaining climate projections over scarcely monitored areas using commonly available data. Secondly, when the simulation is performed at a daily scale (such as for Poli-Hydro in our paper), which is also the scale used for the model comparison, the Degree-Day approach generally gives satisfactory results, especially when tuned with snow cover/depth data [Soncini et al., 2017]. These points were covered in the Introduction to the original manuscript, P2L1-11, and were rearranged into the revised manuscript, P2L24-29.

Finally, the choice of the two models was motivated by the fact that both Alpine3D and Poli-Hydro are well-assessed models that have been used over a large array of conditions and over catchments of largely varying size, with satisfactory accuracy in reproducing stream flows, snow/ice dynamics and cryospheric contributions. An extensive discussion about the aforementioned points have been added in the Introduction of the revised manuscript, P2L37-P4L62.

The authors preferred RMSE, NSE and KGE metrics. However, in a satellite based remote sensing study, the readers would expect spatial pattern evaluation metrics for evaluating spatial snow cover output of the models. The authors should include one of these well established metrics e.g. FSS, EOF, Kappa or SPAEF.

We don't use spatial evaluation metrics for snow cover because the paper is not a satellite-based remote sensing study. The topic of the paper is primarily hydrology rather than snow distribution. Therefore, snow cover represents one single step towards the estimation of discharge (under present conditions and climate change). In our paper, satellite data are only mentioned in Section 4.1.3 as, in the model Poli-Hydro, the validation of the modelled Snow Cover Area against MODIS imagery is one single step towards the calibration of snow Degree-Days. However, the target in the calibration of the model Poli-Hydro is discharge uniquely. In our paper, we refer to the work of [Fuso et al., 2021] for further details concerning the calibration of snow Degree-Days. In [Fuso et al., 2021], the authors extensively focus on the set-up and performances of Poli-Hydro in this sense, which we did not repeat as it was not the main scope of our paper.

Furthermore, satellite snow imagery can only provide binary information about snow cover (i.e., presence or absence). However, the most important parameter involved for discharge estimation is the Snow Water Equivalent, which cannot be derived from satellite imagery, at least for the time being.

Currently the methods section is not well organized. A separate section should solely focus on calibration details in addition to statistical scores (3.2)

Thank you for this suggestion. We agree. In the updated version of the manuscript, we have rearranged the Methods and the Results section. In the Methods section (Section 3) we now give a general model description (Section 3.1) and then individually present and describe each model used (Alpine3D, Section 3.1.1; StreamFlow, Section 3.1.2; Poli-Hydro, Section 3.1.3). Then, the statistical scores used throughout the paper are presented in Section 3.2. In the Results section (Section 4), calibration details are given for each model, along with performance scores (Alpine3D, Section 4.1.1; StreamFlow, Section 4.1.2; Poli-Hydro, Section 4.1.3).

2.3 Section 4

P16L5-10: Calibrated values of many parameters were taken from other studies. Why? Computer source limitation? Apparently this is not an issue as P15L7 indicates that 10k model runs were performed.

Yes, as mentioned in P16L5 and in Table 7 (P18L416 in Table 8, as of the updated version of the manuscript), the model Poli-Hydro is fed with a set of fixed calibration parameters obtained from previous studies over the same areas. To feed the model with plausible parameters verified in previous studies is a common practice in operational hydrology to ease the multi-objective optimization problem. Moreover, many of such fixed parameters are physically based and refer to the soil type, land cover and other "static" properties of the area, which are unlikely to change in the short-term. Therefore, recalibrating all the parameters at once would be a computational effort that is unlikely to bring any significant benefit. This clarification has been added to the updated version of the manuscript, P18L416-420. On the other hand, the 10'000 model runs mentioned in P15L7 (P17L389 in the updated version of the manuscript) relate to the calibration of the model chain Alpine3D+StreamFlow.

Why MC method was used for calibration? and not CMAES, DDS or SCE requiring less model runs. Please justify.

We agree and we thank you for this suggestion. This is indeed an improvement that some of the authors would have wanted to make to the modelling chain Alpine3D+StreamFlow at some point. Albeit other algorithms such as the mentioned ones would certainly require fewer model runs, an increase in accuracy would still be unlikely. For this reason, our opinion is that the effort required to implement the modelling chain with more efficient heuristic algorithms would not bring concrete advantages to the calibrations in this paper. However, we will certainly take this comment in high regard for future work.

Little details were given about model calibration. Instead of taking calibrated values from the literature it would be more robust to apply spatial sensitivity analysis first to select most important parameters affecting snow processes (see doi:10.5194/hess-19-1887-2015). Then a rigorous calibration with a global search algorithm (listed above) should be applied with appropriate objective functions (spatial error metrics) focusing on spatial distribution of snow.

We agree that there was potential to better detail the model calibration procedure, and we tried to improve it in the revision process. In the updated version of the manuscript, Section 4.1 gives details about the calibration of each model: Alpine3d (Section 4.1.1), StreamFlow (Section 4.1.2) and Poli-Hydro (Section 4.1.3, divided into its Snow Melt module and its Flow Propagation module).

Thank you for suggesting this paper. We have read it carefully: the methodology is indeed interesting, and it could be applied to a diverse range of context with possibly appealing results. However, we have several concerns about integrating it in our work.

With respect to the modelling chain Alpine3D-Streamflow, snow cover is not calibrated: only the hydrological part is. We certainly agree that this was not clearly specified in the first version of the manuscript, undoubtedly leading to confusion or lack of clarity in this respect. In the updated version of the manuscript, this is now clearly specified on P17L383.

With respect to Poli-Hydro, we doubt that spatial sensitivity analysis could bring any substantial improvement to the model calibration. The reason is twofold. On the one hand, the fixed parameters taken from literature are not spatially heterogeneous but, indeed, constant. Besides, they are not snow related. On the other hand, in both Poli-Hydro and the model used in the suggested paper, snow melt (i.e., the most relevant contribution to discharge generation in partly-glacierized catchments) is governed by temperature and radiation (Section 3.1.3). Our opinion is that a sensitivity analysis on two variables may not lead to interesting conclusions. At the same time, since the central focus of our paper is the discharge estimation rather than the spatial distribution of snow cover, it would be more interesting to investigate on which parameters the process by which snow melt turns into flow depends. However, as these parameters are physically based and have already been evaluated in previous works on the same study area and over the same time span, our opinion is that to recalibrate them could add further complexities without bringing considerable advantages to the overall quality of our work.

3 Response to Reviewer 2

3.1 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.

Thank you for this suggestion. We agree: Section 1 was not well balanced in the first version of the manuscript, as it dwelt a lot on the object of study

- the model comparison - but hardly at all on the specific choice of the models. In the updated version of the manuscript, we added an extensive explanation and literature review on the reason why these models have been chosen, P2L37-P3L62.

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.

Again, we agree. Section 1 did not sufficiently highlight the innovation of our work compared to previous ones. In the Introduction of the revised version of the manuscript, we tried to put more stress on this aspect: P3L67-69, P3L74-75, P3L86-P4L97.

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.

The Reviewer is right: the cited sentence generalized the case study too much and created ambiguity about the purpose of the paper. However, our impression is that throughout the Discussion, the Climate Change dissertation and the Conclusion (Section 4.2.4, Section 4.3 and Section 5, respectively, in the first version of the manuscript), no particular concept was generalised, but everything was clearly related to the case study specifically. Thus, this sentence has been modified in the new version of the manuscript: P3L88-90.

3.2 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.

Thank you for suggesting this. This subset of model chain outputs has been used in previous studies [Epting et al., 2021] because it captures both high and low climate change signals for air temperature and precipitation [MeteoSwiss, 2018]. In the updated version of the manuscript, this point has been specified in a new paragraph (P9L187-189).

3.3 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.

Thank you for this input. Our opinion is that separated presentation is functional to introduce and describe the characteristics of each model, although we agree with the Reviewer about the fact that a summary of the main differences in terms of structure, hypothesis, parameters should be added to help the reader. For this reason, we added Table 5 to the updated version of the manuscript. Table 5 contains a summary of the main features and differences of the models used for the case study.

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

Equations for statistical scores have been added to the updated version of the manuscript: P16L362-364.

3.4 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.

We agree that this was never clearly stated throughout the first version of the paper. In the updated version, we improved the calibration process description for each model (Section 4.1.1 for Alpine3D, Section 4.1.2 for StreamFlow, Section 4.1.3 for Poli-Hydro). Specifically, for Alpine3D, we specified that no calibration takes place in a strict sense.

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.

Thank you for bringing up this point. While dealing with this, we realised that we reported wrong values of NSE scores for the model Poli-Hydro. In the updated version of the manuscript, we corrected the values (Table 9 and Table 10).

It is certainly true that the performance of Poli-Hydro in the calibration phase is not ideal for Dischma. In the first version of the manuscript, partial explanation was given at the end of Section 4.2.3 in Figure 9 when seasonal statistical performances are analysed. However, we realised that the level of detail was not sufficient and that explanations should have been made even earlier in the face of fairly low statistical performance in the calibration phase. Our interpretation is that such a low score is largely attributable to the spatial computational resolution of 500 m, which we justified throughout Section 2.6.1 of the reviewed manuscript, P11L242-255. Furthermore, we addressed low scores in the calibration phase in a new paragraph in Section 4.1.3 of the reviewed manuscript, P20L429-442. Figure 9 in the first version of the manuscript now serves as validation and does not figure as a new finding.

"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?

The explanation is twofold. On the one hand, there is the different rain-snow threshold temperature at which the two schemes are normally implemented. This point was explained at the very beginning of Section 4.2.1, with reference to Figure 3, on the first version of our manuscript. On the other hand, Poli-Hydro is run at a daily resolution, whereas both versions of Alpine3D are run sub-daily. This means that Poli-Hydro relies on a melting scheme that only considers the average daily temperature, which has repercussions on the melt dynamics, as explained in Section 4.2.4 of the first version of the manuscript.

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?

For both catchments, runoff and discharge predicted by A3D and its hybrid Degree-Day version A3D_{DD} are very similar to each other. If, on the one hand, it is true that A3D_{DD} computes the energy entering the snowpack with a simplified Degree-Day approach, on the other hand, the start-up of this hybrid mode still depends upon a full, multi-layer Energy-Balance scheme. Thus, A3D_{DD} is just performing a simplified computation, but only after benefiting from the full complexity of A3D. This point was covered in Section 4.2.4 of the first version of the manuscript (P26L10-P27L3).

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

We covered this point by adding some comments in Section 4.2.3 of the updated version of the manuscript (P28L523-532).

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

We agree. We have modified the structure of the updated version of the manuscript as follows. All calibration results are now listed in Section 4.1, and the model comparison is made throughout Section 4.2 - for present conditions and climate change. The discussion part is now in a new section (Section 5), divided between present conditions and climate change.

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.

We agree that this sentence in Section 2.3.2 of the first version of the manuscript: *"[...] However, at the daily scale and at longer time scales, streamflows are not largely disturbed overall, and hydrological modelling exercise provides acceptable results [Fuso et al., 2021]. [...]"* is inaccurate and slightly contradictory to our conclusions. We have deleted lines 5-9 of Section 2.3.2 of the first version of the manuscript. In the updated version, we added a new section (Section 2.3.3, P7L142-151) focussing on the river regulation over Mera catchment. In this new section, we presented the issues, we provided the due literature review, and we justified the choices made.

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?

Thank you for pointing this out, we should have specified it. In the updated version of the manuscript, we added a paragraph to explain this point (P33L590-594).

3.5 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?

These suggestions have been taken into consideration and we decided to rephrase and partially rewrite consistent parts of our Conclusions. In the new version of the manuscript, we tried to address the suggestions from the Reviewer from P37L713 to P37L731.

4 Summary of the relevant changes made to the manuscript

The relevant changes made to the manuscript are the following:

- The Introduction (Section 1 in both versions of the manuscript) has been extended and better detailed, as of the suggestions of the Reviewers;
- Figure 1 and Figure 2 in the first version of the manuscript have been merged into a single figure in the updated version (Figure 1);
- According to the suggestions of the Reviewers, Section 2.3.3 has been added to the revised version of the manuscript to introduce the problem of missing regulation information over the Mera catchment;
- In the updated version of the manuscript, Table 5 has been added, to give an overview and summary of the models used and of their differences/similarities, as suggested by the Reviewers;
- In the updated version of the manuscript, Results and Discussion are presented into two separate sections (Section 4 for the Results, Section 5 for the Discussion);
- In the updated version of the manuscript, the presentation of calibration results in Section 4 has been improved and made clearer for all models, as suggested by the Reviewers;
- The Conclusion (Section 6 in the updated version of the manuscript) has been reviewed and improved by taking into account the comments made by the Reviewers.

References

[Beven, 1989] Beven, K. (1989). Changing ideas in hydrology — The case of physically-based models. *Journal of Hydrology*, 105(1):157–172.

[Beven, 1993] Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1):41–51. Research Perspectives in Hydrology.

[Beven, 2006] Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1):18–36. The model parameter estimation experiment.

[Epting et al., 2021] Epting, J., Michel, A., Annette, A., and Peter, H. (2021). Climate change effects on groundwater recharge and temperatures in swiss alluvial aquifers. *Journal of Hydrology X*, 11:100071.

[Fuso et al., 2021] Fuso, F., Casale, F., Giudici, F., and Bocchiola, D. (2021). Future hydrology of the cryospheric driven lake como catchment in italy under climate change scenarios. *Climate*, 9(1).

[MeteoSwiss, 2018] MeteoSwiss (2018). *CH2018, Climate Scenarios for Switzerland, Technical Report, National Centre for Climate Services, Zurich.*

[Schlögl et al., 2016] Schlögl, S., Marty, C., Bavay, M., and Lehning, M. (2016). Sensitivity of alpine3d modeled snow cover to modifications in dem resolution, station coverage and meteorological input quantities. *Environmental Modelling & Software*, 83:387–396.

[Soncini et al., 2017] Soncini, A., Bocchiola, D., Azzoni, R., and Diolaiuti, G. (2017). A methodology for monitoring and modeling of high altitude alpine catchments. *Progress in Physical Geography: Earth and Environment*, 41(4):393–420.