January 27th, 2022

Dear Reviewer,

We thank you very much for all your valuable comments on our manuscript. We are aware that this review has been a significant time investment and therefore especially appreciate your feedback and commitment. During this first revision process, we have addressed all the comments, and the vast majority of them will lead to additions or clarifications in the text. Please find below a detailed answer to all the raised points. We hope that these answers along with the respective proposed modifications of the manuscript will provide the requested level of clarification.

Best regards,

Francesca Carletti, on behalf of the author team.

Reviewer comments are repeated in *gray italic*, author replies in black regular font.

I enjoyed reading the manuscript on snow model comparison and climate change impact assessment in different catchments.

I have several major concerns:

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

The question is appropriate and we probably should have specified it in Section 2.6.1.
For the model Alpine3D over the Dischma catchment, the computational resolution was chosen referring to the work of Schlögl et al., 2016[1]. In this paper, the authors test the effects of Alpine3D input variation on Snow Water Equivalent (SWE) quantification, and a big effort is spent on testing different horizontal DEM resolutions. The authors selected four different resolutions (25, 200, 500, 1000 m) for the DEM grid and land cover data. Results show that downscaling from a horizontal resolution of 500 m to one of 25 m, the relative difference in SWE decreases by only 3% approximately. Considering such findings, we decided that this simplification would have been acceptable for the scopes of our paper. Besides, the focus of our paper is finally the estimation of the discharge, and the benefits of a slightly more accurate SWE quantification risk to be lost in the flow routing process – especially during the calibration. In addition, Alpine3D being a complex model, a 100 m resolution over the ~5 times larger Mera catchment, i.e., a computational cost multiplied by 25, is technically not doable.

For consistency, a resolution of 500 m was kept in Poli-Hydro as well. Given that we had no previous similar sensitivity studies over the Mera, we took your advice and repeated the calibration with a resolution of 100 m using the model Poli-Hydro. This has not led to significant improvements in NSE and PBIAS, in the face of considerably higher computational times (See Table 1).

[1] Schlögl, S., Marty, C., Bavay, M., and Lehning, M.: Sensitivity of Alpine3D modeled snow cover to modifications in DEM resolution, station coverage and meteorological input quantities, Environmental Modelling & Software, 83, 387–396, https://doi.org/https://doi.org/10.1016/j.envsoft.2016.02.017, https://www.sciencedirect.com/science/article/pii/S1364815216300378, 2016.

Table 1: Sensitivity analysis on calibration scores over Mera catchment with Poli-Hydro.

	Res = 500 m			Res = 100 m		
	NSE	**PBIAS**	**Time**	**NSE**	**PBIAS**	**Time**
Mera	0.67	+5.90%	2 min	0.65	-4.29%	5 hrs

Another key point is that, as explained throughout our paper, the Mera catchment is highly exploited by human activities, which brings additional complexity to discharge simulation and risks to overshadow the improved accuracy brought by a finer DEM resolution.

2) 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[2,3,4]. 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.
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 gives satisfactory results, especially when tuned with snow cover/depth data[5].

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 (from high-altitude, heavily cryospheric conditions, to low-altitude, arid or semi-arid areas, with or without snow/ice contributions) and over catchments of largely varying size (from ~10 to ~10000 km2), with satisfactory accuracy in reproducing stream flows, snow/ice dynamics and cryospheric contributions[6,7].

---

[2] Beven, K 1989, 'Changing ideas in hydrology - The case of physically-based models', *Journal of Hydrology*, vol. 105, no. 1-2, pp. 157-172. https://doi.org/10.1016/0022-1694(89)90101-7
[3] Beven, Keith J.. “Prophecy, reality and uncertainty in distributed hydrological modelling.” *Advances in Water Resources* 16 (1993): 41-51.
[4] Beven, K. J. (2006) *A Manifesto for the Equifinality Thesis.* Journal of Hydrology, 320 (1-2). pp. 18-36.
[5] Soncini, Andrea & Bocchiola, Daniele & Azzoni, Roberto & Diolaiuti, Guglielmina. (2017). A methodology for monitoring and modeling of high altitude Alpine catchments. Progress in Physical Geography. 41. 030913331771083. 10.1177/0309133317710832.
[6] Casale, Francesca & Fuso, Flavia & Giuliani, Matteo & Castelletti, Andrea & Bocchiola, Daniele. (2021). Exploring future vulnerabilities of subalpine Italian regulated lakes under different climate scenarios: bottom-up vs top-down and CMIP5 vs CMIP6. Journal of Hydrology: Regional Studies. 38. 100973. 10.1016/j.ejrh.2021.100973.
[7] Michel, A., Schaefli, B., Wever, N., Zekollari, H., Lehning, M., and Huwald, H.: Future water temperature of rivers in Switzerland under climate change investigated with physics-based models, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2021-194, in review, 2021.

3) 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 4.1.2 as, in model Poli-Hydro, the validation of the modelled Snow Cover Area against MODIS[8] 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[9] for further details concerning the calibration of snow degree-days. In Fuso et al., 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 (SWE), which cannot be derived from satellite imagery, at least for the time being.

-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, the model Poli-Hydro is fed with a set of fixed calibration parameters obtained from previous studies over the same areas, whereas another set is calibrated (P16L7). To feed the model with plausible parameters verified in previous studies is a common practice in operational hydrology to ease the multiobjective

---

[8] Snow Cover Extent | Copernicus Global Land Service. Available online: https://land.copernicus.eu/global/products/sce

[9] Fuso, F.; Casale, F.; Giudici, F.; Bocchiola, D. Future Hydrology of the Cryospheric Driven Lake Como Catchment in Italy under Climate Change Scenarios. *Climate* 2021, *9*, 8. https://doi.org/10.3390/cli9010008

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.
On the other hand, the 10’000 model runs mentioned in P15L7 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 is potential to better detail the model calibration procedure, and we will try our best to improve it during the review process. With respect to the model Poli-Hydro, Section 4.1.2 will be enriched with a more detailed calibration description, more carefully divided into a Snow Module (new Section 4.1.2.1) and a Flow Propagation Module (new Section 4.1.2.2). In the new Section 4.1.2.2, we will also explain and justify the choice of previously calibrated physically based parameters.

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 is not clearly specified in Section 4.1.1, undoubtedly leading to confusion or lack of clarity in this respect. We will do our best to cover this lack during the revision process by better detailing Section 4.1.1 into a Snow Module (new Section 4.1.1.1) and a Flow Propagation Module (new Section 4.1.1.2), in line with the aforementioned improvements proposed for the calibration section of the model Poli-Hydro.

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 Berezowski et al., 2015[10], snow melt (i.e., the most relevant contribution to discharge generation in partly-glacierized catchments) is governed by temperature and radiation (see Section 3.1.3 in our paper). 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 snowmelt 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.

[10] Berezowski, T., Nossent, J., Chormański, J., and Batelaan, O.: Spatial sensitivity analysis of snow cover data in a distributed rainfall-runoff model, Hydrol. Earth Syst. Sci., 19, 1887–1904, https://doi.org/10.5194/hess-19-1887-2015, 2015.

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. We will move the calibration part (Section 4.1) to the Methods (Section 3). A separate section focused on calibration results will be then moved to the Results (Section 4).