Reviewer 1

We want to thank Reviewer 1 for her/his constructive comments. Our answers are written in italic letters alongside to the reviewer's comments.

Summary

The authors use a model chain consisting of climate models, a weather generator, and an energy balance snow model to identify dominant uncertainty sources in future changes in snow-water-equivalent and rain-on-snow runoff. They show that changes in ROS events emerge till the end of the century despite large uncertainties while ROS events with substantial snowmelt contributions don't show a clear change signal.

General remarks

The study by Schirmer et al. builds on a complex model chain consisting of climate models, a weather generator and an energy balance snow model to assess the importance of internal variability on the detection of future changes in snow and rain-on-snow runoff events. I think that the combination of different model types to better describe internal variability is a generally a valid approach to determine the importance of internal variability in change assessments of snow-related quantities compared to other uncertainty sources. However, I see a substantial need for clarification regarding the research questions and methodology and think that the approach chosen to decompose uncertainty into different contributors needs refinement. Given the current 'incomplete' methods descriptions, it is difficult to assess the validity of the results. Furthermore, I think that the manuscript would profit from reorganization, i.e. restructuring the methods section following a more logical sequence and from separating the results from the discussion. Finally, the manuscript would in my opinion profit from a visualization of the most important modeling steps and their relationships and from refining figures by adapting color schemes and adding legends. Please find my more detailed review below.

Thank you for your assessment. We agree that the method description needs to be expanded and that the manuscript would benefit from a modelling flowchart, a more chronological sequence with a clearer separation of sections, and a refinement of the figures. For more details, please see our specific responses to your comments below.

Major points

1. Research questions: the research questions are not entirely clear and should be explicitly stated in the introduction. From how I understand the study it is something along the lines of: 'How does the importance of internal variability differ between temperature-driven snow resources and rain-driven rain-on-snow events' and 'When is the time of emergence of changes in snow availability and rain-on-snow runoff.'

We will rework the final paragraph of the introduction to ensure that the research questions are concisely formulated.

2. Introduction: In addition to model-based studies looking at changes in rain-on-snow floods, there are also observation-based studies, which I think should be mentioned in the introduction. E.g. Sikorska and Seibert (2020; 10.1080/02626667.2020.1749761) or Cheggwidden et al. (2020; 10.1088/1748-9326/ab986f).

We agree and add these studies in the introduction.

3. Methods section organization: The methods section does not seem to follow a logical order and could in my opinion be more logically organized by following a ‘chronological’ modeling order. E.g. Area, Climate models, Weather generator, Snow model, Rain-on-snow definition, Change assessment, Uncertainty decomposition.

Our current methods section is structured so that the most important methods for this manuscript are described first (i.e. the weather generator and the snow model). However, we can certainly put the methods in a more chronological order, as suggested, to clarify the methodological flow.

Providing a flowchart linking the most important modeling and analysis steps might enable further improvements in communication. Furthermore, the methods section lacks important methodological detail, which makes it difficult to assess the validity of the results.

We agree that adding a flowchart as well providing more details in the methods section will improve the manuscript, which we will implement as suggested.

4. Glacier retreat: The study region is influenced by a glacier, which affects runoff formation. However, the glacier-related changes in flow are not represented in the modeling chain (l. 86-87). This does not seem to be justified and might explain why melt-influenced changes in ROS events are don’t show up clearly.

Yes, this study region is influenced by a large glacier. However, we do not want to investigate the combined effects of snow and glacier retreat on mean snow water resources or ROS properties, but only to show the effects of snow. For this reason, we have not compared observed runoff measurements in the area with our modelled data. We chose this catchment because of its good data availability and high elevation. The study area is treated as if there were no glacier. Only the altitudes and the meteorological input data play a role for the study area. It can therefore be considered rather as an example area, perhaps even a virtual area, for which the influence of natural climate variability on snow water resources and ROS events is studied as an example. These considerations will be presented in more detail in the next version of the manuscript.

5. Weather generator: The weather generator description (Section 2.3) lacks important detail and it is therefore difficult to assess the validity of the approach. E.g. how does the weather generator use the climate simulations, how does the weather generator work, how is the temporal downscaling performed (l. 122), how are the different variables generated (l. 122), how is the inter-variable consistence (l.124) evaluated?

There are two comprehensive publications, Peleg et al. (2017, 2019), which we cite in this manuscript to answer these questions. We understand that a manuscript should stand on its own, but we also had to consider the readability of this already quite long manuscript. This is especially true since this study is an application to the weather generator among other methods. We have also not gone into detail on energy balance snow modelling or climate modelling. Given the complexity of, for example, how the weather generator works in detail, we have chosen to provide references to these issues, but at the same time to show in broad terms how the weather generator was trained and applied. We make sure that the references for these important questions are clearly visible, but to address you concern will also include more details from cited publications.

6. Snow model and variables: It remains unclear to me how the weather generator output is used to derive different snow-related variables (Section 2.2). Was the analysis performed per grid cell? Which variables were exactly derived? How was the model calibrated (l.95)? And what does the ‘unpublished’ model adjustment (l. 95-96) do?

In the more chronological order of the methods section suggested above we will include more details answering these questions. For now, the results achieved with the snow model are mean values of grid cells, either within the whole catchment or within all ROS affected pixel. This is described in each results section and will be placed also in the method section.

The model is an open-source energy balance snow model, which is publicly available. These types of models are usually not calibrated, and we did not calibrate our model as well. We will also provide more details on the adjustments to the model.

7. Bias correction: Section 2.5. suggests that some bias correction might have been necessary to adjust simulated to observed values. Was such bias correction performed and if so why?

Yes, the climate model chains were bias corrected and downscaled with quantile mapping (see section 2.4). We will state this more clearly, and also why we have chosen to use bias corrected climate model output.

8. Uncertainty partitioning: The uncertainty partitioning procedure described in Section 2.6. does not seem to properly separate internal variability (residuals) from the signal. Or at least I can not see how the different uncertainty components have been decomposed e.g. using a procedure such as the one proposed by Hawkins and Sutton (2009; 10.1175/2009BAMS2607.1). The procedure used to derive climate model uncertainty also seems to encompass internal variability (l.153-154) and the procedure used to derive internal variability also seems to include climate model uncertainty (l.156-157). Furthermore, it would be nice to compute fractional uncertainty contributions that add up to 1, which currently does not seem to be the case.

We follow the uncertainty partition method described by Fatichi et al. (2016), which is significantly different from other partition methods (such as the one described by Hawkins and Sutton that the reviewer mentioned). In our case, the reason to use the first method and not the second is that it is more suitable for the situation where you have the complete internal climate variability - model uncertainty - scenario uncertainty chain (as obtained from the AWE-GEN-2d model). Using Fatichi et al.'s method, the fractional uncertainty contribution cannot be summed to 1 (see their paper for more details). In light of the reviewer comment (and a similar comment given by the second reviewer), we see the need to explain the method and its differences from other uncertainty partition methods in more detail. Accordingly, the relevant section of the manuscript will be extended.

9. Validation: I think that the methods section needs a 'Validation' subsection describing how the different models were evaluated. E.g. how were the validation stations chosen? Which variables were validated, …

We will add a validation subsection in the methods section in order to be clearer on this topic.

10. Rain-on-snow events: how have these events been defined? There is a section called 'rain-on-snow' definition, which does, however, not really explain what you understand by a 'rain-on-snow' event. How is the 'surface water input' computed?

We do not define ROS events, but apply daily pixel-based criteria that then result in a certain size of "contributing area" each day. An "ROS day" can then be defined as a day with a contributing area of a certain size, which may depend on the application or the user. We determined a climate change signal of ROS frequency with variable event sizes (x-axis of Figure 7). The intensity and contribution of snowmelt required a certain size of event (> 1/3 of the total area affected according to pixel-based criteria). Since the reviewers have pointed out that methods and results should be more clearly separated, we will better organize these details and expand this topic.

The "surface water input" is calculated with the energy balance snow model and is the water input available at the ground surface through either snowpack runoff, or rain in case of snow-free conditions, or a mixture of both in case of fractional snowcover. We will add this definition in the methods section.

11. Results: I would clearly separate the results part from the methods section and discussion. Some parts can be moved from the Results to the Methods section (e.g. l. 206-216) and other parts to a newly created Discussion section (essentially everything that compares the study's findings to findings of existing studies). Furthermore, it would be nice if the results section followed a similar structure as the methods section.

As already written above, we will add more details in the methods section, which will also help to separate results from methods. We thank the reviewer for this suggestion.

12. Figure 11: Would be nice to depict the fractional contributions of the individual uncertainty sources. Currently, all of the sources seem to not be clearly separated (probably also related to the issue risen in comment 8).

Yearly mean fractional contributions are summarized in Figure 12. We think that including all daily fractional uncertainties also in Figure 11 would look to busy.

13. Discussion: I miss a discussion of alternative strategies that could be used to quantify the relative importance of internal variability to other uncertainty sources besides weather generators. Recently, quite a few studies have used single model initialized large ensembles (SMILES) for uncertainty decomposition and I think it would be important to mention this (e.g. Maher et al. 2021; 10.5194/esd-12-401-2021 or Lehner et al. 2020; 10.5194/esd-11-491-2020). Other topics I would address in the discussion section include: model deficiencies, comparisons of results with other studies (as distributed throughout the results section), and the generalizability of the findings to other geographical and climatic contexts given that the study relies on one catchment only.

We will add a discussion on other ways to quantify natural climate variability, along with the suggested publications. We note, however, that ensembles derived from other climate models (SMILES, for example) have much coarser spatial and temporal resolutions than ours. Finding other studies that partition the uncertainties for the processes we are simulating at fine scales is therefore not easy. The findings will also be discussed with regard to model deficiencies and generalizability.

14. Figures: Figure design should be improved by using continuous color schemes for continuous variables (e.g. Figure 1c) and by avoiding the use of rainbow color schemes, which are not color-blindness friendly (Figure 8a and 8b).

We agree. As suggested by the reviewer, we will replace the color scheme in Figure 1c with a strict continuous one and for Figure 8 a more color-blindness friendly color scheme.

Furthermore, complete legends should be provided for all figures.

We are not quite sure what is missing. We agree that some figures of the kind like Figure 2, 4, 6 etc. do not have legends. However, all the relevant information is in the caption, which gives the period, the range and the underlying data for the calculation of the range. This would certainly be too much text for a legend. However, we will add simplifications of all the information in a legend (e.g. red: future climate, blue: current climate).

Minor points

L. l. 40-41: I don't think that it is correct to say that 'no studies have previously included internal climate variability in their analysis' as internal climate variability is per definition part of every change impact assessment. However, it might be ok to say that 'the relative importance of internal variability compared to other uncertainty sources has not been previously assessed.'

We will follow the suggestions by the reviewer.

L.58-59: needs rephrasing.

We will rephrase this sentence.

L. 61: 'they' will

Many thanks for finding this error. We will change as suggested.

L. 63: that 'drive runoff'

Many thanks for finding this error. We will change as suggested.

L. 72-73: repeats content already provided in l. 68-69 and can be removed.

Many thanks for finding this error. We will change as suggested.

L. 140: How do you define 'frequency', number of events per year or non-exceedance probability or anything else?

We will add a definition of frequency here, i.e. the number of events per year dependent on an event size larger than a certain threshold, which is equivalent to the yearly exceedance probability.

L. 259-260: Not sure whether the statement that 'the uncertainty range for current climate is, by definition, only determined by natural variability' is true. If the simulations for the current period are run with different climate models (which is as I understand it the case), climate model uncertainty might also be present.

We will clarify in the methods section that the current climate period is NOT affected by climate model uncertainty in our model setup.

L. 279-280: Don't understand this sentence and think it needs rephrasing.

We will rephrase this sentence.

L. 293: Sentence seems incomplete. Would be nice to repeat the pixel-based criteria, which the reader might no longer have present at this stage.

We will do as suggested.