**Response to the Editor**

*Comment: We received the three reviewers' comments that all suggested only minor revisions. Your replies to their comments allow us to proceed with the process towards the revised manuscript. Please, incorporate new material as suggested in the interactive discussion and in your replies to their comments/suggestions/questions.*

*The revised manuscript will be checked by the Editor only. You may also change the title as already discussed with the reviewers.*

Response:

Thank you for the positive assessment and the handling of our manuscript. We incorporated the reviewers' comments as suggested in our replies to the reviewers' comments. The following point-by-point reply to the comments shows in detail which parts were changed/added and the corresponding lines in the revised manuscript. Furthermore, the manuscript was checked for typos etc.

**Point-to-point reply to Reviewers**

Author response to RC1:

We are grateful for the detailed and critical review by Reviewer #1 and we highly appreciate the generally positive assessment of our manuscript. In the following we address the general and specific comments raised and the related changes in the revised manuscript.

*Comment: I appreciated the fact that many different runoff signatures are discussed (actually the title could state that not only extremes are considered) and the results are compared to the existing literature.*

Response: We appreciate this positive feedback. In the title, we wanted to highlight the analysis of annual extremes as this is mostly lacking in previous studies. However, we acknowledge that the title does not span the variety of analysis addressed in the paper. We adapted the title: "Future changes in annual, seasonal and monthly runoff signatures in contrasting Alpine catchments in Austria:"

*Comment: CMIP5 climate projections are used here, which will be soon become obsolete given the CMIP6 simulations already available (the Authors should also discuss this limitation)*

Response: As highlighted in the reply to Reviewer #3, to date CMIP6 simulations are not available as downscaled projections to regional scale using RCMs. The scale of this study is too small to reliably use GCM outputs as climate inputs to our study catchments. However, it is interesting to indicate that as soon as they are available new simulations should be used for similar studies to assess differences in outcomes. We therefore added to Section 4.7: "A new set of GCM simulations is available (CMIP6, Eyring et al., 2016). However, these could not be used in this study due to the importance of coupling the GCM-RCM simulations, which will become available for CMIP6 in future." (l.550-552)

*Comment; I also would have expected a more critical questioning on how well this modelling approach can capture runoff changes (e.g. based on the discussions in Duethmann et al., 2020, HESS).*

Response: Thanks for this comment. We considered and discussed the modelling approach with regard to transient climate. However, it obviously did not become clear enough in the manuscript. Duethmann et al. (2020) found two main causes for the inability of the HBV model to represent runoff in a changing climate: inhomogeneities in precipitation data due to variable number of stations and variations in vegetation dynamics that were not considered. With respect to the first cause, and with the deliberate intention to avoid this limitation, our modelling approach relies on the same set for stations during calibration and evaluation. For EURO-CORDEX simulations the gridded data set was downscaled to the location of these ground stations. For detecting changes in runoff due to climate change, we compare the EURO-CORDEX simulations of the past and the future. Therefore, inhomogeneities in precipitation data do not affect the results. To further assess and illustrate how well the past EURO-CORDEX simulations represent the past climatic and runoff signatures, they were compared to simulations using observations and to observed data (Section 3.2, Fig. 6; Fig. S13-17). The analysis suggests that, overall, the simulations match the runoff signatures relatively well. We further elaborated this section (3.2) in the revised manuscript.

With respect to vegetation dynamics, we acknowledge that feedback of climate change on vegetation dynamics and thus, possibly increased evapotranspiration was not considered. As mentioned by Duethmann et al. (2020), assessing changes in future vegetation dynamics is difficult due to missing future information. Therefore, changes in growing season and land use change remain very uncertain in future. To prevent introducing additional uncertainty in our modelling approach these effects were omitted. This is acknowledged and discussed in the original manuscript (l.520-525). In the revised version we extend this discussion to also include the suggested reference as well as perspectives on the effects of increased temperatures and rising $CO_2$ levels (fertilization effect) on plant transpiration, although the effects are not yet fully understood (e.g., Frank et al., 2015).

"However, in reality parameters such as maximum storage capacity in the unsaturated root zone can change due to for instance vegetation adaptation to changing climate. Moreover, the partitioning of precipitation will likely be affected by changes in vegetation dynamics, such as the likely extension of the growing season, and can significantly affect changes in runoff (Duethmann et al., 2020). Nonetheless, changes in vegetation dynamics due to climate change are not considered in this study due to a lack of understanding of the overall effect (e.g., Frank et al., 2015)." (l.556 ff.)

Due to the known difficulties of models to reproduce changes in transient climate, we intended to make the representation of hydrological process and the model internal dynamics of and feedbacks between these processes as robust as possible to avoid major misrepresentations of the system. To ensure that, we calibrated and evaluated the model simultaneously to eight individual objective functions describing different signatures of flow (Section 2.2.2). In addition, we performed a long-term calibration of 20 years and a model evaluation of 8 to 10 years and we used 300 parameter sets in an ensemble analysis to account for parameter uncertainty. Furthermore, we considered future changes in glacier extent. With this approach we aim to ensure that our modelling approach can capture runoff changes. However, the uncertainties in model structure remains and was not further assessed, although model structure uncertainty can be significant (e.g. Bouaziz et al., 2021; Knoben et al., 2020). This was clarified in the revised version l.543 ff.: "In general, different models with different structures are often not consistent in the results (e.g., Knoben et al., 2020) or their internal dynamics (Bouaziz et al., 2021). This uncertainty in model structure was not assessed here and it would be worthwhile to repeat a similar study using another hydrological model."

*Comment: Regarding the attribution of runoff changes to the causes (precipitation, evapotranspiration and snowmelt), a more quantitative analysis would have made the paper even more interesting*

Response: We agree that a more detailed analysis of the causes of runoff changes would be desirable. To attribute changes to specific causes during runoff events, a tracking of the sources of water throughout the hydrological model would be necessary to disentangle the contribution of snowmelt and precipitation.  Implementing a correct source tracking remains very difficult (Weiler et al., 2018) and would need detailed tracer data and considerable adaptation of the model structure to allow for mixing effects so that fluxes can then be individually tracked through the system (e.g. Hrachowitz et al., 2013).

Attributing each change in input (precipitation, evapotranspiration and snowmelt) to a change in runoff is very difficult since changes and processes interact. Therefore, a quantitative separation of effects is not warranted here with the available data. The main objective was to examine the combined effect of changes in precipitation and temperature, which was assessed by our simulations. To support the analysis, figures of monthly changes in precipitation, temperature and potential evaporation were added in the Supplementary Information (Fig. S19-21).

*Comment: Line 85: what could be the bias introduced by scaling the runoff data, rather than the precipitation data, for consistency with Budyko? Since the Authors are interested in percental changes, wouldn't this choice exacerbate the changes in magnitude for runoff in those catchments?*

Response: First of all, we would like to point out that the scaling of runoff data has no effect on the timing of runoff extremes. A scaling factor for the runoff was derived based on the Budyko framework, so that calibration would be feasible. This scaling factor is not applied to modelled runoff of the past and future generated using the EURO-CORDEX simulations. Thus, the absolute runoff is likely underestimated in simulations. However, the relative change in runoff remains unaffected because it is derived using the past and future modelled runoff from EURO-CORDEX simulations. Absolute changes are likely to be higher than our projections, so that our projections are likely a lower limit of absolute change. We explained this in Section 4.7 line 509-510 in the original manuscript but we further clarified and elaborated on that in the revised manuscript: “Due to scaling of runoff for calibration in these catchments, the absolute runoff is likely underestimated in our simulations. Therefore, our simulations likely represent a lower limit of absolute runoff change. However, the relative change in runoff remains unaffected as it is derived using the past and future modelled runoff from EURO-CORDEX simulations.” (l.533 ff.)

*Comment: Line 104: what is the advantage of transferring the precipitation and temperature data from the EUROCODEX pixels to the ground stations and not working with the pixels themselves?*

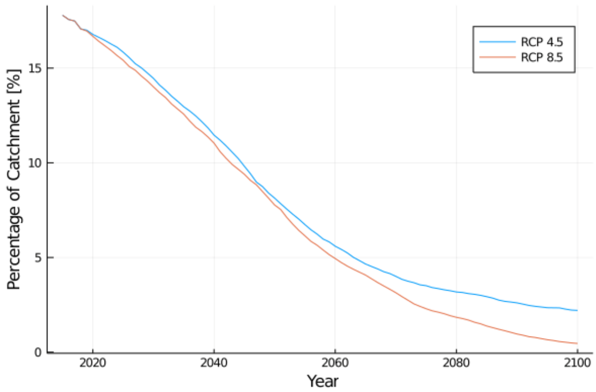
Response: The hydrological model was calibrated using data from ground observation stations for precipitation and temperature by dividing the catchments into different precipitation zones. To ensure as much consistency as possible between the ground data and the EURO-CORDEX data and to avoid a bias introduced by the model parameters that were obtained from calibration to ground data, we decided to transfer the data from the EURO-CORDEX pixels to the ground stations. This was done to keep inhomogeneities in precipitation data (Duethmann et al., 2020) as low as possible, as also discussed above. We clarified the approach and the associated limitations of using point-scale data in Section 4.7 line 546 ff: “Another uncertainty arises from calibrating the model with in situ observed data at point-scale but using projection data for the future. To reduce this limitation, data of the same spatial scale was used, limiting the effect of inhomogeneities in precipitation data on our results.”

*Comment: Line 115: is Table 1 the right table?*

Response: No, the reference is wrong. It was adapted to “(i) one to four precipitation zones per catchment (Fig. 1), (ii) the four HRUs per precipitation zone (Fig. 2) and (iii) individual 200 m elevation zones per HRU (e.g., Roodari et al., 2021).”

*Comment: Line 170: has the glacier extent in the Pitztal a linear decrease till 2100?*

Response: No, the glacier extent is adapted according to results by Zekollari et al. (2019) as mentioned in line 99ff. Below is a figure showing the percentage of the Pitztal catchment covered by glacier, from 2015 to 2100 for RCP 4.5. and RCP 8.5. Differences in glacial extent between the two scenarios are largest at the end of the 21st century.




*Comment: Page 20: the Gailtal catchment seems special, in terms of annual maxima. Apart from the double peak season, what could explain the largest change in magnitude for RCP8.5?*

Response: The Gailtal and Defreggental catchments are located South of the Alpine main ridge and thus under strong influence of moisture circulation from the Mediterranean. However, the Defreggental is located at high elevation so that even under future climate change snow melt will play a role as flood-generating process. The Gailtal catchment is mostly dependent on rainfall as flood generating process as most annual maximum flows occur in autumn. As mentioned in the discussion, timing of floods in southern Austria are strongly influenced by Meridional south-east and south weather regimes (Parajka et al, 2010). To discuss this finding for the Gailtal we added the following in the revised version of the manuscript (l. 472 ff.) “In the Gailtal, the increase in AMF magnitudes is higher under RCP 8.5. A possible explanation is that, particularly under RCP8.5, changes in precipitation intensities or maximum daily precipitation may be higher for the Meridional weather regimes than in the Northern Alps and thus impact the Gailtal, where rainfall is the main flood-generating mechanism.”

*Comment: Line 346: one interesting reference on observed trends in evapotranspiration in Austria and possible causes is Duethmann and Blöschl (2018, HESS).*

Response: Thanks for pointing out this interesting study in Austria, which also identified atmospheric demand as main cause for increase in evapotranspiration in the past 40 years. We added this reference to the statement in line 361: “This slightly lower annual runoff can be attributed to changes in the future partitioning of water fluxes and thus an increased fraction of precipitation to be evaporated due to increased atmospheric demand (cf. Fig. 7a). Increasing atmospheric demand has also been identified as the main driver for increasing evaporation in Austria in the past (Duethmann and Blöschl, 2018).”

*Comment: Section 4: the section contains a lot of attribution statements. Are these statements confirmed by simulations, e.g., increase in precipitation keeping temperature* *unchanged and vice-versa? Or is it just an expert interpretation of the simulations with all variables changing?*

Response: Simulations were only performed with all variables changing. However, mean changes in monthly precipitation, snow melt and potential evaporation were analyzed. Based on this analysis, expert interpretations about possible attributions were made. We clarified this in the revised manuscript (l. 376 ff.): "Simulations were performed by changing all variables simultaneously. The attribution of changes to single variables in the discussion is therefore based on expert interpretations of the results and mean changes in monthly precipitation, temperature, snowmelt and potential evaporation (Fig. S19-21)."

*Comment: Section 4.7: this section could be linked to the hypotheses for potential causes of the divergence between observed and simulated discharge changes in Duethmann et al. (2020, HESS).*

Response: We added a linkage to the section as already explained in the answer to a previous comment. "However, in reality parameters such as maximum storage capacity in the unsaturated root zone can change due to for instance vegetation adaptation to changing climate. Moreover, the partitioning of precipitation will likely be affected by changes in vegetation dynamics, such as the likely extension of the growing season, and can significantly affect changes in runoff (Duethmann et al., 2020). Nonetheless, changes in vegetation dynamics due to climate change are not considered in this study due to a lack of understanding of the overall effect (e.g., Frank et al., 2015)." (l.556 ff.)

Regarding the model structure uncertainty, we would add l.543 ff.: "In general, different models with different structures are often not consistent in the results (e.g., Knoben et al., 2020) or their internal dynamics (Bouaziz et al., 2021). This uncertainty in model structure was not assessed here and it would be worthwhile to repeat a similar study using another hydrological model."


Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., De Niel, J., Dewals, B. J., Drogue, G., Grelier, B., Melsen, L. A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A. H., Willems, P., Savenije, H. H. G., and Hrachowitz, M.: Behind the scenes of streamflow model performance, Hydrol. Earth Syst. Sci., 25, 1069–1095, https://doi.org/10.5194/hess-25-1069-2021, 2021.

Duethmann, D. and Blöschl, G.: Why has catchment evaporation increased in the past 40 years? A data-based study in Austria, Hydrol. Earth Syst. Sci., 22, 5143–5158, https://doi.org/10.5194/hess-22-5143-2018, 2018.

Frank, D., Poulter, B., Saurer, M. et al. Water-use efficiency and transpiration across European forests during the Anthropocene. Nature Clim Change 5, 579–583, https://doi.org/10.1038/nclimate2614, 2015

Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., and Soulsby, C.: What can flux tracking teach us about water age distribution patterns and their temporal dynamics?, Hydrol. Earth Syst. Sci., 17, 533–564, https://doi.org/10.5194/hess-17-533-2013, 2013.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A.: A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, *56*(9), https://doi.org/10.1029/2019WR025975, 2020

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., ... & Blöschl, G.: Seasonal characteristics of flood regimes across the Alpine–Carpathian range. *Journal of hydrology*, *394*(1-2), 78-89, https://doi.org/10.1016/j.jhydrol.2010.05.015, 2010.

Weiler, M., Seibert, J., & Stahl, K.: Magic components—Why quantifying rain, snowmelt, and icemelt in river discharge is not easy. *Hydrological processes*, *32*(1), 160-166, https://doi.org/10.1002/hyp.11361, 2018.

Zekollari, H., Huss, M., and Farinotti, D.: Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble, The Cryosphere, 13, 1125–1146, https://doi.org/10.5194/tc-13-1125-2019, 2019.

Author response to RC2:

We highly appreciate the relevant feedback and the positive assessment of our manuscript by Reviewer #2. In the following we address the specific comments raised and the incorporation into the revised manuscript.

1. *Data: The authors should explain the reason of using Thornthwhaite method (one of the simplest methods) for daily evaporation estimation since the choice of the method affects the results.*

   Response: Due to data restrictions it was not possible to use a potential evaporation method beyond a temperature approach. For deciding on the calculation methods, Thornthwhaite and Hargreaves method (Hargreaves and Samani, 1985) were compared to values from the Hydrological Atlas Austria and the PhD thesis of Kling (2006). Whereas the Hargreaves method overestimated potential evaporation in the Austrian Alps, the Thornthwhaite method was more in line with literature values and was therefore chosen. In the revised manuscript we added "The Thornthwhaite method compared well with published estimates of potential evaporation in the Austrian Alps (BMLFUW, 2007; Kling, 2006), while the Hargreaves method (Hargreaves and Samani, 1985), another temperature-based method, overestimated potential evaporation." (l.91 ff)

2. *Data: Authors could give at least a basic description of both applied scenarios RCP4.5 and RCP8.5 (e.g. moderately optimistic scenario RCP4.5, worst-case climate change scenarios RCP8.5).*

   Response: In the revised manuscript, we added a brief description of the scenarios: "RCP 4.5 is an intermediate pathway where emissions are partly reduced, yielding 4.5 Wm-2 radiative forcing by the year 2100. RCP 8.5 represents a pathway with increasing greenhouse gas emissions and no mitigation measures." (l. 109ff)

3. *Methods: Please, elaborate the decision for the selection of the period 2071-2100 and not the period 2041-2070 for the investigation of the influence of climate change on future runoff extremes. The uncertainty of the results for such a distant future (2071-2100) is much greater and therefore confidence in the results is much lower.*

   Response: Although uncertainties in distant future are larger, the effects of climate change are more pronounced and can thus be more easily detected. To cope with uncertainties, an ensemble of projections is used and uncertainty ranges are given in the results.

   In the revision, we added: "Although uncertainties are larger at the end of the century than for the mid-century, the stronger climate change signal enhances detection of potential changes due to climate change." (l.166 ff)

4. *Results, Figure 5 caption: It could be useful for readers to add colours to the explanation in the brackets: "Red and grey lines represent the mean flow regime….. "*

Response: Agreed. This was added.

5. *Results: Section 3.3.2: I would suggest to change the order of the words in the title (timing and magnitude) to follow the structure of the section. The same goes to some other sections (i.e. 3.3.3, 4.3, 4.4)*

   Response: Agreed. We changed the title of Section 3.3.2 and 4.3 to "Annual maxima (timing and magnitude)" and of Section 3.3.3 and 4.4 to "Annual minima (timing and magnitude)".

6. *Discussion, lines 339-340: The statement that the increase in projected future precipitation compared to the past is in contrast to other climate projections for Austria is not elaborated enough. What are the possible reasons? Did other authors (e.g. Stanzel and Nachtnebel, 2010) use the same scenarios for the same catchments/regions?*

   Response: This is indeed a very interesting point. Stanzel and Nachtnebel (2010) used A1B, A2 and B1 scenarios with the REMO-UBA RCM for entire Austria. These are different scenarios and models than the ones used in this study. This can be a possible reason. Another reason might be that previous studies focused on other catchments. Unfortunately, all studies that reported mean changes in precipitation in Austria where not run with the EURO-CORDEX ensemble. However, Smiatek et al. (2016) performed a climate analysis of the EURO-CORDEX ensemble for the Alpine region and found that the majority of simulations indicate summer and winter season precipitation increase, which would be in line with the projections used in our studies. They also highlight the uncertainty that remains regarding precipitation projections in the Alpine region. Therefore, the reason for the differences is likely the differences in GCM/RCM combinations used for the studies. To clarify this, we extended the statement to:
   "The increase in projected future precipitation in Austria is in contrast to results of previous studies that are not based on the EURO-CORDEX ensembles, which suggested no change or a decrease in precipitation (Stanzel and Nachtnebel, 2010; Goler et al., 2016). However, our findings are consistent with the results of an analysis of the EURO-CORDEX ensemble for the Alpine region as reported by Smiatek et al. (2016)."

7. *Discussion: In my opinion the section 4.5 is redundant. The societal impacts were not evaluated in the research; therefore, the discussion about the topic is not relevant.*

   Response: The aim of this section was to give a brief overview of the possible societal relevance related to the results of this study. However, we agree that the societal impacts were not explicitly evaluated. We would nevertheless find it important to provide perspectives on the potential relevance of the findings. We therefore removed this section, but to briefly discuss the implications of the changes in the individual indicators in the associated sections 4.1-4.4:

   Line 354: "The increase in annual runoff in future, may have a positive impact on hydropower generation. Nevertheless, seasonal changes can lead to decreased energy production in summer and autumn and increased energy production in winter and spring. Management schemes of hydropower production may need to be adapted to such changing seasonal water availabilities, which could potentially be realized by storing seasonal melt water in artificial basins (Farinotti et al., 2019).

Adaptation measures are likely to be higher for RCP 8.5 due to larger seasonal changes." (l. 371 ff.)

Line 385: “The changes in monthly runoff could lead to a mismatch between water supply and water demand as mountain regions of the Alps are classified as supportive for the lowlands (Viviroli et al., 2007). However, the Alps are identified as basins where present water demands can also be met in 2060 (Mankin et al., 2017). Therefore, water scarcity due to changes in runoff dynamics in the Alps seems unlikely (Immerzeel et al., 2020).” (l.410 ff.)

"This leads to less predictability in the timing of future flood events." (l. 441)

"The increase of magnitudes of maximum flows may locally entail the need to carefully review flood risk assessments and safety of hydraulic structures designed for lower flood estimates." (l. 487 ff.)

8. *Discussion: Section 4.6 (Climate model uncertainty) should be a part of the next section Uncertainty & limitations, where all other uncertainties are discussed.*

   Response: We agree that Section 4.6 also deals with uncertainties. We separated the two sections as the uncertainty in climate modelling chains was evaluated by analysis, whereas the uncertainties and limitations in Section 4.7 where not explicitly assessed. We therefore kept the two sections separate. However, to avoid confusion about the content of each section, we renamed Section 4.7 “Caveats & limitations”.

9. *Discussion, line 495: “….in the largest increases in magnitudes”. Of what? Annual minimum and maximum flows?*

   Response: Yes, added “in the largest increases in magnitudes of annual minimum and maximum flows” for clarification.

10. *Discussion, line 500: It would be useful for readers to define the model 10 more precisely.*

    Response: We added “model 10 (HadGEM2-ES r1i1p1 CCLM4-8-17)” for clarification

The technical corrections were addressed.

Hargreaves, G. H. and Samani, Z. A. (1985): Reference Crop Evapotranspiration from Temperature, Appl. Eng. Agric., 1, 96–99, https://doi.org/10.13031/2013.26773

Kling, H. (2006). *Spatio-temporal Modelling of the Water Balance of Austria*.

Smiatek, G., Kunstmann, H., and Senatore, A. (2016), EURO-CORDEX regional climate model analysis for the Greater Alpine Region: Performance and expected future change, *J. Geophys. Res. Atmos.*, 121, 7710– 7728, https://doi.org/10.1002/2015JD024727

Author response to RC3:

We highly appreciate the detailed feedback on and positive assessment of our manuscript. Below we provide detailed replies to the individual comments and the incorporation into the revised manuscript.

*Comment:* ***Areal meteorological inputs*** *There are several limitations concerning meteorological forcings. This is briefly discussed at l. 502-208 but it could be discussed earlier in the text. First, the typical problem with hydrological applications in mountainous areas is that weather stations are mostly located in plains, typically below 1000 m, while most of the area covered by the catchments is above. In addition to the fact that point measurements in space can misrepresent areal values, the problem is that there is generally a strong relationship between precipitation (and temperature of course) and altitude (see section 3.2 in Ménégoz et al., 2020), these altitudinal gradients being also dependent on the meteorological situations (Gottardi et al., 2012). Reanalysis datasets provided on a regular grid usually take these gradients into account, and the same kind of gradients could be applied to your interpolated data. Without this kind of corrections, I do not see how a correct water balance can be obtained. Could the authors comment on that point?*

Response: Yes, we completely agree that the correct representation of precipitation in mountainous areas remains difficult. It is true that the precipitation stations used for calibration are below the mean catchment elevation. However, the use of global (elevation) correction factors remains similarly problematic, as these can be spatially temporally very dynamic (and thus very uncertain). In any case, in most of the study catchments, the data provide a long-term water balance that is broadly closing and thus was assumed to be plausible, as shown in Figure 7a in the original manuscript. In the catchment where this was not the case, we applied a lumped scaling factor to close the long-term water balance. To refer to the limitations of meteorological forcing earlier in the text, we added a sentence to Section 2.1 “As shown in Fig. 1, precipitation stations are located in the valleys of the catchments at elevations below the mean catchment elevation. However, in most of the study catchments, the data provide a long-term water balance that is broadly closing and thus was assumed to be plausible.” (l. 84 ff.)

*Comment:* ***Bias-correction*** *It is very briefly mentioned at l. 104 that the climate simulations are bias-corrected using scaled distribution mapping. I would appreciate more details about the method proposed by Switanek et al. (2017) and applied in this study. For example, what is the distribution applied to the positive observed precipitation values? Is it a gamma distribution? It is not clear to me what we can expect concerning the correction of extreme values either. Looking at Figure 6, I was puzzled by the mismatch between observed and monthly runoff when climate simulations are used as inputs. It is acknowledged at l. 216 that there could be an “underestimation of temperature in these catchments in the climate simulations”. I understand that the bias-correction is not performing very well then, is that correct? If it is the case, I think it should be discussed in more depth.*

Response: The distribution applied by Switanek et al. (2017) is a gamma distribution (added in line 108). It is possible that another distribution could perform better for the extremes. However, this study is interested in examining both low and high flows. The gamma distribution can be used to bias correct across all values. If the extremes were poorly bias-corrected, where there is some systematic over- or underestimation, then this would be reflected in how well the observed skewness is represented. Switanek et al. (2017) show that the skewness is pretty well modeled through scaled distribution mapping, at least much better than with standard quantile mapping approaches. Bias correction was performed over 1961-2010 but the comparison between observations and simulations was made over a shorter time period (Fig. 6, S13-S17):

For the simulations with climate simulations, the runoff data used is the same as the model period (1981-2010). However, measured runoff is mostly available only from 1986 onwards, so the observed data as well as modelled data forced with observations spans the period

1986-2010. For precipitation and temperature data, a 30-year time period with available observational data was used (1983-2012 for most catchments). In the revision, we aligned time periods in the comparisons of Figure 6, S13-S17 to 1986-2010 to be consistent and eliminate this as potential cause for the observed mismatches. As expected this alignment did not change the results of the comparison. Since the RCMs do not align in time with observations, any sub period will invariably be somewhat different from the observed distribution. The difference in mean annual temperature between simulations and observations in the past is much lower than the difference in mean annual temperature between past and future simulations. This study focuses on the projected changes, so even if the model slightly under- or over-predicts precipitation or temperature in the calibration period, we expect the results still to be valid, because the difference between past observations and simulations are smaller than projected future changes. This has been clarified in the revision Section 3.2 l. 221 ff.

As can be seen in Figure 6, S16, S17 in the top right plot, the climate simulations tend to underestimate monthly mean temperatures in high elevation catchments in spring and summer. This may partly explain the mismatch between observed and monthly runoffs simulated with climate simulations due to a later onset of the melt season, in particular at higher elevations. This is acknowledged in Section 4.7 l.517 and has been further clarified in the revised manuscript:

"This is likely to be related to the underestimation of temperature in these catchments in the climate simulations, which delays runoff due to later snowmelt (Fig. 6). Since bias correction was performed over a longer time period (1961-2010) than this comparison (Fig. 6; Fig. S16-17) and since RCMs do not align in time with observations, any subperiod will invariably somewhat differ from the observed distribution." (l. 227 ff.)

*Comment: **Climate model uncertainty** Section 4.6, dedicated to climate model uncertainty, could be improved. First, as indicated in Table 2 of the manuscript, different GCM / RCM combinations are used in EURO-CORDEX. However, at l. 493-495, it seems that these pairs of climate models are considered as different models (e.g. "model 10"). It must be understood that the different GCMs and RCMs have their own structure, parametrization and, as a consequence, effects on the simulated variables. It is well described in papers dedicated to the partitioning of the different uncertainties (Déqué et al., 2012, Christensen and Kjellström, 2020). The study by Evin et al., 2021 clearly shows the individual effects of each GCM and RCM on the mean seasonal changes of precipitation and temperature in EURO-CORDEX ensembles (my apologies for citing my own work).*

Response: We thank the reviewer for pointing out the wrong reference to the model pairs as "models". We revised it and refer to it as "GCM/RCM combination" in the revised manuscript. We acknowledge that that GCMs and RCMs introduce different uncertainties which are combined when using a combination of GCM-RCM. The aim of Section 4.6 was to briefly describe whether certain GCM/RCM combinations are responsible for the most extreme changes across all catchments, to assess whether largest changes in runoff can be attributed to a specific climate input used or whether these changes cannot be easily attributed to a specific GCM/RCM combination (which was the case). To make clear that an assessment of the individual uncertainties of each GCM and RCM to the results was not realized due to its difficulty, as also mentioned in Evin et al. (2021), we acknowledged this in the revised manuscript: "Assessing the individual uncertainties of GCMs and RCMs used, may yield different results (e.g., Evin et al., 2021), but was not performed here as differences are expected to be small compared to the combined GCM-RCM assessment approach that we opted for."

*Comment:* ***Other uncertainties*** *In section 4.7, other types of uncertainties could be discussed. The hydrological model can have a huge impact and the bias-correction / downscaling methods can also have an important influence (Lafaysse et al., 2014).*

Response: Thanks for this comment. To address it, we added the following "In general, different models with different structures are often not consistent in the results (e.g., Knoben et al., 2020) or their internal dynamics (Bouaziz et al., 2021). This uncertainty in model structure was not assessed here and it would be worthwhile to repeat a similar study using another hydrological model." (l.543 ff.)

We will further add: "Another source for uncertainty is the bias-correction method applied to the climate simulation data. Although bias-correction certainly improves RCM, the choice of the bias-correction method can impact the results (Teutschbein & Seibert, 2012)." (l.552 ff.)

*Comment:* ***Climate projections*** *In the discussion, I think it could be interesting to indicate that CMIP6 simulations are now available but cannot be used for this kind of applications considering that GCM outputs are particularly misrepresented in mountainous areas (I must contradict reviewer #1 here). CMIP6 simulations will probably be downscaled dynamically in the next few years and RCMs represent a real added-value in these areas (Rummukainen 2016). In addition, a few RCMs are now able to represent convective processes and are expected to improve the representation of the precipitation in future climate projections (e.g. CNRM-AROME, Fumière et al., 2020), in particular the "localized convective high-intensity summer rainstorms" indicated at l. 505.*

Response: Thanks for pointing this out. We think it fits well in Section 4.7. and therefore added: "A new set of GCM simulations is available (CMIP6, Eyring et al., 2016). However, these could not be used in this study due to the importance of coupling the GCM-RCM simulations, which will become available for CMIP6 in future." (l.550-552)

*Comment: - Abstract: l. 5: I would add "two emission scenarios:" before RCP 4.5 and RCP 8.5 for the reader who does not necessarily know these scenarios.*

Response: Thanks for pointing this out. It was added.

*Comment: - Abstract: l. 15: "Minimum annual runoff…" I guess this result is still obtained with RCP 8.5, is that correct?*

Response: No, this result represents the mean changes of each catchment for both emission scenarios combined. However, the larger changes are obtained with RCP 8.5. We changed the sentence for clarification: "In the future, minimum annual runoff occur 13–31 days earlier in the winter months for high-elevation catchments, whereas for low-elevation catchments a shift from winter to autumn by about 15–100 days is projected with generally larger changes for RCP 8.5."

*Comment: - Figure 3: I suggest adding a reference to Table 2 in order to remind the meaning of the different objective functions.*

Response: We assume the reviewer refers to Table 4 and not Table 2. We adapted the caption of Figure 3: "Mean model performance of the best 300 parameter sets for the calibration and evaluation periods. Objtot shows the overall model fit, Table 4 gives a description of the objective functions, * indicates the catchments that use eight years of evaluation instead of ten."

*Comment: - l. 235: missing space after "year."*

Response: This was changed.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A.: A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, *56*(9), https://doi.org/10.1029/2019WR025975, 2020

Switanek, M. B., Troch, P. A., Castro, C. L., Leuprecht, A., Chang, H.-I., Mukherjee, R., and Demaria, E. M. C.: Scaled distribution mapping: a bias correction method that preserves raw climate model projected changes, Hydrol. Earth Syst. Sci., 21, 2649–2666, https://doi.org/10.5194/hess-21-2649-2017, 2017.

Teutschbein, C., & Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of hydrology*, *456*, 12-29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.