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

*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 will add a sentence to Section 2.1 l.82 "As shown in Fig. 1, precipitation stations are located in the valleys of the catchments at elevations below the mean catchment elevation. Therefore, precipitation data might not be representative for the whole catchment."

*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. 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 (Figure 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 will align 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. 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.

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 will be further clarified in the revised manuscript.

*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 will revise it and refer to it as "GCM/RCM combination". 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 can not 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 will acknowledge and discuss this (l. 499): "Extremes in changes for different catchments and emission scenarios can often not be traced back to a single GCM/RCM combination. Assessing the individual uncertainties of GCMs and RCMs used, may yield different results (e.g. Evin et al., 2021). However, such a detailed analysis was not done here as its assessment remains non-trivial and was thus considered to be out of the scope of this work." We will further clarify and discuss this and add the suggested reference in the revised manuscript.

*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 will further discuss the impacts of the model choice and add the following in line 515: “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).”

In line 520 we will 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).”

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. l. 520 “A new set of GCM simulations is available (CMIP6, Eyring et al., 2016). However, it 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.”

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 will be 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 suggest to change the sentence for clarification: “In 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 will adapt 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 will be 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.