We would like to thank Anonymous Referee #2 (AR2) to write this review. Below, we will reply to the points made by AR2, with the comments from AR2 in black, and our response in blue.

General comments:

Buitink et al. showed the relative importance of changes in temperature, evaporation and precipitation on changes in discharges from the 1980s to the 2010s using the dS2 model in the Rhine river basin. The manuscript reads well and has little grammatical errors, but the structure and methods could use work to help readers understand the simulations. Information on the methodology is greatly missing, which causes readers to speculate how to interpret the overall conclusions. Also greatly missing are comparisons of this work to other modeling studies. Based on my review, I would suggest major revisions before publication is merited.

Thanks for writing this review. We agree with the comments made, and believe this will improve the quality of the manuscript.

Major comments:

1. Section structure:

The methods section is too brief. Much of the model descriptions are contained in the supporting information and should be moved to the main text. The swapping of variables is confusing to me. How can you realistically change temperature only affecting snow processes or evaporation? Wouldn't both processes be affected by changing the temperature? If the goal of the paper is to simulate the hydrologic response to temperature-driven changes in evaporation and snow processes, specific details on the snow processes being simulated need to be included. The term 'snow processes' is used throughout the paper, but it is unclear which snow processes are simulated. Glacier melt is considered a snow process? Increased melt from glaciers is attributed to changes in discharge later in the manuscript. Is it possible to separate the effects from snowmelt and glacier melt? The discussion section should be separated from the conclusion to compare your results with other studies and address the overall implications from your results better.

This point was also mentioned by the other reviewers. We will move the model descriptions from the supplement to the main manuscript, and add a description on the forcing swapping (see below). We understand the confusion regarding our usage of the term "snow processes", as it is indeed a term grouping snowfall, and melt from snow and ice: we will clarify this in the next version. We will try to separate the effects cause by snowmelt and glaciermelt, as this is already partially visible in Fig. 5a (blue pixels during the second and third periods). We agree that a comparison with other studies is currently lacking, and we will add this in the next version of the manuscript.

2. Figures:

Figure 2 has a lot of results, but it is difficult to interpret due to the small size of the individual panels. Panels 2a, 2e, and 2f would benefit from being stretched out to see the results in more detail.

We agree that Fig. 2 contains a lot of information. We have decided to split the figure into 2 figures, and replace Fig. 2d with an improved version to visualise the method behind forcing-swapping (see below)

Additional comments on Figure 2:

- The colors in 2a do not match the legend. I assume the darker lines are for simulated and the lighter lines are observed, but clarification would be nice.

  This is correct, and we will clarify this in the next version of the manuscript.

- It appears that the sum of the difference in model simulations in 2e during February are cut off by the y-axis limits?

  This is correct and we will fix this.

– Why are there gaps in 2f? Does this mean that the forcing variables failed to explain any fraction?

These are no gaps, but show that the forcing variables failed to explain the difference. We understand the confusion, and hope that our new description and figure on this method (see below) clarifies the results in Fig. 2f.

– It is confusing why the cumulative effects for forcing variables are shown in 2g, but the absolute effects are generally discussed in the text. This is how the 11%, 19%, and 18% from the abstract were calculated, correct?

– All of the significant differences in monthly discharge shown in Fig. 2c are larger for the 1980s compared to the 2010s, yet the sum of the differences from the forcing variables results in positive discharge differences during March and December in Fig. 2e. How is this possible?

This is caused by the fact that the sum of the differences failed to explain the real 2010s discharge (see also the low fraction explained during March). This can again be related to the "forcing interactions"

– The dotted grey line (sum of diff.) in Fig. 2e during early March is +500 m3/sec, but the solid grey line (2010s) shows a negative discharge difference for the same time, can you explain this? Similarly, during early May.

Again the result of interactions between the forcing variables. For example, during March, the sum of differences expects higher discharge values, while 2010s was indeed lower. As the forcing swapped runs failed to capture the interaction between precipitation and temperature (snowfall and melt, evaporation and other storage related processes), it also failed to get close to the 2010s values.

Occasionally text does not seem to align with results from the Figures:

– Snow depth decreases for the majority Europe are reported in line 26 and shown in Figure 1e from ERA5 data, so why is the discharge difference from modified P in Figure 2d positive?

The data in Fig. 2d is synthetic and not based on real results, at it was an effort to conceptualize the calculations performed in Fig. 2e and f. As mentioned earlier, we will remove this figure with an improved conceptual figure.

– In lines 86-87 you write "Both variables are correctly represented, and show similar variability as the observations, even at hourly timescale." I would argue that snow storage is poorly simulated as the maximum simulated snow storage/height is twice as much as the observed maximum snow storage/height. This needs to be explained more. Could the positive discharge difference due to Modified P be due to the simulated maximum snow storage being twice as high as the observed maximum snow storage in Fig. 3b?

There is a difference between the plotted variables: the colored line (simulation) shows the snow storage in snow water equivalent (SWE), where the black line shows the observations as snow height. A direct translation from snow height to SWE is difficult due to snow compaction. This is why we only focus on dynamics.

– In lines 101-102 you write "During spring, this simulation shows higher discharge values resulting from increased snow melt." It is confusing which specific months you are referring to, but in Figs. 4b and 4c modified TSnow does not appear to have a positive effect on discharge. But during low flow conditions (Fig. 4d), TSnow has a positive effect on discharge?

In lines 101-102 we refer to the months March–May, were the simulation with modified Tsnow shows higher discharge values (as visible in Fig. 2e). In this experiment, we replace the temperature time series of 1980s with the time series from 2010s. In Fig. 4 we simply increase the temperature of the 1980s time series. In the three sub-panels, we do not focus on the period March–May, but on the grey highlighted regions in Fig. 4a. Despite this, the increased discharge during March–May is still visible in Fig. 4a. In the next version of the manuscript, we will make sure that we better clarify the periods we are referring to.

– In line 155: "Discharge between these two periods was significantly different for 8 out of 12 months". 10/12 boxes in 2b are full colored, representing significant differences.

The text is wrong, this should indeed be 10 out of 12 months.

- Results are often grouped by months, but then seasonal changes are discussed in the text. This is confusing for the reader to speculate which particular months you are referring to. For instance, the third shaded period in Figure 4 is referred to in the text as the "late summer low flow" period (line 120). But this time period aligns with the end of September through the beginning of November, not typically thought of as late summer. I would suggest referring to the results based on the monthly changes to remove this ambiguity.

  Thanks for this point, and we will clarify this in the next version of the manuscript.

3. Comparison between other studies

Much work has already been conducted on simulated effects of temperature changes on hydrologic response. It would be nice to see a comparison of your results to some of these studies and why your results agree or disagree from theirs. This appears to be completely missing. You list six studies that investigated these effects in lines 28-29, but do go on to discuss their results or compare yours at all.

This is a valid argument: we will add a section to compare our results to similar studies.

Climate model simulations in western North America indicate that the fraction of meltwater volume produced at high snowmelt rates is greatly reduced in a warmer climate (i.e. "Slower snowmelt in a warmer world", Musselman et al., 2017). Additionally, model simulations suggest slower snowmelt decreases streamflow production ("Snowmelt rate dictates streamflow", Barnhart et al., 2016). But, in lines 161 – 163 you write "With higher temperatures, increased melt from glaciers and snow packs can offset the discharge reduction from enhanced evaporation over the majority of the year" and that "these results can be interpreted for the many different basins around the globe depending on both rain- and snowfall". These sentences are confusing and incredibly misleading. Your remarks make it seem like increased snowmelt and glacier melt offsets reduction from evaporation and results in an inconsequential effect on discharge. But in Figures 2b and 2c you show significantly lower discharge for 10/12 months and annually for the 2010s compared to the 1980s.

We agree that these sentences can be misleading, and will rephrase those. With "offset" we mean that the change resulting from Tsnow make the discharge less severe, but not that it can fully compensate for the increased evaporation.

Minor corrections:

Line 14: Higher temperatures have been shown to lead to slower snowmelt rates ("Slower snowmelt in a warmer world", Musselman et al. 2017, Nature Climate).

We will discuss this in a section comparing our work with other studies.

Lines 18-21: Changes in discharge are most likely to be strongly affected by changes in precipitation. It's probably best to focus on the runoff ratio (discharge/precipitation). Runoff ratios (or runoff efficiency) were found to be mostly unchanged in snow-covered areas of the western U.S. despite increasing temperatures and decreased snow fractions ("Warming is Driving Decreases in Snow Fractions While Runoff Efficiency Remains Mostly Unchanged in Snow-Covered Areas of the Western United States"; McCabe et al., 2018; Hydrometeorology).

We will discuss this in a section comparing our work with other studies.

Line 23: "Europe has experienced significant changes in evaporation, snow depth and streamflow over the last decades". Citation needed. Were all of the changes negative? Which decades?

We refer to the relevant studies in the next studies of this paragraph.

Lines 24-25: "Their study shows that both changes in precipitation and evaporation had considerable effects on the streamflow." Did they observe negative changes for both precipitation and evaporation? How do changes in rain compare to changes in snow?

Line 26: "showed that snow depth decreased over the majority of Europe". From when to when?

They showed this to be a trend since the 1950s.

Lines 37-38: "for example, the study by Mastrotheodoros et al. (2020) took more than 6 × 105 CPU hours". What resolution did they use?

A higher resolution of 250x250 m. This will be mentioned to put the difference in a correct perspective, but the difference in resolution alone cannot explain the much larger difference in computer time.

Lines 42 - 44: “This study investigates the hydrological response to temperature-driven changes in evaporation and snow processes, testing our main hypotheses that both seasonal changes in snowmelt and enhanced evaporation will aggravate low flows, and that the changes will increase with temperature under realistic warming.” Is snowmelt the only snow process tested in this study? If so, I would change the terms “snow processes” to “snow melt”.

Also changes in snowfall. We understand the confusion, and will define "snow processes" in the next version.

Lines 53-54: “The Rhine basin was selected because the climate and basin heterogeneity are representative for north-western Europe and many other basins globally”. It seems like a stretch to suggest the Rhine basin is representative of many basins globally

We will rephrase this.

Line 80: “higher flows during late winter in the 2010s”. Be specific about the months. It seems that from 2b discharge is lower for all months in the 2010s. I do not see higher flows during late winter in the 2010s?

This is indeed incorrectly written, we will correct this.

Lines 101-103: “The higher temperatures of the 2010s also resulted in lower discharge values in the first few months of the year. During spring, this simulation shows higher discharge values resulting from increased snow melt.” Again, it is confusing which months you are referring to. I would assume that the first few months of the year are late winter. These two sentences seem contradictory. First you say that higher temperature (affecting snowmelt) resulted in lower discharge, then you say higher temperature (affecting snowmelt) resulted in higher discharge. Was the increase in spring discharge due to an increase in the snowmelt rate of the volume of snowmelt?

We understand the confusion, and will clarify this, including specific references to the months.

Lines 107-108: “The explained fraction is lowest during spring and late summer.” Be specific about which months. Are you referring to the gaps in Fig. 2f?

Yes, we are referring to the gaps, we will clarify this.

Line 120: “late summer low flow”. Be specific about which months. The late summer period aligns with the end of September through early November, more typical of fall/autumn. Figure 4b: The effects from modified TEvap in Figure 4b are hidden by the combined effects.

We will clarify this.

Line 155: “Discharge between these two periods was significantly different for 8 out of 12 months.” Is this simulated or observed? In Fig. 2c you show all but two months being significantly different?

Simulated, and the 8/12 should indeed be 10/12.

Lines 162-163: “With higher temperatures, increased melt from glaciers and snow packs can offset the discharge reduction from enhanced evaporation over the majority of the year.” This interpretation is opposite of previously published studies (i.e. slower snowmelt in a warmer world) and needs clarified.

We will clarify this in the next version.

1 Methods - Forcing swap

In the first experiment, we aim to understand how each forcing variable can explain the resulting changes in discharge, and their relative importance. To perform this, we setup the experiment according to the conceptual overview presented in Fig. 1. In order to investigate how temperature influences evapotranspiration and snow processes separately, we perform model runs in which the total temperature change is splitted into temperature effects on evapotranspiration ("Changed T_{evap}") and snow processes ("Changed T_{snow}"). In addition, another run is performed with only changes in P ("Changed P"), so that these individual runs can be compared to a run where all changes in forcing are enabled (2010s forcing). The resulting simulated discharge is compared to the 1980s run, to determine the discharge change. In this way, we can evaluate the relative impact of each forcing variable on the discharge.

We sum the discharge changes of the three forcing swapped runs, to obtain $\mathrm{Sum}\Delta$:

$$\mathrm{Sum}\Delta = \Delta \mathrm{Q_P} + \Delta \mathrm{Q_{T\ snow}} + \Delta \mathrm{Q_{T\ evap}} \tag{1}$$

where $\Delta \mathrm{Q_x}$ represents the discharge difference of the forcing swapped simulations. We can use this $\mathrm{Sum}\Delta$ to study how well it explains the 2010s run, by comparing it to $\Delta \mathrm{Q_{2010s}}$. We hypothesise that when $\mathrm{Sum}\Delta$ is equal to $\Delta \mathrm{Q_{2010s}}$, the effect of the forcing is additive, and together explain all differents. We will refer to this as the direct effects. In the case of a discrepancy between $\mathrm{Sum}\Delta$ and $\Delta \mathrm{Q_{2010s}}$, this can be attributed to interaction between the three forcing components. We will refer to this as indirect effects. We define $\mathrm{Sum}\Delta$ to have explanatory value when it has the same sign as $\Delta \mathrm{Q_{2010s}}$. We calculate the contribution of the direct effects (ϕ) using the following equation:

$$\phi = \begin{cases} \frac{\min(\mathrm{Sum}\Delta, \Delta \mathrm{Q_{all}})}{\max(\mathrm{Sum}\Delta, \Delta \mathrm{Q_{all}})}, & \text{if } \mathrm{sign}(\mathrm{Sum}\Delta) = \mathrm{sign}(\Delta \mathrm{Q_{all}}) \\ 0, & \text{if } \mathrm{sign}(\mathrm{Sum}\Delta) \neq \mathrm{sign}(\Delta \mathrm{Q_{all}}) \end{cases} \tag{2}$$

This value can then be used to calculate the relative (direct) contribution of each forcing variable, using the following equation:

$$\phi_\mathrm{x} = \frac{\mathrm{abs}(\Delta \mathrm{Q_x})}{(\mathrm{abs}(\Delta \mathrm{Q_P}) + \mathrm{abs}(\Delta \mathrm{Q_{T\ snow}}) + \mathrm{abs}(\Delta \mathrm{Q_{T\ evap}}))} \cdot \phi \tag{3}$$

where $\Delta \mathrm{Q_x}$ should be replaced by $\Delta \mathrm{Q_P}$, $\Delta \mathrm{Q_{T\ snow}}$, or $\Delta \mathrm{Q_{T\ evap}}$.

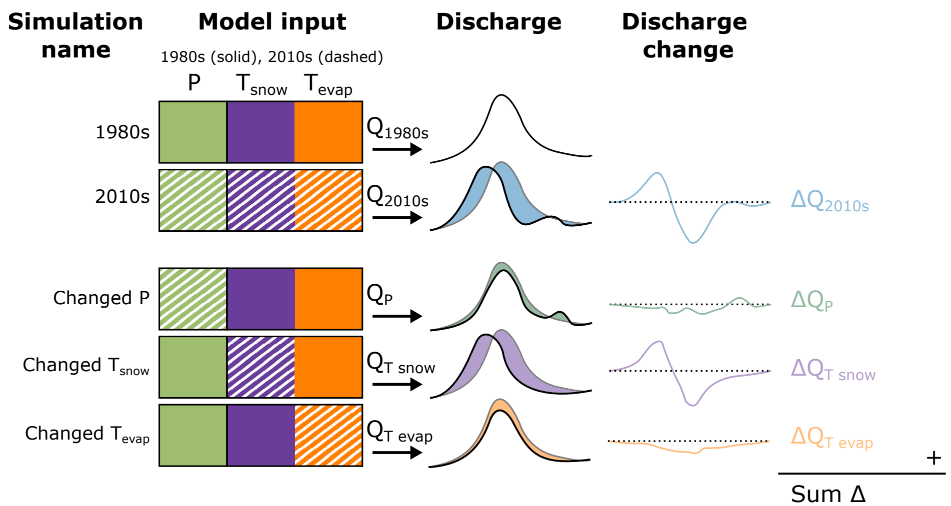

Figure 1. Conceptualization of the forcing swap experiment, showing the different simulations (rows) and steps in the analysis (columns). The different forcing variables are visualized as colored blocks, where the solid and dashed boxes indicate forcing data from the 1980s and the 2010s, respectively.