We would like to thank Anonymous Referee #1 (AR1) to write this review. Below, we will reply to the points made by AR1, with the comments from AR1 in black, and our response in blue.

General comment

Buitink et al. simulate the Rhine River Basin for two 10-year time slices (1980s and 2000s) using the dS2 model and ERA5 data (0.25x0.25). An additional soil moisture model was added in order to attain values of actual evapotranspiration. ERA5 data was interpolated to a 4x4 km grid using bilinear interpolation. They assess changes in discharge between the two time slices and attribute differences in runoff to differences in precipitation, evapotranspiration and snowmelt. In general, it is an interesting study and has the potential to become a valuable contribution to hydrological research. However, I see several major issues regarding the analysis and the text that need to be addressed before it can be considered for publication.

Thanks for your overall positive evaluation of our manuscript, and the suggestions for improvement.

Major comments

Comment 1: Structure text

One major issue I see is that sections in the manuscript are mixed up or even missing. Often, model and method description are located in the result section. The results section is mixed with the discussion. In general, the manuscript lacks important details. I think a more detailed method description, where each step of the analysis conducted is described, needs to be added. Furthermore, model set-up and model components need to be described better. This is crucial to understand you experiments. Please provide more information on the dS2 model and the snow routine. An additional section 'study area and data' could be good to better introduce the Rhine Basin. In my opinion, large parts of the supplementary material can be moved into the actual manuscript.

We understand the confusion caused by the structure of the text, as this comment is also discussed by AR2 and AR3. It was an effort to write a short and concise manuscript, but we now see that we need to provide more explanation to improve readability of the manuscript. We propose to move the model description from the supplement to the main text. We will also add information to describe and introduce the Rhine basin. Additionally, we will add headers to clearly separate the different experiments performed in this study. We will extend the methods description to improve the explanation of the forcing swap variables (see below).

Comment 2: Swapping variables

I still have troubles to understand your approach of swapping forcing variables in order to attribute changes. You try to asses the contribution of the three factors: 1) changes in precipitation, 2) changes in snowmelt due to changes in temperature and 3) changes in evapotranspiration due to changes in temperature. If I understand correctly, you swap individual forcing variables (temperature or precipitation, respectively) between the two 10-year time slices. You run the 1980s with temperature data from the 2000s, for example. So the temperature in the first week of August 2005 becomes the temperature in the first week of 1985, right?

This is correct. We agree in hindsight that a proper description of this experiment is lacking, hence we will add a paragraph in the methods section to explain this experiment. The original conceptual figure of 2d is not enough to properly describe the workflow. Therefore, we propose to add a new conceptual figure together with explanations and equations (added at the end of this reply). We hope this clarifies the experiment.

This seems to be a very rough approach and I am not sure if this is a good idea. Also the performance of this approach seems quite poor. For parts of the year you can not at all explain variations or even expect an opposite trend (in February, for example, strong overestimation and in March even an opposite trend). The three factors you investigate do not really explain the variations in discharge, I think.

Figure 2a shows that the full model is able to correctly simulate the discharge in both periods (1980s and 2010s). As this is the same calibrated model, the only changes between these simulations are in the forcing data (precipitation and temperatures). The three factors (forcing) that are varied are the only things that change, so any change unexplained by the sum of the individual factors is not poor model performance, but reflects complexity and non-linearity in the overall response. We feel that the use of "unexplained" was probably falsely suggestive of poor model performance, so the terminology will be changed in the revision.

All variations that you can not explain, you attribute to 'interactions'. To me, these 'interactions' are not clear. I do not know what you have in mind here. Can you explain more detailed?

With "interactions" we try to describe the (nonlinear) interplay between the forcing variables. For example: precipitation can either fall as snow or rainfall, depending on the temperature. A change in temperature will therefore also influence the precipitation dynamics.

It looks like that a model only using the factor 'changes in precipitation' alone would explain changes in discharge better than the approach with the three factors (Fig. 2e). Please check.

Precipitation does seem indeed to have a very large influence, which is not unexpected given the precipitation differences between the two periods.

What about changes in evapotranspiration due to changes in precipitation? What about changes in snowmelt due to changes in precipitation?

These are also present, and this is a nice example of what we refer to as "interactions". We hope that the new method section (see below) will clarify this.

Comment 3: Attribution

In the abstract you write that variation can be 'explained by the changes induced by snow (11%), evaporation (19%) and precipitation (18%), while 52% was driven by combination of these variables." This bases on results presented in Fig. 2 panels e, f and g, I assume. How do you calculate those percentage values? In February and April, for example, you have a negative variation in discharge between the two 10-year slices of around 1000 m3 /s. This variations can be explained by variation in rainfall to 90-100%, right?

These percentages are indeed based on the results in Fig. 2e–g. As mentioned earlier, we understand the confusion and hope that a better explanation will clarify our workflow. Rainfall can indeed explain large fractions of the change, but we also need to take the changes induced by temperature (both evaporation and snow) into account, as these also change during between the two periods. This is the reason why we did not calculate the total contribution of only precipitation changes, but always take these changes with reference to the combined changes (P + Tevap + Tsnow).

Comment 4: Evapotranspiration

Are you using evaporation or evaporation + transpiration = evapotranspiration? You mention the Penman-Monteith equation to get potential evaporation. Isn't it reference evapotranspiration you get? Do you calculate this? Or is it ERA5 data directly? You show a map of the potential evaporation (Fig. 1c). Can you also show a map of the calculated actual evaporation? Please give more information on the soil hydraulic data used. If I understand correctly, you assume one value of the rootzone depth for the entire basin? Is this value also constant over time? What about vegetation cover?

We do indeed calculate reference evapotranspiration, based on ERA5 data. We will add a map to validate the calculated evapotranspiration by comparing it to GLEAM data. We will move information on the model and data from the supplement to the main manuscript. We do indeed assume one value of rootzone over the entire basin, constant in both space and time.

Comment 5: Temperature 'scenarios'

In my opinion, the different approaches you use need to be explained better. Your swapping approach and the temperature 'scenarios'. Please explain in the method section. In those temperature 'scenarios', you simple add e.g. 2 ∘C to the hourly temperature time series? This again seems a very simple approach. In what way do those increases influence your model components and does this reflect 'real-world' processes? In Fig. 5 b your results indicate that increasing temperatures will rise evaporation in winter, in turn, decreasing discharges. According to my knowledge, evapotranspiration does barely play any role in winter (low radiation, low temperatures, no plants). All studies I know hint at an increase in discharge in winter, e.g. due to liquid instead of solid precipitation. Can you check this and maybe show your simulated evapotranspiration values for winter for this temperature 'scenarios'? Your model shows a linear response of evapotranspiration to temperature. How is evapotranspiration affected by changes in temperatures in the first place? Isn't it an input to your model?

We will clarify the different experiments in the next version of the manuscript. The discharge reduction visible in Fig. 5b can indeed be attributed to increased evaporation, but also a change in water storage. With a change in evaporation (which has indeed the largest influence during summer), water stored in the catchment will be reduced. This reduction continues to affect discharges during winter, as less storage means less discharge production. The change from solid to liquid evaporation is not included in the Tevap run, as this is part of the Tsnow simulation, where we do indeed see an increase in discharge during the months November–May.

These results does seem to show a rather linear response to temperature, although there appears to be a slight curvature in some responses (Tsnow in Fig. 4b, Tevap and Tboth in Fig. 4d). As these processes are threshold processes (e.g. snow fall and melt), you would expect a non-linear response. We expect that this becomes more visible when reaching more extreme ends of the hydrological spectrum (droughts and/or floods), but this is not clearly visible from our analysis.

Comment 6: Validation, Calibration

Why do you use two point measurements of this small pre-Alpine catchment to validate? Isn't there hundreds of snow gauges? Where is this snow gauge exactly? At what elevation? You compare the simulated snow in your 4x4km cell with the point measurement? Why do you simulate in an hourly resolution? Why not daily? For the routing? Can't you use daily data for calibration and validation (even if you initially run hourly)? At the end you aggregate to monthly values anyway. I still do not see the use of the hourly temporal resolution.

We choose to use an hourly resolution to ensure a correct link between space and time, as the choice of hourly resolution can have a large effect on model output (Melsen et al., 2016). For example, an hourly resolution can capture the snow melt dynamics during a single day, which is impossible at the daily timestep.

We validate the model with these point observations, as they are also hourly observations. With this validation, we validate our model in time, on variables which we did not optimize. This proves to give more confidence that the model is able to correctly simulate the fluxes important for our study. As far as we know, most snow observations are daily records of snow depth rather than continuous records at hourly resolution. In the revision we will provide more details about the experimental Rietholzbach catchment were the observations were made. We will add a spatial validation as well, by comparing the snow cover with observed snowcover, and by comparing our evapotranspiration with GLEAM evapotranspiration.

Further comments

Page 1 Line 6: "Increased temperature scenarios show that seasonal changes in snowdynamics could offset a fairly constant negative change in relative runoff induced by evaporation, but not during the melt season." I do not understand. What seasonal changes are you talking about? Where can I see this? What figure?

This is visible in Fig. 4, where the purple line is above the 1980s simulatino for the majority of the year. We will rephrase this sentence.

Page 2 Line 14: "higher snowmelt rates" Lower snowmelt rates? https://www.nature.com/articles/nclimate3225

Thanks for this reference. We will discuss this work in a new discussion section to place our study in existing work.

Page 2 Line 48: "high spatial and temporal resolutions ensure that small scale variability is accounted for." I do not see how you account for small scale variability. The basic input data is very course (30x30 km) and only interpolated linearly to a 4 km grid. Strong spatial variability over short distances in the Alps, for example, is not captured.

Here we refer to temporal variability (e.g. snow melt patterns throughout the day). We will rephrase this.

Page 2 Line 54: "representative" I do not get why it is 'representative'. Representative for other large river basin in Europe? Which ones? As you mention, it is a very heterogeneous basin, so this large heterogeneous basin actually is not representative for any sub-region in the basin, as they do not have this heterogeneity. Please check this again. Maybe I just don't get it.

This is indeed misleading, and we will rephrase this.

Page 2 Line 54: "north" Northern Europe is roughly north of the southern coast of the Baltic Sea. It is more Central Europe, I think.

We will rephrase this.

Page 3 Line 63: "bilinear interpolation" Why do yo use bilinear interpolation? Are there other ways to better address the spatial variability in the basin? Why do you interpolate on a 4x4 grid?

We chose to use a 4x4 km grid to strike a balance between input data, spatial variability and run times. We decided to interpolate using bilinear interpolation to avoid adding additional uncertainty. We are aware that this method also adds uncertainty, but so does running a hydrological model at 0.25x0.25 degree resolution.

Page 3 Line 72: "yearly average precipitation sums decreasing from 1146 mm to 1066 mm" How does this go along with mentioned 'intensification of the hydrological cycle'? Is this only decadal variability or a long-term trend? Can you compare to other 10-year slices?

This can indeed be attributed to decadal variability. As the focus of our manuscript is not on precipitation changes, we decided not to go further into the cause and effect of precipitation changes.

Page 4 Figure 1: "climatic changes" I am not sure weather it is appropriate here to call in climatic changes. You only compare to 10-year time slices only twenty years apart. Usually climatic changes are assessed over longer time periods comparing at least 30-year time slices. Particularly precipitation is subject to strong decadal variability. I don't think it is appropriate to attribute the differences in precipitation that you show to climate change.

With climatic we refer to hydro-meteorological changes. We will rephrase this.

Page 4 Line 92: "Given that dS2 is not calibrated on these variables, and the difference in spatial scale of the input data, this shows that dS2 is able to correctly simulate evaporation and snow processes." This validation using only two point measurements does not really convince me. Are there other ways to better assess the performance in space? Maybe MODIS snow cover maps?

We see this a temporal validation, but we will extend our validation to a spatial validation on observed snow cover and evaporation from GLEAM.

Page 6 Line 121: "confirming our hypothesis" Your hypothesis was a linear response of discharge to temperature change?

The sign of the change was inline with our hypothesis, we will clarify this.

Page 6 Line 132: "increased snowmelt" Isn't it more liquid instead of solid precipitation?

A combination of more melt and liquid precipitation. We will change this.

Page 7 Figure 4: " Panel a shows the yearly average discharge under a 2.5° C " Just for me to understand. First, you simulate the 1980s with your normal input data. Then you add 2.5°C on the temperature time series and re-run the model? So the increase in temperature strongly increases (linearly) the evaporation and hence discharge decreases? Does it makes sense that this effect is the same throughout the year?

This description is correct. See our reply below Comment 5 for an explanation of this behaviour.

Page 7 Line 133: "glaciers in the Alps" There are glaciers in your model?

Glaciers are represented in dS2, although they do not have a separate module, but rather are pixels with excessive amounts of snow.

Page 8 Line 38: "As expected, the majority of the basin is controlled by the change induced by a change in evaporation (84–97%). As a result, the mean discharge is reduced by $\pm 18\%$." Do you think that his is only in you model or is this effect the same in the 'real-world'? Any seasonal differences?

There are indeed seasonal differences. The results presented here are annual values, and are not the same for every season. We expect this also to be realistic, as we already see a decrease in discharge between the 1980s and the 2010s (partly due to changes in precipitation as well).

Page 8 Line 146: "interannual variability" What about decadal variabilities?

As we are mostly focussing on the changes in temperatures, it is known that the difference between the periods can be attributed to changes in climate.

Page 8 Line 147: "effect of different temperatures" Also strong differences in precipitation between the two time slices!

This is true, but we mainly focus on the changes in temperature.

Page 8 Line 149: "downscaling method" A linear interpolation is better than downscaling? Isn't the simple bilinear interpolation you use adding a lot of errors and uncertainties throughout the basin?

We don't argue that linear interpolation is the better method, just that it has less degrees of freedom.

Page 8 Line 159: "With higher temperatures, increased melt from glaciers and snow packs can offset the discharge reduction from enhanced evaporation over the majority of the year." Where can I see this in your result figures? Please explain better what you mean by 'snow driven changes'.

We understand the confusion around the "snow driven changes" and "snow processes" terms. We use this to group all snow and ice related processes (snow fall, melt, etc), and we will make sure this is properly explained in the next version of the manuscript. This "offset" is visible in Fig. 4a, where the Tsnow line shows higher discharge values than the 1980s simulation for the majority of the year.

Page 9 Line 168: "Enhanced melt will offset the negative trend caused by the increased evaporation, until the frozen water storages are depleted." Where do you show this in your study?

This is visible in Fig. 4a, where the Tsnow line shows higher discharge values than the 1980s simulation for the majority of the year. We understand that we do not directly show the frozen water storages, and will clarify this.

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 $\Delta\mathrm{Q_{T\ evap}}$.

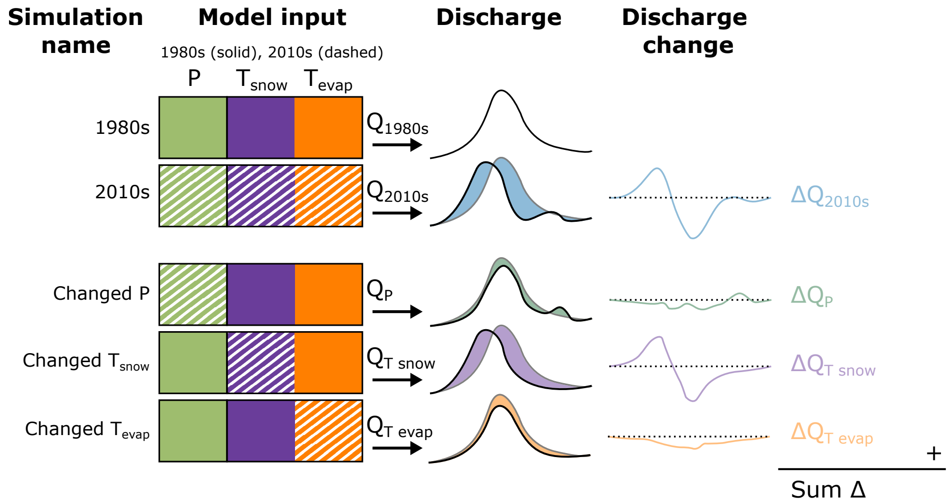

Figure 1. Conceptualization of the forcing swap experiment, showing the different simulations (rows) and steps in the analysis (columns). The different forcing variables are visualized as colored blocks, where the solid and dashed boxes indicate forcing data from the 1980s and the 2010s, respectively.

References

215 Melsen, L. A., Teuling, A. J., Torfs, P. J. J. F., Uijlenhoet, R., Mizukami, N., and Clark, M. P.: HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models, Hydrology and Earth System Sciences, 20, 1069–1079, https://doi.org/10.5194/hess-20-1069-2016, http://www.hydrol-earth-syst-sci.net/20/1069/2016/, 2016.