Thank you for taking the time to review our paper and for your constructive comments. We numbered the comments and provide a response to each below in blue.

1) Main finding and figures: I agree with reviewer #1 that the text at some places in the results and discussion sections read clear and sound, but the figures do not always support the conclusions or findings. As mentioned by reviewer 1 especially the conclusion on the temporal distribution of SWI is not easy to extract from the figures. Figure 4 and 5 show SWI and discharge, but because the years differ in so many aspects (ratio of rain and snowmelt, timing of SWI, variability of SWI, Q), it is hard to tell which process caused the discharge response from the timeseries, i.e. to see a clear link between temporal distribution and discharge. The text describes these different aspects, but how to generalize these results more? Maybe some measures related to the timing of the center of volume for rainfall and snowmelt, antecedent conditions before spring or number and timing of melt/rainfall events could give some insights. Could also some measure on spatial and temporal distribution be combined? Probably the authors know best how they drew this particular conclusion and could use that to focus on that aspect in the results/figures more explicitly.

We recognize that our figures were more descriptive of the results, and that a clearer presentation of the findings should be included in the figures. In our response to reviewer 1, we suggested adding a heatplot (Fig. R1.2), which based on your feedback, we have now updated to also include the fraction of annual SWI occurring in each season (Fig. R2.1, below). In this new figure, we evaluate how different metrics of the (temporal) distribution of SWI are linked to both annual stream discharge and the day that the stream dried up. We think that different aspects of the temporal distribution of SWI are better represented in this figure. Because the figure highlights statistically significant and insignificant correlations, the reader can more easily connect the discharge variables with SWI magnitude and timing variables and determine that the snowfall fractions do not correlate significantly with discharge or dry-out date.

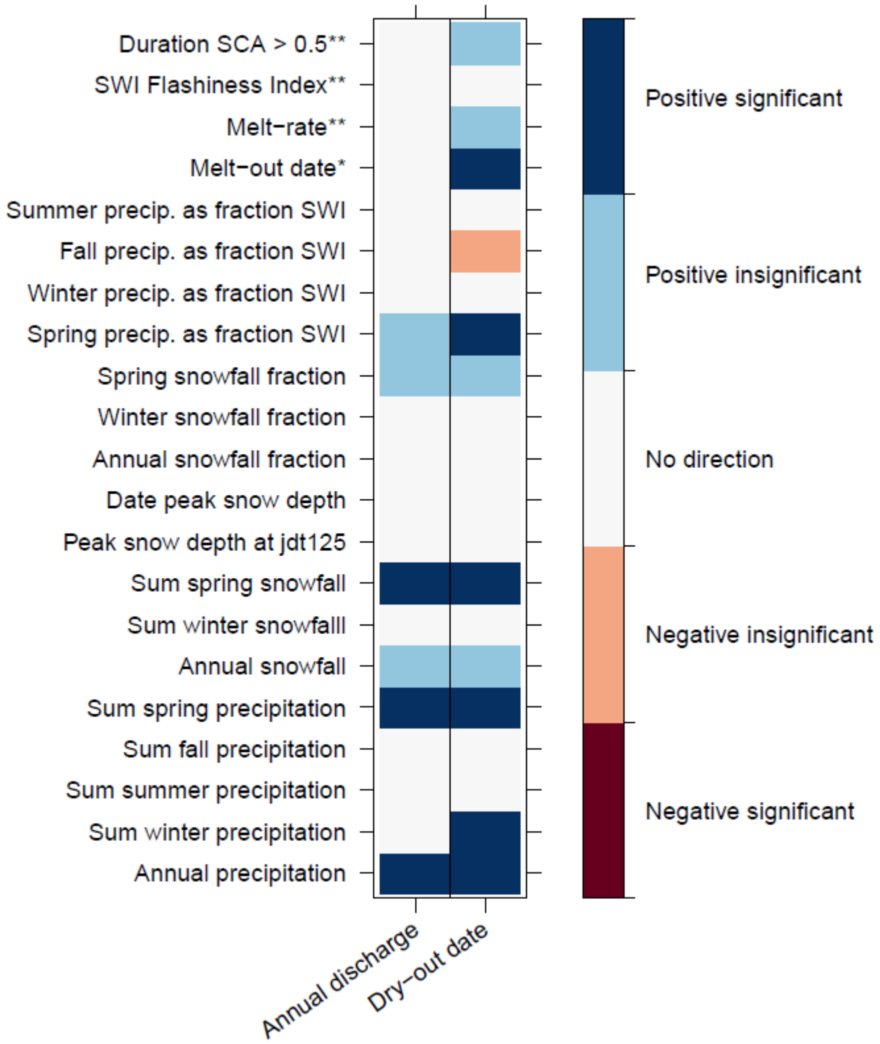

Figure R2.1: Heatplot showing Pearson correlation coefficients (α=0.1) for comparisons between annual discharge, the stream dry-out date and precipitation and snowpack metrics. Significant correlations are marked in dark red (negative) and dark blue (positive), whereas insignificant correlations ($-0.3<R^2<0.3$) are marked in light blue (positive) or light red (negative) and correlations without a direction are marked in white. For most metrics, the comparison is based on the 2004-2014 data record (n=11 years). The comparison with the melt-out date (marked with one asterisk) is based on the simulated years (n=5) and the years for which satellite imagery was available (2016-2019, n=4; which totals to n=9). For the SWI flashiness index, the melt rate, and the number of days when at least half the catchment was snow-covered (marked with two asterisks), we used only the years that were simulated (n=5).

2) Study setup: While going through the manuscript I was wondering why only four years were selected. Because of the many processes that influence the discharge signal, a larger sample of years may have provided stronger evidence how processes relate, i.e. avoid that for example the dry year that was analyzed had many rainfall events. From the data description it is a bit unclear to me what the maximum possible amount of years could have been for analyzing. The decision may have to do with the runtime of the model? At least I would expect some description how the selected years deviate from the mean hydro-climatology of the catchment. Maybe the discussion/limitations section could elaborate on the selection of the years and the intertwined processes when looking at observations and possibilities for future model experiments, isolating some of these aspects (for which discharge would needed to be simulated as well) – but this last point as the authors see fit.

We selected four years because setting up and running the model was a non-trivial task. Also, we aimed to focus on differences in the distribution of SWI and stream discharge for years that had different snowfall ratios and total water inputs and therefore, we selected strongly contrasting years from the 11 potential years of record (Godsey et al., 2018). We will describe this rationale in the revised manuscript, and highlight how the selected years differed from the long-term average: each year's precipitation, snowfall fraction and air temperature is included in Table R2.1 (see below), which will be included in the revised supplementary material. We now also summarize how the years were different from the long-term average in the methods section of the revised manuscript, and will include the information from Table R2.1 in Table 1 of the manuscript.

We also included a scatter plot of annual snow fraction and annual precipitation (Fig. R2.2, see below), that shows that the years we chose to simulate contrasted with the other years captured in the dataset. Although 2007 was slightly drier than 2014 and 2006 was slightly wetter than 2011, we chose to simulate 2011 and 2014 because additional weather stations had been installed in 2011. Temperature and humidity data from these additional stations increased model accuracy and snow depth data from these locations was used to validate the model outputs. We will also include this figure in the revised supplementary material.

Table R2.1: Annual precipitation (P, mm), snowfall fractions (SF, -) and air temperature (T_a, °C), as well as % of the mean of the 2004-2014 record (Godsey et al., 2018). Simulated years are printed bold.

	P (mm)	**% P**	**SF (-)**	**% SF**	**T_a (°C)**	**% T_a**
2004-2014	524	100	0.37	100	8.2	100
2004	470	90	0.49	132	8.4	103
2005	**543**	**104**	**0.23**	**63**	**8.2**	**100**
2006	714	136	0.29	78	8.4	103
2007	402	77	0.31	83	9.3	113
2008	465	89	0.45	123	7.4	91
2009	**549**	**105**	**0.49**	**132**	**8.0**	**98**
2010	**531**	**101**	**0.57**	**155**	**6.6**	**81**
2011	**693**	**132**	**0.41**	**111**	**7.4**	**91**
2012	494	94	0.24	64	8.6	105
2013	456	87	0.26	72	8.6	105
2014	**450**	**86**	**0.30**	**82**	**8.6**	**105**

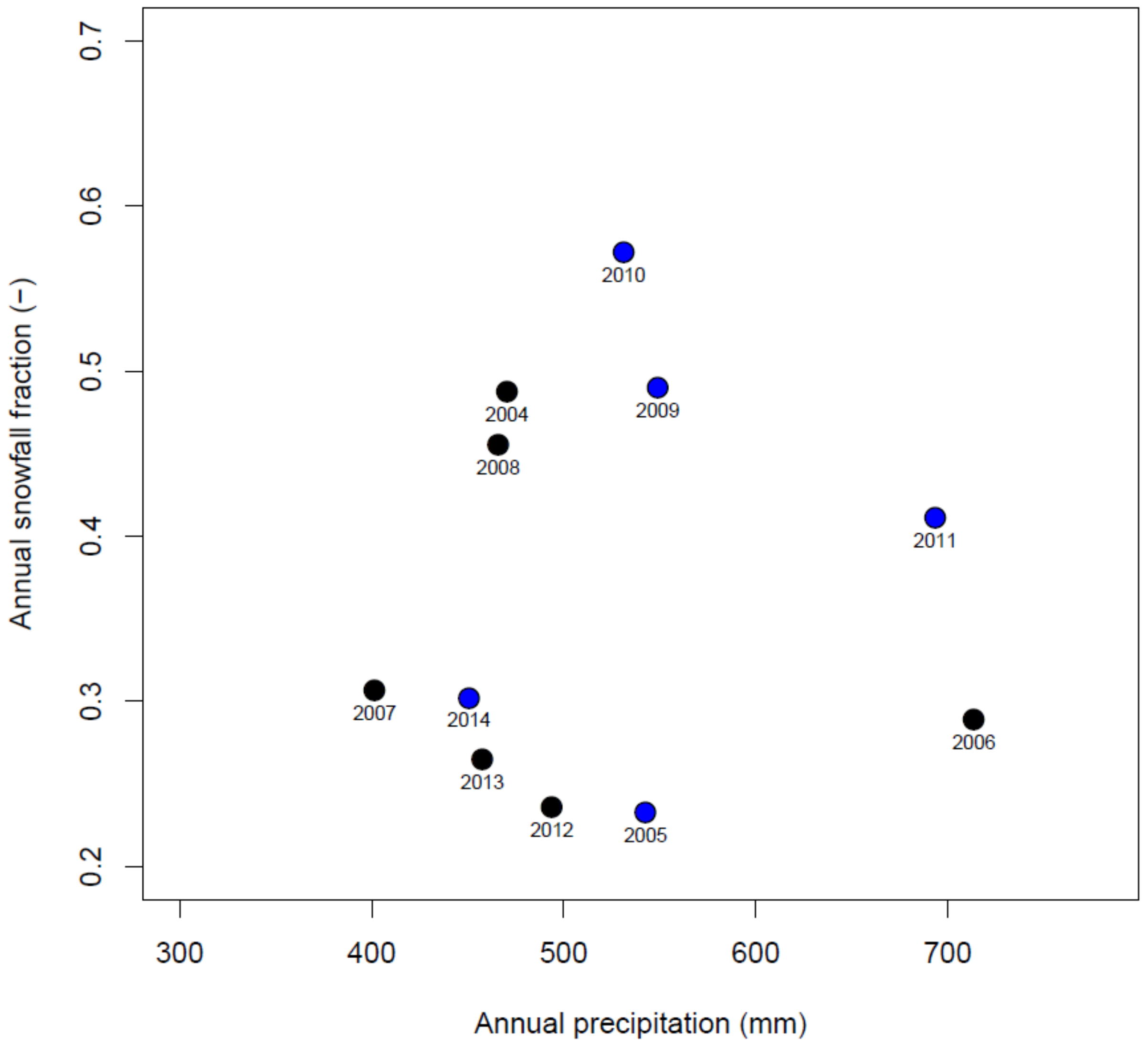

Figure R2.2: Scatterplot of the annual precipitation and snowfall fraction of precipitation at weather station jd125, which is located close to the catchment outlet. Simulated years are shown in blue, other years are shown in black.

3) Argumentation in introduction: Partly related to the comment of reviewer 1 on a better description of the novelty in the introduction, I think that the line of thoughts for this study and the research gap can be better described. In my opinion, the introduction mixes 1) changes in snowmelt generated streamflow, 2) differences between catchments seasonally snow covered and in the rain-to-snow transition zone, 3) rain-to-snow zones as a space-for-time substitution of catchments that are now seasonally snow covered and 4) changes that have occurred in the rain-to-snow transition zone and may occur in the future. Although all of these aspects may be important to put the study into context, I would suggest to clearly identify the research gap (how do yearly variations in rainfall and snowmelt influence discharge, relation with snowfall fraction not yet clear, rain-to-snow zone suitable to analyze 'extremes', i.e. snowy and rainy) and explain the implications for future changes and relations to observed changes in different type of catchments in a more structured way.

Thank you for this comment. We agree that it is important to clearly describe the research gaps and rationale behind the study, and will revise the introduction to emphasize the following points:

1) In contrast to the majority of snow research, this work is conducted in the rain-snow transition zone – a zone that currently covers a significant area of the mountainous western US and might yield insights in the future functioning of areas that are currently seasonally snow-covered.

2) In contrast to other work that often summarizes daily to seasonal responses at watershed/landscape scales, we quantified surface water inputs (SWI) at a high temporal (hourly) and spatial resolution (10-m). These high-resolution SWI estimates allowed us to investigate:

- The spatial variability in snow depths and SWE in a catchment that has a largely intermittent snow cover. In particular, this revealed the importance of snow drifts, even at the rain-snow transition zone.
- The extent to which the temporal distribution of SWI affects stream discharge and stream drying, and how that compares to annual metrics such as snowfall fractions or total precipitation, which are frequently used in larger scale estimations.

In addition, while SWE is frequently used as a summarizing variable for winter precipitation when comparing precipitation to stream discharge, SWI is more directly related to the timing and amount of water resources, and might therefore be an important variable to model in future work addressing similar questions.

4) Methods and data description: Here I missed some details regarding the available data, the model and the choice of years. As indicated above, it is not mentioned how the four climatologically different years were selected. I was also a bit confused by the numbers in table 1, how come that in a rainy year, the SWIsnow is higher than in a snowy year? Are numbers switched here? And without knowing the range of snowfall fractions over a longer time period it is difficult to interpret the values of the different years. It would also be helpful to explain the reasoning and possible hypotheses of selecting rainy and snowy years and wet and dry years. Could temperatures also be given for the years? Regarding the data and model, what is needed as input for the model? And which of the stations do have this data available for which time period.

We recognize that the description of the data, model and selection of years was rather short. We will add more information on the functioning of the model (e.g., how the model calculates snowmelt, how rainwater is handled during rain-on-snow events, how refreezing is represented and how sublimation is considered), as described in the response to reviewer 1, as well as information on the required model input and how that overlaps with availability of data for the different years. We will also describe the selection of the simulated years better with the information described above at comment 2, summarized in Figure R2.2 and Table R2.1. This should provide context for the years that we chose to simulate as well as on the dataset as a whole.

Thank you for pointing out the mistake in Table 1. It should have been 412 mm and 243 mm for SWI from rain in 2005 and 2010 respectively, and 146 mm and 310 mm for SWI from snowmelt in 2005 and 2010, respectively. We will check all values in this table (and the other tables) before submitting the revised manuscript.

Minor and technical corrections:

5) Title + abstract: 'Snowfall fractions' – since you only clarify in the introduction, maybe another term could be used here, e.g. ratio of snowfall to precipitation. Regarding the title, maybe it needs to be adjusted depending on the changes, e.g. temporal distribution and total input? Or specify what is meant with temporal distribution. Stream discharge – Annual (stream) discharge.

We recognize how the title and abstract could be more explicit, and are open to adjusting these after we have made all the changes in the manuscript.

L13 '..spatial and temporal distribution of precipitation' – add phase of precipitation?

Yes indeed, precipitation phase might also impact stream discharge. We will add this to the abstract as "spatial and temporal distribution of precipitation and precipitation phase"

L68 which catchments?

We think that our findings will be most applicable to other small (<10 km^2), semi-arid, mid-elevation, mid-latitude catchments, and will include that as specification. We suggest these catchment characteristics because we think that 1. similarly sized catchments are more likely to have a similar potential for water storage on the surface and in the subsurface, 2. semi-arid, mid-latitude catchments are likely to have a similar vegetation cover 3. mid-latitude, mid-elevation catchments are likely to have a temporal distribution of water inputs that is similar to that in Johnston Draw.

L71-72: on an annual time-scale is this so different, apart from the effects of snow redistribution? Is this something interesting to show for you analyses, i.e. spatial distribution of rainfall and spatial distribution of snowmelt?

Apart from the snow redistribution, distribution of rainfall and snowfall might be quite similar across the catchment, and in both cases increase slightly with elevation. However, snow redistribution causes surface water inputs to differ across the catchment, even at the rain-snow transition zone. We will adapt the text of this paragraph to make sure that is clear for the reader.

When analyzing the data, we visualized SWI inputs from rainfall and snowmelt separately, but refrained from including these figures in the manuscript. We decided this because rainfall amounts were interpolated following a linear orographic gradient derived from precipitation at the gauges at the upper and lower end of the catchment. Because of this, rainfall distribution across the catchment did not reflect any small-scale variations in the spatial distribution of rainfall, other than that caused by elevation. The effects of wind-redistribution of snow were implicitly included in the spatial distribution of snowfall, which was based on the lidar snow depths, and thus included more fine-scale spatial variation. Hence, a direct comparison with the orographic precipitation gradient might overrepresent the differences. Investigating how the spatial distribution of non-redistributed snowfall and rainfall might differ could be achieved by simulating the snowpack with and without wind re-distribution, but we think this is outside of the scope of the manuscript presented here.

L94: ‘However’ – where does this refer to?

This was meant to refer to the difference in catchment wetness that might exist between rainfall versus snowmelt dominated catchments. We recognize how the writing here might have been confusing, and will remove the word ‘however’ so that the sentence now reads: “Rain and snowmelt inputs might result in similar runoff ratios (discharge/SWI) as long as the overall catchment wetness is similar or if the catchment is wet at key locations for water transport.”

L116-117: did increased ET played a role here?

While the long-term analysis of Nayak et al. (2010) does not comment on increased evaporation or transpiration, Seyfried et al. (2011) states that evapotranspiration is most sensitive to increases in PET (implied by increases in air temperature) during ~4-5 weeks each year in which the plants have developed leaves and sufficient water is available in the soil. Before that time, plants use little water, and after that time, the system is strongly water-limited. Hence, although increased plant water use might be important in some systems, we suspect that it might not strongly affect stream discharge in this region. We will include a small summary of this information in section 2 of the revised manuscript.

L195-196: ‘this uncertainty…. Patterns’ – double with few sentences above

This part of the sentence repeats itself because we wanted to highlight the connection between the intra-annual consistency in snowpack patterns introduced above, and the uncertainty related to using the snow-on lidar from only one year, discussed here. To avoid the exact repetition of words, we will change this part of the sentence in L195 to something like: “… might have induced some uncertainty, but this uncertainty is likely to be small given the consistent spatial snow distribution, and was verified in this catchment …”

Section 3.5 How do catchment precipitation and discharge compare? Are there estimations for ET?

Precipitation and discharge are significantly correlated (R^2= 0.6, p-value = 0.005), which is shown in Fig. 6a of the original manuscript. There are no estimations for ET in this catchment. However, there are estimations for a nearby catchment at slightly higher elevation (1930 m) and that receives less precipitation (Upper Sheep Creek; Flerchinger et al., 1998; MAP: 479 mm versus 609 mm for Johnston Draw). Their measurements showed that evapotranspiration depends significantly on precipitation inputs, and amounted to 58% of annual precipitation for a wet year (703 mm precipitation) and 95% for a ‘normal’ year (482 mm of precipitation). Although estimating ET for the years and catchment presented here is beyond the scope of this effort, we will include the information about Upper Sheep Creek in the discussion.

L223 ‘this pattern was masked by the effects of other processes’ – what is meant here? In general in the results section it would be helpful to indicate better when observations or when simulations are described.

Here we mean the snow redistribution processes, and we further detail measurements in which this process can be observed in L223-228. We agree that being specific is helpful, and we will change the text to “…the snowpack distribution was also affected by wind-driven redistribution of snow. For instance, the snow depths at jdt2 ….”

L236-237 ‘differential melt-out patterns’ – what was compared for that?

We compared the simulated persistence of the snowpack with the persistence of the snow-covered area from the satellite imagery. This comparison showed that the areas that were simulated to be snow-covered longer were also snow-covered longer in the satellite imagery. We recognize that we did not explain that very clearly in the manuscript and will add this to section 4.2 of the revised manuscript.

L267 ‘As a result, average daily SWI rates were higher’ – as a result of what?

We meant to say here that average daily SWI rates were higher as a result of higher snowfall and lower rainfall inputs. We will adapt this reference in the revised manuscript so that this is clear.

L274 ‘whereas roughly 30% of SWI….’ – are delays taken into account, or is meant here the comparsion between SWI from month x to month y and discharge from month x to month y? Are the events where Q is higher than SWI also of interest?

This refers to the comparison of SWI in period x and discharge in period x, and no delays are taken into account. We will clarify this in the methods section of the revised the manuscript. Events where Q is higher than SWI are definitely of interest, but these are not further explored in this manuscript because we did not investigate discharge generation during individual events.

L279 Have you tried plotting % of SWI translated into discharge against temperature (annual, or during growing season?)

We did not do this, but find it an interesting suggestion! Plotting runoff efficiency as discharge/precipitation vs. mean air temperature shows that they are weakly and not significantly correlated (R^2= -0.43, p-value=0.217; Fig. R2.3). Perhaps, this corroborates that evapotranspiration is water-limited in this system rather than energy-limited. We will allude to this small additional analysis in the revised manuscript.

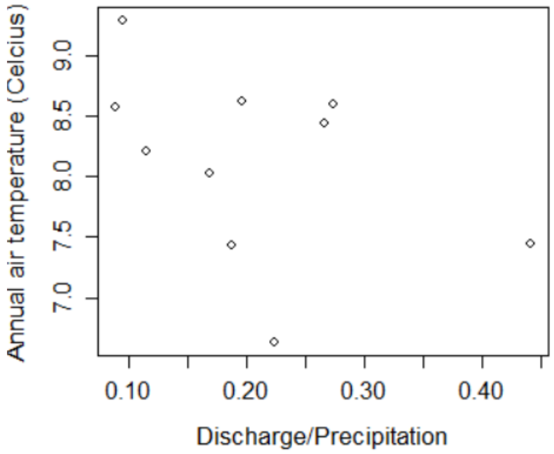

Figure R2.3: Annual air temperature (°C) versus runoff efficiency (discharge/precipitation, mm mm^{-1}).

We also investigated if the runoff ratio was related to precipitation or SWI on the seasonal scale (i.e., spring, summer, fall winter). We found that Q/P was weakly correlated to air temperatures during each season, with the summer period yielding the highest correlation (R^2=-0.54, p-value=0.08). We assume that temporal offsets between snowfall and snowmelt might have led to low correlations in spring and winter. We found strong but insignificant relationships when calculating the efficiency as Q/SWI (up to R^2=-0.72, p-value=0.169, also during the summer season), and suspect that the insignificance of this relationship is likely due to the low number of observations (n=5). Together, these results suggest that temperature likely influences runoff efficiency in the warmer season, but has little effect in the cooler season.

Section 5 – the subsections have no numbering

Thank you. We will number the subsections in the discussion in the revised manuscript.

L357 'This highlights the importance of the temporal distribution of SWI' – also the importance of total water input?

Definitely! Although this was not emphasized in the initial version of the manuscript, we agree that this can be added as a conclusion. We will support this conclusion with the additional heat plot (Figure R2.1) and emphasize this finding in the text.

L360 'events' – throughout the manuscript when using 'event' please check if it is clear why event is meant? Precipitation, rainfall, snowmelt, discharge?

Thanks for pointing this out. We meant 'event' as 'rainfall or snowmelt event' (i.e., an event related to SWI). We now specify the type of event at each of the eight occurrences in the manuscript (e.g., it is written explicitly in the text as precipitation event, rain-on-snow event, snowmelt event…).

L369 'catchment' – sub-catchment?

Yes, 'Treeline' refers to a sub-catchment rather than the entire Dry Creek catchment. We will update this in the revised manuscript.

Discussion on simulated snow depts – could it be extended with a description of the reasons for varying performance for individual years and maybe a hypothesis how such 'bad' simulated years potentially could have influenced the results?

Thank you for this suggestion. We think that a summary of the extensive discussion in our reply to reviewer 1 would be a good addition to the current discussion on simulated snow depths, and aim to include that in the revised manuscript.

In short,

- weighted-average wind directions were similar between most years (246-272° for 2009, 2010, 2011 and 2014), but differed slightly in 2005 (202°).
- we suspect that the combination of a higher snow density (stronger cohesion of snow particles) and lower wind speed (less energy for transport) in 2011 compared to 2009 might have led to less wind-redistribution of snow in that year. This effect would have been exacerbated compared to 2014 because snowpacks in 2014 were much shallower. Since NSE values are based on squared errors, the divergence between the simulated and observed higher snow depths in 2011 would have resulted in a relatively lower performance in that year.

- The snow density, wind speed and wind direction values in 2005 suggest that perhaps, the 2005 simulations might diverge the most from the lidar-derived snow observation in 2009. However, these potential differences will have gone unnoticed because there was only location that recorded snow depths in that year, for which the model performed relatively well (NSE: 0.83).

As to how this might have influenced the results: For years in which the actual snow redistribution was less strong than simulated, snow drifts might have been overrepresented, resulting in a later simulated melt-out date of the snowpack. If the snow redistribution was overestimated in some years, but underestimated in others, this might have resulted in either a stronger or weaker relationship between the snowpack melt-out date and the stream dry-out date.

L419 '..., which influences' – should it be, which may influence? As for example one of your conclusions is that the spatial distribution of SWI stays rather stable over time?

Thank you for this suggestion. Indeed, although we might expect that effects on the spatial distribution could be more severe if snow redistribution patterns are also affected, that is not shown in our work. We will adapt the phrase to "may influence".

L428 -429 Could a short explanation/hypothesis be added why Q was much higher in 2010?

We think that a short explanation can be a nice addition to the conclusions, and will change this part of the conclusions as follows "Despite similar annual SWI (553 vs. 557 mm), snowy 2010 had about twice as much stream discharge as rainy 2005. This is likely related to a higher fraction of SWI occurring in spring 2010 (46%) than in spring 2005 (32%)."

We also checked if the fraction of precipitation occurring in spring was related to annual stream discharge or runoff efficiency, and found a statistically significant positive relationship with the stream dry-out date (R^2= 0.58 p-value=0.06), and a positive, statistically insignificant relationship for annual discharge; R^2=0.43 p-value=0.18). We now include these findings in the heatplot shown in Fig. R2.1.

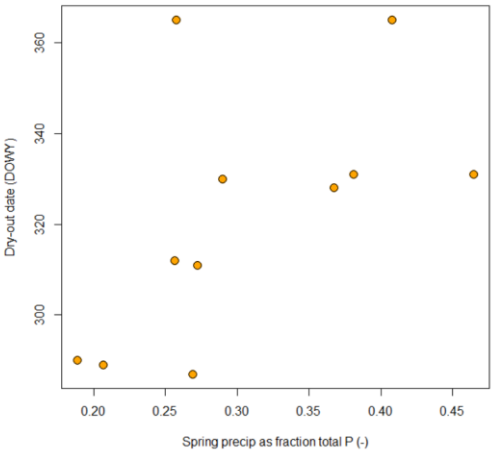

Figure R2.4: Scatter plot of spring precipitation as a fraction of total precipitation and the stream dry-out date (R^2= 0.58 p-value=0.06).

Figure 2e – what do the light coloured pixels mean? Was there no snow cover in the simulations while there was around 0.5 in the satellite observations? Because of the comparison of different years?

The light-colored pixels indicate that the time that an area was simulated to be snow-covered was lower than 0.25 (0-0.25). Indeed, that estimation differs from the fraction of time that these areas were snow-covered based on the satellite imagery (0.5). This difference might be due to a small difference in the mean annual air temperature, which was a bit higher in 2009 than in 2019 (8.0°C versus 6.7°C), which could have resulted in faster melt-out. We will highlight this difference in the text of section 4.2.

For all figures it may be good to not only indicate the year but also its characteristic (i.e. snowy, rainy, wet and dry) in the figure itself instead of the legend.

Thank you for this suggestion. In earlier versions of the figures we indeed included the characteristic of each year in the panel titles, and will re-introduce that for the figures in the revised manuscript.