Dear reviewer,

We thank you for your time spent reviewing our article, and for the comprehensive and constructive comments. Please find our responses (in blue) to your specific comments (in black) below.

Emphasizing novelty of the work

In my opinion, the novelty of the study should be better described. I agree that the focus on the rain-snow transition zone is important and particularly novel, but I would encourage authors to better highlight research gaps and how the study goes beyond to what has been done in the past. Therefore, some additional justification can be added to introduction section (e.g., after research questions).

We are thankful for your comment on clarifying the novelty of our work, and will certainly address this in a revised version of the manuscript. We anticipate elaborating on the following two aspects in a short section after the research questions:

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.

If you have further points in mind that can be emphasized, we welcome your reply.

SWI and model description

Although authors used frequently applied iSnobal/AWSM model, which is well enough described in the literature, it would be good to provide the reader with more specific information about generating snowmelt runoff, which is specifically important for SWI calculation. For example, how does the model calculate snowmelt? For rain-on-snow situations, is the rainwater directly added to SWI at the specific time or is it temporarily stored and delayed in the snowpack? Does model account for refreezing? Does model consider sublimation from snowpack and canopy interception? These details are not fully described in the current manuscript, but I think they might help the reader with better understanding of how the SWI were calculated.

We recognize that our description of SWI was rather brief, and that adding more information on how SWI is generated in the model will be helpful for the reader without referring to other articles. We will add this information to the revised manuscript and will be sure to cover the topics mentioned in your comment. Below you find the answers to the questions posed in your comment, which hopefully gives some insight into what a description might include in the revised manuscript.

1) How does the model calculate snowmelt?

iSnobal solves each component of the energy balance equation for each model time step using the best available estimations of forcing inputs. Melt occurs in a pixel when accumulated input energy is greater than the energy deficit (i.e. cold content) of the snowpack.

2) Is the rainwater directly added to SWI at the specific time or is it temporarily stored and delayed in the snowpack?

Rainfall is only directly accounted as SWI when it occurs over bare ground. During rain-on-snow events, it is included in the energy and water balances. If the energy deficit in the snowpack is exceeded, this results in snowmelt and thus, SWI, but it will be counted as "SWI from snowmelt", because it did not occur over a bare ground pixel.

3) Does the model account for refreezing?

Yes. In order for runoff (i.e., SWI) to occur, the accumulated melt and liquid water content from the previous hour must exceed a defined threshold. Otherwise, the sum of current hour melt and previous hour liquid water content will be carried over into the next hour. If that hour's input energy conditions are negative, that liquid mass is refrozen into the column.

4) Does [the] model consider sublimation from snowpack and canopy interception? Yes and no. Sublimation from the snow surface is computed and combined with the other mass loss processes (evaporation and condensation) as a model output term. As for canopy interception, iSnobal only predicts snowpack on the ground and is not a comprehensive land surface model. Interception must be handled a priori when developing the model forcing input. Although not accounting for the latter introduces

some uncertainty, we expect this to be small with the vegetation types in Johnston Draw.

Single lidar observation and poor model performance in WY2011

L 197: As authors correctly stated, the use of only one lidar survey to describe the snowpack spatial distribution for all study years brings some uncertainty. I see the point that the topography is the main control of snowpack variability. Nevertheless, the meteorological controls might be important as well, such as wind speed and direction influencing snow redistribution and accumulation on leeward sites of slopes. What is the prevailing wind direction? And was it same for all years during snowfall events (and thus likely causing same snowpack distribution)? I would like to see a bit more discussion related to the topic.

We investigated these points further (see below), and will add this discussion to the text of the revised manuscript.

Table 2, Fig. S4: The model performance for north-facing stations and in the "Upper region" (Table 2) in the water year 2011 is relatively poor when comparing simulated and observed SWE values. In addition, even for one single station, simulations for some years are well enough, while this is not the case for another years (e.g., jdt1 and jdt4). Is there any explanation for both temporal and spatial differences in model performance? How confident are observed SWE data for individual stations?

We agree that using a single lidar survey observation raises the question if this observation is truly representative of the snowpack distribution during all years. To address this concern, and to further investigate differences in model performance between years, we now calculated the average wind directions, wind speeds and snow densities for all events during which the snowfall fraction was higher than 0.2 (i.e., 20%) in each year. We used the observational wind speeds and directions from wind-exposed station jdt124, which is located close to the top of the catchment, because this station is most representative of wind along the ridge/scour zone. We computed the averages by considering the impact of each event equally, but also by calculating a weighted average based on the amount of precipitation and snowfall fraction (Table R1.1). We also included a summary of the wind speeds and directions during the 2004-2014 data record for the entire period and during storms and storm-free periods (Figure R1.1).

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 might explain the strong underpredictions of snow depths at north-facing and high-elevation sites (jdt3, jdt4, jdt5 and jdt124b). 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.

Table R1.1: Average and weighted average of snow densities (Density, simulated) and wind speed (W_s, observed) and direction (W_d, observed) during events with an average snowfall fraction of more than 0.2 for each water year.

WY	Average			Weighted average		
	Density (kg m^{-2})	W_s (m s^{-1})	W_d (°)	Density (kg m^{-2})	W_s (m s^{-1})	W_d (°)
2005	124	4.1	187	162	4.8	202
2009	102	5.6	245	102	6.5	252
2010	24	6.6	269	45	8.1	272
2011	117	5.5	232	122	5.7	246
2014	115	6.0	258	126	6.1	266

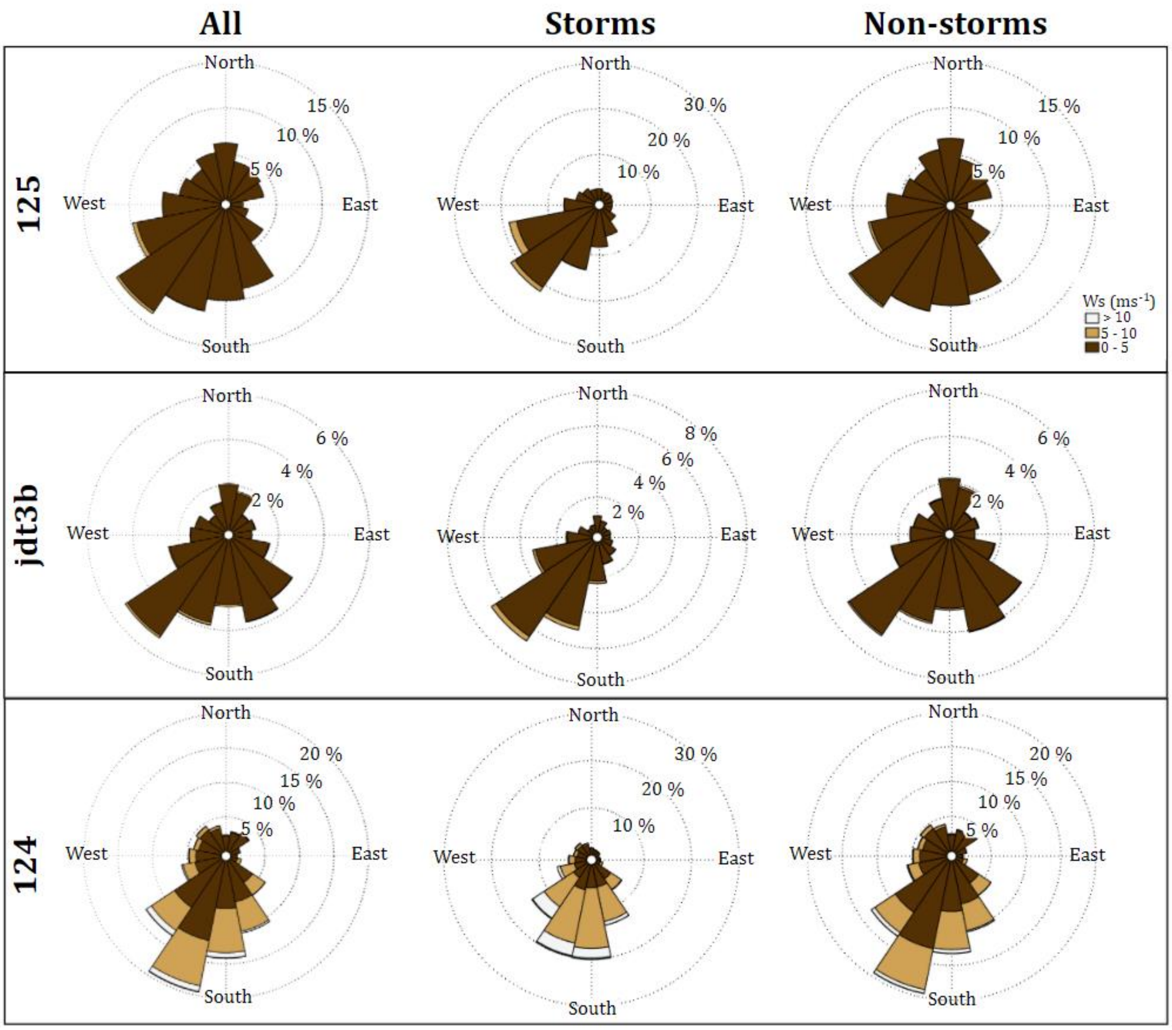

Figure R1.1: Wind roses for stations jd125 (near the catchment outlet), 124 (near the ridge) and jdt3b (a mid-elevation station on the south-facing slope), compiled with data from 2004-2014 (Godsey et al., 2018). The left-hand column includes all measurements, whereas the center and right-hand column only include measurements during storms and storm-free periods, respectively. White indicates higher (> 10 m s^{-1}), orange intermediate (5-10 m s^{-1}) and brown lower (0-5 m s^{-1}) wind speeds.

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). We think that these additional values give more insight into differences between years, and we will describe these in the revised manuscript. The varying performance for simulations at lower stations (jdt125 and jdt1) remains unsolved. We suspect that this might be related to inaccuracies in calculating the phase of precipitation, which would most strongly affect lower elevations at which the phase shifts more often from rain to snow.

We would like to emphasize that despite the low performances for some years and locations, the normalized snow depths were largely acceptable (only five out of 40 year/location-combinations had an NSE_{norm} value below 0.5), which lends confidence that the simulation of ablation and accumulation processes in the model is reasonable.

Regarding the question of how much certainty we have in SWE observations at individual stations: firstly, because only snow depths are available and not SWE, we know that differences in snow density could introduce mismatches between the observed and simulated depths (explained in the manuscript in L385-388). However, in one year at one station, the predicted snow depths were up to 30 cm lower than the observed snow depths, which clearly exceeds estimated offsets due to snow density. Secondly, differences might be introduced because the footprint of the sensor and the cell-size of the model don't match. As installed, the sonic depth sensors have a footprint of ~1-3 m, whereas the simulated snow depths reflect a 10-m resolution grid cell. This is also mentioned in the manuscript L383-385.

Yearly snowfall fractions as a metric

The conclusion that the snowfall fraction is not correlated to annual runoff or day of stream drying is certainly important, but maybe not such surprising. The snowfall fraction does not contain the information about total amount of snowfall, but only its relation to the total amount of precipitation. It means, that a year with high snowfall fraction is not necessarily the year with overall high snowfall. Therefore, it would be maybe interesting to select more characteristics describing the snow conditions in different years (such as amount of snowfall during cold season, annual maximum SWE, amount of snowmelt in spring etc.) to better show whether or not the cold season snowfall could positively influence the stream drying compared to the same amount of rain. Perhaps, the results can be shown in some table (heatmap) of paired correlations between individual characteristics.

We agree that the total amount of snowfall is important in addition to the snowfall fraction and will include this in a revised manuscript. We also appreciate your suggestions for other snowpack and precipitation characteristics to consider in addition to the annual snowfall fraction. Indeed, winter, spring, and the sum of winter and spring snowfall can give a more nuanced analysis of timing of precipitation, and also fits well with our comparison to timing of SWI. Including several metrics might indeed be efficiently shown in a heat map, and we explored this for several additional metrics in Figure R1.2. For the years in which we have sufficient data available, we will also further investigate the relationships between SWE and the melt-out date and melt-rates, such as in Trujillo and Molotch (2014), and potentially include these as variables to explain stream discharge and/or drying. For now, we calculated the (linear) melt rate based on the amount of time it took to completely melt the snow pack from 40% snow coverage (i.e., SCA = 0.4), and represented snow-coverage as the amount of days that the catchment was covered more than 50% (SCA > 0.5). We also included a flashiness index for SWI (Richards-Baker Flashiness Index; Baker et al., 2004). Scatter plots for significant correlations are shown in Figure R1.3.

Because snowfall fraction at the rain-snow transition varies widely from year to year and is one of the most visible manifestations of hydrological changes in this zone that may affect aboveground storage, we would still like to give extra attention to this metric. Although it may be "unsurprising" to some, others may expect it to directly impact stream discharge and/or stream drying, especially in temporally coarser models (e.g., Berghuijs et al., 2014; an analysis that relies on the Budyko water balance framework) or long-term analyses (e.g., Irannezhad et al., 2014).

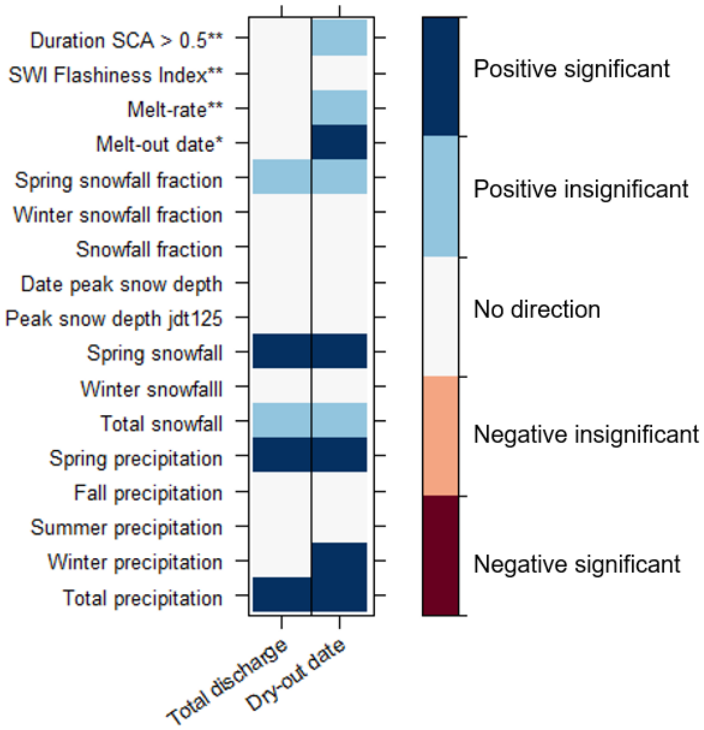

Figure R1.2: Heatplot showing Pearson correlation coefficients (α=0.1) for comparisons between total discharge, the stream dry-out date and precipitation and snowpack metrics. Significant correlations are marked in dark red (negative) and dark blue (positive), whereas insignificant correlations and correlations without a direction ($-0.3<R^2<0.3$) are marked in light blue (positive), light red (negative) and white, respectively. 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), and for the SWI flashiness index, the melt rate and the number of days when at least half the catchment was snow-covered we used only the years that were simulated (n=5, marked with two asterisks).

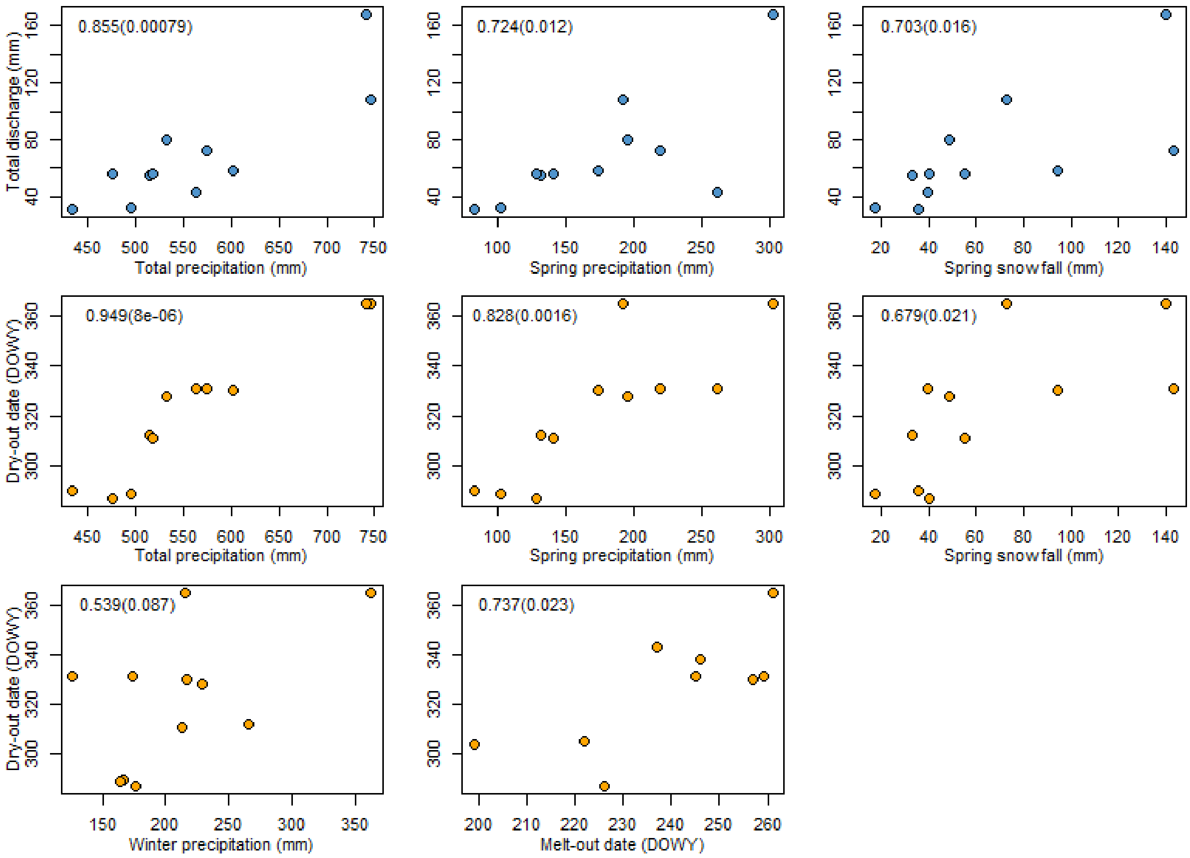

Figure R1.3: Scatter plots of statistically significant comparisons between precipitation and snowpack metrics and total discharge (blue circles) and stream drying (orange circles). Pearson correlation coefficients (R^2) are given at the top right of each panel, with the corresponding p-value in brackets.

Memory effect

L 286-288: This part would maybe deserve a bit more attention since it touches the important issue of catchment storage and its “memory effect”. I found this partial analysis interesting (despite the fact that results did not confirm an effect of “previous water year precipitation”). Therefore, I suggest some extension of the related text.

We appreciate this suggestion and will include more information about the (lack of) memory effect we found in Johnston Draw. Because the stream dries at the outlet in 16 of the 18 years from 2003-2020, it may be difficult to detect any effect of the previous water year precipitation from surface flow data alone. Unfortunately we do not have any groundwater level data to further investigate if and how the memory effect is reflected in subsurface water storage. The frequent stream drying and high potential evaporation rates in this semi-arid, high desert system do suggest that the water that is accessible to plants will be used in the growing season, reducing any memory effect from the shallow, ‘active’ subsurface storage.

Proper SWI accounting to compare with drying

L 297-300: For day of stream drying, would it make more sense to account for sum of SWI preceding the day of stream drying instead of annual sum of SWI?

Thank you for this suggestion. Indeed, any SWI that occurs after the stream dries out cannot have any effect on the date of stream drying. In this catchment, however, the dry summers usually result in very little additional SWI (2.0%, 0.2%, 1.7%, and 0% of annual SWI for 2005, 2009, 2010 and 2011, respectively), meaning the impact of having used annual SWI during these years did not lead to different conclusions. In 2014, the difference was larger (16.5% of annual SWI occurred after the stream dried), which is partly because the dry-out date of the stream was ~1.5 month earlier than in the other years (13 July vs. 25-26 August). We will adapt the metric in the revised manuscript.

Figures that better illustrate main findings

Although, I found the reasoning presented in results and discussion sections correct, the supporting illustrations are, in my opinion, less informative and I am not sure whether they fully support all the results and interpretation. For example, one of the main conclusions is that temporal distribution of SWI is more important than its total amount. While I agree with that, it is difficult to me to clearly see this in figures which mostly shows only time series (Figs. 4 and Fig.5). I do not have any clear suggestion how to make figures more informative and supporting the results, but I would encourage authors to reconsider their illustrations and perhaps add another figure which would better show how the timing of SWI influence the runoff response.

Thank you for this remark. It is for us very important that we capture our main findings in the figures and appreciate your comment to make us aware of this. We will have another look at the figures and brainstorm about how to better represent our finding regarding the timing of SWI and runoff response. Perhaps, a combination with the heat plot suggested earlier (see Figure R1.2 for a preliminary version) might be a helpful way to visualize this.

Technical corrections

L 116: The decrease in streamflow should be expressed in mm/decade to be comparable with other characteristics.

Thank you for pointing this out. We now divided the trend (-0.75 * 10^6 m^3/decade) by the surface area (54.44 km^2) and will include the new trend as -13.8 mm/decade in the text of the revised manuscript.

L 138: “stage height-discharge relationship”. Maybe more common term “rating curve” would be better.

We don't hold any preference between the two terms since we are familiar with both, but will adapt it to rating curve in the revised manuscript to avoid confusion with any future readers.

L 193: "Trujillo et al. (2019, manuscript in preparation)". As it seems from references, this paper has been already published.

The Trujillo et al., (2019) reference refers to an AGU abstract, which has been presented at the 2019 AGU Fall Meeting. The corresponding manuscript is still in preparation. We will add "AGU Fall Meeting" to the bibliography so that this is clear. Also, there should have been a semicolon after 2019. The reference now reads "2019; (in preparation)".

Fig. 6a: The annual discharge is related to the precipitation at jdt125 climate station. Why not to show catchment mean precipitation instead? If I understood correctly, the model interpolates stational data to a catchment scale using some kind of elevation dependency. Therefore, to show catchment precipitation in Fig. 6a makes more sense to me to make it better comparable to catchment runoff.

We used precipitation at the climate station rather than simulated precipitation so that we could include additional years in the dataset (2004-2014) without having to run the model for the additional years. Catchment-average precipitation for the years that we did model was linearly related (R^2: 0.93) to the precipitation at jdt125 (Figure R1.4). That precipitation at this lowest elevation station is slightly lower than the simulated catchment-average precipitation, based on four stations, is not surprising, since precipitation increases with elevation. All in all, the strong correlation indicates that using precipitation at this station is not expected to lead to a different interpretation. We will include Figure R1.4 (below) in the supplementary material, and refer to it in the text and/or caption of Figure 6 in the revised manuscript.

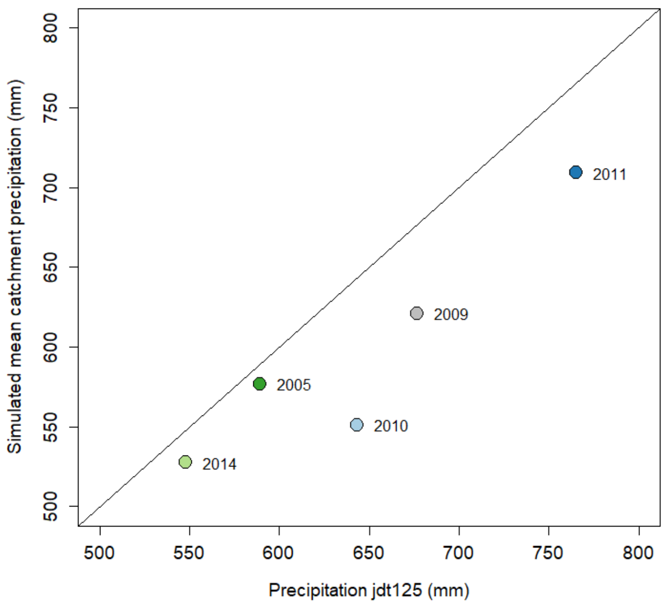

Figure R1.4: Precipitation at jdt125 (the low elevation precipitation gauge) versus the simulated mean catchment precipitation for the years that were modeled.

Fig. 6b: What the triangles represent? Maybe, there is a mistake in the figure as they represent “other years”, but different symbol is used in the legend.

This was a mistake in the legend. The diamonds in the legend (other years, 2016-2019) should have been reversed triangles. We will adapt this in the revised manuscript, and will check the legends and symbols in all figures before resubmitting.

On behalf of all authors,

Leonie Kiewiet

References

Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W.: A NEW FLASHINESS INDEX: CHARACTERISTICS AND APPLICATIONS TO MIDWESTERN RIVERS AND STREAMS, J Am Water Resources Assoc, 40, 503–522, https://doi.org/10.1111/j.1752-1688.2004.tb01046.x, 2004.

Berghuijs, W. R., Woods, R. A., and Hrachowitz, M.: A precipitation shift from snow towards rain leads to a decrease in streamflow, Nature Clim. Change, 4, 583–586, https://doi.org/10.1038/nclimate2246, 2014.

Godsey, S. E., Marks, D., Kormos, P. R., Seyfried, M. S., Enslin, C. L., Winstral, A. H., McNamara, J. P., and Link, T. E.: Eleven years of mountain weather, snow, soil moisture and streamflow data from the rain–snow transition zone – the Johnston Draw catchment, Reynolds Creek Experimental Watershed and Critical Zone Observatory, USA, Earth Syst. Sci. Data, 10, 2018.

Irannezhad, M., Marttila, H., and Kløve, B.: Long-term variations and trends in precipitation in Finland, Int. J. Climatol., 34, 3139–3153, https://doi.org/10.1002/joc.3902, 2014.

Trujillo, E. and Molotch, N. P.: Snowpack regimes of the Western United States, Water Resour. Res., 50, 5611–5623, https://doi.org/10.1002/2013WR014753, 2014.