Review of paper hess-2019-82 “**The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient”** by Keith S. Jennings and Noah P. Molotch

Jono Conway

This paper presents a systematic evaluation of the impact of precipitation phase partitioning on modelled snowfall and snowpack evolution. Multi-year datasets from 11 stations across 5 locations in the western United States are used to drive simulations with a sophisticated snowpack physics model. The effect of parameter and algorithm choice are assessed for a range of commonly used parameterisations. The authors relate the modelled sensitivity to average climate characteristics. Snowfall and maximum accumulated snow in warmer maritime locations with high precipitation and winter temperatures between 0 and 4 degrees Celsius are found to be most sensitive to precipitation partitioning, while snowfall in colder inland locations are found to be less sensitive.

The manuscript is well written and with good figures and a clear systematic structure. It addresses a topic of high interest and relevance internationally. However, there are some areas that should be addressed before the paper could be accepted for publication.

Major comments

While the paper is framed as a comparison of methods used to partition precipitation, the results mainly reflect the range of Ta thresholds used (0 to 3 C) rather than the choice of parameterisation. This is in part due to the use the range metric on results that are generally are bounded by the two extreme Ta thresholds. The abstract and conclusions should reflect this (i.e. being explicit about choice of parameter values and/or parameterisation rather than using the ambiguous term “method”. If the authors wish to make general statements, then using “precipitation partitioning” would be more appropriate. It is well established the Ta alone is a poor predictor of precipitation phase, so to really compare methods, those that perform poorly against observed SWE (e.g. Ta0 and Ta3) should be removed from the analysis. This would highlight the differences induced by using different parameterisations that have a sound physical basis. If the Ta0 and Ta3 options are to be retained, then further justification for their inclusion should be given in the methods section. The dependence of the results (especially the range of Ta with a large range in modelled snow) on the range of Ta thresholds used should also be discussed. Perhaps the use of a standard deviation or similar metric rather than a range metric would put the focus on the choice of parameterisation. Further specific comments address this issue.

Timing and magnitude of SWE ranges seem mainly related to snowfall and accumulation, whereas as range of melt rate does not have high sensitivity or clear relation to climate. This should be clearer in the abstract and conclusions.

The abstract and conclusions need to highlight the novel aspects of the results presented here and provide clearer recommendations for future research. While the analysis is comprehensive, the result is not entirely new and, in my opinion, there are other results in the paper that could (and should) be highlighted in addition to the main result that the relative differences are largest in maritime snowpack. For example, the fact that using threshold or ranges for Ta (for the same 50% crossover) do not produce large differences in the snowpack, or that partitioning choice has little effect on snowmelt rate and the effects are dominated by snowfall. At present, the authors recommendations for future researchers are unclear.

The use of multiple linear regression is probably not appropriate here, but if retained should be presented and discussed more fully.

Specific comments (page-line)

1-15 please be clear the study modelled non-vegetated snowpacks only.

4-3 Given that they form a key part of the results, please include average values for Tw and Td in Table 1.

7-26 The large bias in LWin is concerning – perhaps the influence of vegetation on the measurements whereas LWin is modelled for non-vegetated location? This should be discussed when presenting the validation results in Figure A2.

17-8 “80.1% of the variance in annual snowfall fraction standard deviation” – the figure caption and methods describes this as the “range in annual snowfall fraction” – please clarify which it is and correct.

18-1 Figure 6 and 7 – given that the range in snowfall fraction is driven primarily by the two extreme air temperature threshold methods (Ta0 and Ta3) these results are presumably quite sensitive to the choice of the Ta thresholds? Please discuss and if possible show the sensitivity of the results to the choice of threshold.

18-6 Looking at the figure, it seems that a multiple linear regression may not be appropriate. There seems to be two groupings – highly sensitive warm and wet locations, less sensitive drier locations that span both warm and cold locations. Also, given that the equation is not presented nor used further, and the issues discussed with extrapolating the equations, please consider removing the regression. If it is retained, please present the equation and display contours of predicted values on Figure 8 so that the reader can visualise the predicted relationships.

19-10 The validation results presented in the appendix should be included in the results or methods section, especially as they form part of the discussion, rather than simply an intermediate methodological step.

20-3 “In that context, one can consider the RegBi model as a baseline given its top rank in a Northern Hemisphere precipitation phase method comparison”. Please describe and discuss the results presented here (figure A1) that seem to show similar performance for a range of methods that incorporate humidity information. The discussion as it is not balanced and does not accurately reflect the results presented. Please revise.

20-11 “a referendum.” This does not seem an appropriate term – please revise. You could either give an expert view based on the results presented here, or cite others work.

20-24 “Therefore, our use of a single model may overestimate or underestimate the sensitivity of snow cover evolution to precipitation phase method at certain sites and points in time.” This statement is very broad - more effort is needed to quantify and discuss the uncertainty of the model simulations.

21-23 “These large variations in snow cover evolution were likely due to the combined effect of reduced frozen mass entering the snowpack and subsequent changes to the snowpack energy balance”. More detailed results are needed to support this statement. For example, the change in snowfall mass and albedo could be shown to illustrate the importance of the direct and indirect effects on snowpack mass balance.

22-3 “In this context, the precipitation phase methods that produced more rainfall affected snow cover evolution not just through reduced frozen mass but also through changes to the snowpack energy budget.” These results are not shown here (they could be?) so this statement is speculation. Please revise.

22-25 “winter and spring average Ta values (0°C–4°C) that lead to the greatest uncertainty in rain-snow partitioning,” I would argue that the uncertainty is not in the actual rain-snow partitioning, but rather due to the use of an inappropriate parameterisation (only Ta) which requires a wide range of parameter tuning. Please revise.

23-1 Please mention that no clear relationship was found for snowmelt rate in the conclusions – this is still a key result and an important caveat to the earlier statement that “precipitation phase method introduced significant variability into simulated snow accumulation and melt”.

23-30 How was the r2 calculated here? the average r2 of hourly SWE/snowdepth or something else? Please include in the text and figure caption.

24-1 Given the poor performance of some methods (Ta0 ,Ta3, Tr0) should they be excluded from the analysis? If not, further discussion is needed.

24-6 “at all stations.” Given that SWE and snowdepth are only presented for some sites in Figure A2, I presume not all sites contribute to averages here? Please list the sites that contribute to each of the SWE and snowdepth validation statistics in the text or caption.

Figure A2 – why is the snowdepth bias 0 for the JD sites?

Editorial comments:

10-6 “daily Ta and RH” do you mean ”daily average Ta and RH”?

16-5 “not computed because for” -> “not computed for”