Dear Dr. Markus Hrachowitz,

Please find included in this file a list of the major changes made to the manuscript along with our responses to the reviews by Drs. Schaefli and Conway and a tracked changes version of the revised manuscript.

Per your recommendations on major revisions we have substantially increased the text and analysis devoted to the observational SWE and snow depth data as well as adding material on transient snow.

We are thankful for the time invested by you and the reviewers. We feel your suggestions have substantially improved the revised manuscript.

Best regards,
Keith Jennings and Noah Molotch
2019-07-16

Changes made to manuscript (for minor text edits as well as figure and section renumbering, please see tracked changes manuscript):

1. Updated abstract text per reviewer recommendations.
2. Updated introduction text per reviewer recommendations.
3. Added wet bulb and dew point temperature data to Table 1 per Dr. Conway's recommendation.
4. Added more information on longwave radiation and model validation to Sect. 3.1 per reviewer recommendations.
5. Added motivation for using the different T_a thresholds per reviewer recommendations.
6. Added section 3.4 on identifying and deviating from an optimized threshold per Dr. Schaefli's recommendation.
7. Removed multiple linear regression from Sect. 3.5 per Dr. Conway's recommendation.
8. Moved model validation to Sect. 4.1 from Appendix and updated the text and figures. Note: the original snow depth validation performed at JD stations was done in cm. This was changed to mm to be consistent with the other sites.
9. Added new paragraph to Sect. 4.4 and that includes further text on SWE/depth observations along with how transient snow is particularly affected by method choice in terms of snow cover duration. Also added Fig. 8 to display this information visually.
10. Added Sect. 4.5 and Fig. 9 per Dr. Schaefli's recommendation.
11. Changed Fig. 12 to remove multiple linear regression and add loess surface.
12. Rewrote Sect. 5.1 and updated other discussion sections per review recommendations.
13. Updated conclusions per reviewer recommendations.
14. Added model, code, and plot color ramp info to "Code and data availability" section.
15. Added "Competing interests" section.
16. Added Table S2 and Figs. S12 and S13 to the supplement per Dr. Schaefli's recommendation. Note: Figs. S12 and S13 are different from R1 and R2 in the response to Dr. Schaefli because our original analysis included snow-free periods. These were removed and the snowfall frequency curves were recalculated.

We thank Dr. Schaefli for their thoughtful review. Our responses can be found in blue throughout the following text. Please note, tables and figures specific to this response document are given with the prefix R (for example, Table R1 in the comment below). Tables and figures in the manuscript are referred to by numbers only.

This well written paper analyzes a key question for snow hydrology, which is the im- pact of precipitation phase algorithms on snow water equivalent (SWE) modelling in different climates. The paper studies four more or less different methods of precipita- tion phase computation (each with different portioning parameters) and assesses the impact of the methods on different snow accumulation and melt metrics, obtained with the model SNOWPACK at five different locations in the US. The methods are based on temperature thresholds and on bilinear regression. The analysis gives an answer to the general question of how important it is to carefully choose the precipitation phase method for different climates.

A drawback of the study is that it is purely simulation-based and does not use observed SWE data to push the study further. In fact, with the observed SWE data and SNOW- PACK, it might have been possible to estimate actual daily or hourly snow accumulation amounts and compute best parameter values for the studied precipitation phase meth- ods at the selected stations. This way, it would have been possible to judge how critical deviations from these best estimates would be at the different sites. In other words, this would allow to answer questions like "how critical is it to have a 1°C error in the air temperature threshold at a warm site as opposed to a cold site"? "How important is it to use dew point or wetbulb temperature at warm sites versus at cold sites?"

We agree this is a drawback of both this study and many other snow modeling research projects. There are, unfortunately, scant direct observations of precipitation phase in mountain regions. One of the few studies we are aware of that uses observations of precipitation phase—in this case snow board measurements—showed rain-snow partitioning errors can lead to significant biases in modeled snow accumulation at a site with a maritime climate similar to the HJ Andrews and Southern Sierra (Wayand et al., 2017). There is also evidence suggesting an optimized air temperature threshold varies throughout the snow season (Storck et al., 2002), meaning no single air temperature threshold (or range) would be applicable across sites and times.

At our study sites, there are no direct observations of precipitation phase, but we were interested in pursuing your question further. Table R1 below shows the optimized rain-snow air temperature threshold using four different data sources for each station. The second and third columns (Map and Obs.) correspond to data from earlier work that examined the spatial variability of rain-snow partitioning across the Northern Hemisphere (Jennings et al., 2018). The methods, quoted from the paper, are as follows:

> *"To construct a spatially continuous 50% rain–snow T_s* [air temperature] *threshold product across the Northern Hemisphere, we applied the optimized bivariate model to the MERRA-2 gridded reanalysis dataset*[63,64]*. Hourly 2 m T_s, specific humidity (q), P_s, and precipitation data were accessed from 1980 through 2007 and summarized to a daily time step. RH was calculated from the MERRA-2 data using an empirical equation as a function of q, P_s, and T_s. Daily snowfall probability was then simulated for each grid cell using the bivariate model when precipitation was greater than 1 mm and T_s fell within the range of −8 to 8 °C. We then calculated the 50% rain–snow T_s threshold by fitting the hyperbolic tangent to binned estimates of snowfall frequency per MERRA-2 grid cell using Eq. 1."*
>
> *"We classified precipitation reports as either rain or snow using the World Meteorological Organization precipitation phase categories described in detail in Dai*[40,61]*. Precipitation amounts were not included in the dataset and we removed sleet as well as potential mixed-phase observations from the analysis because the relative proportions of solid and liquid precipitation during such events were not reported (i.e., it was impossible to quantify the amount of precipitation falling as snow versus rain). The classification of precipitation events was then used to quantify the rain–snow frequency per 1 °C T_s bin from –8 to 8 °C at each station. In other words, if there were 100 total precipitation observations from 1 to 2 °C, 75 of which were snow, the snowfall frequency in that bin would be 75.0%. We then calculated the 50% rain–snow T_s threshold for each station using the approach of Dai*[40]*, where a sigmoidal curve is fit to observations of snowfall frequency per 1 °C T_s bin from –8 to 8 °C using a hyperbolic tangent function:*
>
> $$T_{50} = \frac{\tanh^{-1}\left(\frac{F}{a} + d\right)}{b} + c \tag{1}$$
>
> *where T_{50} equals the 50% rain–snow T_s threshold (°C), F equals snowfall frequency (in this case 0.5, dimensionless), and a, b, c, and d are the fitting parameters (dimensionless)."*

The fourth and fifth columns in Table R1 use changes in SWE and snow depth to estimate a rain-snow air temperature threshold. We used a modified version of the approach of Rajagopal and Harpold (2016) to predict precipitation phase by designating a daily increase of SWE or snow depth as snowfall and a zero change or decrease as rainfall when precipitation was greater than

2.54 mm and SWE or snow depth was greater than 0 mm. As with the Map and Obs. methods detailed above, we then binned snowfall frequency per 1°C air temperature bin (Figures R1 and R2) and computed the rain-snow air temperature threshold using Eq. 1 above. The SWE approach yielded values that approximated the Map and Obs. methods, but the depth-derived values were significantly lower. We would thus argue that this method was not appropriate for our purposes, although previous work has shown it to reasonably estimate precipitation phase at subdaily time scales (e.g., Marks et al., 2013; Zhang et al., 2017).

Table R1. Optimized rain-snow air temperature thresholds for each station in the study using four different data sources: 1-Map) The spatially continuous threshold map from Jennings et al. (2018) created using reanalysis data from MERRA-2 and the bivariate binary logistic regression model; 2-Obs.) The observed threshold from the closest meteorological station (Jennings et al., 2018); 3-SWE) The threshold inferred from changes in SWE at each study station (Fig. R1); 4-Depth) The threshold inferred from changes in snow depth at each study station (Fig. R2). An NA indicates there were insufficient data to estimate the threshold from SWE and/or snow depth.

	Optimized rain-snow air temperature threshold (°C)			
Station	Map	Obs.	SWE	Depth
HJA-CEN	1.19	1.12	1.29	-0.24
HJA-VAN	1.19	1.12	0.8	-0.84
HJA-UPL	1.19	1.12	-0.4	-0.81
SSC-LWR	1.7	1	NA	0.14
SSC-UPR	1.7	1	0.87	-0.34
YOS-DAN	2.21	2.78	NA	NA
JD-125	2.25	1.25	NA	-0.97
JD-124b	2.25	1.25	NA	-1.91
JD-124	2.25	1.25	NA	0.41
NWT-C1	2.84	2.34	3.57	NA
NWT-SDL	2.84	2.34	NA	NA

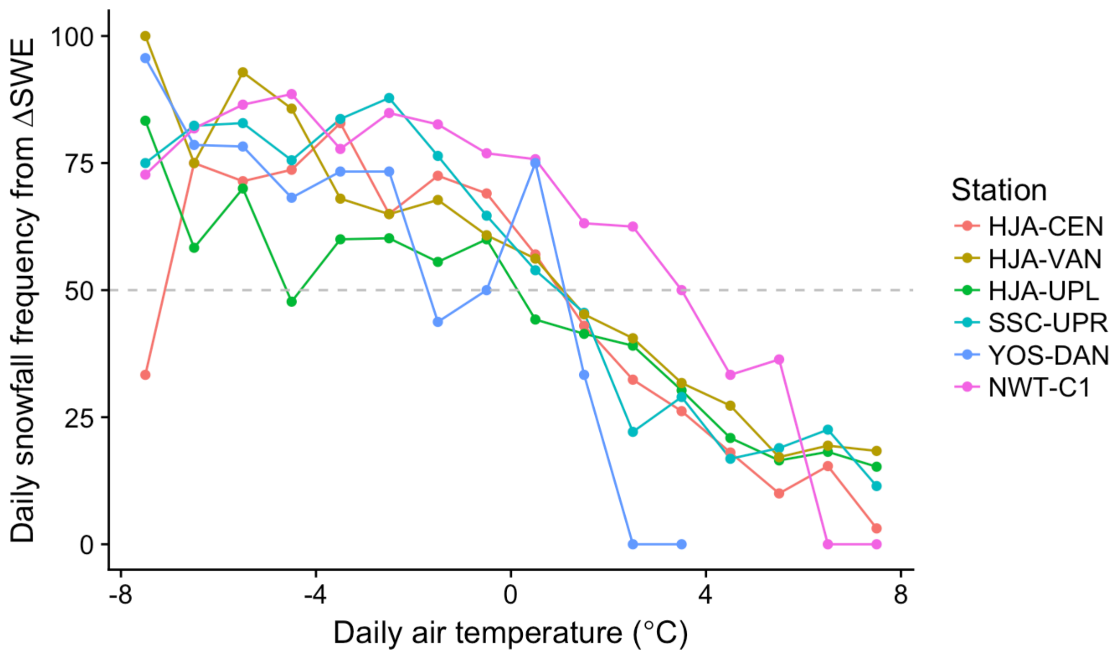


Figure R1. Snowfall frequency per 1°C air temperature bin as computed from SWE data. On days with precipitation > 2.54 mm, an increase in SWE was designated as a snowfall event, while a zero change or decrease in SWE was designated as rainfall.

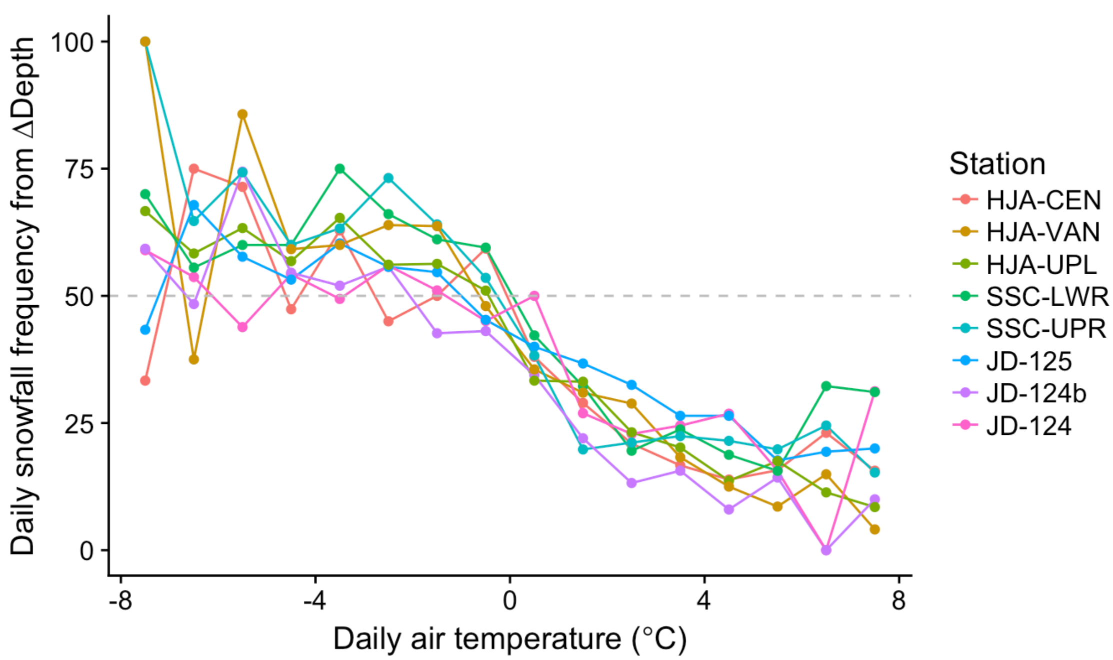


Figure R2. Snowfall frequency per 1°C air temperature bin as computed from snow depth data. On days with precipitation > 2.54 mm, an increase in snow depth was designated as a snowfall event, while a zero change or decrease in snow depth was designated as rainfall.

Returning to the question of "how critical is it to have a 1◦C error in the air temperature threshold at a warm site as opposed to a cold site," we analyzed the effect of deviating by 1°C from the mean threshold as calculated from the Map and Obs. columns in Table R1. In this context we rounded to the nearest integer degree to be consistent with our thresholds, giving the HJA stations a 1°C threshold, SSC a 1°C threshold, YOS a 2°C threshold, JD a 2°C threshold, and NWT a 3°C threshold. Because we did not include a 4°C air temperature threshold in our phase methods, we could only analyze a negative deviation at NWT. In Table R2 below, we present the mean peak SWE, peak SWE day of water year (DOWY), and snow cover duration (SCD) using the optimized air temperature threshold (center column, abbreviated Thresh.), the optimized threshold minus 1°C (left column, Thresh - 1°C), and the optimized threshold + 1°C (right column, Thresh + 1°C). Consistent with our findings in the paper, the warm maritime HJA and SSC stations are profoundly affected by deviations from the optimized threshold. Differences at these sites produced by deviating by ±1°C from the optimized thresholds range between 141 and 403 mm for peak SWE, 1 and 16 d for peak SWE DOWY, and 9 and 29 d for SCD. Compare this to 1 to 10 mm for peak SWE, 0 to 1 d for peak SWE DOWY, and 1 to 5 d for SCD at the YOS and NWT stations. The consistent story is again that threshold choice makes a much larger impact at a warm site relative to a cold one.

Table R2. Mean peak SWE, peak SWE DOWY, and SCD at the study stations using an optimized air temperature threshold as well as -1°C and +1°C deviations from the threshold.

	Mean peak SWE (mm)			Mean peak SWE DOWY (d)			Mean SCD (d)		
Station	Thresh - 1°C	Thresh.	Thresh + 1°C	Thresh - 1°C	Thresh.	Thresh + 1°C	Thresh - 1°C	Thresh.	Thresh + 1°C
HJA-CEN	414.2	528.9	611.8	128	142	144	144	159	172
HJA-VAN	564.3	645.3	726.6	132	134	145	164	173	184
HJA-UPL	984.2	1165.5	1387.3	160	166	173	190	202	210
SSC-LWR	401.5	535.6	624.6	154	161	162	137	146	151
SSC-UPR	508.4	585.4	649.7	154	155	155	142	147	151
YOS-DAN	668.8	677.8	678.8	169	169	170	206	209	209
JD-125	77.2	89.6	99.9	116	116	116	75	83	97
JD-124b	180.2	191.7	203.8	124	126	127	122	131	134
JD-124	72.3	81.5	87.3	128	115	115	78	81	92
NWT-C1	400.1	406.7	NA	204	204	NA	224	229	NA
NWT-SDL	914	914.6	NA	225	225	NA	240	241	NA

For the final question, "How important is it to use dew point or wetbulb temperature at warm sites versus at cold sites?", we would argue the best practice is to use a humidity-based temperature metric at all sites. Such methods better represent precipitation and produce better model outcomes (e.g., Ding et al., 2014; Harder and Pomeroy, 2013, 2014; Harpold et al., 2017;

Jennings et al., 2018; Marks et al., 2013). The bivariate binary logistic regression model, which performed best relative to other methods when compared to precipitation phase observations in a previous study (Jennings et al., 2018), produced snow cover metrics similar to the optimized threshold at most stations. It produced mean peak SWE, peak SWE DOWY, and SCD biases (relative to the optimized threshold) of -18.0 mm, 0.5 d, and -1.9 d, respectively.

Please note, we have not added the above material to the manuscript yet because it is consistent with the findings already presented in the submitted document. If you find this material worthy of inclusion, please let us know and we can add it as either supplementary material or as an appendix.

This having said, the study is nevertheless worth publishing and interesting for the readers of HESS. Below some general and detail comments.

General comments

I would not say that a study tests 12 different methods if only a few methods are tested with different parameter values; this oversells the study in the abstract. I would in fact say that the study tested four different methods: based on air temperature (with different 50% thresholds and different transition ranges, some of the ranges being 0), based on dew point and wet bulb temperature and based on binary regression.

Fair point. We have updated the text (see response to detailed comments below) to say we tested 5 different methods (counting the range as a different method than the threshold because the former produces mixed precipitation and the latter does not).

A key analysis of the paper is the one of "Climatic controls on precipitation phase method sensitivity".(section 4.4); it analyzes how the results vary with air temperature. Air temperature sensitivity is, however, built into each method in a different way. In the case of daily snowfall fraction: the fact that it shows the highest standard deviation for air temperatures between 0 and 4 C simply expresses the fact that several methods use thresholds in this range. The result would look different if the thresholds were between -2 and 2 C. This should be better reflected in the discussion of of the results.

Correct, the variability is tied to the methods themselves. However, we think it is important to present this information because the methods we used are based on empirical relationships (air temperature thresholds and ranges, dew point temperature thresholds), physical principles (wet bulb threshold to approximate hydrometeor temperature (Harder and Pomeroy, 2013)), and statistical relationships (the binary logistic regression models). A threshold of -2°C would likely widen the range of variability but it would have no empirical, physical, or statistical relationship to precipitation phase partitioning except in some extremely rare, unique cases. Furthermore, this comment was similar to the feedback from Dr. Jono Conway, who noted the range in variability was likely produced by the extreme air temperature thresholds and ranges (T_{a0}, T_{ar0}, and Ta_3). To respond to his comment, we removed these methods and re-performed the analysis and the finding was the same (please see Figure R2 in our response to Dr. Conway). Even limiting the analysis to the most representative methods, the variability stays highest between 0°C and 4°C.

Additionally, it is essential to point out this air temperature range of variability for two reasons:

1. Areas most "at risk" to the snow-rain transition due to climate warming have seasonal air temperatures near and slightly above freezing (e.g., Nolin and Daly, 2006)
2. 0°C to 4°C is also the air temperature range where precipitation phase methods perform the worst (Ding et al., 2014; Jennings et al., 2018).

Thus, we have a compounded problem in that we are concerned with snow-to-rain shifts in areas with seasonal air temperatures where precipitation phase partitioning is most uncertain and our available methods exhibit downgraded performance. Given that we showed these areas (i.e., winter and spring air temperatures above freezing) also express the greatest sensitivity in terms of peak SWE magnitude and timing, plus snow cover duration, we think it is necessary to include this information.

In general, the conclusion that precipitation falling in the range 0 – 4 C explains much of the variation observed across the methods comes from the choice of the threshold values. Without actual comparison to observed data, the results are hard to generalize. Why is there no comparison to actual SWE-derived thresholds?

Please see our responses above.

Furthermore, when reading the results section where actual SWE curves are presented for the first time, it is a little disappointing to see that all studied sites show a typical seasonal snow cover with significant accumulation over many weeks. The most sen- sitive sites would typically be the ones where the snow cover might build up several times during the winter.

We should note here that the SWE curves as presented are daily averages (Fig. 4 in submitted manuscript), which has the affect of obscuring transience. As we mentioned in the Study sites and data section, the HJA and JD stations are sometimes transient (p. 4 lines 9-10 through p. 5 line 1, and p.5 lines 21-22) and they are most sensitive to phase method choice in terms of peak SWE magnitude (HJA only) as well as peak SWE timing and SCD (Fig. 5 in submitted manuscript).

Detailed comments

- The abstract does not mentioned what types of methods have been tested nor whether they have been compared to reference data or which method performed best

Yes, that is an oversight on our part. We have changed the abstract to note:

"The methods in this study included different permutations of air, wet bulb, and dew point temperature thresholds, air temperature ranges, and binary logistic regression models."

We have also added a line saying:

"Compared to observations of snow depth and SWE, the binary logistic regression models produced the lowest mean biases, while high and low air temperature thresholds tended to respectively overpredict and underpredict snow accumulation."

- Introduction: it would have been interesting to shortly discuss how /where pre- cipitation phase is actually observed; as far as I am aware of, actual precipitation phase observations are crucially missing at most places.

Good point. We have added a line to the first paragraph of the Introduction:

"Complicating matters is the fact precipitation phase is rarely observed in mountain regions on a continuous bases over long time scales."

- Introduction: the manuscript focuses its discussion on snow-hydrological models. How do meteorological forecast models determine the limit (elevation) of snow fall? Completing the literature review with this respect would complete the picture

This is covered in discussion (p. 21 lines 4-17) and not necessary for the Introduction as we do not utilize any atmospheric model methods in this work.

- P. 2: "In general, warmer sites are more sensitive to precipitation phase method selection in terms of annual snowfall fraction variability, though it is less certain how this variability translates into divergences in simulated snow accumulation and melt. " This statement is given without reference. In what is the apparently previously known result different from your own findings?

Text changed to: *"This previous work has shown, in general, warmer sites are more sensitive…"*

in order to clearly connect the statement with the published literature in the previous line.

- Study sites: It might be useful to know the variability of the daily air temperature around the seasonal mean (ie. the anomalies, obtained e.g. by fitting a sine curve to air temperature as in the work of Woods, 2009. It is this variability that will tell something about the probability of switching from accumulation to melting conditions and about a site sensivitiy to the chosen temperature threshold.

This sounds similar to the point raised by Nayak et al. (2010), who showed the effects of switching from sub-freezing to freeze-thaw diurnal cycles on snowpacks at Reynolds Creek. It is clear fluctuations above and below freezing having important effects on snow cover energetics. However, we are unclear as to what new, relevant information such data would provide to the current study. Perhaps we are misunderstanding the comment, so please clarify if so.

- Methods: it is not clear at this stage that all stations always show a seasonal snow cover (significant accumulation over several weeks), which is important for the concept of “peak SWE” to be meaningful

It is noted in the Study sites and data section (p. 4-5) for each location whether seasonal snowpacks develop or not.

- the current definition of snowmelt rate is probably over sensitive to spurious shifts from a primary to a secondary SWE peak, which could reduce the melt dura- tion sensibly; how could this measure be made more robust? Similar comment applies to the peak SWE date that is discussed in the results section. Is this measure useful? Minor modifications of SWE accumulation can switch the SWE peak date between a spurious primary or secondary peak (Figure 4 suggest that stations with two peaks might exist, but I might be mistaken).

We noted on p. 15 (lines 3-6): *“We found the greatest differences in peak SWE dates were generally simulated on years with low/transient snow cover. In these cases, late-season precipitation was simulated as rain by the low T_a thresholds and snow by the high T_a thresholds, meaning an early SWE maximum was recorded as the peak in the former case and a late SWE maximum in the latter case.”* Given that peak SWE timing is an important measure of melt onset in the western US, we find it is necessary to highlight the variability in this metric as produced by different phase methods. Our finding indicates research on future changes to snowmelt timing is affected by modeling decisions on assigning precipitation phase.

Regarding snowmelt rate, we present the seasonal melt rate or ablation slope (e.g., Trujillo and Molotch, 2014) because of the importance of the spring snowmelt freshet to streamflow generation in many mountainous areas of the western US. However, we admit this overlooks the important winter contributions of snowmelt to groundwater and streamflow in maritime and transient snow environments. Switching the analysis to include all days when snowmelt was > 0 mm, we found marginal differences across the precipitation phase methods (mean differences were all less than 2.2 mm d^{-1}, which is less than the nominal precision of the SNOTEL snow pillows in the western US). Looking at daily average melt rate differences between the T_{a0} and T_{a3} thresholds helps illustrate why. Figure R3 below shows that generally T_{a0} produces higher melt rates than T_{a3} early in the snow cover season, while the reverse is true later in the season. Although annual average melt rates exhibit few differences, this figure shows the timing of terrestrial water inputs is important.

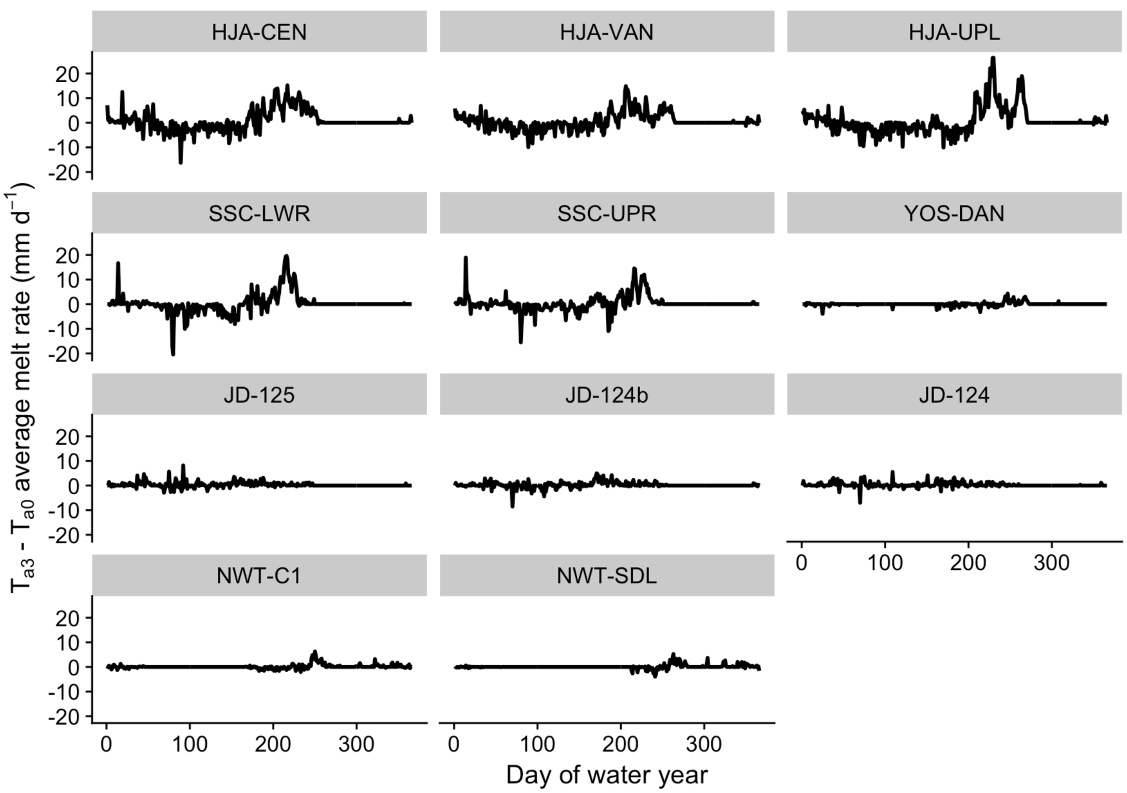


Figure R3. The difference in daily average snowmelt rate between T_{a3} and T_{a0}.

- P. 14 "meaning a significant proportion of water was simulated to have run off using one precipitation phase method versus being stored in the snowpack". This not well formulated since rainfall does not necessarily run off. It can infiltrate and recharge the groundwater.

We agree this was imprecise wording. We have changed this to, *"meaning a significant proportion of water was simulated to have infiltrated or run off using one precipitation phase method versus being stored in the snowpack…"*

- Section 4.4: Here, standard deviations are calculated across the results of all 12 computation methods. Standard deviation does not seem to be a good measure to quantify the variability of values that do not come from an actual sample of a given process but of values pertaining to different methods. (Besides: how are standard deviations obtained? First per method and then averaged over all methods?)

The standard deviation values presented in Section 4.4 and Figure 6 are computed per air temperature and RH bin across all stations and methods as noted in the text. Although standard deviation is an appropriate metric of variability in this context, we redid the analysis using the uncertainty formulation from Harder and Pomeroy (2014). We modified it to be per RH and temperature bin. The result was the same (Figure R4):

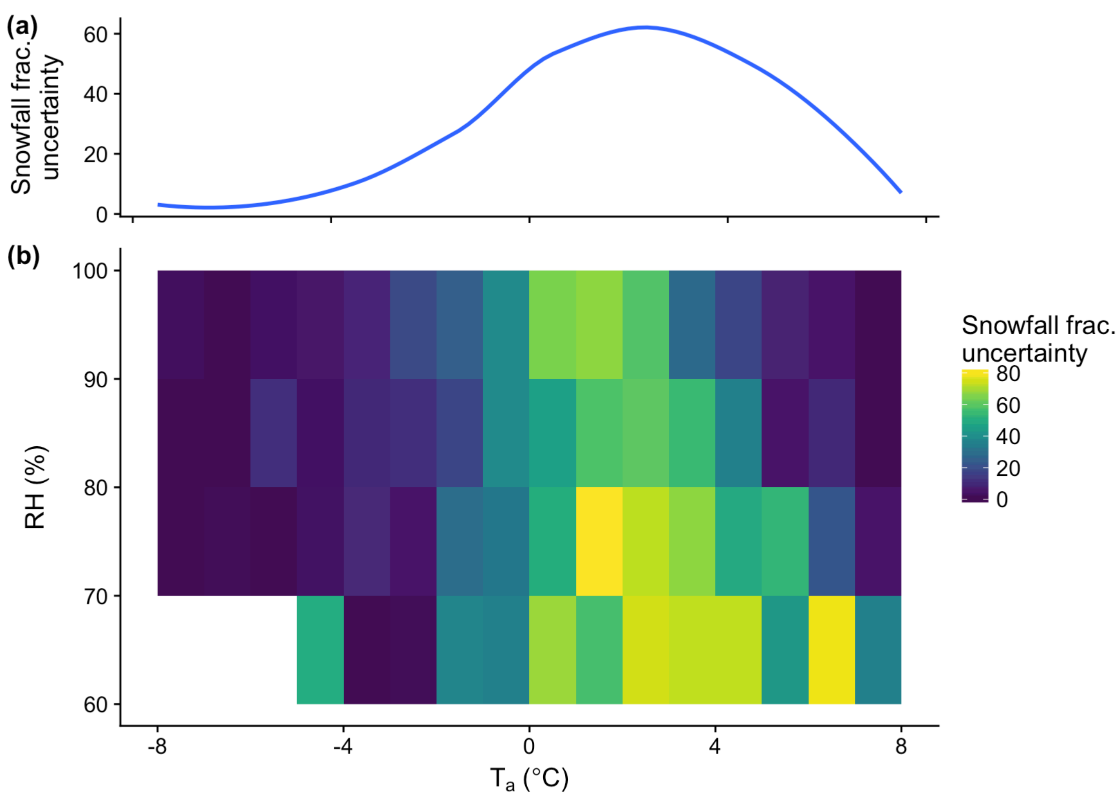


Figure R4. Same as Figure 6 in submitted manuscript, but standard deviation is replaced with the uncertainty metric from Eq. 1 in Harder and Pomeroy (2014).

Screenshot of text from Harder and Pomeroy (2014) showing the uncertainty metric equation:

$$uncertainty = \frac{\sum_{i=1}^{n}(\mathrm{Max}_i - \mathrm{Min}_i)}{n} \quad (1)$$

where Min and Max refer to the lowest and highest values of a model output variable from the 63 model runs, i is the index (time step) of the value and n is the number of values (total time steps). The units of uncertainty are the same as the hydrological variable being considered. The uncertainty and differences between PPMs are summarized using mean values over an entire hydrological year (1 October–30 September).

References

Ding, B., Yang, K., Qin, J., Wang, L., Chen, Y. and He, X.: The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization, J. Hydrol., 513, 154–163, 2014.

Harder, P. and Pomeroy, J.: Estimating precipitation phase using a psychrometric energy balance method, Hydrol. Process., 27(13), 1901–1914, doi:10.1002/hyp.9799, 2013.

Harder, P. and Pomeroy, J. W.: Hydrological model uncertainty due to precipitation-phase partitioning methods, Hydrol. Process., 28(14), 4311–4327, 2014.

Harpold, A. A., Kaplan, M., Klos, P. Z., Link, T., McNamara, J. P., Rajagopal, S., Schumer, R. and Steele, C. M.: Rain or snow: hydrologic processes, observations, prediction, and research needs, Hydrol Earth Syst Sci, 21, 1–22, 2017.

Jennings, K. S., Winchell, T. S., Livneh, B. and Molotch, N. P.: Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere, Nat. Commun., 9, doi:10.1038/s41467-018-03629-7, 2018.

Marks, D., Winstral, A., Reba, M., Pomeroy, J. and Kumar, M.: An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin, Adv. Water Resour., 55, 98–110, doi:10.1016/j.advwatres.2012.11.012, 2013.

Nayak, A., Marks, D., Chandler, D. G. and Seyfried, M.: Long-term snow, climate, and streamflow trends at the Reynolds Creek Experimental Watershed, Owyhee Mountains, Idaho, United States: CLIMATE TRENDS AT RCEW, Water Resour. Res., 46(6), n/a-n/a, doi:10.1029/2008WR007525, 2010.

Nolin, A. W. and Daly, C.: Mapping "at risk" snow in the Pacific Northwest, J. Hydrometeorol., 7(5), 1164–1171, 2006.

Rajagopal, S. and Harpold, A. A.: Testing and Improving Temperature Thresholds for Snow and Rain Prediction in the Western United States, JAWRA J. Am. Water Resour. Assoc. [online] Available from: http://onlinelibrary.wiley.com/doi/10.1111/1752-1688.12443/full (Accessed 23 August 2016), 2016.

Storck, P., Lettenmaier, D. P. and Bolton, S. M.: Measurement of snow interception and canopy effects on snow accumulation and melt in a mountainous maritime climate, Oregon, United States, Water Resour. Res., 38(11), 5–1, 2002.
Trujillo, E. and Molotch, N. P.: Snowpack regimes of the Western United States, Water Resour. Res., 50(7), 5611–5623, doi:10.1002/2013WR014753, 2014.
Wayand, N. E., Clark, M. P. and Lundquist, J. D.: Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations, Hydrol. Process., 31(2), 349–363, doi:10.1002/hyp.11002, 2017.
Zhang, Z., Glaser, S., Bales, R., Conklin, M., Rice, R. and Marks, D.: Insights into mountain precipitation and snowpack from a basin-scale wireless-sensor network, Water Resour. Res., 53(8), 6626–6641, doi:10.1002/2016WR018825, 2017.

We thank Dr. Conway for their insightful review. Our responses can be found in blue throughout the following text. Please note, tables and figures specific to this response document are given with the prefix R (for example, Table R1 in the comment below). Tables and figures in the manuscript are referred to by numbers only.

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.

We should note here that at YOS-DAN, NWT-C1, and NWT-SDL, the T_{d1} threshold produced greater annual snowfall fractions than T_{a3}. Thus, although T_{a3} produced higher snowfall fractions at the remaining sites, this effect was by no means universal. Additionally, as mentioned in the Introduction (p. 3 lines 17-18) and further highlighted in our response to Dr. Schaefli, a 3°C T_a threshold is appropriate in upland continental areas of the western US (e.g., NWT) where snowfall is more common at warmer temperatures than in other locales. Our concern is that many land surface and hydrologic models use spatially uniform air temperature thresholds to partition precipitation phase, so we argue that it is essential to incorporate thresholds that cover the range of observed rain-snow partitioning air temperatures for our study sites (1°C to 3°C). And, despite decades of evidence showing its inefficacy, the T_{a0} threshold and T_{ar0} range are still commonly employed to partition precipitation phase. For example, the widely used VIC macroscale hydrologic model assigns precipitation phase with a default -0.5°C to +0.5°C temperature range, centered on 0°C (https://github.com/UW-Hydro/VIC/; accessed 2019-05-20). We therefore included this method, if only to provide more evidence that it underpredicts snowfall, snow accumulation, and snow cover duration.

In order to further address the need to use the whole range of air temperature thresholds, we have added new text to the Methods (Sect. 3.2):

"T_a thresholds were chosen to represent the spatial variability of rain-snow partitioning in the western United States, where values of approximately 1°C are common near the Pacific Coast, increasing towards 3°C in the Rocky Mountains (Jennings et al., 2018). Additionally, despite significant literature showing its poor performance (e.g., Jennings et al., 2018; Marks et al., 2013), we included a 0°C T_a threshold in the analysis because it is still widely used in observational and model-based hydrologic studies."

In an early draft of this manuscript, we analyzed standard deviations in addition to the ranges presented in the submitted version. The story remained the same: warm maritime sites were greatly impacted by precipitation phase method choice, while cold sites were not. This is illustrated in Figure R1 below:

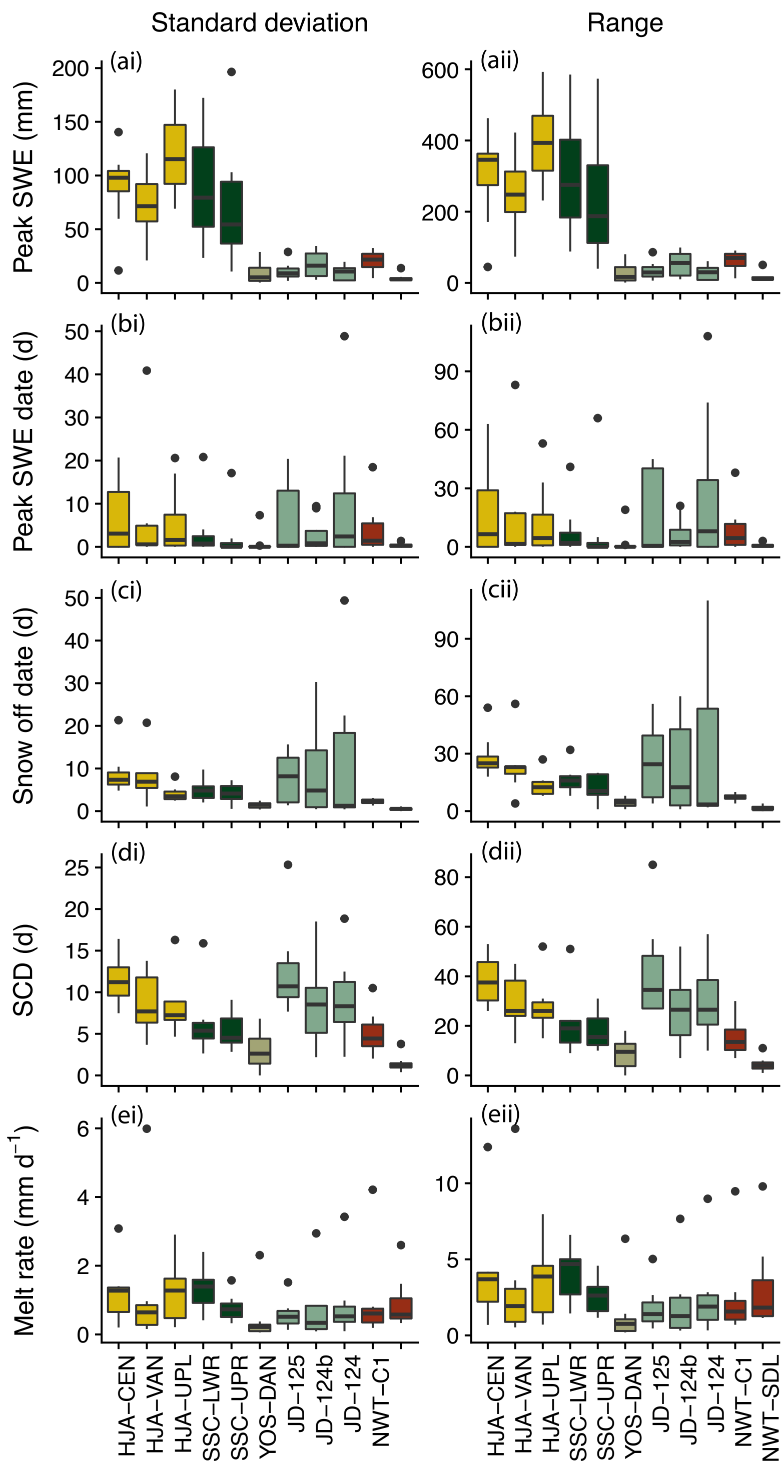


Figure R5. The annual standard deviation (left column) and range (right) in simulated peak SWE (ai,aii), peak SWE date (bi,bii), snow-off date (ci,cii), snow cover duration (di,dii), and melt rate (ei,eii) due to precipitation phase method selection at the study stations.

Regarding the semantics of "method" versus "precipitation partitioning" versus "parameterization," "precipitation phase method" is commonly used to describe modeling and empirical approaches to discriminating between rain and snow (e.g., Harder and Pomeroy, 2014; Harpold et al., 2017). We will leave as is.

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.

We added to text to the abstract noting this finding:

"Average ranges in snowmelt rate were typically less than 4 mm d^{-1} *and exhibited little relationship to seasonal climate."*

And to the conclusion:

"In contrast to the marked differences in peak SWE, melt onset, and snow cover duration between the warm and cold stations, ranges in snowmelt rate exhibited little relationship to seasonal climate."

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.

For the ranges, we added text to Discussion Sect. 5.1:

"In the course of this work we found negligible differences between T_{a0} *and* T_{r0} *as well as between* T_{a1} *and* T_{r1} *in terms of annual snowfall fraction (Fig. 5) and model performance (Fig. 3). This suggests the ranges and the mixed-phase precipitation they produced provided little further information on precipitation phase at the hourly model time scale relative to the thresholds. However, it should be noted there is relatively little quantitative data on the frequency and solid-liquid proportions of mixed-phase events (e.g., Yuter et al., 2006). Work from the Torino region of Italy showed mixed-phase events are relatively few compared to all-rain and all-snow events (Avanzi et al., 2014), while research in a maritime climate indicated*

mixed-phase events can be quite frequent (Wayand et al., 2016). Thus, future work would benefit from further explorations of the frequency of mixed-phase events and model representations thereof at multiple time scales."

For snowmelt rate, it is not that the effect is small (Table 5 shows relative differences between 11.5% and 235.5%), it is that the metric showed no relationship to seasonal climate. Please see our response to the comment above for the extra material we added on snowmelt rate.

For novelty/implications, we changed the final line of the abstract to:

"This study shows care should be taken when selecting a precipitation phase method as the variability introduced to snow accumulation and melt will likely propagate into simulated streamflow and land surface albedo, particularly at the warmer fringes of the seasonal snow zone."

Regarding future directions, suggestions were given in the original manuscript (p. 8 lines 15-16, p. 20 lines 25-27, p. 21 lines 1-3, p. 22 lines 16-19, p. 23 lines 13-16). Given the multiple lines devoted to this topic and the further additions noted in this response, we find no further recommendations are needed.

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

Please see our response to the specific comment below on this topic.

Specific comments (page-line)

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

We have reconfigured the abstract to note these were point simulations with no canopy cover.

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

Added to 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.

We added more text expanding upon the bias in Sect. 3.1:

"*At the HJA stations, we bias-corrected the LW_{in} estimate based on one year of LW_{in} observations from HJA-VAN that showed a -56.9 W m^{-2} wintertime bias, which may have been related to site vegetation conditions. This was significantly larger in magnitude than the bias found in the Unsworth and Monteith (1975) estimate by Flerchinger et al. (2009), suggesting its performance is more spatially variable than previously noted. This finding also underscores the need for enhanced monitoring of the radiation budget at snow modeling sites (Lapo et al., 2015; Raleigh et al., 2015, 2016).*"

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.

Yes, good catch. We have changed the text to "*…annual snowfall fraction range*" to match the figure.

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.

Yes, Figure 6 is designed to illustrate the effect of threshold/method choice on daily snowfall fraction. The data presented are standard deviations, which minimizes the effect of the extreme T_a thresholds. However, we were curious how removing T_{a0}, T_{ar0}, and T_{a3} would affect the analysis and we found it made little difference (Figure R2 is nearly identical to Figure 6 in the submitted manuscript). Except for 1 outlier at -19.5% all differences in the standard deviations for the T_a and RH bins are between -7% and +5%, with a mean difference of -1.3% (computed by subtracting the SD for the analysis with all methods included from the analysis with T_{a0}, T_{ar0}, and T_{a3} removed).

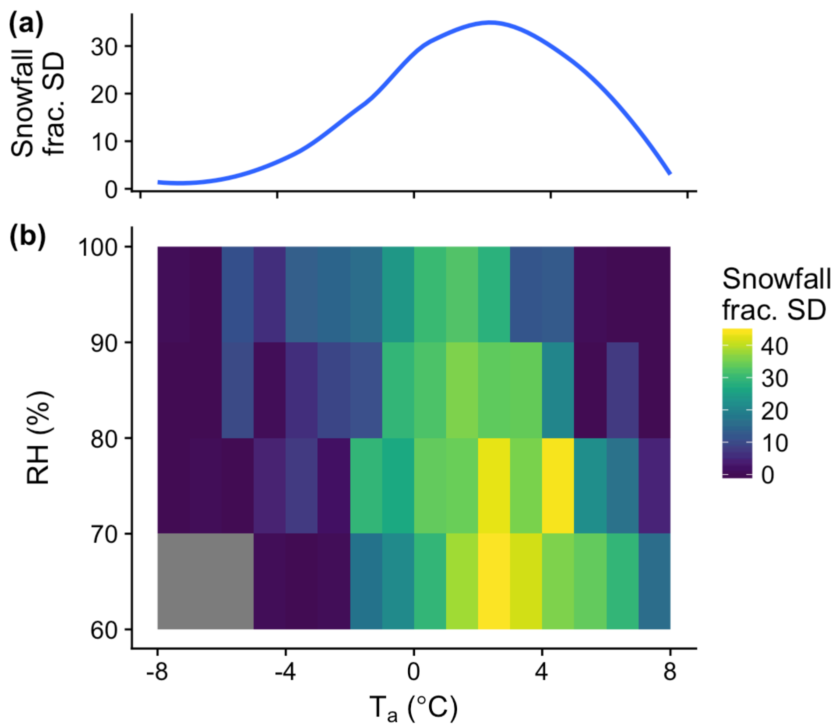


Figure R2. Same as Figure 6 in submitted manuscript but with T_{a0}, T_{ar0}, and T_{a3} removed from the analysis.

Given this analysis held up to the removal of the three least physically representative thresholds, we find the inclusions of Figures 6 and 7 along with the associated text to be appropriate.

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.

We have decided use a loess function to create a smooth surface presented behind the station data (new Fig. 10 shown below). This, we believe, more clearly shows the clustering of low peak SWE ranges at the colder and/or low precipitation sites and high peak SWE ranges at the maritime sites without introducing the statistical pitfalls of multiple linear regression. We also edited the text to remove the linear regression statements.

"We next evaluated how sensitivity in peak SWE was related to seasonal climate. In this case, warmer T_a and increased PPT were both associated with greater ranges in the peak SWE simulated by the different precipitation phase methods (Fig. 10). This meant the maritime sites

HJA and SSC had the greatest sensitivity to precipitation phase method due to their relatively warm T_a and high PPT values. Conversely, moderate PPT values and lower T_a led to minimal sensitivity at the cold continental NWT stations and the cold maritime YOS-DAN station. Again, the effect of T_a on sensitivity was manifest in the data. In high snowfall years at NWT-SDL, Dec–May PPT approached that of the low Dec–May PPT years at HJA and SSC. However, despite the increased PPT at NWT-SDL, the range in peak SWE predicted by the different precipitation phase methods remained low."

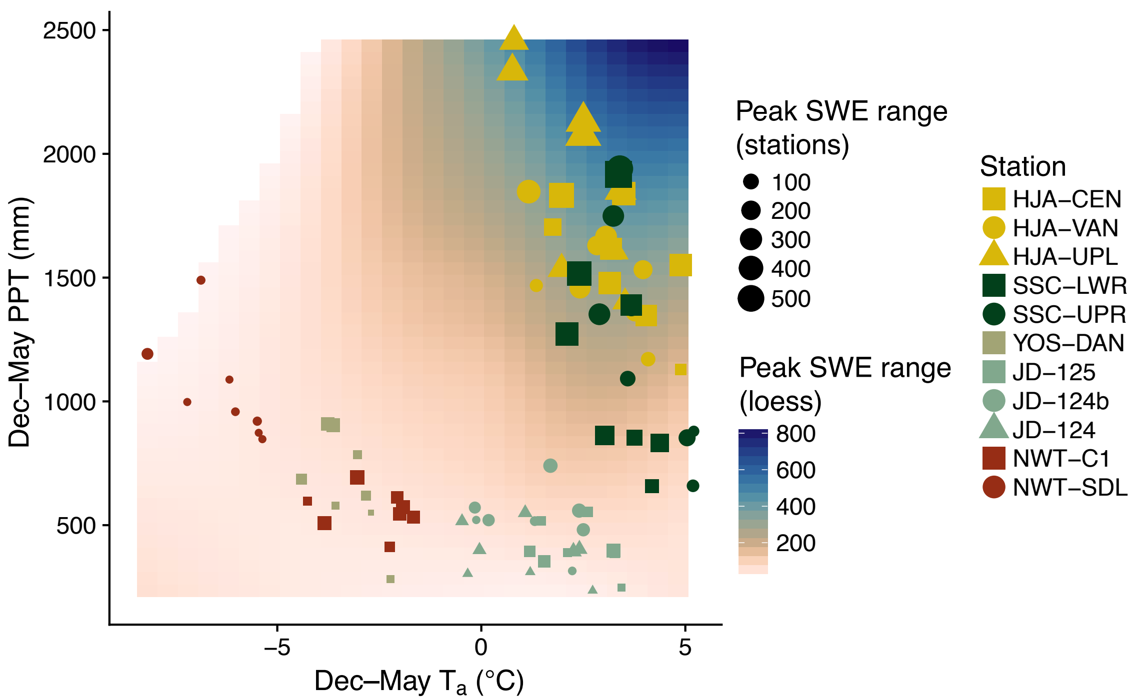


Figure 10. Range in annual peak SWE as simulated by the different precipitation phase methods at the 11 study stations. Each point represents one simulation year at a given station and larger points correspond to larger differences in maximum minus minimum peak SWE. The background shading corresponds to ranges in peak SWE predicted by a loess function fit to the station data. **[Please note, this figure has changed from 8 in the submitted manuscript to 10 in the revised version because we moved the validation figures from the appendix to the results as per the recommendation below.]**

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.

We have moved this from the Appendix to be the first results section:

We have also added material in the Methods detailing how validation was performed (appended to the end of Sect. 3.1):

"To validate model output, we compared simulated SWE and snow depth to observations at our study stations. SWE was observed at all HJA stations, SSC-UPR, YOS-DAN, and both NWT stations, while snow depth was observed at all HJA stations, both SSC stations, and all JD stations. All SWE data were derived from automated snow pillow measurements except for NWT-SDL, which was acquired through manual snow pit observations (Williams, 2016). Similarly, automated ultrasonic snow depth sensors produced all snow depth data. Comparisons were made at the daily time scale when either simulated or observed SWE or snow depth were > 0 mm. This was done to prevent artificial enhancement of objective function values during periods when snow cover was absent."

We have also edited discussion Sect. 5.1 to reflect these changes and to incorporate feedback from the two subsequent comments.

"In this work we showed that the selection of a precipitation phase method produces varying degrees of variability in modeled snow accumulation and melt at our study stations. The different methods also expressed variable performance relative to observations of SWE and snow depth, with the binary regression models, Reg_{Bi} and Reg_{Tri}, as well as the T_{a1} threshold producing the lowest biases (Fig. 3). Previous observational work has shown that, in general, methods incorporating humidity information outperform T_a-only methods when it comes to predicting precipitation phase (Harder and Pomeroy, 2013; Jennings et al., 2018; Marks et al., 2013; Ye et al., 2013). The Reg_{Bi} method, which predicts phase as a function of T_a and RH, exceeded all other methods in partitioning rain and snow in a Northern Hemisphere precipitation phase method comparison (Jennings et al., 2018). Our study showed that Reg_{Bi} also typically produced simulations of SWE and snow depth that had low biases relative to observations (Fig. 3) and led to snow cover evolution metrics that were neither extremely high nor low compared to the other methods examined in this work. This finding is complemented by the performance of other humidity-based metrics, which produced average SWE and snow depth biases between -19.2 mm and 25.1 mm, and -64.1 mm and 45.0 mm, respectively.

This is in contrast to the T_a thresholds and ranges, which produced the largest magnitude biases. Notably, the four worst performers were the T_{a0}, T_{ar0}, T_{a2}, and T_{a3} methods, with the former two underpredicting snow accumulation and the latter two overpredicting. Across our study sites, the only T_a methods that performed well relative to observations were the T_{a1} threshold and T_{ar1} range. These modeling results confirm again that including humidity information, whether it be in the form of a binary logistic regression model. T_w, or T_d, offers advantages over a T_a-only *method. It is important to note again that we chose methods that covered the range in rain-snow partitioning T_a values across our study domain or that included humidity information. The only methods not falling into this category were T_{a0} and T_{ar0}, which were chosen because they are still employed as default methods in some models and studies. Although there are some small geographic regions where such a threshold or range may be appropriate (Jennings et al., 2018), they are unsuitable for many locations and should not be used for large-scale studies.*

In the course of this work we found negligible differences between T_{a0} and T_{r0} as well as between T_{a1} and T_{r1} in terms of annual snowfall fraction (Fig. 5) and model performance (Fig. 3). This suggests the ranges and the mixed-phase precipitation they produced provided little further information on precipitation phase at the hourly model time scale relative to the thresholds. However, it should be noted there is relatively little quantitative data on the frequency and solid-liquid proportions of mixed-phase events (e.g., Yuter et al., 2006). Work from the Torino region of Italy showed mixed-phase events are relatively few compared to all-rain and all-snow events (Avanzi et al., 2014), while research in a maritime climate indicated mixed-phase events can be quite frequent (Wayand et al., 2016). Future work would therefore benefit from further explorations of the frequency of mixed-phase events and model representations thereof at multiple time scales.

Despite the analyses presented in this work, it is important to note that uncertainties in forcing data, model structure and parameters, as well as a lack of precipitation phase observations prevent this research from being able to unequivocally identify a "best" precipitation phase method for snow modeling. However, as noted above, including humidity information improves the prediction of precipitation phase relative to observations and generally increases model performance. Our primary aim in this research was to quantify how snow simulations were affected by the choice of precipitation phase method across a climatic gradient. We did not create optimized model setups at each site, but rather kept model setup consistent in order to compare the sensitivity of phase partitioning without introducing other uncertainties. Thus, the low r^2 and higher bias values at HJA-VAN, NWT-SDL, and JD-124 (Fig. 4) could likely be improved with model tuning, but we did not pursue such an approach."

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.

Please see edited discussion Sect. 5.1 above.

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.

Please see edited discussion Sect. 5.1 above.

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.

This statement is broad because model intercomparisons say little about the effect of precipitation phase method selection. For example, SnowMIP2 used different precipitation phase methods at different sites (Rutter et al., 2009). Thus it is still unknown how model selection and phase partitioning methods interact (i.e., would a temperature index model be more affected than a physics-based model?). We stand by our statement but have clarified with some extra text:

“Given this variable performance and differences in snow model structure and physics, it is possible that some models may be more or less sensitive to the choice of a precipitation phase method. Our use of a single model may overestimate or underestimate the sensitivity of snow accumulation and melt to precipitation phase method selection. Future research should therefore focus on how model choice affects the sensitivity of simulated snow cover evolution to precipitation phase method.”

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.

There are 10 citations in the previous lines detailing how rain vs. snow affects the snowpack energy balance. We include this as a discussion because a full treatment of the energy balance data is outside of the scope of this already fairly long manuscript.

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.

See response to comment above. We also added a qualifier to the sentence and followed it with a future research line:

"In this context, the precipitation phase methods that produced more rainfall likely affected snow cover evolution not just through reduced frozen mass but also through changes to the snowpack energy budget. Further observational and modeling research is warranted to evaluate how rain versus snow affects snowpack energetics."

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.

As we showed in our response to a comment above, the range of uncertainty stays the same when removing T_{a0}, T_{ar0}, and T_{a3}. Furthermore, the difficulty of predicting precipitation phase and the resulting uncertainty at temperatures slightly above freezing is a well known phenomenon in both hydrology (Ding et al., 2014; Harpold et al., 2017; Jennings et al., 2018) and atmospheric science (Ralph et al., 2005; Stewart et al., 2015).

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".

Added per recommendation on an earlier comment.

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.

Added to Methods Sect. 3.1 as noted above.

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.

Please see earlier comments on the T_{a0}, T_{ar0} and T_{a3} methods.

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.

Please see new results text above.

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

That is an artifact of the data at JD. The low overall snow depth produced low absolute biases. We added this text to the figure caption (please note, this figure and section has been moved to Results 4.1 per recommendation on an earlier comment):

"Note: in panel (c) the low mean biases for JD snow depth are due to small observed snow depth values at the site. Mean relative biases at these stations were 35.4% (JD-125), 3.8% (JD-124b), and 35.7% (JD-124)."

Editorial comments:

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

Yes, line changed to "daily average T_a and RH"

16-5 "not computed because for" -> "not computed for"

Redundant "because" has been removed.

References

Avanzi, F., De Michele, C. and Ghezzi, A.: Liquid-solid partitioning of precipitation along an altitude gradient and its statistical properties: An Italian case study, Am. J. Clim. Change, 2014 [online] Available from: http://file.scirp.org/Html/7-2360069_43990.htm (Accessed 24 August 2016), 2014.
Ding, B., Yang, K., Qin, J., Wang, L., Chen, Y. and He, X.: The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization, J. Hydrol., 513, 154–163, 2014.
Flerchinger, G. N., Xaio, W., Marks, D., Sauer, T. J. and Yu, Q.: Comparison of algorithms for incoming atmospheric long-wave radiation, Water Resour. Res., 45(3), doi:10.1029/2008WR007394, 2009.
Harder, P. and Pomeroy, J.: Estimating precipitation phase using a psychrometric energy balance method, Hydrol. Process., 27(13), 1901–1914, doi:10.1002/hyp.9799, 2013.
Harder, P. and Pomeroy, J. W.: Hydrological model uncertainty due to precipitation-phase partitioning methods, Hydrol. Process., 28(14), 4311–4327, 2014.
Harpold, A. A., Kaplan, M., Klos, P. Z., Link, T., McNamara, J. P., Rajagopal, S., Schumer, R. and Steele, C. M.: Rain or snow: hydrologic processes, observations, prediction, and research needs, Hydrol Earth Syst Sci, 21, 1–22, 2017.

Jennings, K. S., Winchell, T. S., Livneh, B. and Molotch, N. P.: Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere, Nat. Commun., 9, doi:10.1038/s41467-018-03629-7, 2018.
Lapo, K. E., Hinkelman, L. M., Raleigh, M. S. and Lundquist, J. D.: Impact of errors in the downwelling irradiances on simulations of snow water equivalent, snow surface temperature, and the snow energy balance, Water Resour. Res., 51(3), 1649–1670, 2015.
Marks, D., Winstral, A., Reba, M., Pomeroy, J. and Kumar, M.: An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin, Adv. Water Resour., 55, 98–110, doi:10.1016/j.advwatres.2012.11.012, 2013.
Raleigh, M. S., Lundquist, J. D. and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrol. Earth Syst. Sci., 19(7), 3153–3179, doi:10.5194/hess-19-3153-2015, 2015.
Raleigh, M. S., Livneh, B., Lapo, K. and Lundquist, J. D.: How Does Availability of Meteorological Forcing Data Impact Physically Based Snowpack Simulations?, J. Hydrometeorol., 17(1), 99–120, doi:10.1175/JHM-D-14-0235.1, 2016.
Ralph, F. M., Rauber, R. M., Jewett, B. F., Kingsmill, D. E. and others: Improving short-term (0-48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP workshop, Bull. Am. Meteorol. Soc., 86(11), 1619, 2005.
Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H. and others: Evaluation of forest snow processes models (SnowMIP2), J. Geophys. Res. Atmospheres, 114(D6), doi:doi/10.1029/2008JD011063/, 2009.
Stewart, R. E., Thériault, J. M. and Henson, W.: On the Characteristics of and Processes Producing Winter Precipitation Types near 0°C, Bull. Am. Meteorol. Soc., 96(4), 623–639, doi:10.1175/BAMS-D-14-00032.1, 2015.
Unsworth, M. H. and Monteith, J. L.: Long-wave radiation at the ground I. Angular distribution of incoming radiation, Q. J. R. Meteorol. Soc., 101(427), 13–24, 1975.
Wayand, N. E., Stimberis, J., Zagrodnik, J. P., Mass, C. F. and Lundquist, J. D.: Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA, J. Geophys. Res. Atmospheres, 121(17), 9929–9942, 2016.
Williams, M.: Snow cover profile data for Niwot Ridge, Green Lakes Valley from 1993/2/26 - ongoing, weekly to biweekly, [online] Available from: http://niwot.colorado.edu/index.php/data/data/snow-cover-profile-data-for-niwot-ridge-and-green-lakes-valley-1993-ongoi, 2016.
Ye, H., Cohen, J. and Rawlins, M.: Discrimination of Solid from Liquid Precipitation over Northern Eurasia Using Surface Atmospheric Conditions*, J. Hydrometeorol., 14(4), 1345–1355, 2013.
Yuter, S. E., Kingsmill, D. E., Nance, L. B. and Löffler-Mang, M.: Observations of Precipitation Size and Fall Speed Characteristics within Coexisting Rain and Wet Snow, J. Appl. Meteorol. Climatol., 45(10), 1450–1464, doi:10.1175/JAM2406.1, 2006.

The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient

Keith S. Jennings[1,2,3,4] and Noah P. Molotch[1,2,5]

[1]Geography Department, University of Colorado Boulder, 260 UCB, Boulder, Colorado 80309, USA
[2]Institute of Arctic and Alpine Research, University of Colorado Boulder, 450 UCB, Boulder, CO 80309, USA
[3]Department of Geography, University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557, USA
[4]Desert Research Institute, 2215 Raggio Parkway, Reno, NV 89512, USA
[5]NASA Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Correspondence to: Keith S. Jennings (keithj@unr.edu)

Abstract. A critical component of hydrologic modeling in cold and temperate regions is partitioning precipitation into snow and rain, yet little is known about how uncertainty in precipitation phase propagates into variability in simulated snow accumulation and melt. Given the wide variety of methods for distinguishing between snow and rain, it is imperative to evaluate the sensitivity of snowpack model output to precipitation phase determination methods, especially considering the potential of snow-to-rain shifts associated with climate warming to fundamentally change the hydrology of snow-dominated areas. To address these needs we quantified the sensitivity of simulated snow accumulation and melt to rain-snow partitioning methods at sites in the western United States using the SNOWPACK model without the canopy module activated. The methods in this study included different permutations of air, wet bulb, and dew point temperature thresholds, air temperature ranges, and binary logistic regression models. Compared to observations of snow depth and SWE, the binary logistic regression models produced the lowest mean biases, while high and low air temperature thresholds tended to overpredict and underpredict snow accumulation, respectively. Relative differences between the minimum and maximum annual snowfall fractions predicted by the different methods sometimes exceeded 100% at elevations less than 2000 m in the Oregon Cascades and California's Sierra Nevada mountains. This led to ranges in annual peak snow water equivalent (SWE) typically greater than 200 mm, exceeding 400 mm in certain years. At the warmer sites, ranges in snowmelt timing predicted by the different methods were generally larger than 2 weeks, while ranges in snow cover duration approached 1 month and greater. Conversely, the three coldest sites in this work were relatively insensitive to the choice of a precipitation phase method with average ranges in annual snowfall fraction, peak SWE, snowmelt timing, and snow cover duration less than 18%, 62 mm, 10 d, and 15 d, respectively. Average ranges in snowmelt rate were typically less than 4 mm d^{-1} and exhibited little relationship to seasonal climate. Overall, sites with a greater proportion of precipitation falling at air temperatures between 0°C and 4°C exhibited the greatest sensitivity to method selection, suggesting the identification and use of an optimal precipitation phase method is most important at the warmer fringes of the seasonal snow zone.

Keith Jennings 7/1/2019 2:01 PM
Deleted: modeled

Keith Jennings 7/1/2019 2:04 PM
Deleted: research

Keith Jennings 7/1/2019 2:00 PM
Deleted: Simulations using the physics-based SNOWPACK model and 12 different precipitation phase methods indicated maritime sites were the most sensitive to method selection.

Keith Jennings 7/1/2019 2:08 PM
Formatted: Superscript

Keith Jennings 7/16/2019 7:48 AM
Deleted: .

Keith Jennings 7/16/2019 7:48 AM
Formatted: Not Highlight

Keith Jennings 7/1/2019 2:07 PM
Deleted: These findings have large implications for modeled snowpack water storage and land surface albedo at the warmer fringes of the seasonal snow zone.

1 Introduction

One of the most prominent impacts of climate warming has been a shift from snow to rain in temperate and cold regions across the globe (e.g., Knowles et al., 2006; Trenberth, 2011), a trend that is expected to continue with further increases in air temperature (Bintanja and Andry, 2017; Klos et al., 2014; O'Gorman, 2014; Safeeq et al., 2015). In order to assess how this change affects global hydroclimate, researchers have employed snow models, hydrologic models, and land surface models of varying degrees of complexity (e.g., Barnett et al., 2005). One trait many of these models share is the partitioning of rainfall and snowfall based on a spatially uniform air temperature threshold or a range between two thresholds with a linear mix of liquid and solid precipitation in between. Recent work has called into question this simplistic treatment of precipitation phase (Feiccabrino et al., 2015; Harpold et al., 2017b) because of the pronounced spatial variability of rain-snow partitioning (Jennings et al., 2018b; Ye et al., 2013). Complicating matters is the fact precipitation phase is rarely observed in mountain regions on a continuous basis over long time scales.

The use of a spatially uniform air temperature threshold is seemingly logical given the strong temperature-dependence of precipitation phase. Observational work has shown that precipitation is primarily solid at temperatures at and below the freezing point (Auer Jr, 1974; Avanzi et al., 2014; United States Army Corps of Engineers, 1956) and that the probability of snowfall decreases following a sigmoidal curve as air temperature increases above 0°C (Dai, 2008; Fassnacht et al., 2013; Kienzle, 2008). However, the point at which the sigmoidal curve crosses 50% snow probability (i.e., the 50% rain-snow air temperature threshold) has been shown to vary significantly across the Northern Hemisphere (Jennings et al., 2018b). Thus, a single air temperature threshold, or range, cannot accurately represent precipitation phase partitioning across large spatial extents (Raleigh and Lundquist, 2012). Part of this variability can be ascribed to relative humidity as recent work has shown snowfall is more probable at a given air temperature in more arid conditions (Froidurot et al., 2014; Gjertsen and Ødegaard, 2005; Jennings et al., 2018b). Surface air pressure also affects phase partitioning, but to a lesser degree than air temperature and humidity, with snowfall more common at higher air temperatures when surface pressure is lower (Ding et al., 2014; Jennings et al., 2018b; Rajagopal and Harpold, 2016).

Given the secondary controls exerted by humidity and surface pressure on the probability of rain versus snow, precipitation phase methods have been developed to leverage this information into more accurate rain and snow predictions. These methods include dew point temperature thresholds (Marks et al., 2013; Ye et al., 2013), wet/ice bulb temperature thresholds (Anderson, 1968; Harder and Pomeroy, 2013), and binary logistic regression equations that predict the probability of snow as a function of various meteorological quantities (Froidurot et al., 2014). In general, methods incorporating humidity better predict precipitation phase than air temperature-only methods relative to observations across the Northern Hemisphere (Jennings et al., 2018b), likely due to their better representation of the hydrometeor energy balance (Harder and Pomeroy, 2013; Harpold et al., 2017b). Furthermore, the spatial variability of phase partitioning is reduced when using humidity information in addition to air temperature (Ye et al., 2013).

This wide variety of precipitation phase methods leads to variations in snowfall fraction—the percentage of annual precipitation that falls as snow—approaching 30% or greater when applied to station meteorological data and reanalysis products (Harpold et al., 2017c; Jennings et al., 2018b; Raleigh et al., 2016). This previous work has shown warmer sites are generally more sensitive to precipitation phase method selection in terms of annual snowfall fraction variability, though it is less certain how this variability translates into divergences in simulated snow accumulation and melt. To that end, Harder and Pomeroy (2014) showed that precipitation phase method selection can produce ranges in annual peak SWE and snow cover duration of 160 mm and 36 d, respectively. However, this work only examined relatively cold research basins in Canada and did not consider the warmer mid-latitude, maritime climates that have been shown to be most "at risk" to the effects of climate warming on snow accumulation (e.g., Nolin and Daly, 2006). Similarly, other researchers have found higher air temperature thresholds generate greater annual peak SWE and increased snow accumulation during storm events at individual sites and basins (Fassnacht and Soulis, 2002; Mizukami et al., 2013; Wayand et al., 2017; Wen et al., 2013).

We are therefore left with the question of how the sensitivity of modeled snow accumulation and melt to precipitation phase method selection varies across sites with different climatic characteristics. Considering over 1 billion people worldwide rely on mountain snowpacks for water resources (Barnett et al., 2005; Mankin et al., 2015), it is essential that models accurately represent precipitation phase partitioning as well as snowpack water storage and snowmelt timing. Furthermore, snowpacks are highly reflective relative to bare ground, meaning simulated snow cover duration has a significant effect on modeled land surface albedo. These issues are further compounded when future warming-driven changes to snow accumulation and melt are taken into consideration, particularly if precipitation phase method selection induces uncertainty approaching that of the warming signal. Thus, it is necessary to quantify the baseline uncertainty in snow cover evolution due to the choice of a precipitation phase method, and then evaluate how the uncertainty relates to seasonal climate at a diverse selection of sites.

The western United States offers a unique opportunity to perform such research for several reasons. One, the region includes both maritime and continental climates. Two, the region expresses a wide range of 50% rain-snow air temperature thresholds, increasing from ~1°C near the Pacific Coast to over 3°C in the Rocky Mountains (Jennings et al., 2018b). And three, model forcing and validation data are freely available through publicly funded networks. In the research presented herein, we simulate eight years of snow cover evolution using different permutations of 5 precipitation phase methods at sites that span a climatic gradient from warm maritime to cold continental to answer the following research questions:

1. What is the sensitivity of annual snowfall fraction and modeled snow accumulation and melt due to precipitation phase method selection?
2. How is the sensitivity controlled by air temperature, relative humidity, and precipitation?

Keith Jennings 5/21/2019 10:38 AM
Deleted: In general,

Unknown
Field Code Changed

Unknown
Field Code Changed

Keith Jennings 7/2/2019 2:49 PM
Deleted: 12

Keith Jennings 7/2/2019 2:50 PM
Deleted: with the goal of answering

2 Study sites and data

We selected sites in the western United States (Fig. 1) with long-term forcing and validation data that represented a range of snow conditions from transient snow with rain-on-snow and midwinter melt events to cold, deep seasonal snowpacks with little midwinter snowmelt. For this work, three stations at the HJ Andrews Experimental Forest were used to represent warm, maritime snowpacks. The two stations at the Southern Sierra Critical Zone Observatory (CZO) also have warm, maritime climates, but seasonal snowpacks develop more consistently. The final maritime site is Dana Meadows in Yosemite National Park; however, this site consistently develops deep seasonal snowpacks due to considerably colder winter air temperatures than the other two maritime sites. The semi-arid Johnston Draw site forms part of the Reynolds Creek Experimental Watershed and is located in the intermountain transition zone between maritime and continental climates. Finally, the two stations at Niwot Ridge are representative of cold continental locations. More information on the sites is presented in the text below and in Table 1.

Figure 1: The western United States showing the 5 study sites. Details on the stations at each site along with their meteorological characteristics can be found in Sect. 2 and in Table 1.

Keith Jennings 7/15/2019 9:19 AM
Deleted: are detailed

Keith Jennings 7/15/2019 9:25 AM
Deleted: the following paragraphs

The HJ Andrews Experimental Forest (HJA), located in western Oregon, is part of the Long Term Ecological Research (LTER) network. We focused on the three meteorological stations with long-term forcing and validation data: Cenmet (HJA-CEN), Vanmet (HJA-VAN), and Uplmet (HJA-UPL). Due to its lower elevation, the HJA-CEN site only develops seasonal snowpacks during some winters, but is otherwise transient. HJA-VAN and HJA-UPL typically develop seasonal snowpacks, but snow is transient in some years. Winter melt and rain-on-snow events are common throughout HJA (Harr, 1986; Jennings and Jones, 2015; Mazurkiewicz et al., 2008; Perkins and Jones, 2008). This site represents a typical maritime climate within the rain-snow transition zone.

The Upper (SSC-UPR) and Lower (SSC-LWR) Providence Creek stations in the Southern Sierra CZO (SSC) are within the maritime zone of California's Sierra Nevada mountains and generally develop seasonal snowpacks. Reported annual snowfall fractions range between 20% and 60%, and rain-on-snow events can occur at both stations (Hunsaker et al., 2012). SSC-UPR and SSC-LWR can be either rain- or snow-dominated depending on the climate of a particular year (Hunsaker et al., 2012). This site represents maritime climates in the seasonal snow zone where winter melt events are frequent but snow cover persists throughout the winter.

The Dana Meadows station (YOS-DAN) is located within California's Yosemite National Park and is part of the Yosemite Hydroclimate Network (Lundquist et al., 2016). YOS-DAN receives significant winter precipitation, which produces snowpacks several meters deep due to cold winter temperatures (Lundquist et al., 2016; Rice et al., 2011). Although YOS-DAN has a maritime climate, annual snowfall fraction can exceed 90% (Lundquist et al., 2016) because of the station's high elevation and strongly seasonal precipitation. Winter melt makes up a relatively low proportion of annual snowmelt at this elevation (Rice et al., 2011).

Johnston Draw (JD) is a sub-watershed within the larger Reynolds Creek Experimental Watershed, which is part of the CZO network in southwestern Idaho. Reynolds is within the rain-snow transition zone (Nayak et al., 2010) and has a semi-arid intermountain climate, bridging the divide between maritime and continental. We focused our simulations on three stations with co-located meteorological and snow depth measurements: 125 (JD-125), 124b (JD-124b), and 124 (JD-124). Previous work has shown average annual snowfall fraction ranges from 39% at the lower station to 53% at the highest (Godsey et al., 2018). Similar to the HJA stations, seasonal snowpacks develop at the Johnston Draw stations in some years, but are transient in others. Due to high wind speeds and complex terrain, snow patterns vary across sites from year to year (Godsey et al., 2018). Additionally, winter melt and rain-on-snow events occur throughout the Reynolds Creek Experimental Watershed (Marks et al., 2001; Marks and Winstral, 2001).

The Niwot Ridge LTER (NWT) in Colorado's Rocky Mountains has a cold continental climate (Greenland, 1989) with previously reported annual snowfall fractions ranging between 63% and 80% (Caine, 1996; Knowles et al., 2015). The C1 station (NWT-C1) is in the subalpine area of NWT and Saddle (NWT-SDL) is situated above treeline in the alpine. Winter melt and rain-on-snow events are rare at both stations, particularly at NWT-SDL. High winter wind speeds are responsible for significant spatial variation in snow depth at NWT-SDL (Erickson et al., 2005; Litaor et al., 2008), while a dense stand of lodgepole pine reduces the effect of wind on snow cover evolution at NWT-C1.

Unknown
Field Code Changed

Unknown
Field Code Changed

Unknown
Field Code Changed

Unknown
Field Code Changed

Unknown
Field Code Changed

Keith Jennings 7/3/2019 9:30 AM
Deleted: not

Unknown
Field Code Changed

Table 1. Station information plus average annual and December/January/February (DJF) climatic conditions (T_a = air temperature, T_w = wet bulb temperature, T_d = dew point temperature, RH = relative humidity, VW = wind speed, and PPT = precipitation) for the 8 years of the study period (WY2004–WY2011).

Site	Station	Code	Elev. (m)	Annual T_a (°C)	Annual T_w (°C)	Annual T_d (°C)	Annual RH (%)	Annual VW (m s^{-1})	Annual PPT (mm)	DJF T_a (°C)	DJF T_w (°C)	DJF T_d (°C)	DJF RH (%)	DJF VW (m s^{-1})	DJF PPT (mm)	DJF PPT* (%)
HJ Andrews	Cenmet	HJA-CEN	1020	7.5	5.1	3.6	81.2	1	2308	1.7	0.2	-0.8	86.3	1	957	41.5
	Vanmet	HJA-VAN	1275	7	4.2	2	76.8	1.2	2259	1.3	-0.6	-2.7	80.4	1.3	956	42.3
	Uplmet	HJA-UPL	1295	6.5	3.8	1.8	77.3	0.8	2841	0.7	-1.1	-2.9	81.6	0.8	1133	39.9
Southern Sierra CZO	Lower Providence	SSC-LWR	1753	8.4‡	4.7	1.7	68.3	0.9	1538	1.3	-0.7	-2.5	79.5	0.6	821	53.4
	Upper Providence	SSC-UPR	1981	9.1	4.4	-0.6	57.4	1.2	1613	2.3	-1.0	-5.9	63.7	0.9	878	54.4
Yosemite Nat. Park	Dana Meadows	YOS-DAN	2987	1.4	-2.5	-8.6	55.6	1.3	811	-5.5	-8.0	-13.3	62.7	1.4	468	57.7
Johnston Draw	125	JD-125	1508	7.9	3	-1.7	57.6	1.7	586	-1.5	-3.6	-5.9	74.3	1.7	217	37
	124b	JD-124b	1778	6.8	2.3	-2.2	59.2	1.8	718	-2.1	-4.2	-6.6	74.5	1.9	301	41.9
	124	JD-124	1804	6.9	2.1	-2.9	56.8	4.4	580	-2.2	-4.3	-7.0	72.5	5.3	198	34.1
Niwot Ridge	C1	NWT-C1	3022	2.6	-1	-5.5	60.8	2.7	917	-6.3	-8.8	-13.3	62.3	4.1	216	23.6
	Saddle	NWT-SDL	3528	-0.7	-3.7	-7.6	64.3	8.5	1483	-9.9	-11.6	-14.8	71.4	11.7	592	39.9†

*Column corresponds to percentage of annual precipitation that falls during DJF.
‡Average T_a values are cooler at SSC-LWR than SSC-UPR due to differences in vegetation and physiography at the two stations (M. Safeeq, personal communication, 20 June 2018).
†High DJF precipitation percentage likely due to gage overcatch reduction factors. The alpine precipitation gage sees significant overcatch due to blowing snow (Williams et al., 1998) and reduction factors were developed relative to observed changes in the NWT-SDL snow pit SWE (Jennings et al., 2018a).

3 Methods

3.1 Model setup, forcing data preparation, and validation

We used the one-dimensional, physics-based SNOWPACK model (Bartelt and Lehning, 2002; Lehning et al., 2002a, 2002b) to evaluate the sensitivity of snow cover evolution to various precipitation phase methods. SNOWPACK is forced with air temperature (T_a), relative humidity (RH), wind speed (VW), incoming shortwave radiation (SW_{in}), incoming longwave radiation (LW_{in}), and precipitation (PPT) at an hourly or longer time step. Part of our motivation for using SNOWPACK, in addition to the model's consistent performance in snow model studies (Etchevers et al., 2004; Rutter et al., 2009) and extensive validation (Jennings et al., 2018a; Lehning et al., 2001; Lundy et al., 2001; Meromy et al., 2015), was that it offers the user the option to include precipitation phase as part of the forcing data. In this scheme, the user can identify

Keith Jennings 7/3/2019 9:30 AM
Deleted: ... [1]
Keith Jennings 7/2/2019 9:15 AM
Formatted: Subscript
Keith Jennings 7/2/2019 9:15 AM
Formatted: Subscript
Keith Jennings 5/23/2019 2:15 PM
Formatted: Normal, Don't keep with next
Keith Jennings 5/24/2019 11:40 AM
Deleted: and
Unknown
Field Code Changed

a time step as all-snow (0) or all-rain (1), or a mix of precipitation (decimal values between 0 and 1). Further details on the precipitation phase methods implemented in this study are provided in Sect. 3.2 below.

We ran SNOWPACK at an hourly time step and kept model setup nearly identical across the sites in order to make the precipitation phase sensitivity results as comparable as possible. The only changes made to model setup were the meteorological measurement heights (Table S1), which were provided as part of the various forcing datasets. In some cases, this approach overlooked important changes to the snow accumulation and melt processes (e.g., snowfall interception, enhancement of incoming longwave radiation) caused by forest cover, notably at the HJ Andrews site and, to a lesser extent, NWT-C1. However, we wanted the simulations to represent snow cover evolution without introducing the confounding hydrologic effects of interception and model representation thereof, meaning the canopy module for SNOWPACK was not activated at any of the sites. We acknowledge properly representing snow-forest interactions is critical to modeling snow in many basins (Lehning et al., 2006; Rutter et al., 2009) as tree cover exerts important controls on snow accumulation and melt (Dickerson-Lange et al., 2017; Lundquist et al., 2013; Roth and Nolin, 2017). Future work should therefore examine how model representations of both vegetation and precipitation phase interact to produce uncertainty in modeled SWE.

Where possible, we relied on quality control and infilling methods from the dataset creators given their familiarity with meteorological processes at their respective sites. At HJA, the provided data were quality controlled, but not serially complete. We first infilled data with instruments at different heights located at the same station when those measurements were available. We used linear regressions from the other stations to fill all other missing data. For the SSC stations, we performed an additional quality control routine based on Meek and Hatfield (1994) in order to clean up spurious data points. We then infilled missing data by regressing the two SSC stations. All other datasets were serially complete and we performed no further quality control or infilling procedures.

Additionally, none of the sites had LW_{in} measurements available for the entirety of the study period. We used the empirical estimates of LW_{in} provided with the NWT and YOS-DAN datasets to force SNOWPACK. At NWT, LW_{in} was estimated as a function of T_a, RH, and SW_{in} using the approaches of Angström (1915), Dilley and O'Brien (1998), and Crawford and Duchon (1999) as detailed in Jennings et al. (2018a). LW_{in} was estimated at YOS-DAN (Lundquist et al., 2016) using the equations presented in Prata (1996) and Deardoff (1978). For the other sites, we used the empirical Unsworth and Monteith (1975) formulation that is included with the forcing data preprocessor MeteoIO (Bavay and Egger, 2014). At the HJA stations, we bias-corrected the LW_{in} estimate based on one year of LW_{in} observations from HJA-VAN that showed a -56.9 W m^{-2} wintertime bias, which may have been related to site vegetation conditions. This was significantly larger in magnitude than the bias found in the Unsworth and Monteith (1975) estimate by Flerchinger et al. (2009), suggesting its performance is more spatially variable than previously noted. This finding also underscores the need for enhanced monitoring of the radiation budget at snow modeling sites (Lapo et al., 2015; Raleigh et al., 2015, 2016). No bias corrections or additional methods were examined at the JD and SSC stations.

To validate model output, we compared daily simulated SWE and snow depth to observations at our study stations. SWE was observed at all HJA stations, SSC-UPR, YOS-DAN, and both NWT stations, while snow depth was observed at

Unknown
Field Code Changed

Unknown
Field Code Changed

Keith Jennings 7/15/2019 9:27 AM
Deleted: the former

Unknown
Field Code Changed

Keith Jennings 7/15/2019 9:27 AM
Deleted: the latter

Unknown
Field Code Changed

Unknown
Field Code Changed

Unknown
Field Code Changed

all HJA stations, both SSC stations, and all JD stations. All SWE data were derived from automated snow pillow measurements except for manual snow pit observations at NWT-SDL (Williams, 2016), while automated ultrasonic snow depth sensors produced all snow depth data. Comparisons were made at the daily time scale when either simulated or observed SWE or snow depth were > 0 mm. This was done to prevent artificial enhancement of objective function values during periods when snow cover was absent.

3.2 Precipitation phase methods

We evaluated a selection of precipitation phase methods found in the literature, including the more typical T_a thresholds and ranges as well as methods incorporating humidity (Table 2). The T_a thresholds were chosen to represent the spatial variability of rain-snow partitioning in the western United States, where values of approximately 1°C are common near the Pacific Coast, increasing towards 3°C in the Rocky Mountains (Jennings et al., 2018). Additionally, despite significant literature showing its poor performance (e.g., Jennings et al., 2018; Marks et al., 2013), we included a 0°C T_a threshold in the analysis because it is still widely used in observational and model-based hydrologic studies. For the T_a, dew point (T_d), and wet bulb (T_w) thresholds, precipitation was designated as all-rain when the temperature was warmer than the threshold and all-snow when cooler than or equal to the threshold. When using the T_a ranges, a linear mix of precipitation phase was given when T_a fell within the range during precipitation with all-rain above the warmer threshold and all-snow below the cooler threshold. The binary regression methods (Froidurot et al., 2014; Jennings et al., 2018b) computed the probability of snow (p_{snow}) as a function of T_a and RH (Reg_{Bi}, Eq. 1) and as a function of T_a, RH, and surface pressure (P_s, Reg_{Tri}, Eq. 2). Precipitation was set to be all snow when p_{snow} >= 0.5 and rain when p_{snow} < 0.5:

$$p_{snow} = \frac{1}{1 + e^{(-10.04 + 1.41T_a + 0.09RH)}} \tag{1}$$

$$p_{snow} = \frac{1}{1 + e^{(-12.8 + 1.41T_a + 0.09RH + 0.03P_s)}} \tag{2}$$

Each of the study sites included RH as part of their meteorological observations, but only the HJA and JD stations had observations of T_d, while no sites had long-term T_w measurements. To keep precipitation phase methods constant across the sites, we calculated T_d (Alduchov and Eskridge, 1996) and T_w (Stull, 2011) as empirical functions of T_a and RH. The empirical formulation tracked observed T_d at JD with an r^2 of 1.0 and a slight cool bias of -0.3°C. There were no observations on which to validate the T_w estimates, but Stull (2011) shows biases typically < 1.0°C.

It should be noted that although this work pursues a wide variety of precipitation phase methods, it is not wholly comprehensive. For example, some models fit a sigmoidal curve between two thresholds when assigning precipitation phase in a T_a range (e.g., Fassnacht et al., 2013; Kienzle, 2008; Leavesley et al., 1995). However, we did not include this method because it should produce little variability in annual snowfall fraction relative to the linear T_a ranges if a uniform distribution of T_a and precipitation is assumed within the T_a range. Additionally, models of cloud microphysics are increasingly used to

Keith Jennings 7/2/2019 2:58 PM
Formatted: Subscript

Keith Jennings 7/2/2019 2:59 PM
Formatted: Subscript

Keith Jennings 7/3/2019 9:36 AM
Formatted: Justified, Indent: First line: 0.5", Line spacing: 1.5 lines

Keith Jennings 7/3/2019 9:35 AM
Deleted: air temperature

Keith Jennings 7/3/2019 9:35 AM
Formatted: Subscript

Keith Jennings 7/16/2019 7:51 AM
Deleted:

Keith Jennings 7/15/2019 9:31 AM
Deleted: temperature

simulate precipitation phase. The wide variety of microphysics schemes available suggests that a critical examination of these methods should be made, as well. However, such an analysis is beyond the scope of the current work.

Table 2. Details on the precipitation phase methods used in this work. The temperature value for each threshold method is given in the "Rain-snow threshold" column. The "All-snow threshold" and "All-rain threshold" columns respectively give the T_a values below which all precipitation is snow and above which all precipitation is rain for the T_a range methods. The regression models compute phase as a function of meteorological conditions (Eqs. 1 and 2) during precipitation and are not associated with a threshold value. Due to a large variety of precipitation thresholds and ranges (Feiccabrino et al., 2015; Harpold et al., 2017b; Jennings et al., 2018b), the citations are listed if the values are approximate.

Category	Method	Rain-snow threshold (°C)	All-snow threshold (°C)	All-rain threshold (°C)	Citation(s)
T_a threshold	T_{a0}	0.0	NA	NA	(Jennings et al., 2018a; Lehning et al., 2002b*; Lynch-Stieglitz, 1994; Rajagopal and Harpold, 2016; Wen et al., 2013)
	T_{a1}	1.0	NA	NA	
	T_{a2}	2.0	NA	NA	
	T_{a3}	3.0	NA	NA	
T_a range	T_{ar0}	NA	-0.5	0.5	(Cherkauer et al., 2003; Tarboton and Luce, 1996; United States Army Corps of Engineers, 1956; Wayand et al., 2016; Wigmosta et al., 1994)
	T_{ar1}	NA	-1.0	3.0	
T_d threshold	T_{d0}	0.0	NA	NA	(Marks et al., 2013; Zhang et al., 2017)
	T_{d1}	1.0	NA	NA	
T_w threshold	T_{w0}	0.0	NA	NA	(Anderson, 1968; Harder and Pomeroy, 2013; Marks et al., 2013)
	T_{w1}	1.0	NA	NA	
Binary logistic regression	Reg_{Bi}	NA	NA	NA	(Froidurot et al., 2014; Jennings et al., 2018b)
	Reg_{Tri}	NA	NA	NA	

*The SNOWPACK default is a 1.2°C T_a threshold.

3.3 Evaluating the effect of precipitation phase method selection on snowfall fraction and simulated snow cover evolution

For water years (WY, 1 October of the previous calendar year to 30 September) 2004–2011, we simulated snowpack accumulation and melt at the 11 stations using the SNOWPACK model. Each station had a total of 12 unique model runs corresponding to the different precipitation phase methods. All forcing data and model setup remained the same across the runs at each site except for the precipitation phase method. For each site and for each of the different precipitation phase methods we quantified the average annual snowfall fraction, peak SWE magnitude, the timing of peak SWE, snowmelt rate, and snow cover duration (Fig. 2). For this work, snowmelt rate is computed as the daily average snowmelt rate between peak SWE timing and the first day where SWE = 0 mm. Snow cover duration is the total number of days when simulated SWE is greater than zero. For each of the sites we present the average simulated quantities noted above as well as the range and relative differences of snow metrics associated with the different precipitation phase methods. In this work, relative difference is defined as the percentage difference between the maximum and minimum snow metric value (e.g., if T_{a0} produced a minimum peak SWE of 200 mm and T_{a3} produced a maximum peak SWE of 400 mm, the relative difference

Keith Jennings 7/3/2019 9:36 AM
Deleted: ... [2]

Unknown
Field Code Changed

Unknown
Field Code Changed

Keith Jennings 7/15/2019 9:33 AM
Formatted Table

Unknown
Field Code Changed

Keith Jennings 7/15/2019 9:33 AM
Formatted Table

would be 100%). Stations with greater variability in their snow cover evolution metrics were considered to be more sensitive to the choice of precipitation phase method.

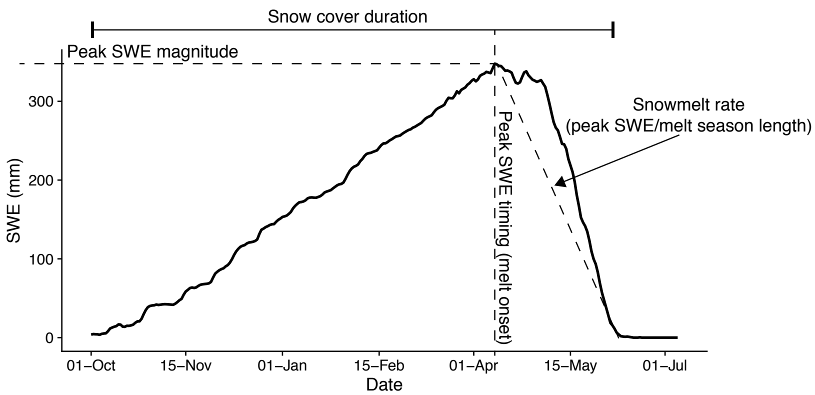


Figure 2. Example niveograph showing seasonal snow cover evolution, adapted from Trujillo and Molotch (2014).

3.4 Computing the effect of deviating from an optimized rain-snow T_a threshold

We were interested in identifying an optimized rain-snow T_a threshold for our study sites despite a lack of direct precipitation phase observations. We did this through the use of several data sources: 1) the spatially continuous rain-snow T_a threshold map from Jennings et al. (2018b); 2) the observed rain-snow T_a threshold (Jennings et al., 2018b) from the measurement location closest to each study site; 3) changes in observed SWE; and 4) changes in observed snow depth at each site to infer snowfall. For 3 and 4, we used a modified version of the approach of Rajagopal and Harpold (2016) to predict precipitation phase by designating a daily increase of SWE or snow depth as snowfall and a zero change or decrease as rainfall when precipitation was greater than 2.54 mm and SWE or snow depth was greater than 0 mm. We then binned snowfall frequency per 1°C T_a bin and computed the rain-snow T_a threshold using the hyperbolic tangent equation of Dai (2008). Thresholds from 1, 2, 3, and 4 were then arithmetically averaged and rounded to the nearest integer value to produce an optimized rain-snow T_a threshold for each study site, consistent with the studied threshold values in Table 2. Following this step, we evaluated the effect of deviating by ± 1°C from the optimized threshold by quantifying differences in peak SWE, peak SWE timing, snow cover duration, and snowmelt rate across the selected thresholds.

3.5 Evaluating the relationships between climate and snow cover sensitivity

In addition to quantifying the variability introduced by the different precipitation phase methods, we evaluated the control exerted by daily meteorology and seasonal climate on snow cover evolution sensitivity at our study sites. We first examined how daily average T_a and RH introduced variability into simulated snowfall fraction. We did this by grouping all

Keith Jennings 7/2/2019 9:22 AM
Formatted: Heading 2, Space After: 0 pt, Line spacing: single, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian

Keith Jennings 7/2/2019 9:20 AM
Formatted: Subscript

Keith Jennings 7/2/2019 9:33 AM
Formatted: Subscript

Keith Jennings 7/2/2019 9:33 AM
Formatted: Subscript

Unknown
Field Code Changed

Keith Jennings 7/2/2019 9:33 AM
Formatted: Subscript

Unknown
Field Code Changed

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/15/2019 9:47 AM
Formatted: Subscript

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto, Subscript

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto

Keith Jennings 7/2/2019 9:34 AM
Formatted: Subscript

Keith Jennings 7/2/2019 9:33 AM
Formatted: Font color: Auto, English (US)

Keith Jennings 7/2/2019 9:16 AM
Deleted: 4

daily meteorological conditions in 1°C T_a bins from -8°C to +8°C and 10% RH bins from 60% to 100% on days with precipitation. We then calculated the standard deviation in daily snowfall fraction within each bin across all sites and methods. Those results were used to determine the T_a range that produced the greatest standard deviation in daily snowfall fraction. Next, we computed the proportion of December through May (Dec–May, i.e., winter and spring) precipitation that fell within that T_a range at each site for each simulation year and used that percentage to predict annual snowfall fraction range with ordinary least squares regression. Finally, we evaluated how Dec–May T_a and precipitation produced variability in peak SWE by plotting peak SWE range as a function of the two meteorological quantities and plotting a predictive surface with a loess function.

4 Results

4.1 Model validation

Figure 3 displays the mean bias and r^2 values for the different precipitation phase methods relative to observations of SWE and snow depth, aggregated across all stations with a given validation measurement. In terms of mean bias, the binary regression models, Reg_{Bi} and Reg_{Tri}, as well as the T_{a1} threshold provided the best performance with average values between 3.1 mm and 9.1 mm compared to observed SWE and between 4.9 mm and 14.5 mm compared to observed snow depth. Conversely, the T_{a0}, T_{a2}, T_{a3} thresholds and the T_{ar0} range provided the worst performance with T_{a2} and T_{a3} overpredicting and T_{a0} and T_{ar0} underpredicting snow accumulation by upwards of 100 mm relative to observed SWE and 200 mm and greater relative to observed snow depth. There was relatively little divergence in r^2 values across the methods, with differences of only 0.07 and 0.08 between the maximum and minimum average r^2 values for SWE and snow depth, respectively. The lowest r^2 values were produced by the T_{a0} threshold and T_{ar0} range, while T_{d1}, T_{w1} and the higher T_a thresholds produced the highest values.

Keith Jennings 7/3/2019 9:39 AM
Deleted: quantified

Keith Jennings 7/16/2019 7:52 AM
Deleted: PPT

Keith Jennings 7/3/2019 9:39 AM
Deleted: controlled

Keith Jennings 7/3/2019 9:39 AM
Deleted: at our study sites by computing a multiple linear regression with

Keith Jennings 7/3/2019 9:39 AM
Deleted:

Keith Jennings 7/3/2019 9:39 AM
Deleted: acting as the predictor variables

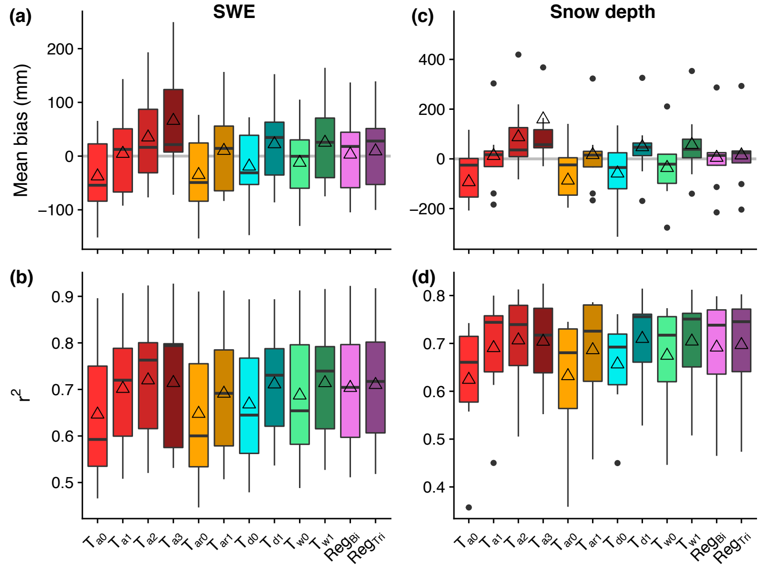


Figure 3. Mean bias (top row) and r^2 (bottom row) values for the SNOWPACK simulations relative to observed SWE (a,b) and snow depth (c,d). The boxplots show the median, interquartile range, minimum, maximum, and outlying values for each objective function for the different precipitation phase methods at all stations. The open triangles indicate the mean objective function value for that precipitation phase method at all stations.

Figure 4 presents model performance relative to observed SWE and snow depth at the different sites. Mean biases were lowest at the NWT stations and at SSC-UPR relative to SWE observations and at the JD stations and SSC-LWR relative to snow depth observations. Average r^2 values were between 0.65 and 0.91 for SWE except at NWT-SDL (0.52) and HJA-VAN (0.51), and 0.61 and 0.79 for snow depth except at JD-124 (0.46).

Keith Jennings 5/22/2019 3:40 PM
Formatted: Caption, Don't keep with next

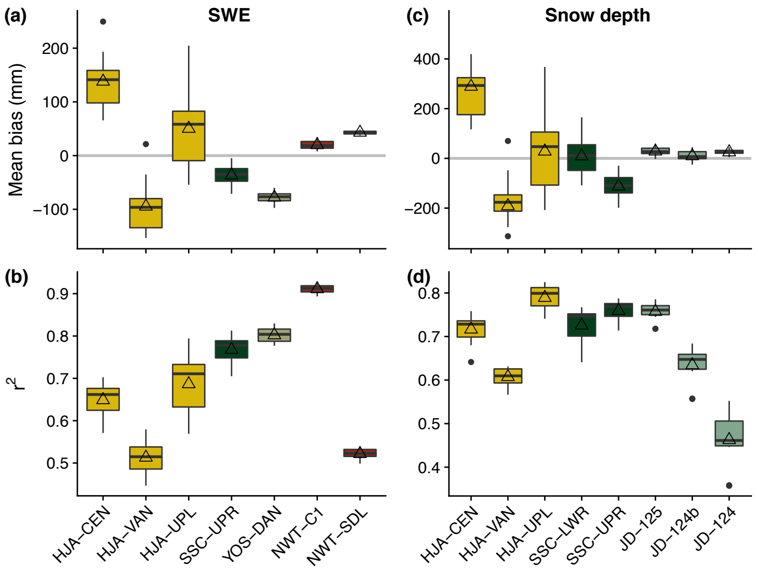


Figure 4. Mean bias (top row) and r^2 (bottom row) values for the SNOWPACK simulations relative to observed SWE (a,b) and snow depth (c,d). The boxplots show the median, interquartile range, minimum, maximum, and outlying values for each objective function for the different precipitation phase methods at a given station. The open triangles indicate the mean objective function value for all precipitation phase methods at that station. Note: in panel (c) the low mean biases for JD snow depth are due to small snow depth values at the site. Mean relative biases at these stations were 35.4% (JD-125), 3.8% (JD-124b), and 35.7% (JD-124).

Keith Jennings 5/22/2019 3:41 PM
Formatted: Caption, Don't keep with next

4.2, Mean simulated snow cover properties

Keith Jennings 5/22/2019 3:42 PM
Deleted: 1

The study locations showed significant differences in simulated snow accumulation and melt. Values presented in Table 3 were computed by taking the mean and standard deviation of the given snow metric using all 12 simulations at each station, where each simulation corresponded to a different precipitation phase method. Mean peak SWE ranged from 73.1 mm at JD-124 to 1146.1 mm at HJA-UPL. The date of peak SWE, or melt onset, also displayed large variability with values ranging from 24 January at JD-125 to 13 May at NWT-SDL. Melt rates were all greater than 10 mm d^{-1} during the ablation season except for the JD stations and the greatest melt rates were simulated at HJA-UPL and NWT-SDL. Snow cover duration was greatest at NWT-SDL at 241.1 d, while snow cover was simulated for less than 3 months, on average, at JD-125 and JD-124.

Table 3. Mean snow cover evolution metrics for the 11 stations. Each mean and standard deviation was calculated across all water years and all precipitation phase methods.

	Peak SWE (mm)		Peak SWE date		Melt rate (mm d^{-1})		SCD (d)	
Station	**Mean**	**SD**	**Mean**	**SD (d)**	**Mean**	**SD**	**Mean**	**SD**
HJA-CEN	522.7	252.9	16-Feb	22.0	15	3.9	158.4	28.2
HJA-VAN	643.1	305.9	14-Feb	22.2	14.5	3.2	173.1	27.9
HJA-UPL	1146.1	469.9	14-Mar	23.0	24.9	7.3	201.1	22.4
SSC-LWR	531.9	160.1	8-Mar	19.0	17.6	3.6	145.6	27.8
SSC-UPR	617.9	298.8	5-Mar	26.6	17.6	6.0	149.2	35.6
YOS-DAN	674.4	236.7	18-Mar	17.5	10.9	4.1	208.2	40.3
JD-125	83.4	46.5	24-Jan	28.5	4	1.5	78.1	31.5
JD-124b	177.5	87.6	1-Feb	25.8	5.7	2.5	122.4	23.9
JD-124	73.1	35.0	2-Feb	31.4	3.5	2.8	77.6	30.7
NWT-C1	407.2	78.5	22-Apr	10.8	11.9	2.8	225.3	19.2
NWT-SDL	915	234.2	13-May	10.0	24.4	10.1	241.1	14.9

4.3 Effect of precipitation phase method on simulated snowfall fraction

Average annual snowfall fraction (all methods, all years) ranged from 32.3% at the HJA-CEN station to 92.4% at the YOS-DAN station (Table 4, Fig. 5). In this case, more strongly seasonal precipitation at YOS-DAN (Table 1) produced a higher annual snowfall fraction than NWT-SDL, despite the former station's warmer average T_a. YOS-DAN and NWT-SDL also had the lowest ranges at 10.1% and 10.3%, respectively, suggesting precipitation phase method selection was less important relative to the other stations. Conversely, the range in annual snowfall fraction simulated by the different methods was greater than 18% at all remaining stations, reaching a maximum of 32.3% at SSC-LWR. For all sites except YOS and NWT, relative differences were greater than 30%. In some years at HJA, SSC, and JD, the relative difference between the minimum annual snowfall fraction and the maximum exceeded 100%, meaning the methods producing the most snow simulated more than double the annual snowfall fraction of those producing the most rainfall. The greatest relative difference in annual snowfall fraction of 126.9% was simulated at HJA-CEN, more than 10-times greater than at YOS-DAN and NWT-SDL.

Keith Jennings 7/15/2019 9:52 AM
Deleted:

Keith Jennings 5/22/2019 3:42 PM
Deleted: 2

Keith Jennings 7/3/2019 10:25 AM
Deleted: 3

Keith Jennings 7/15/2019 9:54 AM
Deleted: These two stations

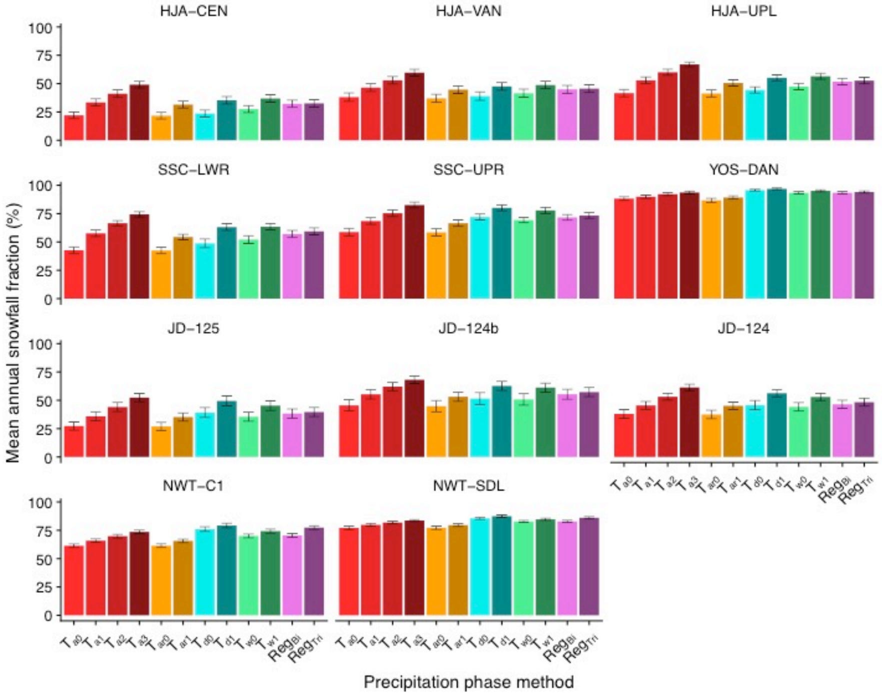


Figure 5. Mean annual snowfall fraction at the 11 study stations for the different precipitation phase methods. The whiskers represent the standard error of annual snowfall fraction for the 8 simulation years. For this plot and all subsequent figures showing the station data, the maritime sites are shown in the top two rows, the intermountain site is in the third row, and the continental site is in the bottom row.

Keith Jennings 5/22/2019 3:40 PM
Deleted: 3

Keith Jennings 7/2/2019 2:51 PM
Deleted: 12

Table 4. Statistics for average annual snowfall fraction computed using the different precipitation phase methods across all simulation years at the 11 study stations. The range was calculated by subtracting the lowest average annual snowfall fraction from the highest average annual snowfall fraction at each station. The relative difference was then computed as the range divided by the minimum and multiplied by 100%. T_{a0} and T_{ar0} typically produced the lowest average annual snowfall fractions, while T_{a3} and T_{d1} led to the highest average annual snowfall fractions.

	Annual snowfall fraction (%)		
Station	**Average**	**Range**	**Relative difference**
HJA-CEN	32.3	27.4	126.9
HJA-VAN	45.5	22.6	61.1
HJA-UPL	51.8	25.7	62.3
SSC-LWR	56.8	32.3	76.7
SSC-UPR	71.2	25.0	42.8
YOS-DAN	92.4	10.1	11.6
JD-125	39.1	26.0	97.1
JD-124b	55.7	23.2	51.8
JD-124	47.9	23.9	63.6
NWT-C1	70.4	18.2	29.7
NWT-SDL	82.4	10.3	13.4

4.4 Effect of precipitation phase method on simulated snow accumulation and melt

There were marked differences between the stations in terms of the effect precipitation phase method choice had on seasonal snow cover evolution. Figure 6 presents the simulated mean daily SWE of all simulation years at the study stations along with the difference between the minimum and maximum mean daily SWE produced by the precipitation phase methods. At HJA, SSC, and JD, differences increased throughout the accumulation period, reaching a maximum after peak SWE during the snowmelt season. At NWT and YOS, the differences were typically negligible throughout the accumulation season as cold winter and early spring temperatures produced little divergence in the amount of snowfall versus rainfall simulated by the different methods. At these stations, differences in the mean daily SWE produced by the precipitation phase methods did not appear until approximately the date of peak SWE. Mean daily SWE differences were always less than 90 mm at NWT and YOS, while they sometimes exceeded 200 mm at HJA and SSC. Differences were typically small in magnitude at JD, but were proportionally large due to low mean daily SWE values.

Keith Jennings 7/2/2019 2:51 PM
Deleted: 12

Keith Jennings 7/15/2019 9:56 AM
Deleted: s

Keith Jennings 7/15/2019 9:56 AM
Deleted: site

Keith Jennings 5/22/2019 3:42 PM
Deleted: 3

Keith Jennings 7/3/2019 10:26 AM
Deleted: 4

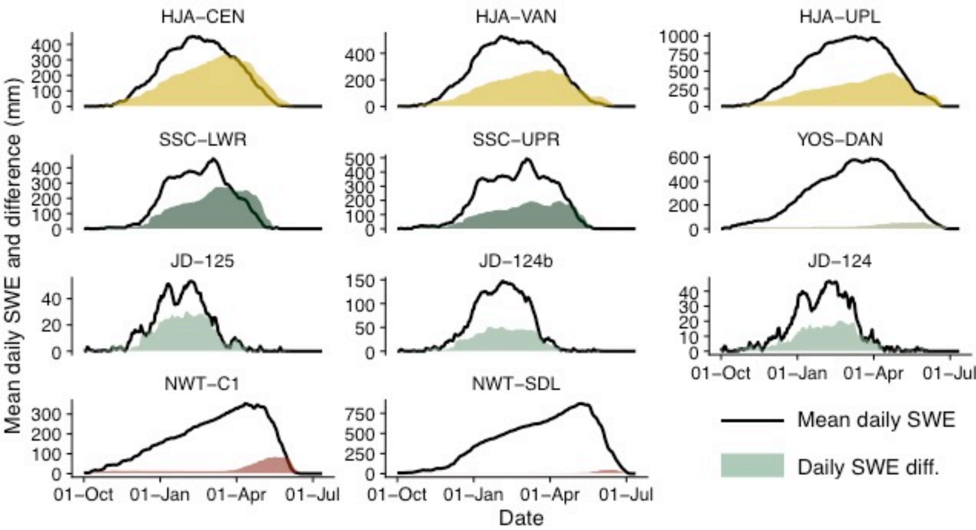


Figure 6. Mean daily simulated SWE (solid black line) and the difference between maximum and minimum mean daily SWE (shading) at the study stations. The mean daily SWE was computed by averaging the simulated SWE on each day for all precipitation phase methods across the simulation years. The difference was calculated by subtracting the minimum mean daily SWE from the maximum mean daily SWE produced by the different precipitation phase methods (mean daily SWE plots broken out by precipitation phase method can be viewed in Figs. S1–S11). The T_{a0} and T_{ar0} methods typically produced the minimum mean daily SWE, while T_{a3} and T_{d1} produced the maximum.

Breaking down the analysis to the individual snow cover evolution metrics reveals more differences in the sensitivity of the sites to precipitation phase method selection (Fig. 7). In terms of peak SWE range, the HJA and SSC stations were most sensitive, with average ranges all greater than 200 mm, exceeding 400 mm in some years (Fig 7a). Conversely, YOS and NWT were relatively insensitive as their average ranges were all less than 65 mm. The largest annual range in peak SWE at the NWT and YOS stations was just 90.8 mm at NWT-C1, which was considerably less than the maximum peak SWE range of 592.5 mm simulated at HJA-UPL. Although the JD stations showed little sensitivity in terms of range with average annual peak SWE differences less than 55 mm, they expressed significant sensitivity when looking at relative differences (Table 5) due to their low mean annual peak SWE (Table 3). Thus, percentage-wise, JD was as sensitive as the two warm maritime sites to the selection of a precipitation phase method. At JD, HJA, and SSC it was common for the relative difference between minimum and maximum modeled peak SWE to be well above 50%, meaning a significant proportion of water was simulated to have infiltrated or run off using one precipitation phase method versus being stored in the snowpack using another method. This is in stark contrast to the 4.0% and 1.8% relative differences at YOS-DAN and NWT-SDL.

Keith Jennings 5/22/2019 3:40 PM
Deleted: 4

Keith Jennings 7/3/2019 10:27 AM
Deleted: 5

Keith Jennings 7/3/2019 10:27 AM
Deleted: 5

Keith Jennings 7/2/2019 2:56 PM
Deleted: meaning a significant proportion of water was simulated to have run off using one precipitation phase method versus being stored in the snowpack

JD and HJA were also sensitive to precipitation phase method selection in terms of peak SWE date (Fig. 7b) with 4 of the 6 stations having average ranges greater than 2 weeks. In some simulation years, peak SWE date ranges exceeded 1 month at HJA, JD, and SSC. We found the greatest differences in peak SWE dates were generally simulated on years with low/transient snow cover. In these cases, late-season precipitation was simulated as rain by the low T_a thresholds and snow by the high T_a thresholds, meaning an early SWE maximum was recorded as the peak in the former case and a late SWE maximum in the latter case. Compared to the other stations, peak SWE date ranges were generally small at NWT-SDL and YOS-DAN with an average range of just 0.8 d at the former and 2.5 d at the latter.

Similar sensitivities were simulated for snow cover duration (Fig. 7c) with the warm maritime sites and JD being the most impacted by precipitation phase method choice. JD-125 had the greatest average range in annual snow cover duration at 42.0 d and all other ranges at JD and HJA were greater than 26.8 d. SSC-LWR and SSC-UPR expressed slightly lower average ranges at 20.9 d and 18.1 d, respectively. NWT-C1 approached the sensitivity of the warmer stations, while NWT-SDL and YOS-DAN were again the least sensitive. Relative differences were greatest at JD (Table 5) because simulated snow cover was typically of a shorter duration compared to the other sites (Table 3). The average relative difference at JD-125 of 120.4% meant that snow cover simulated using the T_{a3} threshold lasted twice as long as snow cover using the T_{a0} threshold. Notably, there was an order of magnitude of difference between JD, HJA, and SSC and YOS and NWT with average relative differences in snow cover duration greater than 10% at the former three sites and less than 10% at the latter two.

Differences among the stations were relatively low for melt rate (Fig. 7d) with the interquartile ranges generally showing some degree of overlap. JD stations had the greatest sensitivity in terms of relative differences (Table 5) due to their low mean annual melt rates, which were an order of magnitude lower than those simulated at the other sites (Table 3). Overall, melt rate at YOS-DAN was the least affected by precipitation phase method selection in terms of range and relative difference. It should be noted here again that the forcing data were kept constant for the different modeling scenarios—only the precipitation phase methods were varied. Thus, any changes to melt rate were caused by shifts in snowmelt timing and by the hydrologic and energy balance impacts of rain versus snow.

To close this section, it is useful to visualize what these differences look like in terms of annual snow cover evolution. Figure 8 shows examples of large ranges in peak SWE (a), peak SWE date and melt rate (b), and snow cover duration (c), while the bottom panel exemplifies a year with little simulated divergence in the snow cover metrics (d). For peak SWE, it is evident that differences in snow accumulation begin with the first snowfall and the SWE simulated by the various precipitation phase methods continues to diverge throughout the accumulation season (Fig. 8a). In the next panel down, simulated SWE is fairly consistent and tracks observed SWE early in the accumulation season before diverging at the onset of winter snowmelt (Fig. 8b). T_{a2} and T_{a3} produce a late peak SWE on 7 April with melt rates greater than 24 mm d^{-1}, while all other methods predict peak SWE to occur 52 d to 53 d earlier with slower melt rates between 15.5 mm d^{-1} and 16.3 mm d^{-1}. In terms of snow cover duration, years with transient snow tended to be most sensitive as Figure 8c illustrates. Modeled snow depth generally follows the observed pattern of accumulation and melt, but the methods diverge in terms of

Keith Jennings 7/3/2019 10:28 AM
Deleted: 5

Keith Jennings 7/3/2019 10:29 AM
Deleted: 5

Keith Jennings 7/15/2019 10:02 AM
Deleted: Finally, d

Keith Jennings 7/3/2019 10:34 AM
Deleted: 5

Keith Jennings 7/15/2019 10:02 AM
Deleted:). Ranges for the different stations were relatively similar to one another

Keith Jennings 7/15/2019 10:55 AM
Formatted: Subscript

Keith Jennings 7/15/2019 10:55 AM
Formatted: Subscript

Keith Jennings 7/15/2019 10:55 AM
Formatted: Superscript

Keith Jennings 7/15/2019 10:57 AM
Formatted: Superscript

how often and for how long they reach 0 mm. Finally, there is little divergence in simulated SWE throughout the entire accumulation season at NWT-C1 with small differences arising only after the date of peak SWE (Fig. 8d).

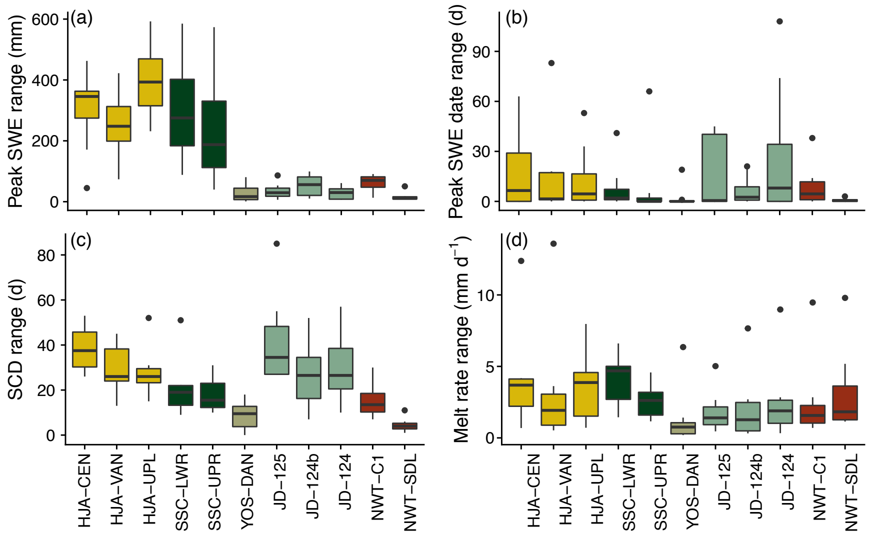


Figure 7. The annual range in simulated peak SWE (a), peak SWE date (b), snow cover duration (c), and melt rate (d) due to precipitation phase method selection at the study stations.

Table 5. Average relative differences in annual peak SWE, snow cover duration, and melt rate at the 11 stations. Relative differences were not computed for peak SWE date because the relative difference value would change depending on if day of year or day of water year were used in the calculation.

	Average relative difference (%)		
Station	**Peak SWE**	**SCD**	**Melt rate**
HJA-CEN	86.6	28.0	33.2
HJA-VAN	55.2	19.6	27.5
HJA-UPL	49.1	15.4	19.6
SSC-LWR	78.6	14.9	26.7
SSC-UPR	43.9	13.4	15.6
YOS-DAN	4.0	4.8	11.5
JD-125	74.6	120.4	220.2
JD-124b	54.7	28.7	47.8
JD-124	71.9	72.4	235.5
NWT-C1	16.9	7.0	26.0
NWT-SDL	1.8	1.9	13.0

Keith Jennings 5/22/2019 3:40 PM
Deleted: 5

Keith Jennings 5/20/2019 3:45 PM
Deleted: because

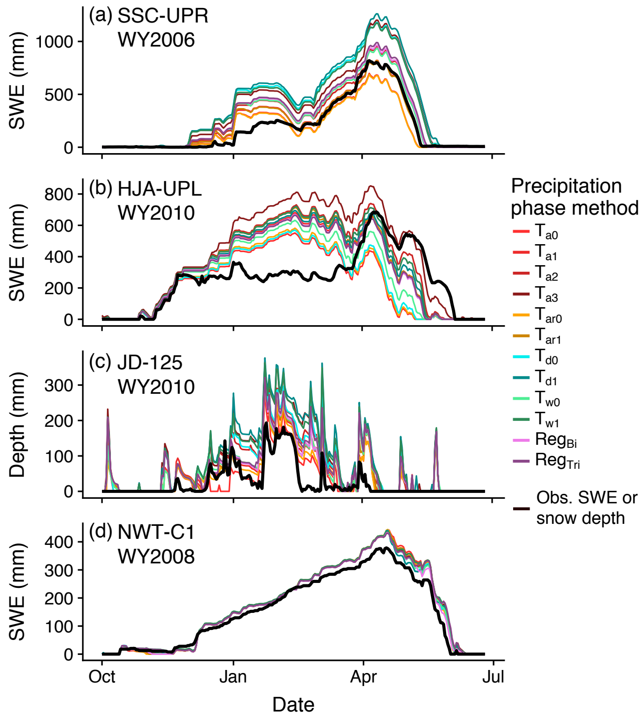


Figure 8. Example simulation years from a selection of stations showing pronounced differences in peak SWE magnitude (a), peak SWE date and snowmelt rate (b), and snow cover duration (c). The bottom panel exemplifies a site and year with little divergence in the studied snowpack metrics (d). In all panels simulated SWE and snow depth are represented by colored lines for the different methods, while observed SWE and snow depth are shown in black.

Keith Jennings 7/15/2019 11:09 AM
Formatted: Caption

4.5 The effect of deviating from an optimized rain-snow T_a threshold

Using the data outlined in Sect. 3.4, we identified optimized rain-snow T_a thresholds of 1.0°C for HJA and SSC, 2.0°C for YOS-DAN and JD, and 3.0°C for NWT (for all snowfall frequency curves and threshold values, please see Figs. S12 and S13, and Table S2). Again, we rounded to the nearest integer value to be consistent with the other T_a thresholds studied in this work. Consistent with our findings in Sect. 4.4, the warm maritime HJA and SSC stations were profoundly affected by deviations from the optimized threshold (Fig. 9). Differences at these sites produced by deviating by only ±1°C from the optimized thresholds range between 141 and 403 mm for peak SWE, 1 and 16 d for peak SWE DOWY, and 9 and 29 d for SCD. Compare this to 1 to 10 mm for peak SWE, 0 to 1 d for peak SWE DOWY, and 1 to 5 d for SCD at the YOS

Keith Jennings 7/2/2019 9:42 AM
Formatted: Normal

Keith Jennings 7/2/2019 2:17 PM
Formatted: Subscript

Keith Jennings 7/2/2019 2:18 PM
Formatted: Subscript

and NWT stations. The consistent story is again that threshold choice makes a much larger impact at a warm site relative to a cold one.

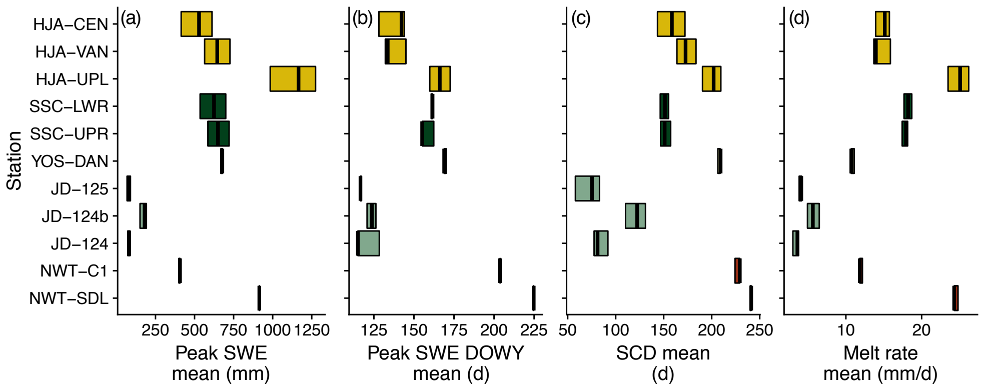


Figure 9. Mean peak SWE (a), peak SWE day of water year (DOWY, b), snow cover duration (SCD, c), and melt rate (d) for the different stations. The width of the bar in each plot represents the spread between the mean snow cover metric value when the model was run with the optimized threshold plus 1°C and the optimized threshold minus 1°C. The center line represents the snow cover metric value when the model was run with the optimized threshold. Note: at NWT, only -1°C from the optimized value of 3°C was evaluated because we did not include a 4°C threshold in this work.

4.6 Climatic controls on precipitation phase method sensitivity

In general, daily snowfall fraction standard deviation was greatest at daily T_a values between 0°C and 4°C (Fig. 10a). RH provided a secondary control, with greater daily snowfall fraction variability at lower RH values (Fig. 10b). Overall, the largest standard deviations in snowfall fraction were simulated at daily RH less than 80% and T_a between 1°C and 3°C. However, it should be noted that 75.2% of all precipitation recorded at the study stations occurred in the 90%–100% RH bin. Therefore, although daily snowfall fraction standard deviations were highest at lower RH values, the majority of the variability in snowfall fraction was an effect of T_a. In this context, the percentage of Dec–May precipitation that fell within the 0°C–4°C T_a range explained 80.1% of the variance in annual snowfall fraction range across the study sites (Fig. 11).

Keith Jennings 7/2/2019 2:49 PM
Formatted: English (US)

Keith Jennings 7/2/2019 2:23 PM
Formatted: Keep with next

Keith Jennings 7/2/2019 2:23 PM
Formatted: Caption

Keith Jennings 5/22/2019 3:42 PM
Deleted: 4

Keith Jennings 7/3/2019 10:53 AM
Deleted: 6

Keith Jennings 7/3/2019 10:53 AM
Deleted: 6

Keith Jennings 5/21/2019 3:49 PM
Deleted: standard deviation

Keith Jennings 7/3/2019 10:54 AM
Deleted: 7

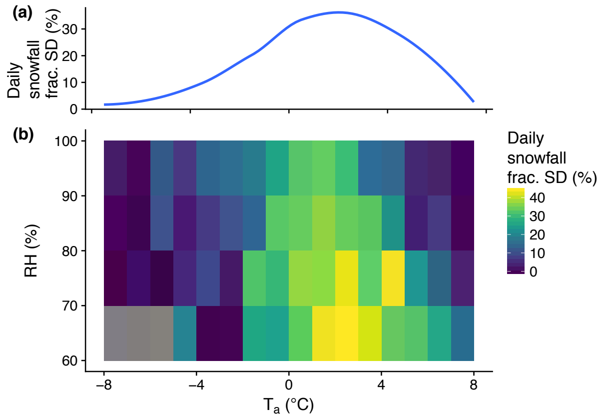


Figure 10. The standard deviation of daily snowfall fraction as a function of T_a (a) and as a function of T_a and RH (b). We binned the meteorological quantities within the ranges shown and calculated the standard deviation of snowfall fraction per T_a bin (a) and T_a/RH bin (b) using simulated precipitation phase from all stations and all methods.

Keith Jennings 5/22/2019 3:40 PM
Deleted: 6

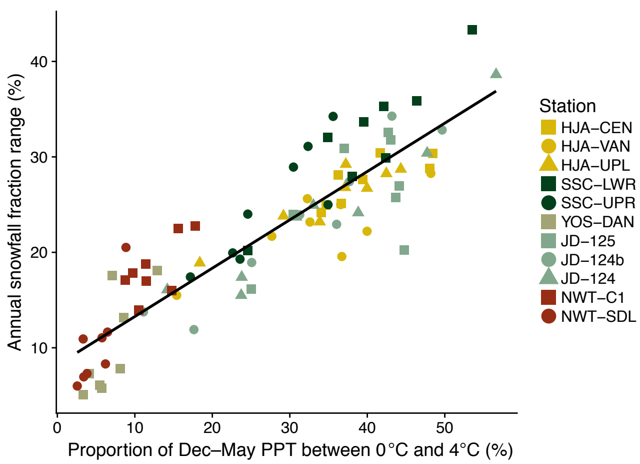


Figure 11. Range in annual snowfall fraction as predicted by the proportion of Dec–May PPT falling between 0°C and 4°C. Each point represents one simulation year at a station identified by the color and shape. The black line of best fit was calculated using ordinary least squares regression (r^2 = 0.80, *p-value* < 0.0001).

We next evaluated how sensitivity in peak SWE was related to seasonal climate. In this case, warmer T_a and increased PPT were both associated with greater ranges in the peak SWE simulated by the different precipitation phase

Keith Jennings 5/22/2019 3:40 PM
Deleted: 7

Keith Jennings 5/23/2019 1:45 PM
Deleted: A multiple linear regression with Dec–May T_a and Dec–May PPT as the predictor variables explained 80.6% of the variance in the range of annual peak SWE at the stations (Fig. 8)

Keith Jennings 5/23/2019 1:45 PM
Deleted: .

methods (Fig. 12). This meant the maritime sites HJA and SSC had the greatest sensitivity to precipitation phase method due to their relatively warm T_a and high PPT values. Conversely, moderate PPT values and lower T_a led to minimal sensitivity at the cold continental NWT stations and the cold maritime YOS-DAN station. Again, the effect of T_a on sensitivity was manifest in the data. In high snowfall years at NWT-SDL, Dec–May PPT approached that of the low Dec–May PPT years at HJA and SSC. However, despite the increased PPT at NWT-SDL, the range in peak SWE predicted by the different precipitation phase methods remained low.

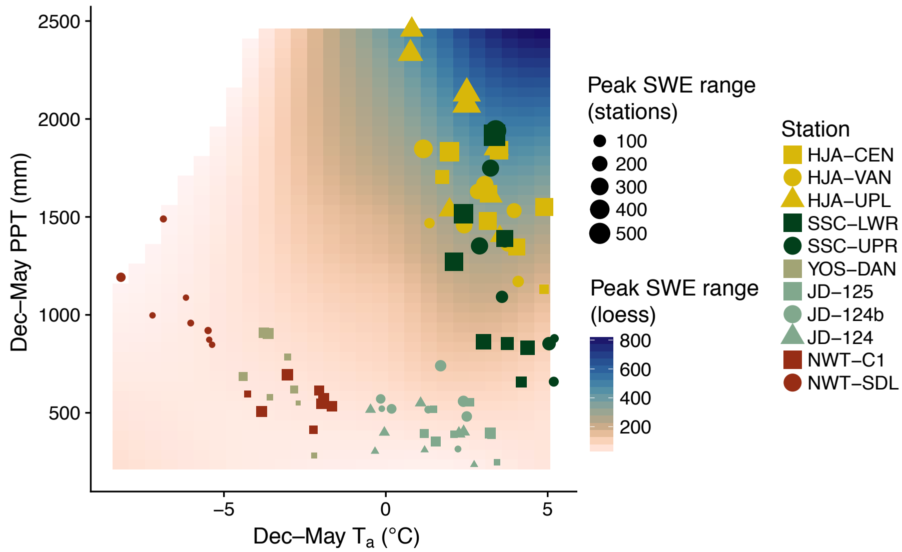


Figure 12. Range in annual peak SWE as simulated by the different precipitation phase methods at the 11 study stations. Each point represents one simulation year at a given station and larger points correspond to larger differences in maximum minus minimum peak SWE. The background shading corresponds to ranges in peak SWE predicted by a loess function fit to the station data.

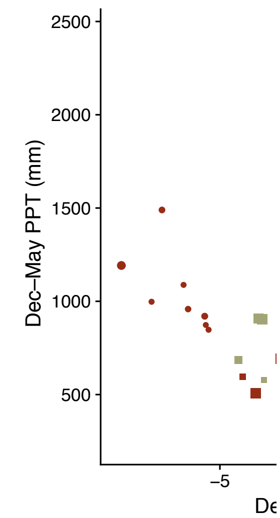
5. Discussion

5.1 A best precipitation phase method?

In this work we showed that the selection of a precipitation phase method produces varying degrees of variability in modeled snow accumulation and melt at our study stations. The different methods also expressed variable performance relative to observations of SWE and snow depth, with the binary regression models, Reg_{Bi} and Reg_{Tri}, as well as the T_{a1} threshold producing the lowest biases (Fig. 3). Previous observational work has shown that, in general, methods

Deleted: Additionally, the multiple linear regression performed here is likely only valid for the range of climatic conditions at our study sites. For example, extrapolating the regression to T_a values above 5°C would indicate greater peak SWE sensitivity for a given PPT value. However, moving towards increasingly warmer T_a would likely ... [3]

Keith Jennings 5/23/2019 1:45 PM



Deleted:

Keith Jennings 5/22/2019 3:40 PM

Deleted: 8

Keith Jennings 5/23/2019 1:45 PM

Deleted: Range in annual peak SWE as simulated by the different precipitation phase methods at the 11 study stations. Each point represents one simulation year at a given station and larger points correspond to larger differences in maximum minus minimum peak SWE. Predicting the peak SWE range as a functi ... [4]

Keith Jennings 5/24/2019 1:48 PM

Deleted: introduces

Keith Jennings 5/22/2019 3:43 PM

Deleted: A1

Keith Jennings 5/24/2019 11:25 AM

Deleted: In terms of mean bias, the binary regression models, Reg_{Bi} and Reg_{Tri}, as well as the T_{a1} threshold provided the best performance with average values between 3.1 mm and 9.1 mm compared to observed SWE and between -1.7 mm and 6.9 mm compared to observed snow depth. Conversely, the T_{a0}, T_{a2}, T_{a3} thresholds and th ... [5]

incorporating humidity information outperform T_a-only methods when it comes to predicting precipitation phase (Harder and Pomeroy, 2013; Jennings et al., 2018b; Marks et al., 2013; Ye et al., 2013). The Reg_{Bi} method, which predicts phase as a function of T_a and RH, exceeded all other methods in partitioning rain and snow in a Northern Hemisphere precipitation phase method comparison (Jennings et al., 2018b). Our study showed that Reg_{Bi} also typically produced simulations of SWE and snow depth that had low biases relative to observations (Fig. 3) and led to snow cover evolution metrics that were neither extremely high nor low compared to the other methods examined in this work. This finding is complemented by the performance of other humidity-based metrics, which produced average SWE and snow depth biases between -19.2 mm and 25.1 mm, and -58.3 mm and 56.7 mm, respectively.

This is in contrast to the T_a thresholds and ranges, which produced the largest magnitude biases. Notably, the four worst performers were the T_{a0}, T_{ar0}, T_{a2}, and T_{a3} methods, with the former two underpredicting snow accumulation and the latter two overpredicting. Across our study sites, the only T_a methods that performed well relative to observations were the T_{a1} threshold and T_{ar1} range. These modeling results confirm again that including humidity information, whether it is in the form of a binary logistic regression model, T_w, or T_d, offers advantages over a T_a-only method. It is important to note again that we chose methods that covered the range in rain-snow partitioning T_a values across our study domain or that included humidity information. The only methods not falling into this category were T_{a0} and T_{ar0}, which were chosen because they are still employed as default methods in some models and studies. Although there are some small geographic regions where such a threshold or range may be appropriate (Jennings et al., 2018b), they are unsuitable for many locations and should not be used for large-scale studies.

In the course of this work we found negligible differences between T_{a0} and T_{r0} as well as between T_{a1} and T_{r1} in terms of annual snowfall fraction (Fig. 5) and model performance (Fig. 3). This suggests the ranges and the mixed-phase precipitation they produced provided little further information on precipitation phase at the hourly model time scale relative to the thresholds. However, it should be noted there is relatively little quantitative data on the frequency and solid-liquid proportions of mixed-phase events (e.g., Yuter et al., 2006). Work from the Torino region of Italy showed mixed-phase events are relatively few compared to all-rain and all-snow events (Avanzi et al., 2014), while research in a maritime climate indicated mixed-phase events can be quite frequent (Wayand et al., 2016). Future work would therefore benefit from further explorations of the frequency of mixed-phase events and model representations thereof at multiple time scales.

Despite the analyses presented in this work, it is important to note that uncertainties in forcing data, model structure and parameters, as well as a lack of precipitation phase observations prevent this research from being able to unequivocally identify a "best" precipitation phase method for snow modeling. However, as noted above, including humidity information improves the prediction of precipitation phase relative to observations and generally increases model performance. Our primary aim in this research was to quantify how snow simulations were affected by the choice of precipitation phase method across a climatic gradient. We did not create optimized model setups at each site, but rather kept model setup consistent in order to compare the sensitivity of phase partitioning without introducing other uncertainties. Thus, the low r^2 and higher bias values at HJA-VAN, NWT-SDL, and JD-124 (Fig. 4) could likely be improved with model tuning, but we

Keith Jennings 5/24/2019 1:22 PM
Deleted: In that context, one can consider t ... [6]

Keith Jennings 5/24/2019 1:49 PM
Formatted: Subscript

Keith Jennings 5/24/2019 1:23 PM
Deleted: as a baseline given its top rank ... [7]

Keith Jennings 5/24/2019 1:32 PM
Formatted: Subscript

Keith Jennings 5/24/2019 1:33 PM
Deleted: The T_{a2}

Keith Jennings 5/24/2019 1:33 PM
Formatted ... [8]

Keith Jennings 5/24/2019 1:33 PM
Deleted:

Keith Jennings 5/24/2019 1:40 PM
Formatted ... [9]

Keith Jennings 5/24/2019 1:33 PM
Deleted: , T_{d1} and T_{w1} thresholds produced greater peak SWE and longer snow cover duration, while the lower thresholds led to less snow accumulation and shorter snow cover duration. Additionally, our model-based study showed that uncertainty in daily snowfall fraction peaked at T_a between 0°C and 4°C (Fig. 4.8), which is the same T_a range reported by Ding et al. (2014) in which precipitation phase methods exhibit degraded performance relative to observations.

Keith Jennings 5/24/2019 1:41 PM
Deleted: a referendum on the…ble to ... [10]

did not pursue such an approach. Additionally, we showed in Sect. 4.5 that deviating from an optimized T_a rain-snow threshold by ±1°C had a much larger effect on simulated snow accumulation and melt at HJA and SSC than YOS and NWT. Such a finding indicates finding an optimal threshold is much more important in areas with winter T_a near freezing.

5.2 Assumptions and limitations

Snow modeling studies are hindered by inherent uncertainties in model structure (Essery et al., 2013; Etchevers et al., 2004; Rutter et al., 2009; Slater et al., 2001) and forcing data (Lapo et al., 2015; Raleigh et al., 2015, 2016). While the research presented herein shows that precipitation phase method should be considered another critical component of model uncertainty, our work was also likely affected by the aforementioned issues in structure and forcing data which can be seen in the variability of model performance at the different sites (Fig. 4). In this work, we used the well-validated, physics-based SNOWPACK model, but past research has shown there is no best snow model and that model performance varies both within and across study sites (e.g., Rutter et al., 2009). Given this variable performance and differences in snow model structure and physics, it is possible that some models may be more or less sensitive to the choice of a precipitation phase method. Our use of a single model may overestimate or underestimate the sensitivity of snow accumulation and melt to precipitation phase method selection. Future research should therefore focus on how model choice affects the sensitivity of simulated snow cover evolution to precipitation phase method.

In addition to the uncertainties introduced by the SNOWPACK model, we used empirical methods to estimate T_d and T_w, which could affect rain-snow partitioning. We were satisfied with the performance of the T_d method as it strongly matched T_d observations from Johnston Draw (Sect. 3.2). However, there were no observations of T_w on which to validate the Stull (2011) method, which was optimized for standard surface pressure and for a range of T_a and RH values. The figures in Stull (2011) show that pressure-induced uncertainty in T_w is generally less than 1°C when RH > 50%. Additionally, the total percentage of precipitation observations falling within the Stull (2011) T_a and RH ranges was between 94.3% and 100% at our stations. Thus, we expect only marginal uncertainty to be introduced by the empirical methods. However, precipitation phase and hydrometeor temperature are strongly related to T_w (Harder and Pomeroy, 2013), suggesting there should be enhanced monitoring of T_w at research sites.

Furthermore, our research only examined methods that partition precipitation phase using surface meteorological quantities such as T_a, RH, and P_s. Atmospheric and climate models can also be used for hydroclimatic simulations either through direct coupling in earth systems models or as forcing data for land surface models. Many such models employ microphysics schemes to assign and track precipitation phase from the formation of a hydrometeor, through various atmospheric layers, to the land surface. For example, the Weather Research and Forecasting (WRF) model (Skamarock et al., 2005) has been used to simulate snow cover accumulation and ablation over large study domains in the western United States when coupled to a land surface model (Ikeda et al., 2010; Musselman et al., 2017a; Rasmussen et al., 2011). WRF has also been used to model the elevation of the rain-snow transition line in order to evaluate which basin areas are receiving solid or liquid precipitation during storm events (Minder et al., 2011). In addition, work from the 5th phase of Coupled Model

Keith Jennings 7/3/2019 10:59 AM
Formatted: Subscript

Keith Jennings 7/3/2019 11:01 AM
Formatted: Subscript

Unknown
Field Code Changed

Unknown
Field Code Changed

Keith Jennings 5/24/2019 11:38 AM
Deleted: A2

Keith Jennings 5/24/2019 2:00 PM
Deleted: Therefore

Keith Jennings 5/24/2019 2:01 PM
Deleted: ,

Keith Jennings 5/24/2019 2:03 PM
Deleted: o

Keith Jennings 5/24/2019 2:02 PM
Deleted: snow cover evolution to precipitation phase method at certain sites and points in time

Keith Jennings 7/15/2019 12:39 PM
Deleted: and

Keith Jennings 7/15/2019 12:40 PM
Formatted: Subscript

Unknown
Field Code Changed

Intercomparison Project (CMIP5) has shown that climate models produce different snowfall fractions due to variations in both climate and precipitation phase method (Krasting et al., 2013). In CMIP5, some models utilize microphysics schemes, while others assign precipitation phase at the land surface using methods similar to the ones presented in this work. Therefore, understanding and quantifying the sensitivity of model output due to precipitation phase method selection is important for both hydrologic and climate modeling studies.

5.3 Physical mechanisms controlling sensitivity to phase method

The warm maritime sites HJA and SSC expressed the largest peak SWE ranges from precipitation phase method selection (Fig. 7). These ranges were typically larger than 200 mm and sometimes exceeded 400 mm with relative differences usually greater than 50%, indicating large uncertainty in snowpack water storage. Additionally, peak SWE date ranges typically exceeded 2 weeks at these stations, meaning the timing of snowmelt onset was also affected by precipitation phase method. 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. For the former, both HJA and SSC had high proportions of precipitation falling between 0°C and 4°C (Fig. 11), which led to wide ranges in annual snowfall fraction (Table 4). The methods producing lower annual snowfall fractions (e.g., T_{a0} and T_{ar0}) generally corresponded to reduced snow cover duration simply because there was less frozen mass to melt. In other words, the energy required to melt the entire snowpack was reduced relative to the methods producing higher snowfall fractions, and the snowpack could be melted over a shorter time period.

Compounding the response of the warm maritime sites is the fact that snow and rain have different fates when they enter a snowpack with resultant effects on the snowpack energy budget. Snowfall can increase snowpack cold content (Jennings et al., 2018a), refresh surface albedo (Clow et al., 2016; Molotch et al., 2004; Molotch and Bales, 2006; Painter et al., 2012; United States Army Corps of Engineers, 1956), and provide dry pore space that must reach field capacity with liquid water before runoff can begin (Bengtsson, 1982; Seligman et al., 2014). Rainfall, conversely, can advect heat to the snowpack (Marks et al., 1998), infiltrate and run off (Harr, 1981, 1986), or be refrozen in the snowpack if there is cold content to be satisfied. In this context, the precipitation phase methods that produced more rainfall likely affected snow cover evolution not just through reduced frozen mass but also through changes to the snowpack energy budget. Further observational and modeling research is warranted to evaluate how rain versus snow affects snowpack energetics.

5.4 Why precipitation phase matters to climate warming simulations

The shift from snow to rain in cold and temperate regions across the globe is expected to continue with further warming. Future air temperature increases will likely produce reduced snowfall fractions (Klos et al., 2014; Lute et al., 2015; Safeeq et al., 2015), lower peak SWE values (Adam et al., 2009), earlier snowmelt onset (Stewart et al., 2004), slower snowmelt rates (Musselman et al., 2017a), and changes to the intensity and location of rain-on-snow events (Musselman et al., 2018). These warming-driven changes will impact both water resources availability (Barnett et al., 2008) and land

Keith Jennings 7/3/2019 11:03 AM
Deleted: 5

Keith Jennings 7/3/2019 11:03 AM
Deleted: 7

Keith Jennings 7/3/2019 11:04 AM
Deleted: (Fig. 6)

Keith Jennings 7/3/2019 11:05 AM
Deleted: was

Unknown
Field Code Changed

Unknown
Field Code Changed

Unknown
Field Code Changed

surface albedo (Déry and Brown, 2007). Most "at risk" to reductions in snowfall fraction and snow accumulation are areas with winter T_a near 0°C (Nolin and Daly, 2006). Concerningly, our work shows it is precisely these areas that have the greatest modeled snow cover accumulation and melt sensitivity to precipitation phase method selection. Compounding the problem is the fact that all precipitation phase methods exhibit downgraded performance relative to observations between 0°C and 4°C (Ding et al., 2014; Jennings et al., 2018b).

Harpold et al. (2017c) showed that future changes to snowfall fraction are moderated or exacerbated by the choice of a precipitation phase method, depending on the area's relative humidity. However, how this uncertainty affects the conclusion of climate change predictions is typically not discussed. In the context of the work presented herein, there should be a focus applied to areas where the baseline variability in peak SWE, snowmelt onset, and snow cover duration due to precipitation phase method approaches or exceeds the simulated change in the associated snowpack properties with warming. In warm maritime climates, research has shown peak SWE may decrease by upwards of several hundred millimeters as warming continues (e.g., Cooper et al., 2016; Leung et al., 2004; Minder, 2010; Musselman et al., 2017b), which is near the range of peak SWE sensitivity values reported in this work. Precipitation phase method selection is also likely to impact simulations of future warm snow droughts where anomalously warm winters are associated with low peak SWE (Harpold et al., 2017b). In addition, snow cover duration variability due to precipitation phase method selection in earth systems models may affect simulations of the snow-albedo feedback, which is the amplification of surface warming due to reduced snow cover (Hall, 2004; Hall and Qu, 2006). As climate warming shifts new areas towards the winter and spring average T_a values (0°C–4°C) that lead to the greatest uncertainty in rain-snow partitioning, our research suggests that uncertainty in future hydroclimatic states will be exacerbated by precipitation phase method selection.

6 Conclusion

In this work we simulated seasonal snow cover evolution using the SNOWPACK model forced with different permutations of 5 precipitation phase methods at 11 study stations spanning a climatic gradient from warm maritime to cold continental. We found the choice of a precipitation phase method affected model performance and introduced significant variability into simulated snow accumulation and melt. Overall, the binary logistic regression models produced the lowest mean biases, while high and low air temperature thresholds tended to overpredict and underpredict snow accumulation, respectively. Warm maritime sites were the most sensitive to method selection with relative differences in annual snowfall fraction near and above 100% and ranges in peak SWE typically greater than 200 mm, exceeding 400 mm in certain years. At these sites the different methods produced ranges in snowmelt timing and snow cover duration that were generally longer than 2 and 3 weeks, respectively. Conversely, the YOS-DAN and NWT-SDL stations exhibited the lowest sensitivity to precipitation phase method selection with relative differences in annual snowfall fraction between 11.6% and 13.4%. Peak SWE ranges were typically less than 30 mm for these two stations, while average snowmelt onset date ranges were only 0.8 d and 2.5 d at YOS-DAN and NWT-SDL, respectively. In contrast to the marked differences in peak SWE, melt onset, and

Keith Jennings 7/2/2019 2:51 PM
Deleted: 12

snow cover duration between the warm and cold stations, ranges in snowmelt rate exhibited little relationship to seasonal climate. Additionally, we found deviating by ±1°C from an optimized T_a rain-snow threshold had relatively little effect on simulated snow cover evolution at NWT and YOS compared to the larger sensitivity at HJA, SSC, and JD.

The spatially variable sensitivity of snow cover evolution was primarily a result of climatic differences between the stations. Increased Dec–May T_a and PPT were associated with greater peak SWE ranges across the different precipitation phase methods. This meant the maritime sites HJA and SSC, with significant winter and spring PPT, were most affected by precipitation phase method selection. Overall, we found stations with a high proportion of Dec–May PPT falling at T_a between 0°C and 4°C to be more sensitive than those with less PPT in that T_a range. This is troublesome considering climate warming is expected to push new areas in the seasonal snow zone towards winter T_a near 0°C and above. It is therefore critical that future work examine the relationship between the effect of warming on snow cover evolution and the model variability that results from precipitation phase partitioning uncertainty, particularly in areas undergoing a snow-to-rain transition.

Code and data availability

Forcing and validation data can be accessed at the following sites (as of 2019-02-11):

- HJ Andrews LTER: http://dx.doi.org/10.6073/pasta/c96875918bb9c86d330a457bf4295cd9 and http://andlter.forestry.oregonstate.edu/data/ (for sub-daily data)
- Southern Sierra CZO: https://eng.ucmerced.edu/snsjho/files/MHWG/Field/Southern_Sierra_CZO_KREW (Hunsaker et al., 2012)
- Johnston Draw (Reynolds Creek CZO): https://doi.org/10.15482/USDA.ADC/1402076 (Godsey et al., 2018)
- Yosemite Dana Meadows: http://hdl.handle.net/1773/35957 (Lundquist et al., 2016)
- Niwot Ridge LTER: https://doi.org/10.6073/pasta/1538ccf520d89c7a11c2c489d973b232 (Jennings et al., 2018a) and https://doi.org/10.6073/pasta/f62b0a3741737c871958cf7e63c089e0 (Williams, 2016)

SNOWPACK version 3.4.5 was used for this research. Model source code can be accessed at https://models.slf.ch/. For the code used to automatically run the model with the different precipitation phase methods and analyze the output data, please contact the corresponding author. The color palettes used in this manuscript's figures can be accessed online from their respective authors (https://github.com/karthik/wesanderson and http://doi.org/10.5281/zenodo.1243862).

Author contributions

KSJ and NPM designed the study. KSJ performed the analyses and wrote the manuscript. NPM provided feedback and edited the manuscript.

Keith Jennings 7/3/2019 11:16 AM
Formatted: Subscript

Keith Jennings 7/16/2019 8:03 AM
Deleted: Data

Unknown
Field Code Changed

Keith Jennings 7/1/2019 1:53 PM
Formatted: Normal, No bullets or numbering

Keith Jennings 5/22/2019 3:41 PM
Deleted: Appendix A: Model validation ... [11]

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

KSJ was supported by a NASA Earth and Space Science Fellowship (16-EARTH16F-378). The LTER and CZO networks are funded by the United States National Science Foundation. We offer our thanks to Don Henshaw (HJ Andrews LTER) and Mohammed Safeeq (Southern Sierra CZO) for their assistance with the respective datasets. We are also grateful for the other dataset providers for making their data freely and easily accessible. Additionally, we thank Drs. Bettina Schaefli and Jono Conway along with the editor Dr. Markus Hrachowitz for their reviews and feedback that helped improve this manuscript.

References

Adam, J. C., Hamlet, A. F. and Lettenmaier, D. P.: Implications of global climate change for snowmelt hydrology in the twenty-first century, Hydrol. Process., 23(7), 962–972, 2009.

Alduchov, O. A. and Eskridge, R. E.: Improved Magnus form approximation of saturation vapor pressure, J. Appl. Meteorol., 35(4), 601–609, 1996.

Anderson, E. A.: Development and testing of snow pack energy balance equations, Water Resour. Res., 4(1), 19–37, 1968.

Angström, A. K.: A study of the radiation of the atmosphere: based upon observations of the nocturnal radiation during expeditions to Algeria and to California, Smithsonian Institution., 1915.

Auer Jr, A. H.: The rain versus snow threshold temperatures, Weatherwise, 27(2), 67–67, 1974.

Avanzi, F., De Michele, C. and Ghezzi, A.: Liquid-solid partitioning of precipitation along an altitude gradient and its statistical properties: An Italian case study, Am. J. Clim. Change, 2014 [online] Available from: http://file.scirp.org/Html/7-2360069_43990.htm (Accessed 24 August 2016), 2014.

Barnett, T. P., Adam, J. C. and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438(7066), 303–309, 2005.

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., Bala, G., Wood, A. W., Nozawa, T., Mirin, A. A. and others: Human-induced changes in the hydrology of the western United States, science, 319(5866), 1080–1083, 2008.

Keith Jennings 7/3/2019 11:10 AM

Formatted: Bibliography, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Keith Jennings 5/21/2019 3:51 PM

Deleted: Adam, J. C., Hamlet, A. F. and Lettenmaier, D. P.: Implications of global climate change for snowmelt hydrology in the twenty-first century, Hydrol. Process., 23(7), 962–972, 2 ... [12]

Bartelt, P. and Lehning, M.: A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model, Cold Reg. Sci. Technol., 35(3), 123–145, 2002.

Bavay, M. and Egger, T.: MeteoIO 2.4.2: a preprocessing library for meteorological data, Geosci. Model Dev., 7(6), 3135–3151, doi:10.5194/gmd-7-3135-2014, 2014.

Bengtsson, L.: The importance of refreezing on the diurnal snowmelt cycle with application to a northern Swedish catchment, Hydrol. Res., 13(1), 1–12, 1982.

Bintanja, R. and Andry, O.: Towards a rain-dominated Arctic, Nat. Clim. Change, 7(4), 263–267, 2017.

Caine, N.: Streamflow patterns in the alpine environment of North Boulder Creek, Colorado Front Range, Z. Geomorphol. Suppl., 27–42, 1996.

Cherkauer, K. A., Bowling, L. C. and Lettenmaier, D. P.: Variable infiltration capacity cold land process model updates, Glob. Planet. Change, 38(1–2), 151–159, doi:10.1016/S0921-8181(03)00025-0, 2003.

Clow, D. W., Williams, M. W. and Schuster, P. F.: Increasing aeolian dust deposition to snowpacks in the Rocky Mountains inferred from snowpack, wet deposition, and aerosol chemistry, Atmos. Environ., 146, 183–194, doi:10.1016/j.atmosenv.2016.06.076, 2016.

Cooper, M. G., Nolin, A. W. and Safeeq, M.: Testing the recent snow drought as an analog for climate warming sensitivity of Cascades snowpacks, Environ. Res. Lett., 11(8), 084009, doi:10.1088/1748-9326/11/8/084009, 2016.

Crawford, T. M. and Duchon, C. E.: An improved parameterization for estimating effective atmospheric emissivity for use in calculating daytime downwelling longwave radiation, J. Appl. Meteorol., 38(4), 474–480, 1999.

Dai, A.: Temperature and pressure dependence of the rain-snow phase transition over land and ocean, Geophys. Res. Lett., 35(12) [online] Available from: http://onlinelibrary.wiley.com/doi/10.1029/2008GL033295/full (Accessed 12 August 2016), 2008.

Deardoff, J. W.: Efficient prediction of ground surface temperature and moisture with inclusion of a layer of vegetation., J Gerophys Res, 38, 659–661, 1978.

Déry, S. J. and Brown, R. D.: Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback, Geophys. Res. Lett., 34(22), L22504, doi:10.1029/2007GL031474, 2007.

Dickerson-Lange, S. E., Gersonde, R. F., Hubbart, J. A., Link, T. E., Nolin, A. W., Perry, G. H., Roth, T. R., Wayand, N. E. and Lundquist, J. D.: Snow disappearance timing is dominated by forest effects on snow accumulation in warm winter climates of the Pacific Northwest, United States, Hydrol. Process., 31(10), 1846–1862, doi:10.1002/hyp.11144, 2017.

Dilley, A. C. and O'Brien, D. M.: Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water, Q. J. R. Meteorol. Soc., 124(549), 1391–1401, 1998.

Ding, B., Yang, K., Qin, J., Wang, L., Chen, Y. and He, X.: The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization, J. Hydrol., 513, 154–163, 2014.

Erickson, T. A., Williams, M. W. and Winstral, A.: Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States, Water Resour. Res., 41(4), 2005.

Essery, R., Morin, S., Lejeune, Y. and B Ménard, C.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Resour., 55, 131–148, doi:10.1016/j.advwatres.2012.07.013, 2013.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.-J., Essery, R., Fernandez, A. and others: Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project), Ann. Glaciol., 38(1), 150–158, 2004.

Fassnacht, S. R. and Soulis, E. D.: Implications during transitional periods of improvements to the snow processes in the land surface scheme-hydrological model WATCLASS, Atmosphere-Ocean, 40(4), 389–403, 2002.

Fassnacht, S. R., Venable, N. B. H., Khishigbayar, J. and Cherry, M. L.: The probability of precipitation as snow derived from daily air temperature for high elevation areas of Colorado, United States, IAHS-AISH Publ., 65–70, 2013.

Feiccabrino, J., Graff, W., Lundberg, A., Sandström, N. and Gustafsson, D.: Meteorological Knowledge Useful for the Improvement of Snow Rain Separation in Surface Based Models, Hydrology, 2(4), 266–288, doi:10.3390/hydrology2040266, 2015.

Flerchinger, G. N., Xaio, W., Marks, D., Sauer, T. J. and Yu, Q.: Comparison of algorithms for incoming atmospheric long-wave radiation, Water Resour. Res., 45(3), doi:10.1029/2008WR007394, 2009.

Froidurot, S., Zin, I., Hingray, B. and Gautheron, A.: Sensitivity of Precipitation Phase over the Swiss Alps to Different Meteorological Variables, J. Hydrometeorol., 15(2), 685–696, doi:10.1175/JHM-D-13-073.1, 2014.

Gjertsen, U. and Ødegaard, V.: The water phase of precipitation—a comparison between observed, estimated and predicted values, Atmospheric Res., 77(1), 218–231, doi:10.1016/j.atmosres.2004.10.030, 2005.

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 stream flow 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(3), 1207–1216, doi:https://doi.org/10.5194/essd-10-1207-2018, 2018.

Greenland, D.: The climate of Niwot Ridge, front range, Colorado, USA, Arct. Alp. Res., 380–391, 1989.

Hall, A.: The Role of Surface Albedo Feedback in Climate, J. Clim., 17(7), 1550–1568, doi:10.1175/1520-0442(2004)017<1550:TROSAF>2.0.CO;2, 2004.

Hall, A. and Qu, X.: Using the current seasonal cycle to constrain snow albedo feedback in future climate change, Geophys. Res. Lett., 33(3), L03502, doi:10.1029/2005GL025127, 2006.

Harder, P. and Pomeroy, J.: Estimating precipitation phase using a psychrometric energy balance method, Hydrol. Process., 27(13), 1901–1914, doi:10.1002/hyp.9799, 2013.

Harder, P. and Pomeroy, J. W.: Hydrological model uncertainty due to precipitation-phase partitioning methods, Hydrol. Process., 28(14), 4311–4327, 2014.

Harpold, A. A., Dettinger, M. and Rajagopal, S.: Defining snow drought and why it matters, EOS-Earth Space Sci. News, 98, 2017a.

Harpold, A. A., Kaplan, M., Klos, P. Z., Link, T., McNamara, J. P., Rajagopal, S., Schumer, R. and Steele, C. M.: Rain or snow: hydrologic processes, observations, prediction, and research needs, Hydrol Earth Syst Sci, 21, 1–22, 2017b.

Harpold, A. A., Crews, J. B., Rajagopal, S., Winchell, T. and Schumer, R.: Relative Humidity Has Uneven Effects on Shifts From Snow to Rain Over the Western U.S., Geophys. Res. Lett., 44(19), 2017GL075046, doi:10.1002/2017GL075046, 2017c.

Harr, R. D.: Some characteristics and consequences of snowmelt during rainfall in western Oregon, J. Hydrol., 53(3–4), 277–304, 1981.

Harr, R. D.: Effects of clearcutting on rain-on-snow runoff in western Oregon: A new look at old studies, Water Resour. Res., 22(7), 1095–1100, 1986.

Hunsaker, C. T., Whitaker, T. W. and Bales, R. C.: Snowmelt runoff and water yield along elevation and temperature gradients in California's southern Sierra Nevada, JAWRA J. Am. Water Resour. Assoc., 48(4), 667–678, 2012.

Ikeda, K., Rasmussen, R., Liu, C., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J. and Miller, K.: Simulation of seasonal snowfall over Colorado, Atmospheric Res., 97(4), 462–477, 2010.

Jennings, K. S. and Jones, J. A.: Precipitation-snowmelt timing and snowmelt augmentation of large peak flow events, western Cascades, Oregon, Water Resour. Res., 51(9), 7649–7661, doi:10.1002/2014WR016877, 2015.

Jennings, K. S., Kittel, T. G. F. and Molotch, N. P.: Observations and simulations of the seasonal evolution of snowpack cold content and its relation to snowmelt and the snowpack energy budget, The Cryosphere, 12(5), 1595–1614, doi:10.5194/tc-12-1595-2018, 2018a.

Jennings, K. S., Winchell, T. S., Livneh, B. and Molotch, N. P.: Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere, Nat. Commun., 9, doi:10.1038/s41467-018-03629-7, 2018b.

Kienzle, S. W.: A new temperature based method to separate rain and snow, Hydrol. Process., 22(26), 5067–5085, doi:10.1002/hyp.7131, 2008.

Klos, P. Z., Link, T. E. and Abatzoglou, J. T.: Extent of the rain-snow transition zone in the western US under historic and projected climate, Geophys. Res. Lett., 41(13), 4560–4568, 2014.

Knowles, J. F., Harpold, A. A., Cowie, R., Zeliff, M., Barnard, H. R., Burns, S. P., Blanken, P. D., Morse, J. F. and Williams, M. W.: The relative contributions of alpine and subalpine ecosystems to the water balance of a mountainous, headwater catchment, Hydrol. Process., 29(22), 4794–4808, doi:10.1002/hyp.10526, 2015.

Knowles, N., Dettinger, M. D. and Cayan, D. R.: Trends in snowfall versus rainfall in the western United States, J. Clim., 19(18), 4545–4559, 2006.

Krasting, J. P., Broccoli, A. J., Dixon, K. W. and Lanzante, J. R.: Future Changes in Northern Hemisphere Snowfall, J. Clim., 26(20), 7813–7828, doi:10.1175/JCLI-D-12-00832.1, 2013.

Lapo, K. E., Hinkelman, L. M., Raleigh, M. S. and Lundquist, J. D.: Impact of errors in the downwelling irradiances on simulations of snow water equivalent, snow surface temperature, and the snow energy balance, Water Resour. Res., 51(3), 1649–1670, 2015.

Leavesley, G. H., Stannard, L. G., Singh, V. P. and others: The precipitation-runoff modeling system-PRMS., Comput. Models Watershed Hydrol., 281–310, 1995.

Lehning, M., Fierz, C. and Lundy, C.: An objective snow profile comparison method and its application to SNOWPACK, Cold Reg. Sci. Technol., 33(2–3), 253–261, doi:10.1016/S0165-232X(01)00044-1, 2001.

Lehning, M., Bartelt, P., Brown, B., Fierz, C. and Satyawali, P.: A physical SNOWPACK model for the Swiss avalanche warning: Part II. Snow microstructure, Cold Reg. Sci. Technol., 35(3), 147–167, 2002a.

Lehning, M., Bartelt, P., Brown, B. and Fierz, C.: A physical SNOWPACK model for the Swiss avalanche warning: Part III: Meteorological forcing, thin layer formation and evaluation, Cold Reg. Sci. Technol., 35(3), 169–184, 2002b.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M. and Zappa, M.: ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology, Hydrol. Process., 20(10), 2111–2128, 2006.

Leung, L. R., Qian, Y., Bian, X., Washington, W. M., Han, J. and Roads, J. O.: Mid-century ensemble regional climate change scenarios for the western United States, Clim. Change, 62(1–3), 75–113, 2004.

Litaor, M. I., Williams, M. and Seastedt, T. R.: Topographic controls on snow distribution, soil moisture, and species diversity of herbaceous alpine vegetation, Niwot Ridge, Colorado, J. Geophys. Res. Biogeosciences, 113(G2), 2008.

Lundquist, J. D., Dickerson-Lange, S. E., Lutz, J. A. and Cristea, N. C.: Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling: Forests and Snow Retention, Water Resour. Res., 49(10), 6356–6370, doi:10.1002/wrcr.20504, 2013.

Lundquist, J. D., Roche, J. W., Forrester, H., Moore, C., Keenan, E., Perry, G., Cristea, N., Henn, B., Lapo, K., McGurk, B., Cayan, D. R. and Dettinger, M. D.: Yosemite Hydroclimate Network: Distributed stream and atmospheric data for the Tuolumne River watershed and surroundings, Water Resour. Res., 52(9), 7478–7489, doi:10.1002/2016WR019261, 2016.

Lundy, C. C., Brown, R. L., Adams, E. E., Birkeland, K. W. and Lehning, M.: A statistical validation of the SNOWPACK model in a Montana climate, Cold Reg. Sci. Technol., 33(2), 237–246, 2001.

Lute, A. C., Abatzoglou, J. T. and Hegewisch, K. C.: Projected changes in snowfall extremes and interannual variability of snowfall in the western United States, Water Resour. Res., 51(2), 960–972, doi:10.1002/2014WR016267, 2015.

Lynch-Stieglitz, M.: The development and validation of a simple snow model for the GISS GCM, J. Clim., 7(12), 1842–1855, 1994.

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y. and Diffenbaugh, N. S.: The potential for snow to supply human water demand in the present and future, Environ. Res. Lett., 10(11), 114016, doi:10.1088/1748-9326/10/11/114016, 2015.

Marks, D. and Winstral, A.: Comparison of snow deposition, the snow cover energy balance, and snowmelt at two sites in a semiarid mountain basin, J. Hydrometeorol., 2(3), 213–227, 2001.

Marks, D., Kimball, J., Tingey, D. and Link, T.: The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: a case study of the 1996 Pacific Northwest flood, Hydrol. Process., 12(10), 1569–1587, 1998.

Marks, D., Link, T., Winstral, A. and Garen, D.: Simulating snowmelt processes during rain-on-snow over a semi-arid mountain basin, Ann. Glaciol., 32, 195–202, 2001.

Marks, D., Winstral, A., Reba, M., Pomeroy, J. and Kumar, M.: An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin, Adv. Water Resour., 55, 98–110, doi:10.1016/j.advwatres.2012.11.012, 2013.

Mazurkiewicz, A. B., Callery, D. G. and McDonnell, J. J.: Assessing the controls of the snow energy balance and water available for runoff in a rain-on-snow environment, J. Hydrol., 354(1–4), 1–14, 2008.

Meek, D. W. and Hatfield, J. L.: Data quality checking for single station meteorological databases, Agric. For. Meteorol., 69(1), 85–109, 1994.

Meromy, L., Molotch, N. P., Williams, M. W., Musselman, K. N. and Kueppers, L. M.: Snowpack-climate manipulation using infrared heaters in subalpine forests of the Southern Rocky Mountains, USA, Agric. For. Meteorol., 203, 142–157, doi:10.1016/j.agrformet.2014.12.015, 2015.

Minder, J. R.: The Sensitivity of Mountain Snowpack Accumulation to Climate Warming, J. Clim., 23(10), 2634–2650, doi:10.1175/2009JCLI3263.1, 2010.

Minder, J. R., Durran, D. R. and Roe, G. H.: Mesoscale Controls on the Mountainside Snow Line, J. Atmospheric Sci., 68(9), 2107–2127, doi:10.1175/JAS-D-10-05006.1, 2011.

Mizukami, N., Koren, V., Smith, M., Kingsmill, D., Zhang, Z., Cosgrove, B. and Cui, Z.: The impact of precipitation type discrimination on hydrologic simulation: Rain–snow partitioning derived from HMT-West radar-detected brightband height versus surface temperature data, J. Hydrometeorol., 14(4), 1139–1158, 2013.

Molotch, N. P. and Bales, R. C.: Comparison of ground-based and airborne snow surface albedo parameterizations in an alpine watershed: Impact on snowpack mass balance, Water Resour. Res., 42(5), doi:10.1029/2005WR004522, 2006.

Molotch, N. P., Painter, T. H., Bales, R. C. and Dozier, J.: Incorporating remotely-sensed snow albedo into a spatially-distributed snowmelt model, Geophys. Res. Lett., 31(3), doi:10.1029/2003GL019063, 2004.

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K. and Rasmussen, R.: Slower snowmelt in a warmer world, Nat. Clim. Change, 7(3), 214–219, doi:10.1038/nclimate3225, 2017a.

Musselman, K. N., Molotch, N. P. and Margulis, S. A.: Snowmelt response to simulated warming across a large elevation gradient, southern Sierra Nevada, California, The Cryosphere, 11(6), 2847–2866, doi:10.5194/tc-11-2847-2017, 2017b.

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., Barlage, M. and Rasmussen, R.: Projected increases and shifts in rain-on-snow flood risk over western North America, Nat. Clim. Change, 1, doi:10.1038/s41558-018-0236-4, 2018.

Nayak, A., Marks, D., Chandler, D. G. and Seyfried, M.: Long-term snow, climate, and streamflow trends at the Reynolds Creek Experimental Watershed, Owyhee Mountains, Idaho, United States: CLIMATE TRENDS AT RCEW, Water Resour. Res., 46(6), n/a-n/a, doi:10.1029/2008WR007525, 2010.

Nolin, A. W. and Daly, C.: Mapping “at risk” snow in the Pacific Northwest, J. Hydrometeorol., 7(5), 1164–1171, 2006.

O’Gorman, P. A.: Contrasting responses of mean and extreme snowfall to climate change, Nature, 512(7515), 416–418, doi:10.1038/nature13625, 2014.

Painter, T. H., Skiles, S. M., Deems, J. S., Bryant, A. C. and Landry, C. C.: Dust radiative forcing in snow of the Upper Colorado River Basin: 1. A 6 year record of energy balance, radiation, and dust concentrations, Water Resour. Res., 48(7), W07521, doi:10.1029/2012WR011985, 2012.

Perkins, R. M. and Jones, J. A.: Climate variability, snow, and physiographic controls on storm hydrographs in small forested basins, western Cascades, Oregon, Hydrol. Process., 22(25), 4949–4964, 2008.

Prata, A. J.: A new long-wave formula for estimating downward clear-sky radiation at the surface, Q. J. R. Meteorol. Soc., 122(533), 1127–1151, 1996.

Rajagopal, S. and Harpold, A. A.: Testing and Improving Temperature Thresholds for Snow and Rain Prediction in the Western United States, JAWRA J. Am. Water Resour. Assoc. [online] Available from: http://onlinelibrary.wiley.com/doi/10.1111/1752-1688.12443/full (Accessed 23 August 2016), 2016.

Raleigh, M. S. and Lundquist, J. D.: Comparing and combining SWE estimates from the SNOW-17 model using PRISM and SWE reconstruction, Water Resour. Res., 48(1), W01506, doi:10.1029/2011WR010542, 2012.

Raleigh, M. S., Lundquist, J. D. and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrol. Earth Syst. Sci., 19(7), 3153–3179, doi:10.5194/hess-19-3153-2015, 2015.

Raleigh, M. S., Livneh, B., Lapo, K. and Lundquist, J. D.: How Does Availability of Meteorological Forcing Data Impact Physically Based Snowpack Simulations?, J. Hydrometeorol., 17(1), 99–120, doi:10.1175/JHM-D-14-0235.1, 2016.

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J. and Yu, W.: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: a process study of current and warmer climate, J. Clim., 24(12), 3015–3048, 2011.

Rice, R., Bales, R. C., Painter, T. H. and Dozier, J.: Snow water equivalent along elevation gradients in the Merced and Tuolumne River basins of the Sierra Nevada, Water Resour. Res., 47(8), W08515, doi:10.1029/2010WR009278, 2011.

Roth, T. R. and Nolin, A. W.: Forest impacts on snow accumulation and ablation across an elevation gradient in a temperate montane environment, Hydrol Earth Syst Sci, 21(11), 5427–5442, doi:10.5194/hess-21-5427-2017, 2017.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H. and others: Evaluation of forest snow processes models (SnowMIP2), J. Geophys. Res. Atmospheres, 114(D6), doi:doi/10.1029/2008JD011063/, 2009.

Safeeq, M., Shukla, S., Arismendi, I., Grant, G. E., Lewis, S. L. and Nolin, A.: Influence of winter season climate variability on snow–precipitation ratio in the western United States, Int. J. Climatol. [online] Available from: http://onlinelibrary.wiley.com/doi/10.1002/joc.4545/full (Accessed 7 June 2016), 2015.

Seligman, Z. M., Harper, J. T. and Maneta, M. P.: Changes to Snowpack Energy State from Spring Storm Events, Columbia River Headwaters, Montana, J. Hydrometeorol., 15(1), 159–170, doi:10.1175/JHM-D-12-078.1, 2014.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W. and Powers, J. G.: A description of the advanced research WRF version 2, National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div., 2005.

Slater, A. G., Schlosser, C. A., Desborough, C. E., Pitman, A. J., Henderson-Sellers, A., Robock, A., Vinnikov, K. Y., Entin, J., Mitchell, K., Chen, F. and others: The representation of snow in land surface schemes: Results from PILPS 2 (d), J. Hydrometeorol., 2(1), 7–25, 2001.

Stewart, I. T., Cayan, D. R. and Dettinger, M. D.: Changes in snowmelt runoff timing in western North America under a "business as usual" climate change scenario, Clim. Change, 62(1–3), 217–232, 2004.

Stull, R.: Wet-bulb temperature from relative humidity and air temperature, J. Appl. Meteorol. Climatol., 50(11), 2267–2269, 2011.

Tarboton, D. G. and Luce, C. H.: Utah energy balance snow accumulation and melt model (UEB), Citeseer. [online] Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.6983&rep=rep1&type=pdf (Accessed 19 August 2016), 1996.

Trenberth, K. E.: Changes in precipitation with climate change, Clim. Res., 47(1), 123, 2011.

Trujillo, E. and Molotch, N. P.: Snowpack regimes of the Western United States, Water Resour. Res., 50(7), 5611–5623, doi:10.1002/2013WR014753, 2014.

United States Army Corps of Engineers: Snow hydrology, US Army North Pac. Div. Portland Or., 1956.

Unsworth, M. H. and Monteith, J. L.: Long-wave radiation at the ground I. Angular distribution of incoming radiation, Q. J. R. Meteorol. Soc., 101(427), 13–24, 1975.

Wayand, N. E., Stimberis, J., Zagrodnik, J. P., Mass, C. F. and Lundquist, J. D.: Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA, J. Geophys. Res. Atmospheres, 121(17), 9929–9942, 2016.

Wayand, N. E., Clark, M. P. and Lundquist, J. D.: Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations, Hydrol. Process., 31(2), 349–363, doi:10.1002/hyp.11002, 2017.

Wen, L., Nagabhatla, N., Lü, S. and Wang, S.-Y.: Impact of rain snow threshold temperature on snow depth simulation in land surface and regional atmospheric models, Adv. Atmospheric Sci., 30(5), 1449–1460, doi:10.1007/s00376-012-2192-7, 2013.

Wigmosta, M. S., Vail, L. W. and Lettenmaier, D. P.: A distributed hydrology-vegetation model for complex terrain, Water Resour. Res., 30(6), 1665–1679, 1994.

Williams, M.: Snow cover profile data for Niwot Ridge, Green Lakes Valley from 1993/2/26 - ongoing, weekly to biweekly, [online] Available from: http://niwot.colorado.edu/index.php/data/data/snow-cover-profile-data-for-niwot-ridge-and-green-lakes-valley-1993-ongoi, 2016.

Williams, M. W., Bardsley, T. and Rikkers, M.: Overestimation of snow depth and inorganic nitrogen wetfall using NADP data, Niwot Ridge, Colorado, Atmos. Environ., 32(22), 3827–3833, 1998.

Ye, H., Cohen, J. and Rawlins, M.: Discrimination of Solid from Liquid Precipitation over Northern Eurasia Using Surface Atmospheric Conditions*, J. Hydrometeorol., 14(4), 1345–1355, 2013.

Yuter, S. E., Kingsmill, D. E., Nance, L. B. and Löffler-Mang, M.: Observations of Precipitation Size and Fall Speed Characteristics within Coexisting Rain and Wet Snow, J. Appl. Meteorol. Climatol., 45(10), 1450–1464, doi:10.1175/JAM2406.1, 2006.

Zhang, Z., Glaser, S., Bales, R., Conklin, M., Rice, R. and Marks, D.: Insights into mountain precipitation and snowpack from a basin-scale wireless-sensor network, Water Resour. Res., 53(8), 6626–6641, doi:10.1002/2016WR018825, 2017.