Referee comments are in black, author responses are in green.

Lute et al. present simulations of snow over a large domain for reasonably long periods and at high spatial resolution. To achieve this, they have made some quite severe approximations to limit computational expense. Neglecting forest effects seems like a limitation for water resource applications in the western US. Simple canopy models, and snow models with a few layers to allow a more physical representation, need not be very expensive. Has code optimization to offset the cost of improved model physics been considered? MATLAB is convenient for parallel computation, but otherwise might not be the best language for this. Having said that, the authors have adopted a particular development strategy and reporting the results is worthwhile.

We appreciate the thoughtful and thorough comments of the referee.

The referee is correct that in the interest of spatial resolution, and temporal and spatial extent, some simplifications were made – which we tried to be transparent about. That said, the SnowClim model presents a much more physically based approach than the temperature-index modeling approaches that are often used when these levels of extent and resolution are required (e.g. SNOW-17, Anderson, 2006; Kraaijenbrink et al., 2021).

Substantial effort was put into code optimization. Model calibration and the model runs for the Western United States were done using parallel computation on high performance computers. A variable spatial resolution was adopted in order to enable the high spatial resolution and extent. As discussed in the manuscript, the barriers to inclusion of more comprehensive process representations include not only computational expense, but also appropriate data. In the case of forest effects, data to inform a vegetation module for the pre-industrial, recent historical, and future time periods is not available. In addition, there is significant uncertainty in future vegetation change stemming from disturbance regimes and species composition pathways which would add complexity and uncertainty to the model. That said, neglecting forest effects is certainly a limitation of the model and represents one of the top areas for future development that could be added to the modeling approach presented. As it is, the SnowClim data represents a first approximation of snow in vegetated landscapes with demonstrated accurate performance at open locations. These points have been incorporated in the revised manuscript.

I was able to access the code and run the MATLAB Online example easily (a quickstart guide might have saved me a couple of minutes).

I suggest some corrections and clarifications, identified by line number:

24

There are too many to list, but the abstract should give some idea of what kind of “snow metrics” are offered.

We appreciate this suggestion and have added some metrics to line 26.

52

Should note that this is earlier but slower snowmelt

Yes, good point. We have modified the sentence to this effect

106

Equation (1) is not the surface energy balance because G is not a surface flux. I would describe it as the energy balance of the snowpack.

Thank you for raising this point, we agree. We'll use 'snowcover energy balance' instead of 'surface energy balance'. We have modified the text to reflect this.

127

Note that separate visible and near-infrared radiation fluxes are not supplied as inputs; the albedo is simply averaged between these bands. The illumination angle dependence should only be applied to the direct-beam component of incoming radiation.

You are correct that separate radiation fluxes are not supplied to the albedo calculation, we have modified the text to reflect this. And yes, the albedo is the average of the near infrared and visible band albedos.

Regarding the illumination angle, the model does not differentiate the direct-beam and diffuse radiation. Instead, the inclination angle adjustment is applied to global radiation. While this is not technically correct, the parameterization in UEB (the source of this albedo function) is calibrated for global radiation. We also note that this simplification increases computational efficiency and recognizes that diffuse radiation provides a minor contribution to most shortwave radiation driven snowmelt.

133

Equation (3) is incorrect. If the emissivity is not 1, the upward longwave radiation includes a reflected fraction of the downward radiation (Kirchoff's law).

Thank you for catching this. We have corrected the equation, recalibrated the model, and recompleted the western US runs with the new equation and parameters.

179

Constant G is not a very common (or realistic) feature of modern models. Etchevers et al. (2002) included an experiment with fixed G as a sensitivity study, but did not report the results.

While constant G is not very common in modern physics based snow models, G is entirely neglected (or you might say subsumed) in empirical (e.g. temperature-index) snow models. We recognize the referee's point however that the manuscript text might be misleading in that it might suggest that the use of a constant value for G in physics based snow models is more common than it actually is. We have added a sentence to the manuscript in section 2.2.5 to clarify this point and more accurately reflect common practice.

Given unlimited computational capacity and time, incorporating a more thorough treatment of the ground heat flux would be ideal. However, this would require dynamic soil temperatures, which can vary significantly in complex terrain and require the addition of a soil temperature module. The ground heat flux also depends on soil properties including texture, water content, frozen/unfrozen status, as well as litter depth, characteristics, and water content. The benefits of modeling a dynamic ground heat flux likely do not justify the added complexity and uncertainty it would require since in most mountain areas the ground heat flux is small compared to other fluxes (Marks and Dozier, 1992; DeWalle and Rango, 2008).

Given that the goals of the SnowClim model were to incorporate the most important elements of physics based models and to include empirical components to enable greater computational efficiency, we feel that the use of a constant value for G is entirely appropriate. A brief statement to this effect is included in the revised manuscript.

189

Iterative solution of the surface energy balance to find snow surface temperature is not the only possibility. Best et al. (2011, https://gmd.copernicus.org/articles/4/677/2011/), for example, linearizes the surface energy balance equation to find a surface temperature solution without the expense of iteration. In either an iterative or linearised solution, the aim is to find a surface temperature that is consistent with the surface energy balance. I don't see how this can be achieved with the pragmatic but ad hoc approach in SnowClim.

We thank the reviewer for highlighting this additional option. We think the linearization approach could be worth considering in a future version of the SnowClim model, but would require thorough evaluation in the context of the existing model. Furthermore, the point referenced in the manuscript describes options for solving for snow surface temperature in single layer models, however it appears that the linearization approach in Best et al., (2011) is only applied to the multilayer model formulation. We note that the single layer UEB model uses a combination of linearization and iteration to solve for snow surface temperature. We have clarified the text to indicate that iterative and linearization approaches are often used, instead of just noting the iterative option.

281

How were the 210 m and 1050 m resolutions selected? How many points are there in this domain?

The spatial resolutions were selected based on several considerations. For ease of implementation, we chose spatial resolutions that were multiples of 30 m, the spatial resolution of the digital elevation model we used. Beyond that, we selected 1050 m based on the assumption that snow conditions could be relatively uniform across this extent given minimal contrast in elevation and solar radiation (our conditions for a location to be modeled at coarse resolution). At the higher end, we selected a spatial resolution that was not so high as to make our spatial and temporal extent goals unachievable.

We modeled 920,605 points at 1050 m resolution and 64,310,454 points at 210 m resolution. We've noted this in the revised manuscript.

330

SNOTEL sites may not be representative of larger areas. Could this calibration skew the model performance?
shttps://agupubs.onlinelibrary.wiley.com/doi/10.1029/2005WR004229

We agree that SNOTEL sites may not be representative of larger areas (just as any point on the landscape may not be representative of the larger area). SNOTEL sites tend to be located in forest clearings which can have distinctly different wind and radiational characteristics compared to closed canopy forests. If there was a consistent difference across sites between SNOTEL SWE and broader area SWE then calibrating at SNOTEL sites could skew the model performance across broader domains.

Essentially any existing observational network could be less than a perfect representation of the broader area and potentially skew the model results. The SNOTEL sites provide the best available calibration dataset in the study region due to length of records, relatively consistent observational equipment and methods across sites, and the breadth of geographic and climatic conditions that they cover. We also note in the manuscript that the SnowClim model can be calibrated to other observational datasets, however we calibrate to SNOTEL in this case for the reasons aforementioned. We've added some text in the calibration methods section (3.3.1) to acknowledge the concern about representativeness and indicate why we chose to use the SNOTEL network for calibration.

362

Any ideas why model performance would improve from 12 to 24 hour timesteps?

We suspect that the improvement in model performance between 12 and 24 hour timesteps for the time aggregation windows shown in Figure 4 is related to how well the aggregation windows capture the diurnal cycle of energy fluxes. For the 12 hour aggregation starting at 0 GMT (shown in Figure 4), solar radiation oscillates between 0 or a low value and a moderately high value, with nothing in between. In contrast, the 24 hour aggregation window starting at 0 GMT (shown in Figure 4) has moderate solar radiation values in every time step.

We found that if we started the aggregation windows a few hours later (e.g. at 3 GMT or later), then the model performance degraded continuously with coarsening temporal resolution. These aggregation windows allow a more even distribution of solar radiation values across time steps. A note regarding this has been added to the revised manuscript.

369

There is no evaluation of the large-scale snow simulations for the western US. MODIS snow cover extent products would have a convenient resolution for this.

We do not think a comparison with MODIS snow cover extent data would be very illuminating given that it is just a binary snow/no snow product. Recent studies (e.g., Garousi-Nejad and Tarboton 2022) have demonstrated several challenges in comparing modeled snow to the MODIS snow cover area product including factors such as forest cover that may yield incorrect estimates from MODIS. While a comparison to another large-scale dataset could be useful, we feel it is beyond the scope of this paper given the evaluations we have already provided.

Garousi-Nejad, I., & Tarboton, D. G. (2022). A comparison of National Water Model retrospective analysis snow outputs at snow telemetry sites across the Western United States. *Hydrological Processes*, 36(1), e14469. https://doi.org/10.1002/hyp.14469

Figure 2

There are a lot of sites to balance in the calibration, but errors up to +/- 50 days in duration and +/- 50% in maximum SWE even after calibration seem large for practical applications. Other datasets used to demonstrate poorer agreement in the discussion were not calibrated to SNOTEL sites.

We agree that the errors are large in a few cases, however, out of 170 sites only 9 sites had maximum SWE errors larger than +/- 50% and only 4 sites had snow duration errors larger than +/- 50 days. Our approach was to calibrate across all SNOTEL sites at once to arrive at the best parameter set across this broad range of snow climates which could then be used to model snow across the western US. If we had calibrated to each site individually we expect the errors would have been drastically smaller, at the expense of potentially having limited utility in modeling both across spatial domains and under different climates.

We also emphasize the errors in the SnowClim dataset relative to SNOTEL sites stem from a variety of factors including uncertainties and errors in data collection at SNOTEL sites (e.g. snow bridging on the snow pillow), model structure, parameterization, and forcing data. Errors in meteorological forcing, namely precipitation, may lead to biases or errors in the resultant modeled SWE that do not reflect deficiencies in the modeling framework. It is difficult to decompose these errors, but previous studies have shown that forcing data can be one of the primary contributors to overall simulation error (Günther et al., 2019).

One of the datasets we compare to (Guan et al., 2013) was not evaluated relative to SNOTEL sites. They used a combination of modeling, remotely sensed, and ground based data to simulate SWE at 6 sites in the Sierra Nevada Mountains of California. One would think that with remotely sensed and ground observation information, their approach would yield much better results than a pure modeling approach that was calibrated across a much wider range of conditions, however it did not.

Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in snowpack simulations—Assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance. Water Resources Research, 55, 2779–2800. https://doi.org/10.1029/2018WR023403

Table 1

Given temperature, pressure and one of dewpoint temperature, relative humidity and specific humidity, the other two can be calculated – they are not all required forcing data.

Yes, good point. We have revised the table to list specific humidity, but not relative humidity or dewpoint temperature since equations for calculating the latter two are available in the SnowClim model.

Table 2

As it is abbreviated as Q, I guess that the WRF mixing ratio output is taken as the SnowClim specific humidity input (they are not the same thing, but will have nearly identical values).

Specific humidity was calculated from the water vapor mixing ratio (see lines 325-326 in the initial manuscript), they were not assumed to be equal.

Table 5

The monthly SWE and snow depth are minimum, mean and maximum.

We have clarified this in Table 5 to better reflect the available data.