https://egusphere.copernicus.org/preprints/egusphere-2025-1264#RC1

https://doi.org/10.5194/egusphere-2025-1264-RC1

We thank Referee #1 for their helpful comments. Our replies to their comments are shown in bold below.

Summary and overarching comments:

This study evaluates the impacts of an alternate SCF parameterization, that calculates SCF as a function of topographic complexity and SWE, in offline 1-degree CLASSIC model simulations. The alternate SCF parameterization tends to improve the accuracy of SCF simulated in CLASSIC during winter months in topographically complex regions. This study also explores the robustness of results to differing metrological forcing sources, which reveals the large impact of metrological forcing in snow simulation accuracy. This study provides a novel and important advancement for the CLASSIC modeling system that seems to allow land model simulations to better capture SCF, and in turn improve land-atmosphere interactions due to the snow-albedo feedback. Overall, the paper is well written and the study will warrant a publication after addressing the comments below.

Thank you for your overall positive review of our manuscript.

I have four overarching critiques for this analysis. (1) A key motivation for improving SCF in model simulations is to enhance simulated albedo. Although the study briefly covers the impacts of SCF on albedo accuracy using the AMBER score, it would be useful to go into more detail on the albedo analysis which is a critical component of this study. (2) MODIS SCF has questionable accuracy, particularly for representing ground SCF which the land model simulates. This point should be more directly addressed with the consideration of other data sources for SCF (e.g., STC-MODSCAG across the western US, see suggestion below). (3) Discrepancies in spatial resolution between reference data used to validate model simulations and the spatial resolution of the model simulations can largely impact results. Please see specific comment below addressing this point. (4) Figure quality should be improved throughout.

Thank you for your suggestions.

(1) A key motivation for improving SCF in model simulations is to enhance simulated albedo. Although the study briefly covers the impacts of SCF on albedo accuracy using the AMBER score, it would be useful to go into more detail on the albedo analysis which is a critical component of this study.

We agree that it would be helpful to provide more details on the impact of the SL12 parameterization on simulated surface albedo in CLASSIC, especially considering the large impact of snow cover on surface albedo. We will include a figure comparing surface albedo simulated by the model runs using the Control and SL12 parameterizations for various regions with observations when revising our manuscript (see figure below in our reply to your Specific recommendations).

(2) MODIS SCF has questionable accuracy, particularly for representing ground SCF which the land model simulates. This point should be more directly addressed with the consideration of other data sources for SCF (e.g., STC-MODSCAG across the western US, see suggestion below).

We agree that SCF derived from satellite optical sensors like MODIS is viewable snow cover from space during cloud-free overpasses (i.e. from above the canopy), while SCF from CLASSIC represents ground-level SCF (including snow cover beneath the canopy).

Thank you for bringing our attention to the STC-MODSCAG data, which provides snow estimate on the ground and is better suitable for evaluating modelled SCF. However, it is currently only available for the western US, a global dataset is required to evaluate model performance in our study. Previous studies have shown that the accuracy of SCF from MODIS is lower than that from MODSCAG (Painter et al., 2009; Stillinger et al., 2023). Evaluating SCF from standard MODIS and STC-MODSCAG with high resolution airborne lidar data in western US, Stillinger et al. (2023) showed that the median bias (RMSE) was -0.071 (0.127) for MODIS and -0.001 (0.120) for STC-MODSCAG across various snow climates. They also showed that the MODIS SCF product exhibited consistent negative bias of around -0.10 under intermediate canopy cover, with the bias increasing with greater snow cover, reaching -0.25 under full snow cover conditions.

In addition, other evaluation studies suggested the accuracy of MODIS snow products was in the range of 88-93%, and dense forests and steep terrain may obscure the MODIS sensor's view of snow-covered ground, resulting in SCF underestimation (Hall et al., 2002; Hall and Riggs., 2021).

We will acknowledge the uncertainties of the MODIS SCF product mentioned above and provide a brief discussion on its impact on our results when revising our manuscript.

Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E., & Bayr, K.J.: MODIS snow-cover products. Remote Sensing of Environment, 83(1–2), 181–194. http://dx.doi.org/10.1016/S0034-4257(02)00095-0, 2002.

Hall, D. K. & Riggs, G. A.: MODIS/Terra Snow Cover Monthly L3 Global 0.05Deg CMG. (MOD10CM, Version 61). Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD10CM.061. Date Accessed 06-19-2025, 2021.

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., and Dozier, J.: Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, Remote Sens. Environ., 113, 868–879, https://doi.org/10.1016/j.rse.2009.01.001, 2009.

Stillinger, T., Rittger, K., Raleigh, M. S., Michell, A., Davis, R. E., and Bair, E. H.: Landsat, MODIS, and VIIRS snow cover mapping algorithm performance as validated by airborne lidar datasets, The Cryosphere, 17, 567–590, https://doi.org/10.5194/tc-17-567-2023, 2023.

(3) Discrepancies in spatial resolution between reference data used to validate model simulations and the spatial resolution of the model simulations can largely impact results. Please see specific comment below addressing this point.

To minimize these issues, we rely on snow courses and airborne gamma measurements because they are more spatially representative than single point measurements (Meromy et al. 2013). Snow courses consist of multiple measurements along a transect several hundreds of metres to kilometres in length that are averaged together to provide a single SWE value. Airborne gamma measurements are averaged across 300 m wide footprints and along 15–20 km long flight lines. In both cases, these measurements better sample the sub-grid-scale variability than a single-point measurement and so are more effective in capturing the larger-scale average. This decision does not fully close the scale difference between observations and gridded product, but it helps substantially.

In addition, analysis in Mortimer et al. (2024) showed that evaluation of gridded products with spatial resolutions ranging from 4km to 1.25° using this type of reference data yielded consistent performance ranking whether evaluated with airborne gamma or snow courses in non-mountain or mountain areas. This means we can make meaningful relative assessments of the gridded product performance.

In this manuscript, our intent is to provide readers with a sense of the relative simulated SWE errors driven by differences in the forcing data. We believe the reference data are appropriate for this purpose. However, at your suggestion, we also looked further into the sampling variability of the bias within a 1x1 degree CLASSIC grid cell. We identified a subset of grid cells containing multiple reference sites with long records. For simplicity, we restricted this demonstration to February. To remove issues related to sampling dates within a month, we only compared reference sites collected on the same date. As Figure AR1 shows, in nearly all cases, the ranking of SWE magnitudes for each of the three products and the reference SWE are similar from year to year. This demonstrates that the relative product errors assessed in the manuscript are likely to be consistent even if temporally sampled less frequently than demonstrated here. In nearly all cases the product SWE also falls outside of the standard deviation of the reference SWE. This demonstrates that the calculated biases presented in the manuscript are likely to be meaningful, even if spatially sampled less frequently than demonstrated here (however in most cases our arguments rely only on the relative product bias anyway). In rare cases (e.g. red box), the choice of reference site will alter the sign of the bias (but still does not alter the relative sense of bias among the three products).

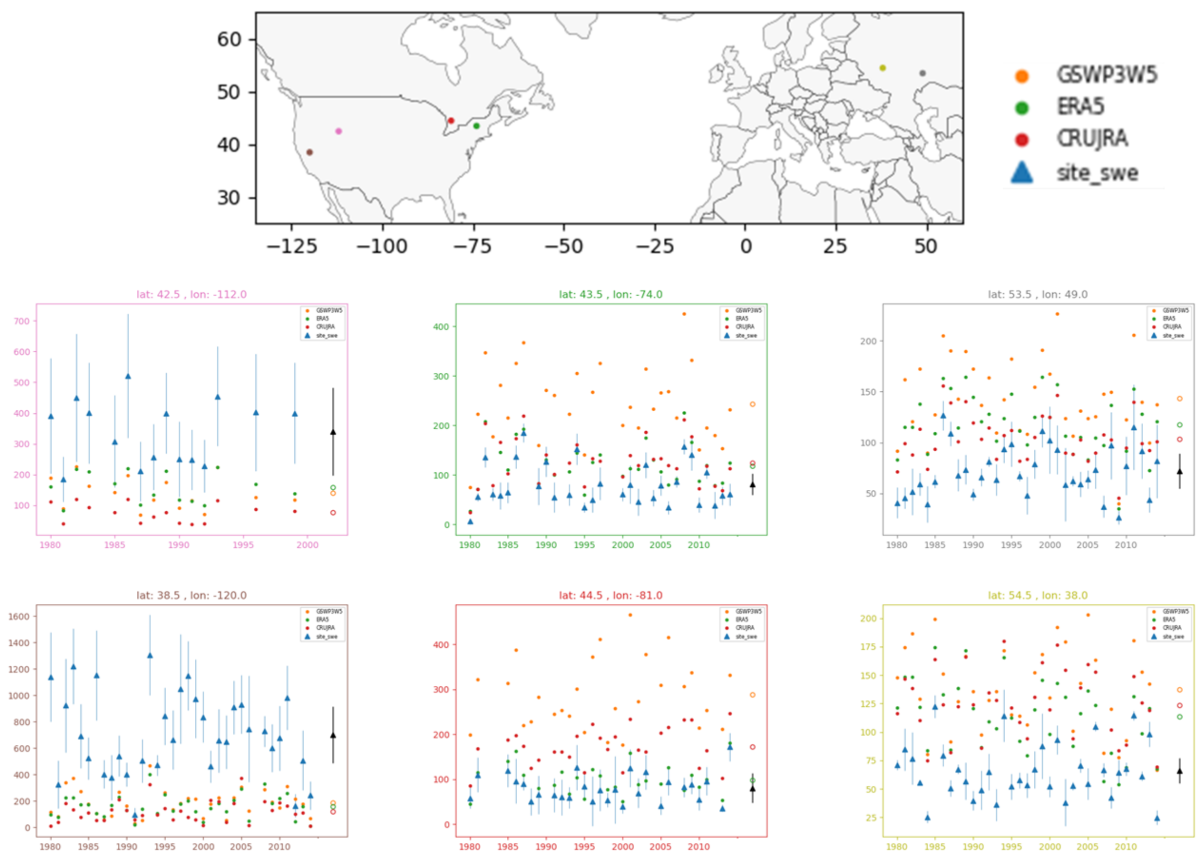

Figure AR1. Mean and standard deviation of reference SWE (blue) for sites measured on the same date within the same model grid cell. When there were multiple dates in the same month, the mean and standard

deviation were calculated for sites measured on the same date and then averaged across the month-year. Dots: modelled SWE for the corresponding grid cell and month. Far-right black line, triangle, and hollow circles show the mean across the time series. For display, only February is shown. Each plot corresponds to one dot on the map (lon/lat listed at top of each plot and colors of plot axes correspond to dots colors on map). Sites in the western US are mountainous, all other are in flat regions as defined in the manuscript.

Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., and Rice, R.: Subgrid variability of snow water equivalent at operational snow stations in the western USA, Hydrol. Process., 27, 2383–2400, https://doi.org/10.1002/hyp.9355, 2013.

Mortimer, C., Mudryk, L., Cho, E., Derksen, C., Brady, M., and Vuyovich, C.: Use of multiple reference data sources to cross-validate gridded snow water equivalent products over North America, The Cryosphere, 18, 5619–5639, https://doi.org/10.5194/tc-18-5619-2024, 2024.

(4) Figure quality should be improved throughout.

We apologize for the poor quality of the figures. The quality was fine in the original Microsoft Word version of the manuscript but deteriorated after converting to the pdf file. We will make sure the figures will all have high quality in the revised manuscript.

Specific recommendations:

Paragraph starting in line 81: Note that some land surface models also consider SCF as a function of snow density and land cover classification (e.g., He et al., 2023). He, C., et al. The community Noah-MP land surface modeling system technical description version 5.0. NCAR Technical Note NCAR/TN-575+ STR, doi: 10.5065/ew8g-yr95, 2023.

Thank you for your suggestion. We will add this point when revising our manuscript.

Section 2.1: please articulate the capacities in which CLASSIC is used, for either research applications or operational modeling.

We agree that it would be nice to include a couple of sentences about the applications of CLASSIC. We will add this when revising our manuscript.

Section 3.1 and Figure 1: Please add information on the calculation of topographic standard deviation. Specifically, what is the resolution of the elevation product which is used to calculate this metric?

This was already provided in the manuscript: "Classification of mountain and flat regions is based on standard deviation of the sub-grid terrain from the ETOPO1 elevation data at 1 arc-minute resolution (NOAA, 2009)."

Section 3.2: Another potential issue with MODIS SCF is not just its accuracy, but also whether its retrieval represents pixel scale SCF or just the ground SCF. Many land models simulate ground SCF, rather than total pixel SCF (e.g., including vegetated fractions of the pixel) and thus a comparison with the MODIS data used here may not be appropriate. The STC-MODSCAG data addresses this issue, and the latest version has available data across the mountainous western US (https://nsidc.org/data/stc_modscgdrf_hist/versions/1#anchor-data-access-tools).

Please consider using these data as an additional reference to evaluate whether the comparisons against MODIS are reliable.

Please see our reply above to your main comments. When revising our manuscript, we will include a summary on the evaluation results of the MODIS SCF data from previous studies to provide some uncertainties on our comparison against MODIS.

Lines 315-316: Simulated snow density is also a source of SCF uncertainty, e.g., Abolafia-Rosenzweig et al. (2024), which could be noted here or in the Discussion. Abolafia-Rosenzweig, Ronnie, et al. "Evaluating and enhancing snow compaction process in the Noah-MP land surface model." Journal of Advances in Modeling Earth Systems 16.2 (2024): e2023MS003869.

Thank you for your suggestion and providing the reference. We will add this point when revising our manuscript.

Section 3.3: These SWE evaluations are likely largely impacted by discrepancies between observed and modelled spatial resolutions. It would be good to emphasize this point further, even in the case of airborne gamma SWE observations. To consider the spatial representativeness of observations, consider comparing time series from in-situ stations contained by the same 1 degree pixel and consider whether there are large discrepancies (e.g., with bias and correlation metrics).

Please see our reply above with Figure AR1. Most of our conclusions about forcing-driven errors are based on the assessed bias.

Also, when observations are measured infrequently (e.g., a few times in a month) are the modelled data screened temporally to match the observational frequency prior to comparison?

The model output was only saved at monthly frequency. How well our date-specific samples will represent a true monthly mean will depend on their distribution over the month of interest. We examine two aspects of this in detail below: lack of snow-free reference measurements and the distribution of measurements within a month.

Despite the challenges highlighted below, we are confident that for the application used in our study, the data reasonably sample the monthly value outside of the shoulder seasons. Owing to the larger uncertainty during the shoulder seasons, in the revised manuscript we will restrict our evaluations with reference data to January-March. Figure 4 and all associated conclusions and discussion will be revised accordingly. Further, we propose to add the following text to the methods in Section 3.3.

"The reference observations do not account for snow-free periods because they are only conducted when there is snow. During the accumulation and ablations seasons, the monthly mean of available reference SWE will therefore often overestimate the true monthly mean value. For this reason, we restrict the comparisons of product SWE with reference SWE to January-March. Additionally, the infrequent sampling of the reference data (Fig. 2 lower left; see also Table 4 in Mortimer and Vionnet, 2025) means that, even when there is continuous snow cover, the monthly value calculated from the available dates with observations may not be representative of the true monthly mean. Investigation of the timing of the in-situ measurements within a month showed that, for the full

domain, the timing of the observations is fairly well distributed across a month. However, this varies regionally and by network with some networks (e.g. Canada) biased towards the beginning of the month and others (e.g. Russia) biased towards the end of the month. We are unable to account for these biases in our analysis. The statistics calculated from comparisons with in situ data are not intended to be used as absolute performance measures. Rather, we are interested in the relative performance of the SL12 parameterization with the three different forcings and over time."

1. Lack of snow-free reference observations

Reference observations are only conducted when there is snow. When we calculated the monthly mean reference SWE we did not account for the snow free period during melt and onset. Thus, if the first (accumulation) or second (melt) half of a month is snow free, the mean reference SWE for that month will be an overestimate of the true mean.

To illustrate, below (Figure AR2) we show the distribution of the bias (reference minus product) for December, February, and May for the western US reference network. In December, there are no reference measurements prior to the 15th so the mean reference SWE calculated from these measurements is not representative of the monthly mean. During the middle of the winter (e.g. February) the sample distribution is concentrated at the beginning and end of the month, capturing the monthly mean (although slightly biased towards the end of the month), and there is no significant trend in the bias versus the day of the month. The problem illustrated in December is not evident in May (melt season) because there are sufficient sites with persistent snow cover. There may, however, be local issues for specific sites that lose snow earlier (since our treatment below lumps data across the western US).

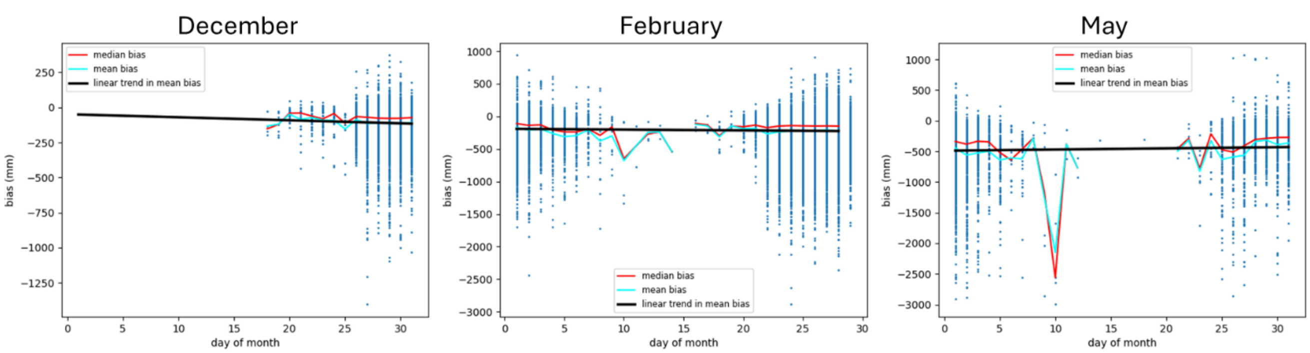

Figure AR2. Bias versus day of the month for reference sites in the NRCS network in the western US. Product bias for matching reference sites (blue dots) with x-axis location corresponds to the day of month of the reference observation and its trend versus the day of the month (black line). Mean and median bias for each day with reference observations in cyan and red, respectively (mean of the blue dots on each day of the month). Horizontal grey dotted line – mean product SWE calculated from the pool of data in the blue dots. For illustration purposes, only the ERA5 forcing is shown.

2. Sample distribution within a month

If reference observations are not evenly distributed across the month this will introduce a bias in the monthly average reference SWE relative to the true monthly value. However, it is challenging to disentangle the timing of the observation from the landcover type and SWE magnitude because different networks, which often cover different snow classes and land cover types, have different sampling schedules. This error is not accounted for in our analysis.

Outside of the accumulation and melt seasons the data as a whole are fairly evenly distributed across a month. However, there are key regional differences because the sampling schedule varies by network (see Table 4 in Mortimer and Vionnet, 2025). Figure AR3, below, shows the number of reference observations from each network in our reference dataset over the full study period. Observations in Finland are centered around the middle of the month. In Russia, they tend to miss the first 5-10days and are biased towards the latter two thirds of the month. In Canada, observations are concentrated at the beginning and middle of the month with a secondary peak at the end of the month. This means the reference mean will tend to be biased towards the end of the month over Russia and the beginning of the month over Canada.

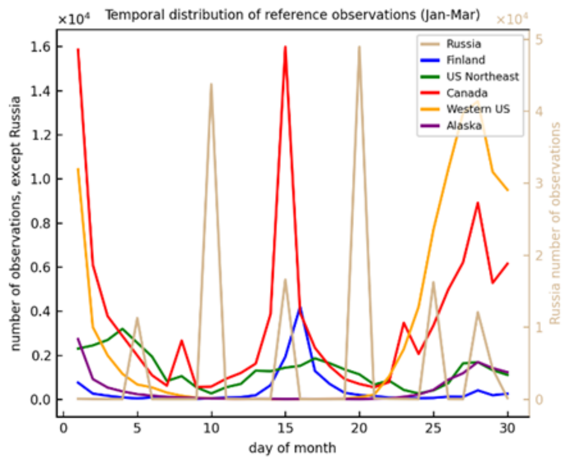

Figure AR3. Number of reference observations by network and day of the month during 1980-2014 for the Jan-March period.

Line 405: Is there truly no feedback in these offline runs? Land models often calculate 2-m air temperature prognostically which could impact SWE. If this is the case for the CLASSIC model, consider re-wording here.

Thank you for noting this. In CLASSIC, 2-m air temperature is not calculated prognostically, but it affects the surface temperature, which may in turn affect SWE through snowmelt. We will reword the sentence when revising our manuscript.

Lines 424-430: Adding more quantitative information here would be useful.

Thanks for your suggestion, we will include more quantitative information when revising our manuscript.

It looks like the simulations tend to underestimate SWE substantially; however, there is a tendency to overestimate winter SCF, largely in the control simulations and modestly in the SL12 simulations. If the SCF scheme is truly accurate at converting SWE or snow depth to SCF then we would expect to see underestimates in SCF. Can this point be added, particularly connecting logic between Sections 4.2 and 4.3?

Thanks for raising this point. We think this "inconsistency" between SWE underestimation and SCF overestimation in the mountain regions is likely due to the following:

During snow accumulation, SCF increases rapidly with snow depth in both the Control and SL12 schemes, as illustrated in the red (Control, only a rough approximation) and cyan (SL12) curves below. SCF reaches 100% in the Control and ~ 80% in the SL12 when snow depth is around 10cm. Snow is usually deep in the mountain regions (e.g. Fig.2d). Though SWE is underestimated in the model, SCF should have reached its maximum value during the peak SWE period (DJF).

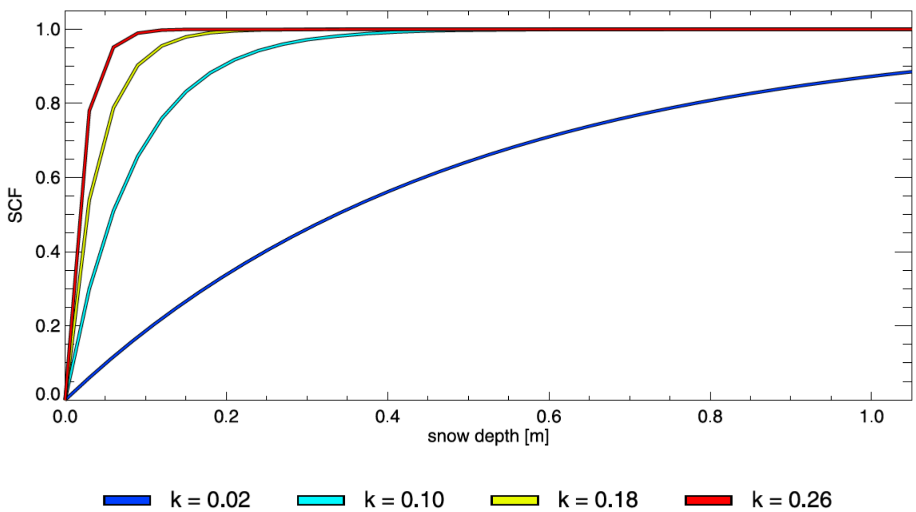

Figure AR4. SCF parameterization for accumulation events. The x axis is snow depth in meters, and they axis is SCF. Colors indicate different values of parameter k_{acc} from equation (1) (Fig. 7 from Swenson and Lawrence, 2012).

In addition, dense forests and steep terrain may obscure the MODIS sensor's view of snow-covered ground, resulting in underestimation (Hall et al., 2002; Marchane et al., 2015). The magnitude of winter SCF overestimation by SL12 in the mountain regions is relatively small (Fig. 5). The mean bias is 0.01, 0.02, and -0.02 for runs forced by CRUJRA, ERA5, and GSWP3-W5E5 (Table 2a), which is within the uncertainty range of the MODIS product. We will add the uncertainties of the MODIS product when revising our manuscript.

A. Marchane, L. Jarlan, L. Hanich, A. Boudhar, S. Gascoin, A. Tavernier, N. Filali, M. Le Page, O. Hagolle, B. Berjamy, Assessment of daily MODIS snow cover products to monitor snow cover dynamics over the Moroccan Atlas mountain range, Remote Sensing of Environment, Vol 160, 72-86, https://doi.org/10.1016/j.rse.2015.01.002, 2015.

Swenson, S. C. and Lawrence, D. M.: A new fractional snow-covered area parameterization for the Community Land Model and its effect on the surface energy balance, J. Geophys. Res.-Atmos., 117, D21107, https://doi.org/10.1029/2012JD018178, 2012.

It would be interesting to consider whether there are significant correlations between SCF biases with topographic complexity in each of the simulations, and in particular highlight if the SL12 scheme reduces or removes this relationship.

Thank you for your suggestion. To investigate whether there are significant correlations between SCF biases and topographic complexity, we made scatter plots (Fig. AR5) between SCF and standard deviation of sub-grid topography during the winter and spring seasons for each of the simulations. Below is an example of the plots for model runs forced by CRUJRA.

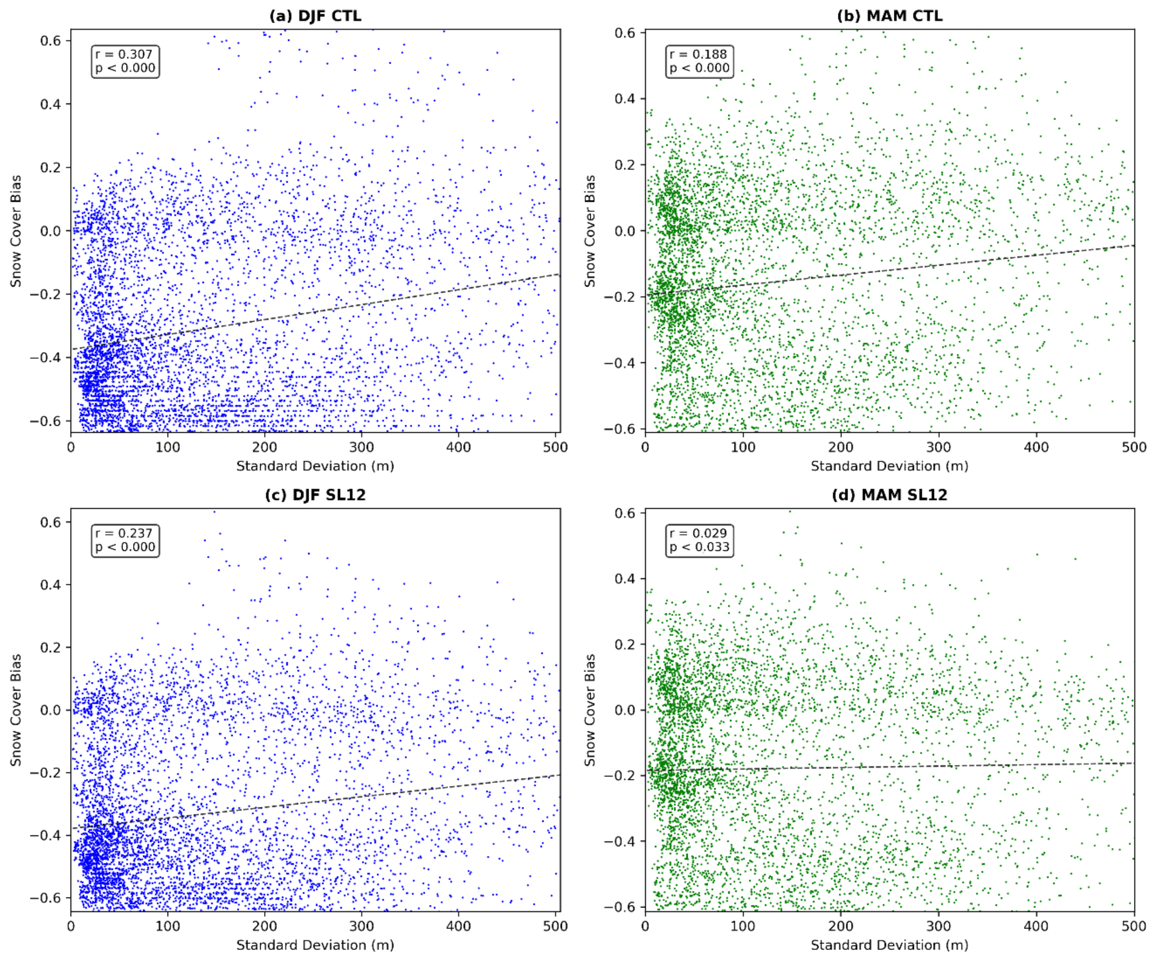

Figure AR5, scatter plots between SCF and standard deviation of sub-grid topography during the winter and spring seasons for model runs forced by CRUJRA.

As expected, there are significant correlations between SCF biases with topographic complexity in all the simulations. The relationship is reduced by the SL12 scheme, especially in the spring. We will mention this in the main text and include a figure in the supplement when revising our manuscript.

Section 4.4: It would be valuable to note whether the albedo biases are consistent with SCF biases (e.g., locations with SCF overestimates have albedo overestimates).

Thank you for your suggestion. We will include the following figure (Fig. AR6) in the revised manuscript. Figure AR6 shows that surface albedo is overestimated by the control scheme, and the overestimation in the mountains is reduced by the SL12 scheme, consistent with the results for SCF.

Note the MODIS surface albedo product does not have shading correction, which tends to lead to underestimation in snow albedo in mountians (Bair et al., 2022). In flat regions, underestimation in MODIS SCF in vegetated regions likely contributed to the underestimation in surface albedo (details can be found in our reply to your main comments above), which at least partly explains the relatively large overestimation in the boreal forest regions.

Bair et al., 2022, https://doi.org/10.5194/tc-16-1765-2022.

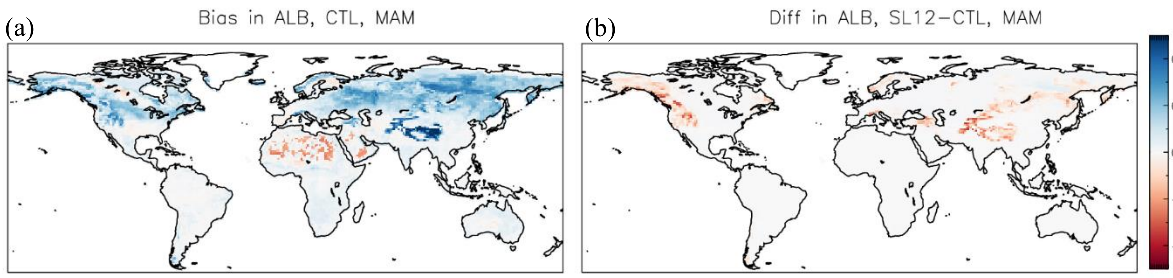

Figure AR6. (a) Surface albedo (ALB) bias in a model run using the Control parameterization in the spring, (b) the difference in ALB between the model runs using the SL12 and CTL parameterizations. The surface albedo from MODIS is used as a reference.

Section 5: here are some potentially useful references for land model SWE biases:

He, Cenlin, et al. "What causes the unobserved early-spring snowpack ablation in convection permitting WRF modeling over Utah Mountains?." Journal of Geophysical Research: Atmospheres 126.22 (2021): e2021JD035284.

Abolafia-Rosenzweig, Ronnie, et al. "Implementation and evaluation of a unified turbulence parameterization throughout the canopy and roughness sublayer in Noah-MP snow simulations." Journal of Advances in Modeling Earth Systems 13.11 (2021): e2021MS002665.

Chen, Fei, et al. "Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study." Journal of Geophysical Research: Atmospheres 119.24 (2014): 13-795.

von Kaenel, Manon, and Steven A. Margulis. "Evaluation of Noah-MP snow simulation across site conditions in the western United States." Journal of Hydrometeorology 25.9 (2024): 1389-1406.

Thanks for providing these helpful references. They will be included in the revised manuscript as appropriate.