We thank Referee #3 for their helpful comments. Our replies to their comments are shown in bold below.

Review of manuscript "Impact of topography and meteorological forcing on snow simulation in the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC)" By Libo Wang et al. This paper discusses the impacts of using an alternative snow cover fraction scheme in the Canadian Land Surface Scheme. The parameterization adopted here was developed by Swenson and Lawrence (2012) and is currently used in other land models (e.g., CESM) and appears more physically realistic than the one currently used in the model here. In particular, the new formulation accounts for topographic effects while the previous one did not. The authors evaluate the new model setup using multiple datasets and satellite observations finding that the performance of the model in the new configuration overall improves. The authors also explore the sensitivity of the (offline) land model forced by three meteorological forcing datasets.

While not presenting a new snow scheme or parameterization but implementing an existing one in a different land model, I think the paper is interesting and well written, and would be of interest to the readership of GMD. Furthermore, the authors perform a comprehensive evaluation of multiple land surface variables which is useful in characterizing the performance of their model forced by multiple reanalysis datasets. Below are my comments on the paper, which the authors should address or respond to before the paper is considered for publication. The style and writing is overall good, although there are a few typos or sentences that could be improved. Please see a list of recommendations below.

Thank you for your overall positive review of our manuscript.

Main comments

Why was this particular parametrization chosen? I understand it is a clear improvement and probably one of the best options in the literature, but I still feel the paper lacks a brief review of what existing options for SCF parameterizations are, and what is used in other snow models. Given the scope of the paper, I think this would be useful to the reader. E.g., I believe adding some detail to the text currently at lines 81 - 87 would improve the discussion and help justify the choice made here.

Thanks for your suggestions. We will expand on the information provided at lines 81-87 when revising our manuscript.

Figure 3 does not seem to have good quality in the pdf. Please improve the quality (e.g., increase resolution, use a larger font, ..). Also, I struggle to fully understand the implications of this comparison: It seems for precipitation ERA5 has a very large (compared to the other datasets) positive bias on mountain regions. I wonder if it would be better to present a measure of relative error rather than the absolute difference in (mm/month) for the entire dataset?

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.

First, we note that the larger bias for ERA5 with respect to the CRU reference data as shown on Figure 3 reflects that both CRUJRA and GSWP3-W5E5 forcing datasets already incorporate CRU reference data for bias-correction (see Section 5.1 for a discussion on why there are still residual biases in these two datasets). Beyond that point, our interpretation of the results is based on the assumption that the relative precipitation amounts are accurate, but that there is uncertainty in how well the CRU reference data reflects precipitation amounts in mountain regions. Figure 3

demonstrates that in a relative sense, ERA5 has more precipitation in mountain regions compared to CRUJRA and GSWP3-W5E5 (this precipitation would typically be snowfall given the seasonal timing and mountain location). Despite this additional snowfall, CLASSIC SWE output forced by ERA5 still has a negative bias in mountain regions (Fig. 4), although less than the larger negative SWE biases identified for both CRUJRA and GSWP3-W5E5 forced output. As stated at lines 614-616 in Section 5.1, we interpret these two apparent contradictions to mean that the CRU precipitation measurements have a sampling bias (supported by Nijssen et al., 2001; Adler et al, 2003; Shi et al., 2017) such that the hemispheric ERA5 precipitation bias in mountain regions is actually smaller in reality than those from CRUJRA and GSWP3-W5E5 (i.e. all forcings have too little precipitation in mountain regions but ERA5 has the closest to accurate amount).

To provide context for the bias plots in Figure 3, we plotted air temperature and precipitation from the three forcing datasets over the NH (a and b) and the HMA (c and d) regions. Figure AR1(b) and (d) show that ERA5 has more precipitation in mountain regions compared to CRUJRA and GSWP3-W5E5. However, it is hard to tell the differences in temperature in the NH regions shown in Figure AR1(a), which were illustrated better in the bias plots shown in Figure 3(a). We will include Figure AR1 in the supplement in the revised manuscript.

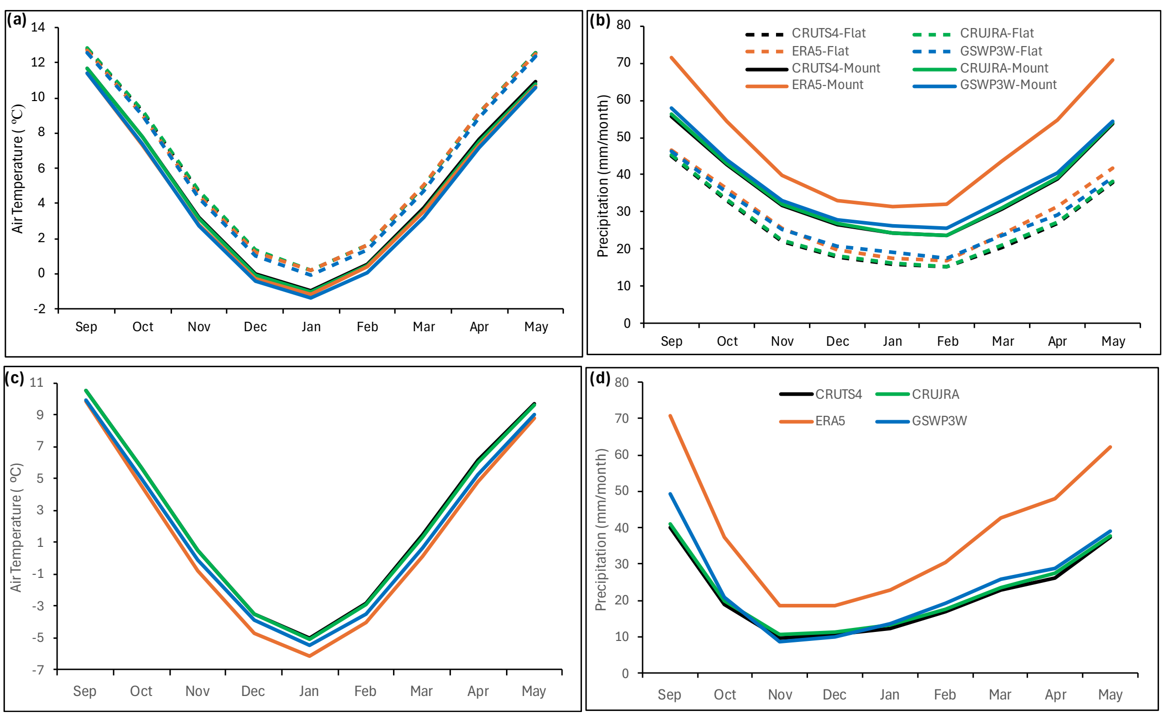

Figure AR1. Monthly mean air temperature (a and c) and precipitation (b and d) in the NH mountainous (solid line) and flat (dashed line) regions (a and b) and the HMA mountainous regions (c and d) over the 1980-2014 period.

L428: "Overall, ERA5-SL12 outperforms the other two model runs with lower bias and better correlation in mountainous regions and it shows similar performance as CRUJRA-SL12 in flat regions" -> Is this in part due to CLASSIC underestimating SWE and ERA5 overestimating winter precipitation compared to other datasets, as shown in Figure 3?

This statement is consistent with the interpretation provided that precipitation in all the forcings is biased low, but that ERA5 precipitation is the least biased and most accurate. This additional precipitation / snowfall, and therefore SWE improves the simulation of other downstream land surface variables in mountain regions (and that ERA5 and CRUJRA forcings lead to comparable performance in flat regions).

Figure 5: As for Figure 3, please also improve the quality / resolution of this figure.

We will improve the quality of all figures in the revised manuscript.

It seems the standard deviation of evolution is an important parameter for the parameterization used here. At what scales / resolution was it computed, and is this consistent with previous applications and with Swenson and Lawrence (2012)? Does this matter?

ETOPO1 elevation data at 1 arc-minute resolution (~1.85 km at the equator; NOAA, 2009) was used to compute the standard deviation of topography in our study. Swenson and Lawrence (2012) did not provide information about the elevation data used in their study.

To assess the impact of DEM resolution, we compared σ_{topo} derived from two DEM datasets: ETOPO1 (1 arc-minute resolution) and ETOPO2022 (15 arc-second resolution, ~500 m). The results show that the differences are limited in extent, primarily concentrated along the mountain edges. We also performed a test simulation using σ_{topo} derived from ETOPO2022 and compared modelled SCF with that from a run using σ_{topo} derived from ETOPO1. The maximum difference was less than 5%.

Thus, the resolution of the DEM data has limited impact on the calculation of sub-grid topographic variability and the simulated SCF.

Figure 2: Could you specify the units of the two colorbars?

We will add units to the colorbars when revising our manuscript.

Minor comments or typos

L64: comma after "Therefore"

Done.

L66: "Snow depth (SND) varies at scales from about 10 to 100 m" reads unclear - could you maybe state more precisely the range of scales at which it can vary? Do you mean "up to" 10-100m? Surely it is variables at larger scales too.

Thanks for noting this. We will modify the sentence when revising our manuscript.

L94: “so that” neither was implemented?

Thanks, we will add that when revising our manuscript.

L131 “each with and without snow cover “ -> do you mean to say that each of these fractions can be partially covered by snow? Please clarify

The four sub-areas are: vegetation over bare soil, bare soil, vegetation over snow, and snow over bare soil. They are calculated based on the fractional coverage of the vegetation categories and SCF. The snow-covered and snow-free areas change dynamically at each time step. We will modify the sentence when revising our manuscript.

L132: “which includes” -> please clarify: does the snow layer include these processes, or the snow scheme?

Thanks for noting this. We meant “snow scheme” here, we will clarify this when revising our manuscript.

L150: “increasing or decreasing this threshold value” -> which threshold value, the minimum snow depth 0.1m? It seems strange this parameter has little effect since it would significantly affect SFC and surface albedo. Could you comment on this?

Yes, the minimum snow depth 0.1m.

The small effect of varying this threshold is mainly because areas with snow depth from 5-10 cm and from 10-15 cm are very limited in the model.

L155: “but also on month of the year” -> would it be more correct to say on the age of the snowpack?

It is not about the age of the snowpack, but whether snow mass is increasing (accumulation) or decreasing (ablation). We will modify the sentence to clarify this when revising our manuscript.

L612: Add comma: “Yet, among …”

Done.