Improving the CLASSIC (v1.8) Snow Model to Better Simulate Arctic Snowpacks

https://doi.org/10.5194/egusphere-2026-492

(Lalande et al., 2026)

Response to RC1

Below are our detailed responses (the original comments from RC1 are shown in *italic black font*, and our answers in blue). Before addressing the reviewer's comments in detail, we list below a few preliminary corrections that were made in the revised manuscript. Line numbers refer to the first submitted versions of the manuscript (egusphere-2026-492.pdf) and the Supplement (egusphere-2026-492-supplement.pdf).

Preliminary corrections

Minor text corrections

- L35–36: we added in the introduction: "*These changes affect the hydrological cycle, infrastructures, and ecosystems (Bokhorst et al., 2016; Bring et al., 2016)****, and may trigger interconnected but asynchronous state shifts among northern ecosystem components, including snowpack, vegetation, permafrost, lakes, and peatlands (Saulnier-Talbot et al., 2024).***"
- L39–40: we changed our formulation "[...], *which greatly affects these fluxes through its impact on soil temperature*" to "***which affects these fluxes through its impacts on soil thermal and hydrological conditions and on gas diffusion between the soil and the atmosphere***" and added the reference of **Mavrovic et al. (2025)** about methane fluxes.
- L86: we replaced "***true** Arctic sites*" with "***representative** Arctic sites*"
- L278–279: we added a citation and slight insights at the end of this sentence: "*as we consider that there is no further snow compaction within the canopy****—vegetation reduces near-surface wind exposure and can therefore limit wind-driven snow redistribution and densification within the canopy (e.g., Lackner et al., 2022), although vegetation–snow–wind interactions may involve more complex relationships that are not explored in the present study.***"
- L582: "*Swamp Angel could trap **some** snow during wind events*" -> "*Swamp Angel could trap **part of the wind-redistributed** snow during wind events*"
- L738: Addition of: "[...] ***and will become increasingly important in a changing climate, inducing Arctic greening (Bayle et al., 2022).***"

- L770–771: we replaced: "*[...] to achieve more consistent simulations of energy, water, and carbon budgets in Arctic environments.*" with "*[...]* ***particularly in the context of ongoing Arctic shrubification and greening, to achieve more consistent simulations of energy, water, and carbon budgets in Arctic environments (e.g., Sturm et al., 2001; Myers-Smith et al., 2011; Bayle et al., 2022).***"

Correction of the normalized MB and RMSE

During final verification of the analysis scripts, an inconsistency was identified in the normalized RMSE and mean bias plots for snow depth and SWE. In Figs. 6, 7, C2, and C3, the individual-site markers for snow depth and SWE (automatic and manual measurements) were inadvertently normalized using the standard deviation of the corresponding model experiment, rather than the standard deviation of the observations. This issue was restricted to the individual-site markers in the normalized panels; the SnowMIP and Arctic site-mean markers, as well as the overall scores, already used the correct normalized metrics. The affected panels and related text have therefore been updated. The impact remains limited overall but is more pronounced at TVC, where the simulated variability was larger and therefore previously led to underestimation of the normalized RMSE and mean bias values. Further details are available in the Appendix at the end of this document.

Reviewer's comments

General comments:

This study evaluates potential snow model enhancements for the CLASSIC model, with a particular focus on Arctic environments. The potential enhancements focus on the thermal conductivity of the surface soil, a revised computation of the temperature at the snow–soil interface, and the addition of a windless exchange coefficient in sensible heat flux calculations. Arctic-specific adaptations include blowing snow sublimation losses, a new snow compaction scheme, and snow thermal conductivity updates. Detailed comparisons between simulations with differing physics representations and observations of snow depth, SWE, snow albedo, surface temperature, and soil temperature indicate that the model enhancements improve model accuracy. However, additional analyses could make these results more robust, comprehensive, and meaningful. In particular, there is value in considering additional evaluation metrics, reporting the statistical significance of results, and more deeply considering whether the model updates are improving accuracy for the right reasons. Overall, this study will be valuable, particularly for understanding potential avenues for improving CLASS snow modeling, but major revisions are recommended before publication.

We thank Reviewer 1 for their helpful and relevant comments. We added most of the clarifications requested by the Reviewer. In particular, we provided more details on the point-scale simulation framework, added a complementary temporal correlation metric, restructured the Results section to better highlight the main findings, and added a new Discussion subsection on the physical consistency of the model improvements across variables. We also clarified several other points to improve the manuscript's clarity. Detailed responses to each comment are provided below.

Specific comments:

Line 124: *Please include a statement noting the potential uncertainties associated with using this rain–snow partitioning threshold.*

Thanks for the suggestion; we have added: "***This temperature-based rain–snow partitioning may introduce uncertainties, especially for mixed-phase events or precipitation occurring near the threshold temperature (e.g., Dai, 2008; Harder & Pomeroy, 2014; Feiccabrino et al., 2015; Jennings et al., 2018; Jennings & Molotch, 2019).***"

Line 137: *Please describe or provide a reference for the quality control and undercatch correction procedures.*

We added the following references:

- Ménard et al. (2019) for the SnowMIP sites, which states: "*six different methods are applied by the seven collecting teams to correct for undercatch: yearly or constant scaling factors, model simulations, matching against snow water equivalent (SWE) or replicate gauges*"
- Domine et al. (2021, 2024) both used the correction of Kochendorfer et al. (2017) at Bylot and Umiujaq
- Dutch et al. (2022) for TVC: "*[...] precipitation was corrected as per Pan et al. (2016).*"

They also describe the methods used to quality control the data.

Section 2.5:

Please provide details regarding the model spatial resolution and how it may impact comparisons with point-based measurements. This is discussed later, but it would be valuable to address it here as well. For example, vegetation and snow fraction at the point scale are binary.

We added the following paragraph:

> "***This study uses CLASSIC v1.8 (Melton et al., 2020) in a 1D point-scale configuration. In practice, strictly point-scale simulations are somewhat idealized, given that the meteorological forcing data and validation measurements cover different areas around the stations, and the spatial footprint of each sensor varies. We therefore prescribed soil properties, vegetation cover, and other site characteristics to be as representative as possible of the area surrounding the measurement sensors. CLASSIC point-scale simulations are treated similarly to single-grid-cell simulations, allowing fractional vegetation cover and separate surface energy and water balances over different subareas, including snow-covered and snow-free fractions. The snow cover fraction parameterization was kept active and follows a linear relationship between 0 and 10 cm snow depth, as further discussed in Sect. 4. However, the presence of snow at the scale of a snow depth sensor is effectively binary; consequently, this introduces uncertainties when comparing model outputs with point-based measurements, particularly during shallow or patchy snow conditions and in wind-affected environments where snow redistribution can generate strong small-scale variability in snow depth. Scanning laser measurements could help improve the spatial representativeness of snow depth observations in future studies (e.g., Picard et al., 2016). To limit these effects, evaluation metrics were computed only when observed snow depth exceeded 10 cm (Sect. 2.6), and further discussion about the uncertainties related to point-scale simulations is covered in Sect. 4.***"

Indeed, in practice, snow depth, snow/soil temperature, etc., are not strictly measured at a single point, so we prescribed boundary conditions to be as representative as possible of the area surrounding the station, including fractional vegetation cover when needed. For example, at Umiujaq, the area around the station is a mix of shrubs and bare ground/moss/lichen surfaces (see Fig. 1c of Domine et al. (2024)); therefore, we prescribed the vegetation cover as 50 % shrubs and 50 % bare ground, with a peat layer to approximate the insulating effect of mosses and lichens.

What uncertainties are associated with the model forcing inputs?
This aspect is already covered in the discussion (Sect. 4.5), but we added this sentence following the forcing description in Sect. 2.5:

> "***Uncertainties in the meteorological forcing inputs can also affect the simulations, particularly at Arctic sites where precipitation, radiation, and near-surface meteorological measurements are more prone to errors due to wind undercatch, frost, polar night conditions, and data gap filling; these uncertainties are further discussed in Sect. 4.5.***"

Section 2.6: *Please include a sentence explaining how the observed snow albedo is not impacted by surrounding vegetation. With a snow depth threshold of 10 cm, it seems that albedo observations may still be influenced by vegetation.*
We agree that albedo can be affected by protuberant vegetation (both in observations and in the model). Those uncertainties are already covered in Sect. 4.6, in complement of Sect. 7 of the Supplement, and we added this sentence in Sect. 2.6:

> "***Observed albedo may still be influenced by protruding vegetation, especially during shallow snow conditions, despite the 10 cm snow depth threshold. However, albedo measurements are available mainly at sites with short or mowed grass, alpine tundra, or little vegetation cover, making them broadly comparable to the simulated snow albedo. Remaining uncertainties may instead arise from limitations in the representation of vegetation in CLASSIC, including possible overestimation of simulated vegetation height and the absence of vegetation bending, as further discussed in Sect. 4.6.***"

Please also include correlation as an evaluation metric to capture how well the model simulates temporal variability at a daily time step.
We actually purposely did not include the correlation in the initial version of our study for the reason described below. Based on your suggestion, we still added Pearson correlation coefficients as a complementary evaluation metric. However, we did not include them in the overall score—nor stress on it—because daily correlations can be difficult to interpret for point-scale snow measurements, especially at Arctic sites where wind-driven redistribution may cause observed snow depth to decrease during or after snowfall events or to increase without concurrent precipitation. We therefore report the correlations as an additional diagnostic in Sect. 2.6 (in addition, we added tables of all metrics in Sect. 11 of the Supplement):

> "***Pearson correlation coefficients were also computed as a complementary diagnostic to assess the temporal agreement between simulated and observed daily time series. These values should be interpreted with caution, as daily simulated time series are not expected***

to exactly follow the observations, especially for snow depth at Arctic sites where wind-driven redistribution can cause local snow depth changes that are not directly related to precipitation events. Therefore, correlations are used only as an additional diagnostic and are not included in the overall score. All metric values are reported in Sect. 11 of the Supplement."

and we added a complementary subsection at the end of the result section:

"***3.4 Temporal correlation***

Pearson correlation coefficients were computed as an additional diagnostic of the temporal agreement between simulated and observed daily time series and are reported in Table S8 of the Supplement. Daily simulated and observed variations are not necessarily expected to be strongly correlated, especially at Arctic sites where wind-driven snow redistribution can produce local snow depth changes not directly related to the precipitation forcing. Correlations are therefore used only as complementary information and did not inform the model development or the ranking of the experiments.

Overall, the correlations support the main conclusions drawn from the RMSE, bias, and score analyses. For automatic snow depth, correlations are generally high at the SnowMIP sites, with values mostly between 0.84 and 0.94 depending on the site and experiment. Lower correlations are found at the Arctic sites, ranging from about 0.49 to 0.78, particularly at Bylot. This is consistent with the stronger influence of local wind-driven snow redistribution in Arctic environments, which can cause point-scale snow depth variations that are not directly related to precipitation events and are not explicitly represented in the 1D simulations. At the SnowMIP sites, the new model developments do not substantially modify the snow depth correlation patterns compared to the DEF experiment, whereas slight improvements are found at some Arctic sites, especially Umiujaq and TVC, where correlations increase from about 0.73 in DEF to 0.77 and 0.78, respectively, after the implementation of the new compaction (COMPAC) and thermal conductivity schemes (CL11) at both sites.

The SWE correlations are also generally high, with values mostly above 0.8 for manual measurements and above 0.9 where automatic SWE measurements are available. Albedo correlations are more moderate, around 0.7 at most sites, with lower values at Sapporo, likely reflecting processes not represented in the model, such as light-absorbing particle deposition, as already discussed. Surface temperature correlations are high at most sites, generally exceeding 0.9, while soil temperature correlations vary more strongly among sites, from about 0.5 to 0.97. Lower soil temperature correlations are found at some sites where soil temperatures remain close to the zero-curtain or where small differences in the timing of snow accumulation, melt, or soil freezing can strongly affect temporal agreement (e.g., at Senator Beck, Sodankylä, and Weissfluhjoch). At TVC, the soil temperature correlation substantially improves from 0.76 in the DEF experiment to 0.97 in the CL11 and ALL_TVC experiments, consistent with the improved representation of snow insulation and soil thermal conditions."

Section 3.1: *It is difficult to determine the primary takeaways while reading through this section. It would be valuable to first present the overarching takeaways and then discuss the details of the individual metrics site by site.*

Thank you for this suggestion. We agree that the Results section is quite dense, making it difficult to identify the main takeaways. We have therefore added a new introductory subsection at the beginning of the Results section to provide an overview of the key findings and to guide readers through the subsequent detailed results:

"3.1 Overview of the main results

Overall, the default version of CLASSIC (DEF) reasonably reproduces the seasonal evolution of snow depth and most variables at the SnowMIP sites (see Sect. 3.2.1 for the site-specific details). The main biases occur at the Arctic sites, where simulated snow depth is overestimated by about 10–20 cm—a large relative bias given the thin Arctic snowpacks—and winter soil temperatures by about 5–10 °C (Fig. 3jj, oo, tt, nn, ss, and xx). This suggests that the simulated Arctic snowpack is too insulating, mainly because the default compaction scheme produces bulk snow densities that are too low for thin, wind-exposed Arctic snowpacks, resulting in snowpacks that are too deep for a given snow mass.

The main improvement for the Arctic sites is the revised snow compaction scheme (Sect. 2.4.2). By allowing thin, wind-exposed snowpacks to reach higher maximum dry-snow densities while having little effect on deeper snowpacks or under weak-wind conditions, this scheme produces simulated bulk snow densities more consistent with Arctic observations. At Bylot, for example, the snow depth bias decreases from about 7.5 cm to nearly zero, while simulated bulk snow density approaches the observed values of about 300 kg m^{-3} (further details are presented in Sect. 3.2.3). This denser and thinner snowpack reduces excessive soil insulation and improves simulated snow, bottom snow, and soil temperatures, with the mean temperature bias at Bylot decreasing from about 10 to 5 °C between the PHYS and COMPAC experiments (which is further enhanced in the CL11 experiment detailed below). Similar improvements are found at TVC, whereas the impact on most SnowMIP sites remains limited, as intended. Lastly, blowing-snow sublimation losses have a smaller impact than the revised compaction scheme but still affect some sites, as further discussed in Sect. 3.2.3

On the other hand, the physics improvements (Sect. 2.3) provide further physical consistency in the snowpack thermal regime and allow the Arctic developments to be built on a more robust basis. The original bottom snow temperature formulation in the DEF experiment tended to overestimate the simulated snow temperature and underestimate the bottom snow temperature, thereby inducing excessive heat storage within the snowpack and biasing temperature gradients at the snow–atmosphere and snow–soil interfaces (see Sect. 3.2.2 for further details). The revised bottom snow temperature formulation addresses these deficiencies. For example, at Col de Porte, the vertically-averaged snow temperature bias decreases from 0.8 to 0.3 °C and the bottom

snow temperature bias from -0.4 to -0.1 °C (e.g., Fig. 4g–j). These new developments do not necessarily improve the performance of all variables and can even degrade simulated soil temperatures when applied alone. However, they are essential for representing the snowpack thermal regime more consistently and for avoiding bias compensation.

The use of the snow thermal conductivity parameterization of Calonne et al. (2011) (Sect. 2.4.3), in combination with the Arctic developments (CL11), further improves simulated snow, bottom snow, and soil temperatures. This suggests that the previous snow thermal conductivity parameterization partly compensated for the bottom snow temperature bias. For example, at Bylot, the winter soil temperature bias increases from 6.81 °C in the DEF experiment to 9.62 °C in the PHYS experiment before decreasing to 0.89 °C in the CL11 experiment (Fig. 5m).

Finally, the addition of the windless exchange coefficient in the formulation of turbulent fluxes (Sect. 2.3.3) has a more targeted effect. It substantially improves simulated snow surface temperatures at the SnowMIP sites, reducing the average RMSE from 3.41 to 1.51 °C and the mean bias from -2.20 to -0.25 °C (Figs. 6k and 7k), while having limited effects on snow depth, SWE, and soil temperatures. More detailed results are presented below."

It could also be valuable to show the results from Figure 3 as a multisite average in addition to the individual site comparisons.

Thank you for the suggestion. We agree that a multisite-average version of Fig. 3 could provide a more synthetic view of the results and would be visually lighter than the full site-by-site figure. However, averaging across sites can mask important site-specific behaviors and compensate for biases between sites. In addition, our Results section is already quite dense, so we do not want to expand it. We hope the above overview of the main results can already help clarify the main takeaways. We also note that multisite averages are already partly provided in the quantitative metric figures (Figs. 6 and 7), where mean values are shown for the SnowMIP sites, the Arctic sites, and all sites together. Although these averages provide more condensed information than seasonal-cycle plots, they already offer a synthetic overview of the model performance across sites.

Why do the simulations show significant snow depth at SNB in July when SWE is near zero?

Thanks for pointing this out. This was actually due to the abrupt end of the observed snow depth data at SNB and SWA when the snow depth approaches zero (see figure below). Therefore, the annual mean day of the year was artificially higher near the meltout period because it excluded the zero values. This also affected simulated values, as they were masked based on observation availability to compare the model and observations over common periods. Therefore, to avoid such a visual artifact in the annual cycle figures, we extrapolated the observed snow depth values at SNB and SWA to zero during the summer and performed linear interpolation for the remaining short periods of missing data. This is done only for annual cycle visualization, not for computing the metrics (which are computed directly on the full daily time series), to avoid including extrapolated information. A similar issue is observed at SOD for snow depth and SWE, but the visual impact is less pronounced; therefore, we did not apply that procedure at that site. We applied this correction to all annual cycle figures and added to the legend the following description:

"At snb and swa only, automatic snow depth observations were extended to zero after melt-out when the observational time series stopped abruptly at zero or near-zero snow depth, while remaining gaps were linearly interpolated; this adjustment is applied only for the annual cycle visualization and not for the computation of the evaluation metrics."

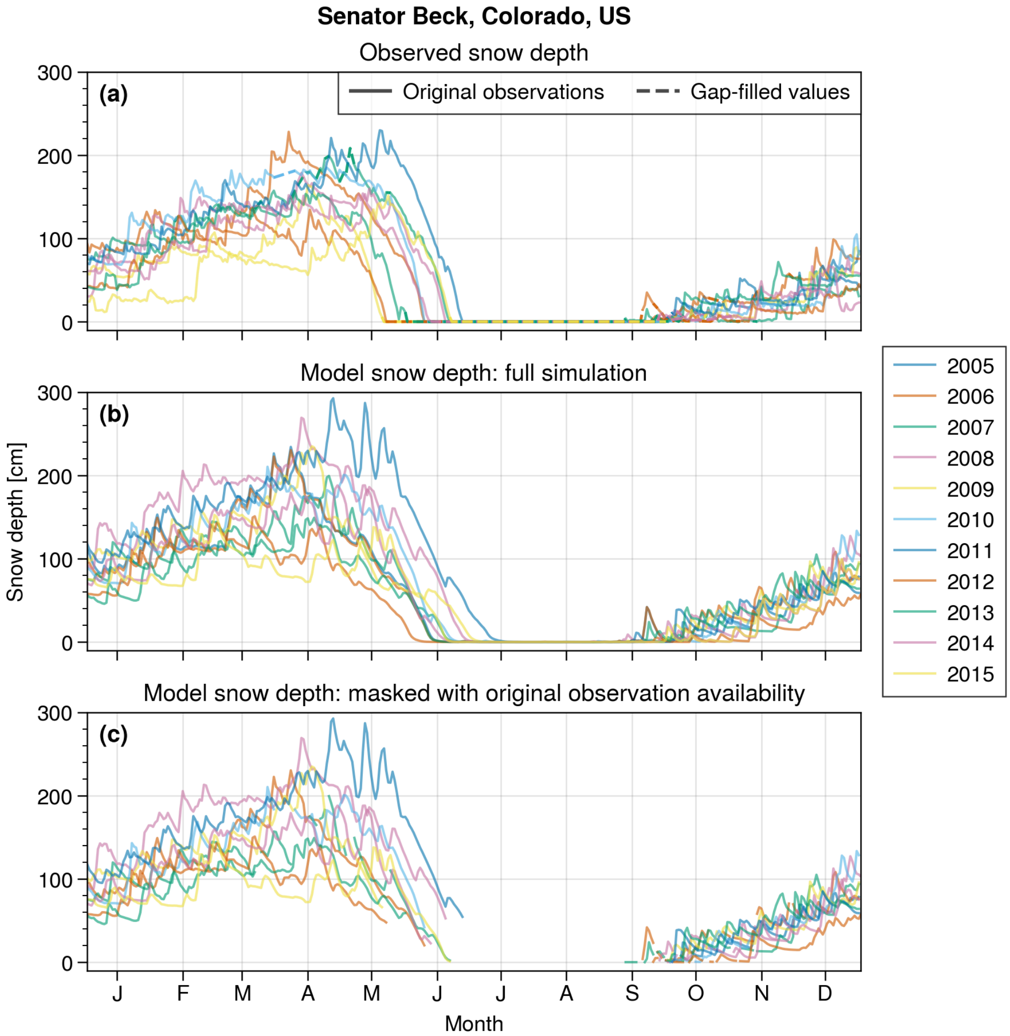

Illustration of the snow depth data treatment at Senator Beck (SNB). (a) Original observed daily snow depth for each year (solid lines), with values added by the gap-filling procedure shown as dashed lines. Long summer gaps (>50 days) were set to zero, while shorter gaps were linearly interpolated. **(b)** Corresponding full simulated snow depth time series. **(c)** Simulated snow depth masked using the availability of the original observations. The comparison between (b) and (c) illustrates how the absence of observed zero values after melt-out can truncate the simulated series and artificially increase the multi-year mean snow depth near the end of the annual cycle. Gap filling was applied only for visualization of the annual cycle and not for the computation of evaluation metrics.

When discussing skill qualitatively (e.g., "good"), please support these statements with skill metrics (RMSE and correlation).

Thank you for the comment. We have revised the Results section to support qualitative statements on model performance with quantitative metrics, including RMSE and correlation, where relevant. For example:

> *"Generally, the default model version (DEF; dark blue line) demonstrates good skills in representing the annual cycle of the snow depth (first column; **mean RMSE of 25 cm and correlation of 0.90 across all SnowMIP sites**) and SWE (second column; **mean RMSE of 40 mm and correlation of 0.95 across all SnowMIP sites**) at most of the SnowMIP sites compared to the observations (black line). However, the DEF experiment overestimates and underestimates the snow depth at the Senator Beck and Swamp Angel sites **with mean biases of 31 and -38 cm**, respectively (panels k and p)."*

Similar metric information has been added throughout the Results section, where qualitative assessments of model skill are discussed. These changes are visible in the revised manuscript.

Please note whether differences in evaluation metrics between simulations are statistically significant.

We agree that statistical tests would be valuable. However, considering that the evaluation metrics combine variables, sites, and periods with different observational uncertainties, forcing uncertainties, and degrees of temporal autocorrelation, it makes that exercise quite delicate—knowing that most uncertainties are actually unknown and/or not characterized. That's why we indicated the Col de Porte ones on some plots as an indicative spread, but it is certainly underestimated at Arctic sites. Bootstrap or a leave-one-site-out method could be an avenue, but that might still not be straightforward: temporal autocorrelation, limited number of years for Arctic sites, non-independent metrics, variable weighting, etc. In addition, the aggregated scores are intended as a diagnostic synthesis of model performance rather than as a statistical estimator with a straightforward uncertainty distribution, even though they are used in the model-tuning process as a practical means. Therefore, the small differences in evaluation metrics should not be overinterpreted, while we placed more emphasis on robust changes that are consistent across several physically related variables, such as simultaneous improvements in snow depth, snow density, snow thermal properties, and soil temperature (see further additions in our manuscript concerning the comments below). We also tempered some affirmations accordingly; for example, we replaced this sentence:

> *"Excluding SCD from the score computation yields scores of 0.311 and 0.317 for the CL11 and ALL_TVC experiments, respectively, at the Arctic sites, and 0.421 and 0.423 when considering all sites. These results highlight the improved performance of the TVC-adjusted compaction parameters for simulating snow depth and soil temperatures."*

by:

> *"Excluding SCD from the score computation yields **comparable scores** for the CL11 and ALL_TVC experiments, with values of 0.311 and 0.317, respectively, at the Arctic sites, and 0.421 and 0.423 when considering all sites. **These small differences should not be overinterpreted as a statistically significant ranking. Rather, they indicate that the two final configurations provide comparable overall performance, with slightly different trade-offs among variables and sites. In particular, the TVC-adjusted compaction parameters (ALL_TVC) improve soil temperatures at TVC but slightly increase warm soil temperature biases at Bylot and Umiujaq, as discussed above. The scores should therefore***

be interpreted as a diagnostic guide, while the robustness of the model developments is assessed primarily from the physical consistency of the changes across snow depth, snow density, snow thermal properties, and soil temperature, as further discussed in Sect. 4.1.*"*

Please discuss the coherence and relationships between changes in skill scores across variables. For example, are the default snow depth and SWE biases correlated? Are changes in the biases of these variables correlated? Do reduced biases in albedo correspond with reduced biases in surface temperature and soil temperature? Are temperature and albedo biases physically consistent with SWE and snow depth biases?

Thanks for pointing this out. We added a whole new section in the discussion that covers those aspects and helps clarify the links between the new developments: "***4.1 Physical consistency of model improvements across variables***", which is available in the updated manuscript.

There is also value in evaluating whether interannual variability is improved in key metrics (e.g., correlation of simulated vs. observed peak SWE, day of disappearance, spring ablation rates, mean annual SWE, etc.).

We agree that interannual diagnostics such as peak SWE, snow disappearance date, spring ablation rate, or mean annual SWE can provide valuable additional information on model performance. However, we did not include these metrics in the present study for several reasons. First, the number of available snow seasons is limited at several sites, particularly at the Arctic sites that are the main focus of this work, which makes interannual correlations for quantities such as peak SWE statistically fragile. Second, some of these metrics can be difficult to interpret at the point scale. For example, the snow disappearance date can differ by several days depending on whether a point snow depth sensor is located on a persistent snow patch or in a nearby snow-free area, especially during patchy melt conditions and when snow depth falls below 10 cm, where vegetation protrusion and local surface heterogeneity become increasingly important. At Bylot, for instance, a board was installed below the snow depth sensor to flatten the vegetation, but this setup may also affect the local surface energy balance and promote earlier melt than in the surrounding area because of imperfect coupling with the soil. Therefore, such point-scale melt-out diagnostics may partly reflect local measurement conditions rather than model performance alone. We have therefore chosen not to further expand the evaluation framework, which already includes several variables, sites, experiments, and complementary metrics. We nevertheless agree that these interannual diagnostics are relevant and could be useful in future studies, especially with longer Arctic time series and/or spatially distributed snow measurements that better capture local snow cover variability.

Section 4.1: *How do errors between the LSM and site observations compare with those reported in previous studies (e.g., Lin et al., 2025; Abolafia-Rosenzweig et al., 2022,2024; Chen et al., 2014; etc.)?*

https://doi.org/10.1175/JHM-D-24-0082.1; https://doi.org/10.1029/2022MS003141; https://doi.org/10.1029/2023MS003869; https://doi.org/10.1002/2014JD022167

Thank you for this suggestion. We agree that these studies provide useful context for interpreting snow model errors in land surface models. However, we also note that a direct quantitative comparison of model errors is not straightforward, because most of these studies focus on different domains, model structures, and target processes, including multilayer Noah-MP simulations over mountainous western U.S. sites, often with different forcing constraints, evaluation metrics, and observational networks. In contrast, our study focuses on a single-layer version of CLASSIC evaluated at SnowMIP and Arctic sites, with particular emphasis on the physical consistency among snow depth, SWE, snow density, snow thermal properties, and soil temperature. We therefore primarily discuss our results in relation to previous studies focused on Arctic snow model development and evaluation (e.g., Barrere et al., 2017; Gouttevin et al., 2018; Royer et al., 2021; Lackner et al., 2022; Woolley et al., 2024). Nevertheless, we have also added the studies suggested by the reviewer at specific points in the Discussion where they are directly relevant:

> *"These improvements related to the compaction scheme did not substantially impact the SWE, which is driven by other processes such as precipitation forcing, sublimation losses, and snow redistribution.* ***Indeed, snow depth and density errors can persist even when SWE biases are reduced, highlighting the importance of evaluating snow compaction processes and complementary snowpack variables in addition to SWE alone (e.g., Abolafia-Rosenzweig et al., 2024)****"*

> *"These developments also illustrate why the score alone is insufficient for interpreting model performance****, because compensating errors can mask process-level deficiencies (e.g., Chen et al., 2014).****"*

> "*Further investigation will be needed by including LAPs and/or adapting the albedo scheme to distinguish the contribution of these respective phenomena to those discrepancies, especially since early melt is also observed at other sites (discussed in other sections below).* ***Future developments could therefore target a more explicit representation of snow albedo evolution, including snow aging, grain-size effects, and light-absorbing particles, following both the theoretical basis of snow radiative transfer (e.g., Warren & Wiscombe, 1980) and recent land-surface-model applications (e.g., Ménégoz et al., 2014; Abolafia-Rosenzweig et al., 2022; Lin et al., 2025).***"

In the context of combined uncertainties from point-scale comparisons, forcing uncertainties, and observational uncertainties, are the changes in model skill significant?
We hope that the addition of Sect. 4.1 (described above) in the discussion will help to clarify this point. In addition, we added a sentence in the conclusion to support this idea:

> *"[...] In particular, we introduce a novel wind-dependent compaction scheme (Sect. 2.4.2) that enables the simulated snowpack to perform as well at Arctic sites as at mid-latitude sites.* ***Together with the revised bottom snow temperature formulation and the updated snow thermal conductivity parameterization, this new development provides physically consistent model responses across snow depth, snow density, snow thermal properties, and soil temperature, supporting the robustness of the proposed developments despite remaining forcing, observational, and site-specific uncertainties.*** *[...]."*

Technical correction:

Remove the comma on line 431.
Done.

Appendix

Correction of normalized RMSE and mean-bias plotting

A consistency check of the plotting scripts (https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_ALL_correct_SH.ipynb and https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_PHYS_correct_SH.ipynb) revealed that the normalized RMSE and mean bias values shown for individual sites in the snow depth and SWE panels were not computed with the observed standard deviation but instead with the standard deviation of the simulated values for each experiment. For example, the previous code for the normalized mean bias of automatic snow depth was:

```
axs[1].scatter(
  df_metrics_list[i].loc[('snd_auto')].mb[1:] /df_metrics_list[i].loc[('snd_auto')].stdd[1:],
  marker=markers[i], label=site.name, color=colors[i]
)
```

This has been replaced by the precomputed normalized metric (https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_ALL_correct_SH_v5.ipynb):

```
axs[1].scatter(
  df_metrics_list[i].loc[('snd_auto')].nmb[1:],
  marker=markers[i], label=site.name, color=colors[i]
)
```

Normalized metrics (*nmb* and *nrmse*) were correctly computed with the observed standard deviation (*std*):

```
df_metrics_list[i].loc[(var, exp), 'nmb'] = mb(simu, obs)/obs.std().values.item(0)
```

The same correction was applied to the snow depth and SWE variables, for both automatic and manual measurements (mean bias and RMSE) in Figs. 6, 7, C2, and C3 (panels a–h). The site-mean markers for SnowMIP and Arctic sites were already computed using the correct *nmb* and *nrmse* values, for example:

```
df_gather.loc[('snd_auto')].nmb[1:].apply(np.mean)
df_gather.loc[('snd_auto')].nrmse[1:].apply(np.mean)
```

Therefore, the overall interpretation and the scores are not affected, since the scores were also computed from the correct observation-based normalization. The main visible change concerns the consistency between individual-site markers and the corresponding SnowMIP/Arctic mean markers. The effect is generally small, but is more noticeable at TVC, where the model standard deviation was larger than the observed variability, which previously led to artificially lower normalized RMSE and mean-bias values.

The updated files and figures are:
https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_ALL_correct_SH_v5.ipynb:

- Fig. 6: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_rmse_all_v5_correct_SH.pdf
- Fig. 7: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_mb_all_v5_correct_SH.pdf

https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_PHYS_correct_SH_v5.ipynb:

- Fig. C2: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_rmse_phys_v5_correct_SH.pdf
- Fig. C3: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_mb_phys_v5_correct_SH.pdf

For review traceability, the previous files containing the plotting inconsistency have been retained:
https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_ALL_correct_SH.ipynb:

- Fig. 6: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_rmse_all_v1_correct_SH.pdf
- Fig. 7: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_mb_all_v1_correct_SH.pdf

https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/paper_ac_rank_PHYS_correct_SH.ipynb:

- Fig. C2: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_rmse_phys_v1_correct_SH.pdf
- Fig. C3: https://github.com/mickaellalande/SnowC2-CLASSIC-1D-analysis/blob/main/CLASSIC-1D/img/paper_mb_phys_v1_correct_SH.pdf

References

Abolafia-Rosenzweig, R., He, C., Chen, F., & Barlage, M. (2024). Evaluating and Enhancing Snow Compaction Process in the Noah-MP Land Surface Model. *Journal of Advances in Modeling Earth Systems*, *16*(2), e2023MS003869. https://doi.org/10.1029/2023MS003869

Abolafia-Rosenzweig, R., He, C., McKenzie Skiles, S., Chen, F., & Gochis, D. (2022). Evaluation and Optimization of Snow Albedo Scheme in Noah-MP Land Surface Model Using In Situ Spectral Observations in the Colorado Rockies. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2022MS003141. https://doi.org/10.1029/2022MS003141

Barrere, M., Domine, F., Decharme, B., Morin, S., Vionnet, V., & Lafaysse, M. (2017). Evaluating the performance of coupled snow–soil models in SURFEXv8 to simulate the permafrost thermal regime at a high Arctic site. *Geoscientific Model Development*, *10*(9), 3461–3479. https://doi.org/10.5194/gmd-10-3461-2017

Bayle, A., Roy, A., Dedieu, J.-P., Boudreau, S., Choler, P., & Lévesque, E. (2022). Two distinct waves of greening in northeastern Canada: Summer warming does not tell the whole story. *Environmental Research Letters*, *17*(6), 064051. https://doi.org/10.1088/1748-9326/ac74d6

Bokhorst, S., Pedersen, S. H., Brucker, L., Anisimov, O., Bjerke, J. W., Brown, R. D., Ehrich, D., Essery, R. L. H., Heilig, A., Ingvander, S., Johansson, C., Johansson, M., Jónsdóttir, I. S., Inga, N., Luojus, K., Macelloni, G., Mariash, H., McLennan, D., Rosqvist, G. N., … Callaghan, T. V. (2016). Changing Arctic snow cover: A review of recent developments and assessment of future needs for observations, modelling, and impacts. *Ambio*, *45*(5), 516–537. https://doi.org/10.1007/s13280-016-0770-0

Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mård, J., Mernild, S. H., Prowse, T., Semenova, O., Stuefer, S. L., & Woo, M. -K. (2016). Arctic terrestrial hydrology: A synthesis of processes, regional effects, and research challenges. *Journal of Geophysical Research: Biogeosciences*, *121*(3), 621–649. https://doi.org/10.1002/2015JG003131

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G.-Y., Wen, L., & Yang, Z.-L. (2014). Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study. *Journal of Geophysical Research: Atmospheres*, *119*(24), 13,795-13,819. https://doi.org/10.1002/2014JD022167

Dai, A. (2008). Temperature and pressure dependence of the rain-snow phase transition over land and ocean. *Geophysical Research Letters*, *35*(12). https://doi.org/10.1029/2008GL033295

Domine, F., Lackner, G., Sarrazin, D., Poirier, M., & Belke-Brea, M. (2021). Meteorological, snow and soil data (2013–2019) from a herb tundra permafrost site at Bylot Island, Canadian high Arctic, for driving and testing snow and land surface models. *Earth System Science Data*, *13*(9), 4331–4348. https://doi.org/10.5194/essd-13-4331-2021

Domine, F., Sarrazin, D., Nadeau, D. F., Lackner, G., & Belke-Brea, M. (2024). Meteorological, snow and soil data, CO_2, water and energy fluxes from a low-Arctic valley of Northern Quebec. *Earth System Science Data*, *16*(3), 1523–1541. https://doi.org/10.5194/essd-16-1523-2024

Dutch, V. R., Rutter, N., Wake, L., Sandells, M., Derksen, C., Walker, B., Hould Gosselin, G., Sonnentag, O., Essery, R., Kelly, R., Marsh, P., King, J., & Boike, J. (2022). Impact of measured and simulated tundra snowpack properties on heat transfer. *The Cryosphere*, *16*(10), 4201–4222. https://doi.org/10.5194/tc-16-4201-2022

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

Gouttevin, I., Langer, M., Löwe, H., Boike, J., Proksch, M., & Schneebeli, M. (2018). Observation and modelling of snow at a polygonal tundra permafrost site: Spatial variability and thermal implications. *The Cryosphere*, *12*(11), 3693–3717. https://doi.org/10.5194/tc-12-3693-2018

Harder, P., & Pomeroy, J. W. (2014). Hydrological model uncertainty due to precipitation-phase partitioning methods. *Hydrological Processes*, *28*(14), 4311–4327. https://doi.org/10.1002/hyp.10214

Jennings, K. S., & Molotch, N. P. (2019). The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient. *Hydrology and Earth System Sciences*, *23*(9), 3765–3786. https://doi.org/10.5194/hess-23-3765-2019

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018). Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1–9. https://doi.org/10.1038/s41467-018-03629-7

Lackner, G., Domine, F., Nadeau, D. F., Lafaysse, M., & Dumont, M. (2022). Snow properties at the forest–tundra ecotone: Predominance of water vapor fluxes even in deep, moderately cold snowpacks. *The Cryosphere*, *16*(8), 3357–3373. https://doi.org/10.5194/tc-16-3357-2022

Lin, T.-S., He, C., Abolafia-Rosenzweig, R., Chen, F., Wang, W., Barlage, M., & Gochis, D. (2025). *Improved Snow Albedo Evolution in Noah-MP Land Surface Model Coupled with a Physical Snowpack Radiative Transfer Scheme*. *26*(2), 185–200. https://doi.org/10.1175/JHM-D-24-0082.1

Mavrovic, A., Sonnentag, O., Voigt, C., Lemmetyinen, J., Aurela, M., & Roy, A. (2025). Winter Methane Fluxes Over Boreal and Arctic Environments. *Geophysical Research Letters*, *52*(21), e2025GL118367. https://doi.org/10.1029/2025GL118367

Melton, J. R., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., & Teckentrup, L. (2020). CLASSIC v1.0: The open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 1: Model framework and site-level performance. *Geoscientific Model Development*, *13*(6), 2825–2850. https://doi.org/10.5194/gmd-13-2825-2020

Ménard, C. B., Essery, R., Barr, A., Bartlett, P., Derry, J., Dumont, M., Fierz, C., Kim, H., Kontu, A., Lejeune, Y., Marks, D., Niwano, M., Raleigh, M., Wang, L., & Wever, N. (2019). Meteorological and evaluation datasets for snow modelling at 10 reference sites: Description of in situ and bias-corrected reanalysis data. *Earth System Science Data*, *11*(2), 865–880. https://doi.org/10.5194/essd-11-865-2019

Ménégoz, M., Krinner, G., Balkanski, Y., Boucher, O., Cozic, A., Lim, S., Ginot, P., Laj, P., Gallée, H., Wagnon, P., Marinoni, A., & Jacobi, H. W. (2014). Snow cover sensitivity to black carbon deposition in the Himalayas: From atmospheric and ice core measurements to regional climate simulations. *Atmospheric Chemistry and Physics*, *14*(8), 4237–4249. https://doi.org/10.5194/acp-14-4237-2014

Myers-Smith, I. H., Forbes, B. C., Wilmking, M., Hallinger, M., Lantz, T., Blok, D., Tape, K. D., Macias-Fauria, M., Sass-Klaassen, U., Lévesque, E., Boudreau, S., Ropars, P., Hermanutz, L., Trant, A., Collier, L. S., Weijers, S., Rozema, J., Rayback, S. A., Schmidt, N. M., … Hik, D. S. (2011). Shrub expansion in tundra ecosystems: Dynamics, impacts and research priorities. *Environmental Research Letters*, *6*(4), 045509. https://doi.org/10.1088/1748-9326/6/4/045509

Pan, X., Yang, D., Li, Y., Barr, A., Helgason, W., Hayashi, M., Marsh, P., Pomeroy, J., & Janowicz, R. J. (2016). Bias corrections of precipitation measurements across experimental sites in different ecoclimatic regions of western Canada. *The Cryosphere*, *10*(5), 2347–2360. https://doi.org/10.5194/tc-10-2347-2016

Picard, G., Arnaud, L., Panel, J.-M., & Morin, S. (2016). Design of a scanning laser meter for monitoring the spatio-temporal evolution of snow depth and its application in the Alps and in Antarctica. *The Cryosphere*, *10*(4), 1495–1511. https://doi.org/10.5194/tc-10-1495-2016

Royer, A., Picard, G., Vargel, C., Langlois, A., Gouttevin, I., & Dumont, M. (2021). Improved Simulation of Arctic Circumpolar Land Area Snow Properties and Soil Temperatures. *Frontiers in Earth Science*, *9*(June), 1–19. https://doi.org/10.3389/feart.2021.685140

Saulnier-Talbot, É., Duchesne, É., Antoniades, D., Arseneault, D., Barnard, C., Berteaux, D., Bhiry, N., Bouchard, F., Boudreau, S., Cazelles, K., Comte, J., Corbeil-Robitaille, M.-Z., Côté, S. D., Couture, R.-M., De Lafontaine, G., Domine, F., Fauteux, D., Fortier, D., Garneau, M., … Bêty, J. (2024). Expert elicitation of state shifts and divergent sensitivities to climate warming across northern ecosystems. *Communications Earth & Environment*, *5*(1), 624. https://doi.org/10.1038/s43247-024-01791-z

Sturm, M., Holmgren, J., McFadden, J. P., Liston, G. E., Chapin, F. S., & Racine, C. H. (2001). Snow–Shrub Interactions in Arctic Tundra: A Hypothesis with Climatic Implications. *Journal of Climate*, *14*(3), 336–344. https://doi.org/10.1175/1520-0442(2001)014%3C0336:SSIIAT%3E2.0.CO;2

Warren, S. G., & Wiscombe, W. J. (1980). A Model for the Spectral Albedo of Snow. II: Snow Containing Atmospheric Aerosols. *Journal of the Atmospheric Sciences*, *37*(12), 2734–2745. https://doi.org/10.1175/1520-0469(1980)037%3C2734:AMFTSA%3E2.0.CO;2

Woolley, G. J., Rutter, N., Wake, L., Vionnet, V., Derksen, C., Essery, R., Marsh, P., Tutton, R., Walker, B., Lafaysse, M., & Pritchard, D. (2024). Multi-physics ensemble modelling of Arctic tundra snowpack properties. *The Cryosphere*, *18*(12), 5685–5711. https://doi.org/10.5194/tc-18-5685-2024