Response to Reviewer#1 Comments

We sincerely thank the reviewers for their thorough and insightful evaluation of our manuscript. Their detailed comments and constructive suggestions have been invaluable in improving the quality of this work. We are grateful for the considerable time and expertise they have dedicated to this review. In the following, we address each comment individually. The reviewer's comments are in black, and our responses are provided in blue. Reviewers are kindly requested to refer to the newly revised/track-changed manuscript for the modifications mentioned in the responses.

Comment 1:

To me, the spin up configuration has flaws. The initial condition of soil, including the thermal and hydrological state of soil, especially for the seasonally frozen soil with water phase change in it, are critically important. According to the manuscript, the authors use a historical simulation from 1850 to 1940 forced by GSWP3, which is a different forcing dataset from ERA5. The spin up is basically a transient simulation, so that to me there is no way to judge if the spin up is completed. Authors should provide additional information on whether the thermal state of soil and atmosphere has reached the equilibrium by the end of the spin up, otherwise the simulated soil temperature, as well as the snow accumulation manners, are not comparable to the observation/ERA-land. Furthermore, the author mentioned that the JSBACH simulation is run without the two-phase spin up like the CLM5 simulation did. I am wondering how the initial conditions used by JSBACH and CLM5 differ from each other. The difference in initial condition could make the soil state totally different. Additionally, to my knowledge the GSWP3 data covers the period of 1901-2014 (`https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/atm/datm7/atm_forcing.datm7.GSWP3.0.5d.v1.c170516/TPHWL/`). So how exactly do authors use this for the spin up from 1950 to 1899?

Response:

We sincerely thank the reviewer for this careful and important comment regarding the spin up strategy and initialization procedure. We fully acknowledge that the description in the original manuscript was not sufficiently detailed and may have created confusion regarding the model equilibration procedure. We will substantially revise the manuscript to clarify these aspects based on the following information

- **CLM and JSBACH spin up method:**

  We agree with the reviewer that demonstrating thermal and hydrological equilibrium at the end of spin up is particularly important for seasonally frozen soils, where freeze-thaw processes are sensitive to initial land states. We clarify that CLM5.1 was initialized following the standard two-phase spin up procedure described in Lawrence et al. (2019); Koven et al. (2013), and implemented in the CMCC seasonal prediction system (Gualdi et al., 2020). In the first phase, the model was brought to equilibrium representative of year-1850 conditions using the standard CLM spin up protocol. This phase consisted of 400 years with accelerated decomposition to equilibrate slow carbon and nitrogen pools, followed by an additional 800 years in normal mode. The second phase was a historical run from 1850, which was achieved by integrating over a repeating 20-year period of the GSWP3 forcing dataset (1901-1920) (Lawrence et al., 2019), while prescribing fixed preindustrial boundary conditions, including constant atmospheric CO_2, nitrogen deposition, aerosol deposition, land use, and no wood harvest. This is the standard CLM approach for generating a stable preindustrial land state when meteorological forcing prior to 1901 is unavailable. From 1901 onward, the model was forced with true time-varying GSWP3 data, which in our setup covers the period 1901-1940. To ensure a smooth transition from GSWP3 to ERA5 forcing, we additionally performed 10 cyclical runs using climatological ERA5 forcing for 1941-1950, which reduced potential drift caused by the forcing dataset change.

In contrast, JSBACH was run continuously from 1941 to 2022 without an equivalent multi phase spin up, initialized from JSBACH3 initial-condition files remapped from the Gaussian grid to the ICON R02B06 grid, and forced with ERA5 reanalysis data at a 3-hourly temporal resolution. Previous JSBACH site-level study has considered a spin up period of 30 years sufficient to reach thermal and hydrological equilibrium in frozen ground simulations (Ekici et al., 2014). Since our analysis period begins in 1986, JSBACH had been integrated for 45 years (1941-1985) prior to evaluation, exceeding this threshold and ensuring that the influence of the initial conditions on the analysed variables is negligible. As dynamic nitrogen cycling is not included in either CLM or JSBACH, slowly evolving nitrogen pools do not impose additional spin up requirements.

- **Evidence of equilibrium - supplementary figures:**

  To directly address the concern regarding whether equilibrium was reached, we will add the following figures in the supplementary material showing the temporal evolution of annual mean soil temperature and soil moisture at the deepest soil layer, and annual mean snow depth over a selected region (Figure 1) for the full simulation period 1941-2022 (shown in Figure 2). Figure 2 demonstrate that both CLM and JSBACH achieve a stable equilibrium state well before the analysis period begins in 1986.

  As shown in Figure 2, CLM5 starts the transient simulation in an already stable state, reflecting the effectiveness of the two-phase spin up procedure. For JSBACH, which was initialized directly in 1941 without a comparable multi-phase spin up, an initial adjustment period of approximately 10-15 years is visible at the beginning of the simulation. After this period, the simulated soil temperature and snow depth primarily show interannual variability and longer-term trends related to climate warming.

  We agree that the initial conditions of the two models were different owing to their distinct model structures and initialization frameworks. However, since the analysis period starts in 1986, both models had already been run for several decades before evaluation (45 years from 1941 to 1985). Therefore, the direct influence of the initial conditions on the analysed frozen-ground variables is expected to be small during the evaluation period, as soil thermal and hydrological memory typically dissipates over such timescales (Yang and Zhang, 2016; Koster and Suarez, 2001; Vinnikov et al., 1996). But since it is important for the readers to know it so we are planning to acknowledge that "Differences in spatial resolution, grid-cell heterogeneity, initial state and boundary condition datasets describing soil properties likely contribute to some of the simulated biases." And we have clarified this point in the revised manuscript (Line 344).

Lines (149-164) has been updated in the revised manuscript to give more clarity on the spin up procedure as follows:

For CLM, the BGC-CROP biogeochemistry crop model was activated (not used for JSBACH), and the model was initialized using a standard two-phase spin up procedure (Koven et al., 2013; Lawrence

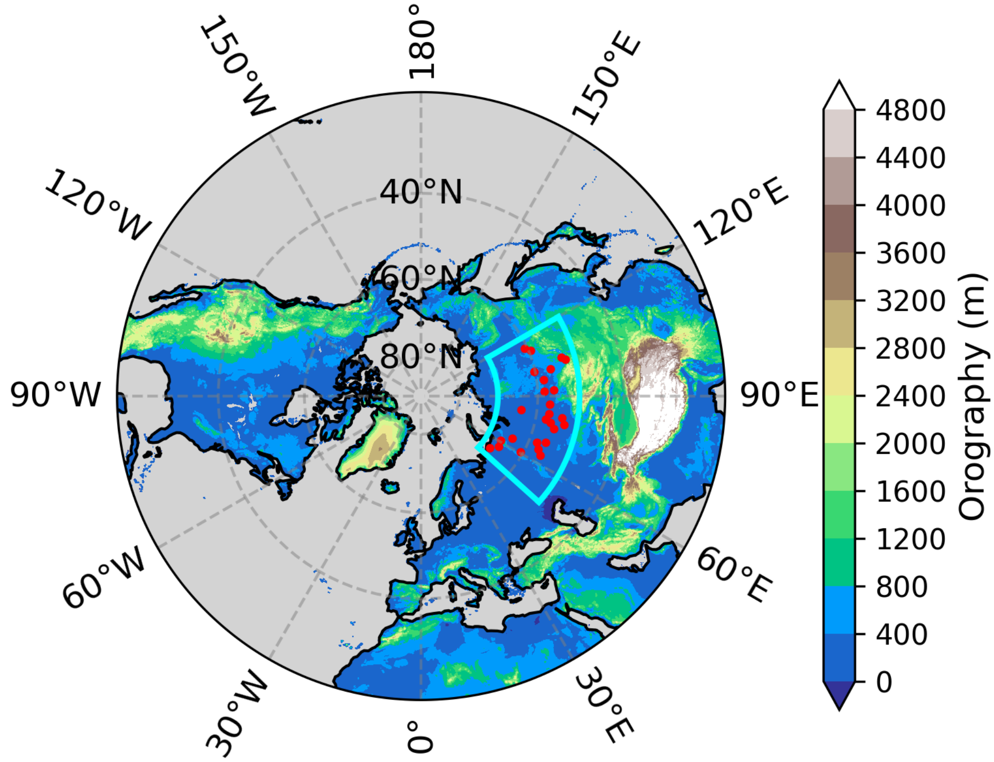


Figure 1: Selected study region for the equilibrium analysis shown by the cyan color box on the orography map.

et al., 2019). First, a preindustrial spin up was performed with 400 years of accelerated decomposition followed by 800 years in normal mode to equilibrate slow carbon and nitrogen pools. Equilibrium conditions for the year 1850 were obtained by integrating over a repeating 20-year cycle (1901-1920) of Global Soil Wetness Project (GSWP3) meteorological forcing under fixed preindustrial boundary conditions. From 1901 onwards, the model was forced with GSWP3 data until 1940, followed by 10 cyclical runs using climatological ERA5 forcing for 1941-1950 to ensure a smooth transition between forcing datasets, with the final spin up state serving as the initial condition for the transient simulation from 1941 to 2022 using 3-hourly ERA5 data. In contrast, JSBACH was run continuously from 1941 to 2022 without an equivalent two-phase spin up. The analysis for both models focused on 1986-2022 to match the availability of observational datasets. Supplementary Figures S1-S2 show the temporal evolution of deep soil temperature, soil moisture and snow depth, confirming that both models reached stable states before the analysis period.

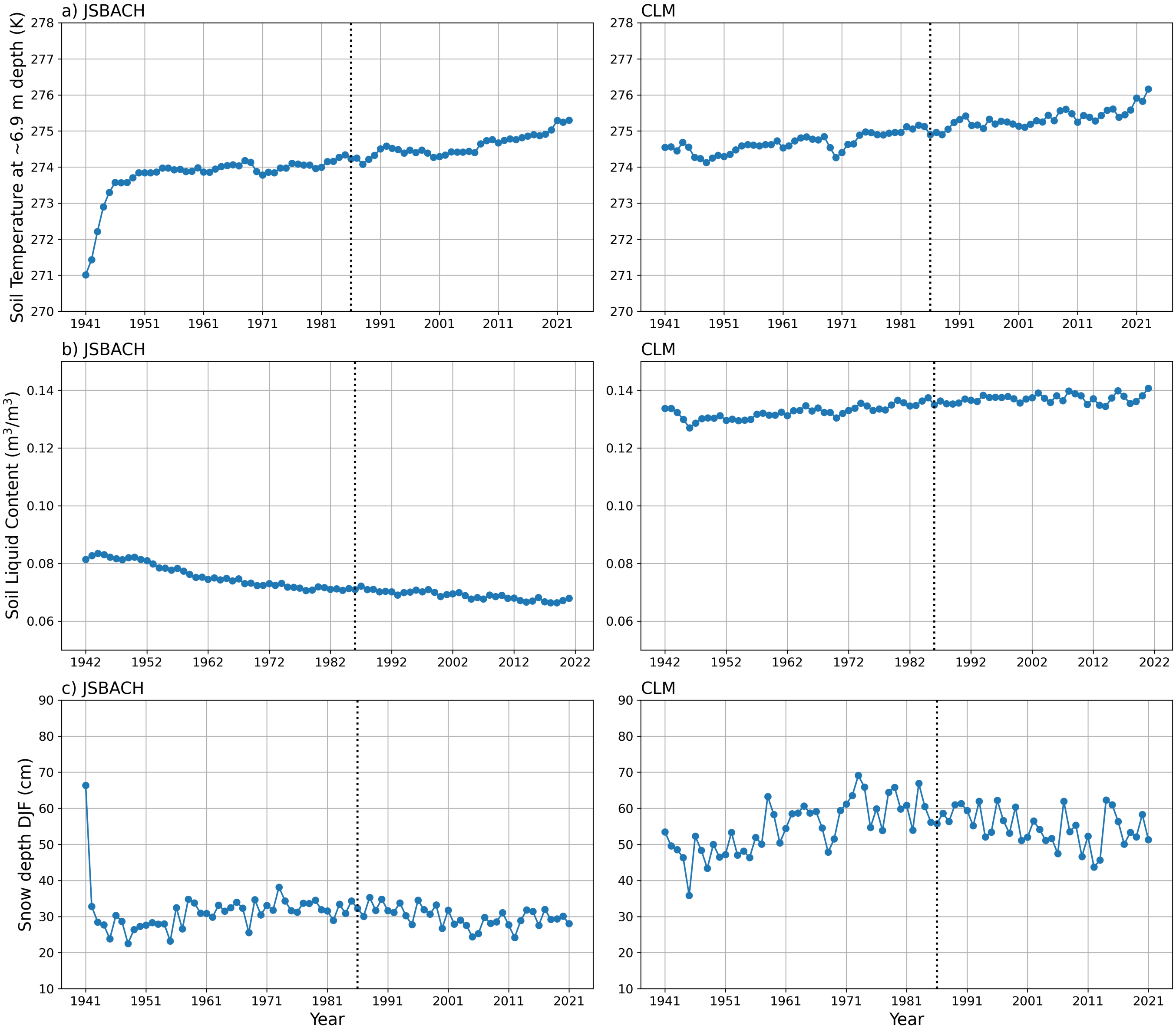


Figure 2: Time series of annual mean (a) soil temperature at 6.9 m depth, (b) volumetric soil liquid content for the whole soil column (m^3/m^3), and (c) snow depth over the regions shown in Figure S1, simulated by JSBACH (first column) and CLM5 (second column) for the period 1941-2022. The dashed vertical line indicates the start of the analysis period (1986).

Comment 2:

According to Table 1, the JSBACH model only has 5 soil layers. To me it is a little bit too few to present the soil temperature profile in permafrost area. Authors should consider add the information of soil layer structure (the depth and thickness of each soil layer) of the two models in the manuscript. Also, the CLM model offers a detailed soil layer structure tuned for permafrost modeling (49 soil layers for 0-9.85 m of soil and 5 bedrock layers). By the way, the definition of permafrost is deviated from that by the International Permafrost Association (the soil with a temperature colder than 0 °C (32 °F) continuously for two or more years). Authors should explain and verify the validity of this definition and how this definition leads to frozen ground extent in the JSBACH model.

Response:

We thank the reviewer for these two distinct and important points. We address them separately below.

We agree that JSBACH uses a relatively coarse vertical discretization (5 soil layers), especially when compared to CLM, which includes a much finer vertical structure (20 soil layers in the soil column extending to 8.6 m depth, plus 5 bedrock layers until about 50 m). Both models were chosen because they are part of Earth System Models (ESMs) that are used for climate projections and climate predictions. As such, their ability to realistically represent seasonally frozen soils is of considerable interest. This comparison provides insight into their respective strengths and limitations, particularly regarding whether a coarser soil vertical discretization can reproduce similar large scale frozen ground patterns as a higher resolution model. Importantly, the aim of this study is not to benchmark one model against the other in terms of structural complexity, but rather to evaluate how two models with fundamentally different configurations, one highly detailed and widely used in frozen ground research and one with a comparatively simpler structure, represent seasonally frozen ground characteristics. Including both models in this study is therefore useful, as it allows us to assess how differences in model complexity and vertical discretization influence the simulation of frozen ground. We will update Table 1 in the revised manuscript to include detailed information on the soil layer structures of both models as follows:

Table 1: Key structural and parametric differences in soil and snow process representations between JSBACH and CLM

Model features	**JSBACH**	**CLM**
Total soil column depth	∼9.8 m	∼49.5 m (bedrock below 8.6 m)
Number of soil layers (default)	5 soil layers	20 soil layers + 5 bedrock layers
Soil layer node depths (m)	0.0325, 0.192, 0.7755, 2.683, 6.984	0.01, 0.04, 0.09, 0.16, 0.26, 0.4, 0.58, 0.8, 1.06, 1.36, 1.7, 2.08, 2.5, 2.99, 3.58, 4.27, 5.06, 5.95, 6.94, 8.03, 9.795, 13.328, 19.483, 28.871, 41.998
Soil layer thickness (m)	0.065, 0.254, 0.913, 2.902, 5.7	0.02, 0.04, 0.06, 0.08, 0.12, 0.16, 0.2, 0.24, 0.28, 0.32, 0.36, 0.4, 0.44, 0.54, 0.64, 0.74, 0.84, 0.94, 1.04, 1.14, 2.39, 4.676, 7.635, 11.14, 15.115
Number of snow layers (default)	5 (hydrologically inactive)	10
Number of PFTs	11	15
Concept of supercooled water	Yes (Niu and Yang, 2006)	Yes (Niu and Yang, 2006)
Soil heat capacity scheme	(De Vries (1953))	(De Vries (1953))
Soil thermal conductivity	(Johansen (1977))	(Farouki (1981))
Bottom boundary condition	Zero flux	Zero flux
Snow density formulation	Temperature dependent	Multi-process: compaction, destructive and constructive, melt metamorphism (Anderson, 1976)
Fresh snow density parameterization	50 kg m^{-3} (constant)	Temperature dependent
Snow thermal conductivity	(Calonne et al. (2011))	(Jordan (1991))

Regarding the permafrost definition, we agree that our definition of Permafrost (PEFT) differs from the standard International Permafrost Association (IPA) definition, which requires ground temperatures to remain below 0°C continuously for two or more consecutive years. We clarify that our classification defines a grid cell as PEFT if, during the 1986-2022 analysis period (37 years), at least one soil layer remains frozen year-round for more than 50% of the years. In practice, this means that a cell must exhibit year-round frozen conditions in at least 19 out of the 37 years. The requirement of ≥19 years with year-round frozen ground ensures that at least one consecutive year of complete freezing occurs (thereby satisfying the baseline condition of the IPA definition), while also capturing the full range of permafrost persistence up to all 37 years being frozen. Therefore, A threshold of >50% guarantees that freeze conditions are climatologically persistent rather than occasional or episodic, so a necessary condition when working with multi-decadal model outputs where interannual variability is high.

Apart from the two-consecutive-year criterion, a similar approach as ours using individual simulation year criteria has been used by Langer et al. (2024) based on a thaw depth parameter, and previous large-scale permafrost assessments have also used alternative diagnostics based on active layer thickness (Dankers et al., 2011), thermal index (Slater and Lawrence, 2013), or threshold-based probabilistic occurrence of permafrost (Gruber, 2012; Burke et al., 2020).

We have revised the manuscript to explicitly clarify the PEFT definition and justify our methodological choice as follows (at Line 261):

Here, PEFT is defined as grid cells where at least one soil layer remains frozen year-round in >50% of the 1986-2022 simulation years (i.e., ≥19 of 37 years), representing persistent frozen-ground conditions in a multi-decadal framework. Although the International Permafrost Association (IPA) conventionally defines PEFT as ground remaining below 0 ^{o}C for two or more consecutive years (Brown et al., 1997), alternative diagnostics are commonly used in gridded model assessments (Steinert et al., 2024; Langer et al., 2024; Gruber, 2012; Slater and Lawrence, 2013; Burke et al., 2020). SFG denotes areas experiencing seasonal freezing for at least 15 days per year.

Comment 3:

Regarding snow depth, it should be noted that it is closely related to how the model deals with solid precipitation. The GSWP3 forcing only has total precipitation (rain+snow), and it is the land surface model that deals with the precipitation partitioning. In this way, other than the snow accumulation modeling, authors should also elaborate on the difference, if any, of precipitation partitioning between the two models in the methodology section.

Response:

We thank the reviewer for pointing this out.

Both JSBACH and CLM5 uses temperature-based linear partitioning methods, but with different threshold temperature limits, as shown below, which we plan to add as an additional row in Table 1. Therefore, we agree that differences in simulated snowfall can arise not only from snowpack physics, but also from the precipitation partitioning schemes used by the two models.

Model features	**JSBACH**	**CLM**
Precipitation (P) partitioning using air temperature (T ^{o}C)	$\text{snow} = \begin{cases} P & T < -1.1 \\ P\,\dfrac{3.3 - T}{4.4} & -1.1 \le T \le 3.3 \\ 0 & T > 3.3 \end{cases}$ rain = P − snow (Wigmosta et al., 1994)	$\text{snow} = \begin{cases} P & T < 0 \\ P\,\dfrac{2 - T}{2} & 0 \le T \le 2 \\ 0 & T > 2 \end{cases}$ rain = P − snow (Lawrence et al., 2019)

Such differences are generally most evident during transition seasons, particularly at the beginning and end of winter, when near-surface air temperature lies within the mixed-phase range between the minimum and maximum threshold temperatures used for rain-snow partitioning. Under these conditions, model specific threshold values can lead to some differences in snowfall fraction, which should be acknowledged in the main text. We have also evaluated the rainfall and snowfall fractions in both models. The plot below shows only small differences in monthly climatological accumulated rainfall and snow fractions over the selected study region (Figure 1) during 1986-2022, indicating that the overall partitioning differences are minor. We will include this supporting figure in the revised material, if needed. We have also added a brief note at Line 141 of the revised manuscript stating:

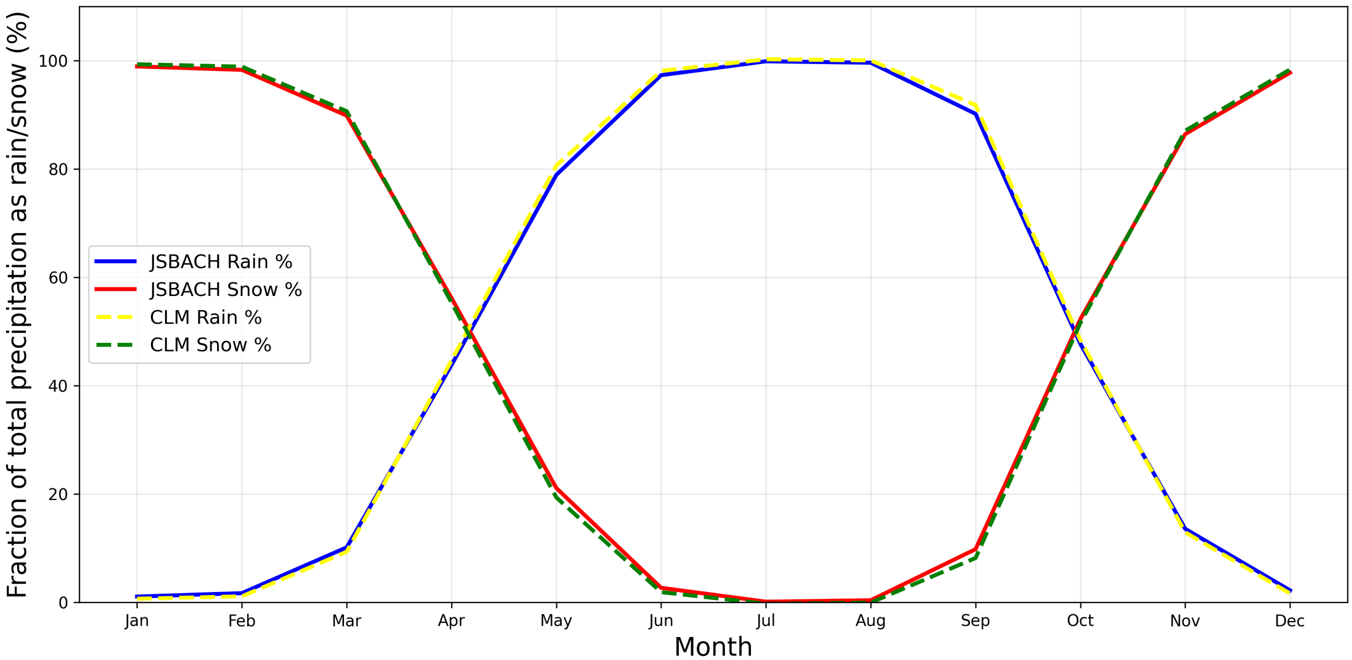


Figure 3: Monthly climatological accumulated rainfall and snowfall fractions averaged over the selected region (Figure 1).

"Since JSBACH and CLM models use different precipitation partitioning schemes (Table 1), small differences in simulated snow related variables are expected, particularly during transitional seasons when near-surface air temperatures fall within the model specific rain-snow threshold range."

Minor comments

Comment 1: Section 3.1 This part should be moved to the methodology section.

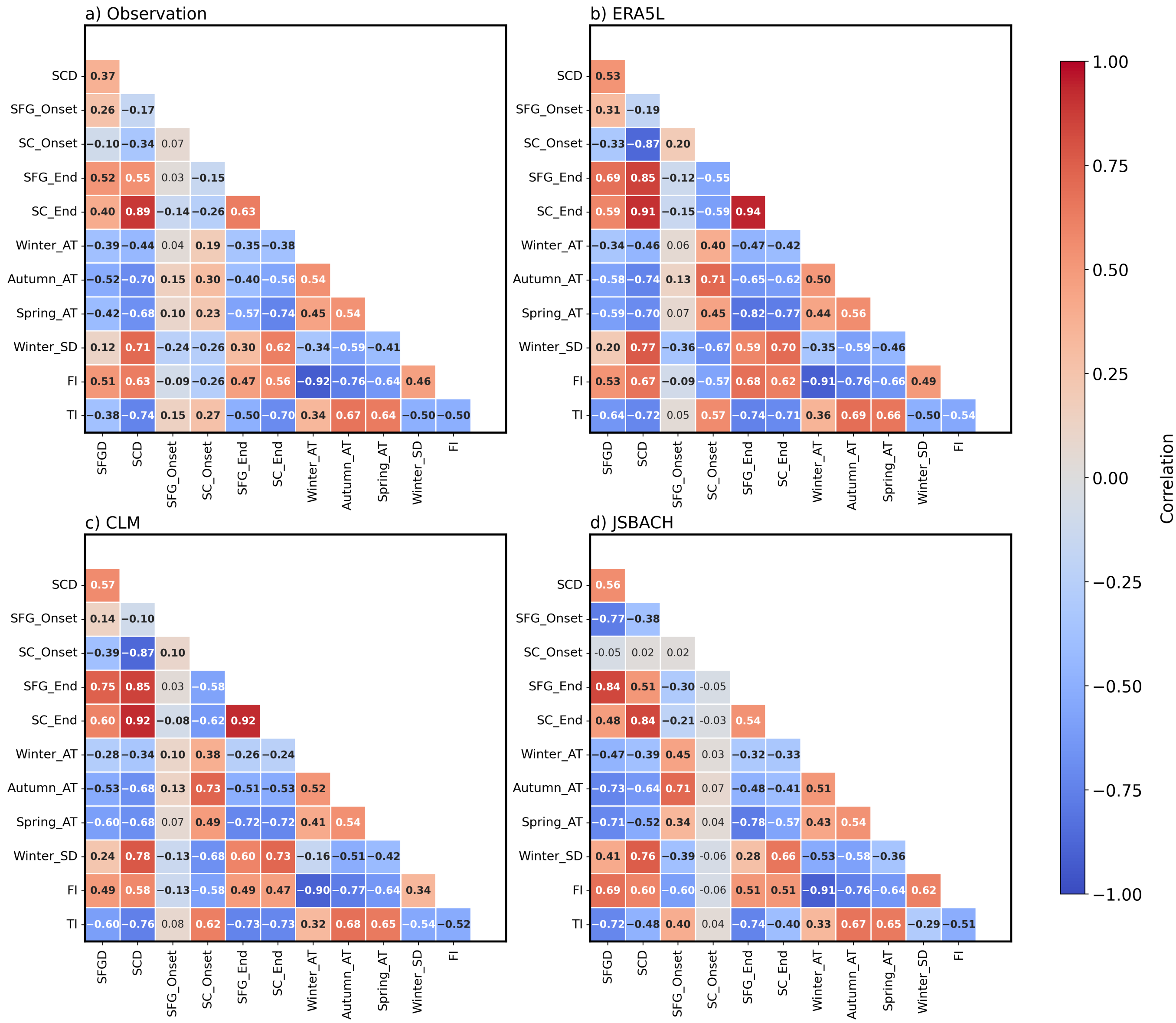
Response:

Following your and the other reviewers' suggestions, we will move this section to the end of the methodology section.

Comment 2: Figure 10: I suggest adding significance test on top of these linear correlation coefficients.

Response:

Thank you for your suggestion. The following significance test plot will replace the previous one in the revised manuscript, with statistically significant values highlighted in bold.



References

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M., and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, Biogeosciences, 10, 7109-7131, https://doi.org/10.5194/bg-10-7109-2013, 2013.

Brown, J., Ferrians, O. J. J., Heginbottom, J. A., and Melnikov, E. S.: Circum-Arctic map of permafrost and ground ice conditions, US Geological Survey, Reston, VA, 1997.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., Van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., Van Den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245-4287, https://doi.org/10.1029/2018MS001583, 2019.

Zhang, T., Heginbottom, J. A., Barry, R. G., and Brown, J.: Further statistics on the distribution of permafrost and ground ice in the Northern Hemisphere1, Polar Geography, 24, 126-131, https://doi.org/10.1080/10889370009377692, 2000.

Gualdi, S., Borrelli, A., Cantelli, A., Davoli, G., del Mar Chavesmontero, M., Masina, S., Navarra, A., Sanna, A., and Tibaldi, S.: The new CMCC operational seasonal prediction system, CMCC Technical Notes, TN0288, https://www.cmcc.it/wp-content/uploads/2020/09/TN0288-csp-09-2020-1.pdf, 2020.

Steinert, N. J., Debolskiy, M. V., Burke, E. J., García-Pereira, F., and Lee, H.: Evaluating permafrost definitions for global permafrost area estimates in CMIP6 climate models, Environ. Res. Lett., 19, 014033, https://doi.org/10.1088/1748-9326/ad10d7, 2024.

Koster, R. D. and Suarez, M. J.: Soil Moisture Memory in Climate Models, J. Hydrometeor, 2, 558-570, https://doi.org/10.1175/1525-7541(2001)002%3C0558:SMMICM%3E2.0.CO;2, 2001.

Vinnikov, K. Y., Robock, A., Speranskaya, N. A., and Schlosser, C. A.: Scales of temporal and spatial variability of midlatitude soil moisture, J. Geophys. Res., 101, 7163-7174, https://doi.org/10.1029/95JD02753, 1996.

Ekici, A., Beer, C., Hagemann, S., Boike, J., Langer, M., and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model, Geoscientific Model Development, 7, 631-647, https://doi.org/10.5194/gmd-7-631-2014, 2014.

Yang, K. and Zhang, J.: Spatiotemporal characteristics of soil temperature memory in China from observation, Theor Appl Climatol, 126, 739-749, https://doi.org/10.1007/s00704-015-1613-9, 2016.

Burke, E. J., Zhang, Y., and Krinner, G.: Evaluating permafrost physics in the Coupled Model Intercomparison Project 6 (CMIP6) models and their sensitivity to climate change, The Cryosphere, 14, 3155-3174, https://doi.org/10.5194/tc-14-3155-2020, 2020.

Dankers, R., Burke, E. J., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, The Cryosphere, 5, 773-790, https://doi.org/10.5194/tc-5-773-2011, 2011.

Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, The Cryosphere, 6, 221-233, https://doi.org/10.5194/tc-6-221-2012, 2012.

Slater, A. G. and Lawrence, D. M.: Diagnosing Present and Future Permafrost from Climate Models, Journal of Climate, 26, 5608-5623, https://doi.org/10.1175/JCLI-D-12-00341.1, 2013.

Langer, M., Nitzbon, J., Groenke, B., Assmann, L.-M., Schneider Von Deimling, T., Stuenzi, S. M., and Westermann, S.: The evolution of Arctic permafrost over the last 3 centuries from ensemble simulations with the CryoGridLite permafrost model, The Cryosphere, 18, 363-385, https://doi.org/10.5194/tc-18-363-2024, 2024.

Wigmosta, M. S., Vail, L. W., and Lettenmaier, D. P.: A distributed hydrology-vegetation model for complex terrain, Water Resources Research, 30, 1665-1679, https://doi.org/10.1029/94WR00436, 1994.

Response to Reviewer#2 Comments

Comment 1:

In the Section 2.2 (Lines 141–147), the manuscript states that CLM uses a two-phase spin-up (preindustrial + historical), whereas JSBACH is run continuously from 1941 to 2022 without equivalent spin-up. This inconsistency may lead to non-comparable initial states, especially for deep soil temperature, soil moisture, and freeze–thaw memory, directly affecting key variables such as maximum freezing depth (MFD) and SFG duration.

Response:

We thank the reviewer for this important and perceptive comment. We apologize that the description of the model spin up procedures in Section 2.2 was insufficiently clear and caused confusion to everyone. Below, we provide a detailed and transparent account of the spin-up and initialization protocols applied to CLM and JSBACH:
We clarify that CLM5.1 was initialized following the standard two-phase spin up procedure described in Lawrence et al. (2019); Koven et al. (2013), and implemented in the CMCC seasonal prediction system (Gualdi et al., 2020). In the first phase, the model was brought to equilibrium representative of year-1850 conditions using the standard CLM spin up protocol. This phase consisted of 400 years with accelerated decomposition to equilibrate slow carbon and nitrogen pools, followed by an additional 800 years in normal mode. The second phase was a historical run from 1850, which was achieved by integrating over a repeating 20-year period of the GSWP3 forcing dataset (1901-1920) (Lawrence et al., 2019), while prescribing fixed preindustrial boundary conditions, including constant atmospheric CO_2, nitrogen deposition, aerosol deposition, land use, and no wood harvest. This is the standard CLM approach for generating a stable preindustrial land state when meteorological forcing prior to 1901 is unavailable. From 1901 onward, the model was forced with true time-varying GSWP3 data, which in our setup covers the period 1901-1940. To ensure a smooth transition from GSWP3 to ERA5 forcing, we additionally performed 10 cyclical runs using climatological ERA5 forcing for 1941-1950, which reduced potential drift caused by the forcing dataset change.

In contrast, JSBACH was run continuously from 1941 to 2022 without an equivalent multi phase spin up, initialized from JSBACH3 initial-condition files remapped from the Gaussian grid to the ICON R02B06 grid, and forced with ERA5 reanalysis data at a 3-hourly temporal resolution. Previous JSBACH site-level study has considered a spin up period of 30 years sufficient to reach thermal and hydrological equilibrium in frozen ground simulations (Ekici et al., 2014). Since our analysis period begins in 1986, JSBACH had been integrated for 45 years (1941-1985) prior to evaluation, exceeding this threshold and ensuring that the influence of the initial conditions on the analysed variables is negligible.
We have also expanded the discussion of spin-up duration and equilibrium considerations along with some plots in the response to **Reviewer #1 (Major Comment 1)**.
We agree that the initial conditions of the two models were different owing to their distinct model structures and initialization frameworks. However, since the analysis period starts in 1986, both models had already been run for several decades before evaluation ($\sim$ 45 years from 1941 to 1985). Therefore, the direct influence of the initial conditions on the analysed frozen-ground variables is expected to be small during the evaluation period, as soil thermal and hydrological memory typically dissipates over such timescales (Yang and Zhang, 2016; Koster and Suarez, 2001; Vinnikov et

al., 1996). But since it is important for the readers to know it so we are planning to acknowledge that "Differences in spatial resolution, grid-cell heterogeneity, **initial state** and boundary condition datasets describing soil properties likely contribute to some of the simulated biases." And we have clarified this point in the revised manuscript (Line 344).
We hope this response and the planned revision satisfactorily address the reviewer's thoughtful comment.

Comment 2:

In the Section 3.1, the study introduces an updated snow densification scheme for JSBACH and then uses only the improved version for intercomparison. I am curious whether CLM and ERA5L have this optimized parametrization. Also, what are the simulated results from the old setups? If it is possible to clearly frame the study as a model development and evaluation paper rather than a pure intercomparison?

Response:

We would like to clarify that the manuscript is primarily intended as an evaluation and intercomparison of seasonally frozen ground characteristics simulated by two widely used land surface models, JSBACH and CLM, together with ERA5L. These models were specifically selected because they differ notably in their vertical soil discretization, snow physics, and parameterization complexity, enabling us to investigate how such structural differences influence the simulation of frozen-ground processes at large spatial scales.
The simulated results using the old JSBACH snow density formulation are documented in detail in the technical report available at the Zenodo repository provided in the Code and Data Availability section of the manuscript (doi: `https://zenodo.org/records/18110147`; file: report_on_improvement_in_JSBACH_final.pdf). We refer the reviewer to this report for a full comparison between the old and updated configurations.
Regarding the snow densification scheme, we would like to clarify that we did not specifically "optimize" JSBACH for this study. During our work, we identified limitations in the earlier snow density formulation, which primarily depended on time evolution, and contributed to the implementation and testing of a physically improved temperature-dependent formulation together with the ICON-Land development team and DWD. The present study therefore evaluates the latest available JSBACH/ICON-Land version rather than a model employing an optimized or best-possible snow parameterization, as further improvements in snow physics are still possible. We have also clarified in Table 1 that the snow-density parameterizations differ among the datasets: JSBACH uses a temperature-dependent snow densification scheme, CLM includes more advanced snow compaction and metamorphism processes, and ERA5L represents snow density evolution through overburden pressure, thermal metamorphism, and retained liquid water following Lynch-Stieglitz et al. (1994).
We agree that the original manuscript structure may have unintentionally suggested a model-development focus. To avoid this confusion, we have decided to move the description of the updated JSBACH snow density scheme from the Results section to the end of the Methodology section in the revised manuscript.

Comment 3:

In the Section 3.3.2, the manuscript attributes many biases primarily to snow insulation effects. The attribution lacks sensitivity experiments to provide a robust evidence.

Response:

We thank the reviewer for this comment. We do relate the simulated soil temperature biases to the insulating capacity of snow, as discussed in the first paragraph of Sect. 3.3.2. Our interpretation is based on the well-established physical understanding that increasing snow depth and snow cover duration enhance the insulating effect of snow by thermally decoupling the soil from colder winter air temperatures, thereby reducing ground heat loss (Zhang , 2005; Zhang et al., 2005). In this study, the inferred relationship between snow insulation and soil thermal biases is therefore based on the simulated snow depth behaviour in each model, particularly whether snow depth is underestimated or overestimated relative to observations.
The JSBACH model development itself provides physically consistent evidence and can therefore be interpreted as a sensitivity experiment. The earlier JSBACH configuration, which used a time evolution based snow density formulation, simulated lower snow depth and exhibited a pronounced cold bias in soil temperature (Appendix Fig. B1; also documented in the Zenodo report- file: report_on_improvement_in_JSBACH_final.pdf). After implementing the updated temperature-dependent snow density parameterization, the simulated snow depth increased and the cold soil temperature bias was substantially reduced, bringing simulated soil temperatures closer to observations.
This comparison between the old and updated JSBACH configurations provides direct evidence that modifying the snow density parameterization and consequently the simulated snow depth resulted in systematic changes in soil temperature consistent with the expected snow insulation effect. Old versus new JSBACH configurations also reveal a clear contrast in the snow insulation effect, as reflected in the TDSA–SD relationship: the old setup (figure below) differs notably from the updated configuration shown in Fig. 7 of the manuscript, highlighting the strong influence of snow representation on simulated soil thermal conditions within the JSBACH model.
The insulating effect can vary considerably depending on snow cover timing and duration which motivated the additional analysis presented in Section 3.3.4, where we examine how snow cover characteristics influence SFG properties. We would like to refer to Sect. 3.1 in the first sentence of Sect. 3.3.2 to clearly establish how the ST biases are attributed to the snow insulation effect. We hope the reviewer finds this clarification satisfactory.

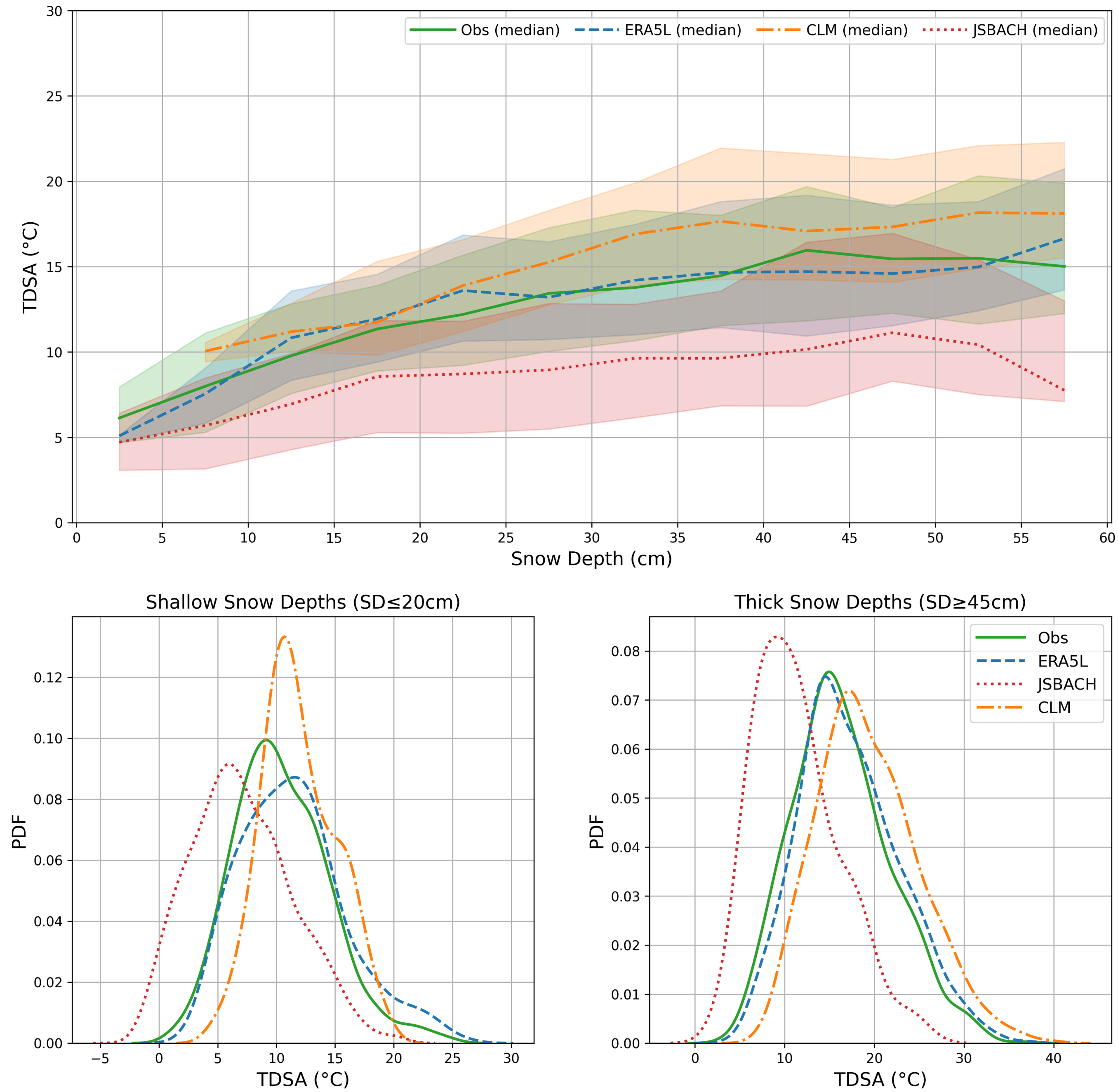


Figure 1: Relationship between wintertime monthly mean SD and soil-air temperature difference (TDSA) across observational and model datasets, shown as median TDSA (dotted lines, 5 cm SD bins) and IQR (shaded) for observations (Obs; green), ERA5L (blue), JSBACH (red), and CLM (orange) in upper panel. Lower panel shows probability density functions (PDFs) of TDSA for shallow (SD $\leq$ 20 cm) and thick (SD $\geq$ 45 cm) snow conditions.

Comment 4:

In the Section 3.3.1, the study compares grid-scale model outputs (30–50 km) with point observations at 26 stations. There is a large mismatch between the two scales. So, how the grids were selected? How the magnitude of the results may be influenced by this mismatch? A discussion may be needed to show the uncertainties.

Response:

We thank the reviewer for raising this important methodological point. We fully agree that a scale mismatch between gridded model output (40–50 km) and point-scale in situ observations is an inherent limitation of this type of evaluation. For the model-to-observation comparison, the gridded model outputs were linearly interpolated to the geographic coordinates of each station. This approach is standard practice in large scale model evaluation studies and allows us to draw conclusions from comparisons between gridded and site-level data (Wang et al., 2016; Luo et al., 2020; Zhang et al., 2005; Damseaux et al., 2024). We acknowledge that this scale mismatch introduces uncertainty, particularly because a single observation site not fully represent the spatial heterogeneity within a model grid cell, including variability in vegetation, snow redistribution, topography, and soil properties.
We agree that the ideal way to quantify the scale-mismatch uncertainty would be to compare the point observations against the spatial variability within a grid cell, e.g., by having multiple stations within the same 40–50 km cell. However, as shown in Table A1, no grid cell in our study area contains more than one station. Therefore, we cannot directly estimate the sub-grid variability of observations or compute a robust uncertainty metric for the mismatch. We therefore cannot formally quantify the representativeness error, and one can only speculate about its magnitude. Despite the unavoidable limitations associated with global model evaluations, the consistent differences observed across multiple stations still provide meaningful information about the systematic strengths and weaknesses of each model, particularly regarding their internal snow and soil process representations. We have also acknowledged and justified this grid-to-point comparison approach in Lines 344–345 of the manuscript.

Comment 5:

In the Section 3.3.4, the manuscript tries to interpret how the factors affect SFG by correlation analysis. In fact, these factors have close correlations, it is preferable to use partial correlation analysis or machine learning–based approaches for attribution analysis.

Response:

We thank the reviewer for this important suggestion. In response, we have decided to add a partial correlation table together with a schematic diagram as suggested by Reviewer #3, summarizing the statistically significant partial correlations between SFG characteristics, snow cover properties, and seasonal thermal drivers across observations and models (Fig. 2). The controlling variables used for each variable pair are listed in Table 1 and were selected based on physically motivated relationships to isolate direct from indirect pathways. To further visualise the relationships presented in Table 2, including both pearson correlation and partial correlation values calculated using the controlling parameters listed in Table 1, we constructed a schematic diagram (Fig. 2) for observations and each model. This diagram helps distinguish genuine physical drivers from spurious inter-correlations.

Variable definitions (for reference):

- SFG_Onset - onset date of seasonally frozen ground at 20 cm depth.
- SFG_End - end date of seasonally frozen ground at 20 cm depth

- SFGD - seasonally frozen ground duration (SFG_End - SFG_Onset)
- SC_Onset - snow cover onset date
- SC_End - snow cover end date
- SCD - snow cover duration (SC_End - SC_Onset)
- Autumn_AT - autumn air temperature
- Winter_AT - winter air temperature
- Spring_AT - spring air temperature
- Winter_SD - winter snow depth

Partial correlations (pc) were computed to remove the influence of shared drivers, particularly Autumn_AT, Spring_AT, and Winter_SD, which co-vary with multiple target variables simultaneously. This approach allows us to determine, for example, whether the relationship between SFG_Onset and SC_Onset is direct or primarily inherited from a common autumn temperature signal.

To illustrate the calculation, we provide an example of the partial correlation between parameters A and B, with C held as the control variable:

$$pc_{AB\cdot C} = \frac{R_{AB} - R_{AC}R_{BC}}{\sqrt{(1 - R_{AC}^2)\,(1 - R_{BC}^2)}} \tag{1}$$

Observation: In the observations, SC_Onset and Autumn_AT show a direct moderate positive relationship (R = 0.3), indicating that warmer autumn temperatures delay snow cover onset. SC_Onset and SFG_Onset exhibits virtually no direct relationship (partial r = 0.02 after controlling for Autumn_AT), indicating that frozen ground onset is not directly driven by snow arrival and is instead likely governed by local factors such as soil properties, vegetation cover, soil moisture, and microclimatic conditions. Similarly, after controlling for SC_Onset, the relationship between Autumn_AT and SFG_Onset is very weak (pc = 0.14), suggesting only limited direct influence of Autumn_AT on soil freezing onset. SC duration is primarily controlled by Winter_SD (pc = -0.52), Spring_AT (pc = -0.50), and Autumn_AT (pc = -0.33). SC_End shows strong dependence on Spring_AT (pc = -0.66) and Winter_SD (pc = +0.52), and a moderate relationship with SFG_End after controlling for spring temperature (pc = +0.37). In contrast, SFG_End is only weakly directly influenced by Spring_AT once the effect of SC_End is removed (pc = -0.20), highlighting that soil thaw timing is mediated through snow cover dynamics rather than temperature alone. In the observations, SFG duration shows a moderate negative relationship with both Autumn_AT (pc = -0.37) and Spring_AT (pc = -0.21) after controlling for SC duration, indicating that colder autumn and spring conditions contribute weakly to longer frozen ground duration independent of snow cover persistence.

ERA5L: In ERA5L, the schematic indicates no direct relationship between SFG_Onset and Autumn_AT. SC_End, however, shows relatively strong relationships with both SFG_End (pc = 0.85) and Spring_AT (pc = -0.69) when the influence of other variables is accounted for. In addition, the direct influence of Spring_AT on SFG_End (pc = -0.42) remains stronger than observed, suggesting that ERA5L overestimates the sensitivity of soil thaw to spring warming. A higher SFG duration-SC duration correlation (R= 0.53) in ERA5L than in observations (R = 0.38)) indicates that the model overestimates the coupling between snow cover duration and frozen ground duration. SFG duration maintains moderate negative relationships with both Autumn_AT (pc = -0.32) and Spring_AT (pc = -0.36) after controlling for SC duration, indicating a slightly stronger sensitivity of frozen ground duration to seasonal warming than observed.

CLM: CLM reproduces the correct directional sign across all major linkages, but SC_Onset's dependence on Autumn_AT is amplified (R = 0.73), suggesting snow onset timing is highly sensitive to autumn temperature conditions. Most critically, the SC_Onset-SFG_Onset relationship collapses to near zero, while the Autumn_AT-SFG_Onset relationship remains weak (pc = +0.09), which may suggest that the relatively high snow depth in CLM could partially thermally isolate the soil, reducing the influence of snow onset timing and autumn air temperature on soil freezing at 20 cm depth. SFG duration shows a stronger dependence on SC duration (R = 0.57) while SC duration is strongly controlled by Winter_SD (pc = +0.69), Spring_AT (pc = -0.52) and Autumn_AT (pc = -0.35), reflecting the dominant role of snow accumulation and melt-season temperature in regulating snow cover duration. Furthermore, the SC_End-SFG_End partial correlation is substantially amplified (pc = 0.84) after controlling for Spring_AT, indicating that snow disappearance strongly regulates soil thaw timing in CLM. Once SC duration and Spring_AT are controlled, Autumn_AT's independent effect on SFG duration weakens to pc = -0.20 in CLM (vs -0.37 in observations), while Spring_AT's independent effect (pc = -0.33) remains slightly stronger than observed (pc = -0.21), suggesting that CLM further suppresses the direct autumn thermal pathway to the soil while retaining an elevated spring signal.

JSBACH: JSBACH exhibits a distinctly different relationship structure, likely reflecting its substantial snow depth underestimation. Unlike observations and CLM, SC_Onset shows virtually no dependence on Autumn_AT (R = 0.07), suggesting that snow accumulation timing in JSBACH is largely disconnected from autumn thermal conditions. The SFG_Onset and Autumn_AT partial correlation surges to pc = 0.71, the strongest signal in the entire network and far exceeding the observed value (pc = 0.14), while SC_Onset and SFG_Onset show no meaningful relationship (pc = -0.04), together confirming that in the absence of adequate snow cover, Autumn_AT drives soil freezing directly without the moderating influence of snow. SFG duration shows a stronger dependence on SC duration (R = +0.56) while SC duration is strongly controlled by Winter_SD (pr = +0.66), Spring_AT (pc = -0.35), and Autumn_AT (pc = -0.27). The SC_End and SFG_End coupling weakens substantially (pc = 0.18 vs 0.37 in observations), while Spring_AT's independent effect on SFG_End remains quite strong even after controlling for SC_End (pc = -0.68), indicating that without adequate snow insulation, spring air temperature forces soil thaw directly rather than through snow cover. In JSBACH, Autumn_AT's independent effect on SFG duration (pc = -0.52) and Spring_AT's independent effect (pc = -0.54) both remain considerably stronger than observed, confirming that seasonal air temperatures control frozen ground duration directly, bypassing snow insulation. In JSBACH, the sign change for SFG_Onset (strong negative dependency on SFGD, R = -0.77) together with the positive dependency of SFG_End on SFG duration (R = +0.84) suggests that both early onset and delayed end contribute to longer frozen ground duration, whereas in other datasets SFG duration shows stronger control by SFG_End.

Table 1: Relationships between seasonal air temperatures, snow depth, snow cover, and seasonally frozen ground characteristics with controlling factors and physical interpretation.

Parameters	**Controlling Factors**	**Remarks**
Autumn_AT × Winter_SD	–	Independent environmental drivers
Winter_AT × Winter_SD	–	Independent environmental drivers
SC_Onset × Autumn_AT	–	Direct autumn temperature control on snow onset
SFG_Onset × SC_Onset	Autumn_AT	Remove shared effect of autumn temperature
SFG_Onset × Autumn_AT	SC_Onset	Isolate direct thermal effect on soil freezing
SCD × Autumn_AT SCD × Spring_AT SCD × Winter_SD	Winter_AT, Spring_AT, Winter_SD Winter_AT, Autumn_AT, Winter_SD Winter_AT, Spring_AT, Autumn_AT	SCD is influenced by seasonal air temperatures and snow depth; controls remove indirect pathways and isolate direct effects.
SC_End × Spring_AT	Winter_AT, Winter_SD	Isolate spring warming effect on snow disappearance
SC_End × Winter_SD	Winter_AT, Spring_AT	Snow depth effect on melt date excluding temperature and thaw influences
SC_End × SCD SC_Onset × SCD	– –	Mathematically linked (duration = end - onset); no control variable required.
SC_End × SFG_End	Spring_AT	Remove shared effect of spring temperature
SFG_End × Spring_AT	SC_End	Direct spring temperature control on SFG onset.
SFGD × SCD	–	Includes the influence of all other factors to show total coupling effect.
SFGD × Autumn_AT	SCD, Spring_AT	Isolates the direct influence of autumn air temperature
SFGD × Spring_AT	SCD, Autumn_AT	Isolates the direct influence of spring air temperature
SFG_End × SFGD SFG_Onset × SFGD	– –	Mathematically linked (duration = end - onset): no control needed.

Table 2: Statistically significant (p value < 0.01) pearson and partial correlations between seasonal air temperatures, winter snow depth, snow cover, and seasonally frozen ground characteristics.

Parameters	**Pearson Correlation (R)**				**Partial Correlation (pc)**			
	Obs	**ERA5L**	**CLM**	**JSBACH**	**Obs**	**ERA5L**	**CLM**	**JSBACH**
Autumn_AT × Winter_SD	-0.59	-0.59	-0.51	-0.58	-	-	-	-
Winter_AT × Winter_SD	-0.34	-0.35	-0.16	-0.53	-	-	-	-
SC_Onset × Autumn_AT	0.3	0.71	0.73	0.07	-	-	-	-
SFG_Onset × SC_Onset	0.07	0.2	0.1	0.02	0.02	0.15	0.004	-0.04
SFG_Onset × Autumn_AT	0.15	0.13	0.13	0.71	0.14	-0.02	0.09	0.71
SCD × Autumn_AT	-0.7	-0.74	-0.68	-0.64	-0.33	-0.4	-0.35	-0.27
SCD × Spring_AT	-0.68	-0.7	-0.68	-0.52	-0.5	-0.5	-0.52	-0.35
SCD × Winter_SD	0.71	0.77	0.78	0.76	0.52	0.62	0.69	0.66
SC_End × Spring_AT	-0.74	-0.77	-0.72	-0.57	-0.66	-0.69	-0.65	-0.51
SC_End × Winter_SD	0.62	0.7	0.73	0.66	0.52	0.61	0.67	0.61
SC_End × SCD	0.89	0.91	0.92	0.84	-	-	-	-
SC_Onset × SCD	-0.34	-0.87	-0.87	0.02	-	-	-	-
SC_End × SFG_End	0.63	0.94	0.92	0.54	0.37	0.85	0.84	0.18
SFG_End × Spring_AT	-0.57	-0.82	-0.72	-0.78	-0.2	-0.42	-0.23	-0.68
SFGD × SCD	0.38	0.53	0.57	0.56	-	-	-	-
SFGD× Autumn_AT	-0.52	-0.58	-0.53	-0.73	-0.37	-0.32	-0.2	-0.52
SFGD × Spring_AT	-0.42	-0.59	-0.6	-0.71	-0.21	-0.36	-0.33	-0.54
SFG_End × SFGD	0.52	0.69	0.75	0.84	-	-	-	-
SFG_Onset × SFGD	0.26	0.31	0.14	-0.77	-	-	-	-

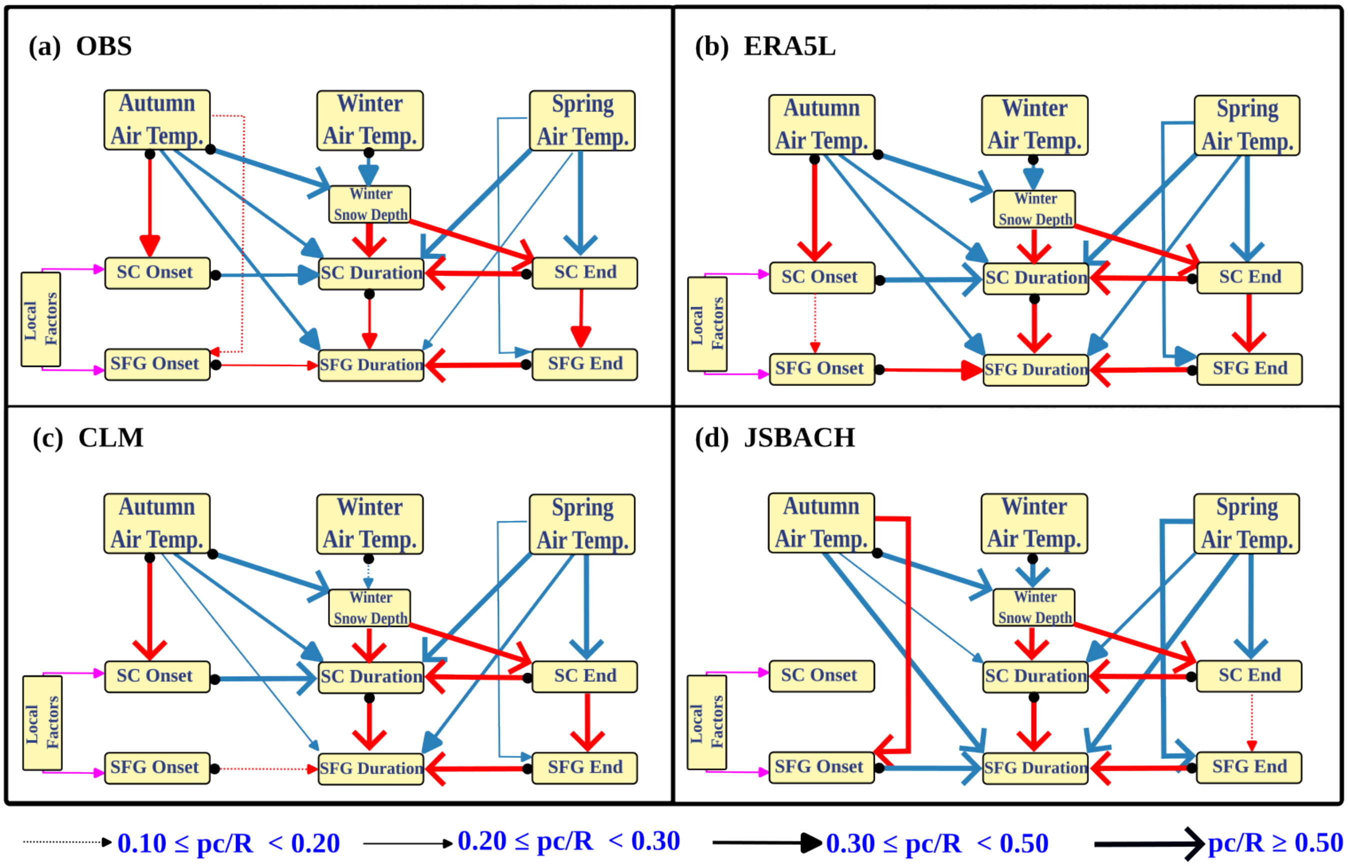


Figure 2: Schematic representation of the statistically significant pearson and partial correlations derived from Appendix Table 2, illustrating the interactions among seasonal air temperature, snow cover, snow depth, and seasonally frozen ground characteristics for (a) OBS, (b) ERA5L, (c) CLM, and (d) JSBACH. Red and blue lines denote positive and negative correlations, respectively, while the magenta-coloured line indicates possible local influences affecting the variables. Filled circles at the arrow origin indicate pearson correlation coefficients; all other values represent partial correlations.

Based on the information we have, we would like to correct our Sect 3.2.4 as follows in the revised manuscript:

Figure 10 shows statistically significant correlation matrices between SFG and snow cover characteristics, seasonal ATs, winter SD, as well as thermal indices for observations and models. To further distinguish direct relationships from shared seasonal influences, partial correlation (pc) analysis was additionally calculated (using `pingouin` package; `https://pingouin-stats.org/generated/pingouin.partial_corr.html`). The resulting partial correlation values are presented in Appendix Table 2, while the corresponding controlling factors used in each analysis are listed in Appendix Table 1. This approach removes the effect of co-varying seasonal controls (autumn AT, spring AT, winter AT and winter SD), allowing a clearer assessment of whether snow cover exerts an independent influence on soil freezing dynamics or whether observed linkages arise primarily through common atmospheric forcing. Figure 2 presents a schematic diagram based on Appendix Table 2, illustrating both pearson correlations (R) and partial correlations (pc) after accounting for shared drivers, thereby highlighting the relative strength and direction of the key interactions among variables.

In observations, SFG onset is not directly controlled by large-scale seasonal conditions. SC_Onset and autumn AT show a direct moderate positive relationship (R = 0.30), indicating that warmer autumn temperatures delay snow cover onset, consistent with a temperature-driven control, yet autumn AT exerts only limited direct influence on SFG onset (pc $= 0.14$), and the mutual partial correlation between SC_Onset and SFG_Onset is negligible (pc $= 0.02$ after controlling for autumn AT). SC duration is primarily controlled by winter SD (pc $= -0.52$), spring AT (pc $= -0.50$), and autumn AT (pc $= -0.33$). SC_End shows strong dependence on spring AT (pc $= -0.66$) and winter SD (pc $= +0.52$), and a moderate relationship with SFG_End after controlling for spring temperature (pc $= +0.37$). In contrast, SFG_End is correlated with the thawing index (R = -0.50) but is only weakly directly influenced by spring AT once the effect of SC_End is removed (pc $= -0.20$), highlighting that soil thaw timing is largely mediated through snow cover dynamics rather than temperature alone. SFG duration shows the strongest association with autumn AT (R $= -0.52$; pc $= -0.37$ after controlling for SC duration and spring AT), with spring AT retaining a moderate independent effect on SFG duration after controlling for SC duration (pc $= -0.21$). Altogether, the observations indicate that snow cover governs thaw timing and, together with the intensity of seasonal air temperature, determines the duration of seasonal ground freezing.

ERA5L and CLM broadly reproduce several of the observed dependencies but often amplify the strength of correlations, suggesting an excessive sensitivity of the simulated soil thermal regime to AT and snow controls. Moreover, both models do not account for the high heterogeneity in local conditions present at observation sites, as land surface properties are spatially smoothed, which likely contributes to the strong dependence of snow cover onset on seasonal variables, particularly autumn AT. In addition, precipitation phase partitioning in the models is governed by grid-scale temperature thresholds, making the onset of snow accumulation primarily temperature-driven. Consequently, local factors that modify the near-surface temperature regime, such as vegetation dynamics, soil thermal properties, microtopography, and surface energy exchanges, cannot be represented as specifically as in station observations. The SC_End-SFG_End coupling is particularly inflated (R > 0.9; pc $= 0.85$ in ERA5L and 0.84 in CLM after controlling for spring AT), far exceeding observations. In ERA5L, spring AT retains a stronger direct effect on SFG_End than observed (pc $= -0.42$), and the SFG duration-SC duration coupling is elevated (R = 0.53 vs. 0.38 in observations), with comparable sensitivity of SFG duration to both autumn AT (pc $= -0.32$) and spring AT (pc $= -0.36$). In CLM, SC_Onset is highly sensitive to autumn AT (R = 0.73), and the dominant role of snow accumulation in regulating SC duration is reflected in strong partial correlations with winter SD (pc $= +0.69$) and spring AT (pc $= -0.52$). CLM weakens the direct influence of autumn AT on soil freezing duration (pc $= -0.20$ compared to -0.37 in observations), while slightly strengthening the independent role of spring AT (pc $= -0.33$ compared to -0.21 in observations). Collectively, the amplified correlations indicate that excessive snow accumulation in ERA5L and CLM produces a tighter-than-observed coupling between snowpack evolution and subsurface freezing, thereby reducing the influence of other environmental controls on SFG variability.

JSBACH on the other hand, exhibits a distinctly different relationship structure, likely reflecting its substantial snow depth underestimation. SC_Onset shows no dependence on autumn AT (R = 0.07), while SFG_Onset is strongly tied to autumn AT (pc = 0.71) with no meaningful SC_Onset-SFG_Onset relationship (pc = -0.04). At the thaw end, the SC_End-SFG_End partial correlation weakens (pc = 0.18 vs. 0.37 in observations), while spring AT's direct effect on SFG_End remains very strong (pc = -0.68), indicating that spring warming forces soil thaw without mediation through snow cover. Consequently, both autumn AT (pc = -0.52) and spring AT (pc = -0.54) exert substantially stronger independent control on SFG duration than observed, bypassing the buffering role of snow insulation. A notable sign reversal in JSBACH where SFG_Onset shows a strong negative relationship with SFG duration (R = -0.77), alongside a strong positive SFG_End-SFG duration relationship (R = +0.84), indicates that both earlier onset and delayed end contribute to longer frozen ground duration, whereas in observations and the other models, SFG duration is primarily controlled through the end date.

The model's limited snow accumulation and relatively shallow snowpack fail to adequately decouple the soil from atmospheric temperature variations, resulting in a more direct soil thermal response to AT. Together, these results highlight that observed SFG characteristics arise from a balanced interaction between snow dynamics and seasonal thermal/freeze energy, whereas models diverge in how they distribute control between AT and snow insulation. ERA5L and CLM exaggerate the thermal decoupling effect of snow, while JSBACH underrepresents it, leading to contrasting biases in the simulated freeze-thaw behavior.

Minor comments

Comment 1: Abstract. SFG should be defined when it first appears.)?

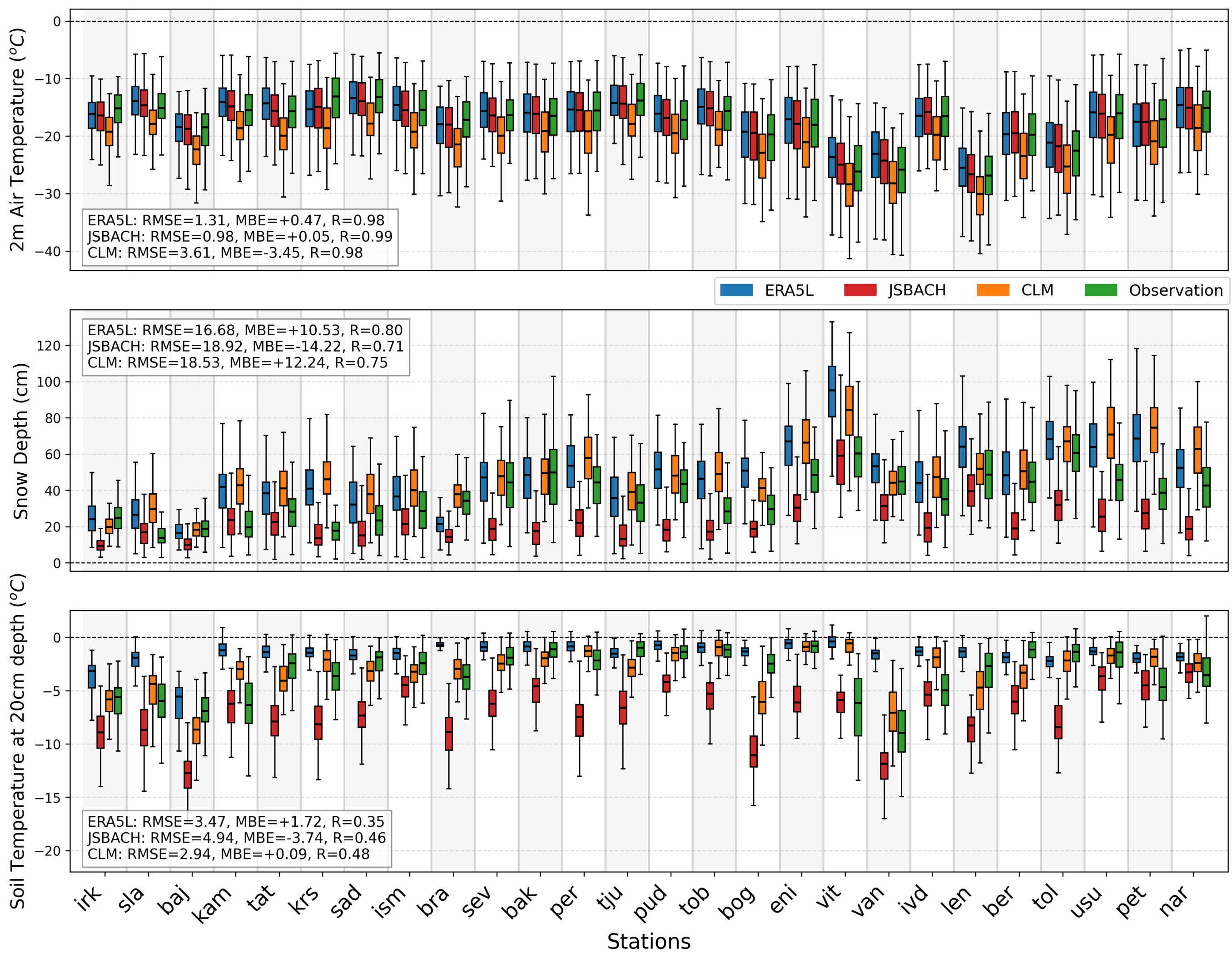
Response: We thank the reviewer for this comment. We apologize for the oversight and will revise the manuscript to define SFG at its first occurrence in the abstract.

Comment 2: In many Figures, the labels (a), (b), and (c)… are not indicated.)?

Response: We thank the reviewer for pointing this out. We will ensure that all figures are updated to include proper panel labels (a), (b), (c), etc. in the revised manuscript.

Comment 3: It is hard to get the core information from the Figure 6. A more concise Figure is needed to summarize the results.)?

Response: We thank the reviewer for this comment. We agree that showing individual sites can reduce readability at first glance. However, we have retained this figure because it provides important information on site-specific variability and allows direct assessment of how snow depth underestimation or overestimation influences model performance at individual locations. We consider this detail important, as it would be lost in a fully aggregated presentation. To avoid any confusion, we have also provided aggregated performance metrics (RMSE, MBE, and R) across all sites, which summarize the overall model behaviour more concisely. To improve clarity, we will revise the figure layout by introducing clearer separation between sites using distinct vertical groupings, as shown in the revised version to be included in the manuscript.



References

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., Van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., Van Den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245-4287, https://doi.org/10.1029/2018MS001583, 2019.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M., and Swenson, S. C.: The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, Biogeosciences, 10, 7109-7131, https://doi.org/10.5194/bg-10-7109-2013, 2013.

Gualdi, S., Borrelli, A., Cantelli, A., Davoli, G., del Mar Chavesmontero, M., Masina, S., Navarra, A., Sanna, A., and Tibaldi, S.: The new CMCC operational seasonal prediction system, CMCC Technical Notes, TN0288, https://www.cmcc.it/wp-content/uploads/2020/09/TN0288-csp-09-2020-1.pdf, 2020.

Ekici, A., Beer, C., Hagemann, S., Boike, J., Langer, M., and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model, Geoscientific Model Development, 7, 631-647, https://doi.org/10.5194/gmd-7-631-2014, 2014.

Yang, K. and Zhang, J.: Spatiotemporal characteristics of soil temperature memory in China from observation, Theor Appl Climatol, 126, 739-749, https://doi.org/10.1007/s00704-015-1613-9, 2016.

Koster, R. D. and Suarez, M. J.: Soil Moisture Memory in Climate Models, J. Hydrometeor, 2, 558-570, https://doi.org/10.1175/1525-7541(2001)002%3C0558:SMMICM%3E2.0.CO;2, 2001.

Vinnikov, K. Y., Robock, A., Speranskaya, N. A., and Schlosser, C. A.: Scales of temporal and spatial variability of midlatitude soil moisture, J. Geophys. Res., 101, 7163-7174, https://doi.org/10.1029/95JD02753, 1996.

Lynch-Stieglitz, M.: The Development and Validation of a Simple Snow Model for the GISS GCM, Journal of Climate, 7, 1842–1855, https://doi.org/10.1175/1520-0442(1994)007%3C1842:TDAVOA%3E2.0.CO;2, 1994.

Zhang, T.: Influence of the seasonal snow cover on the ground thermal regime: An overview, Reviews of Geophysics, 43, 2004RG000157, https://doi.org/10.1029/2004RG000157, 2005.

Zhang, T., Frauenfeld, O. W., Serreze, M. C., Etringer, A., Oelke, C., McCreight, J., Barry, R. G., Gilichinsky, D., Yang, D., Ye, H., Ling, F., and Chudinova, S.: Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin, J. Geophys. Res., 110, 2004JD005642, https://doi.org/10.1029/2004JD005642, 2005.

Guo, D. and Wang, H.: Simulation of permafrost and seasonally frozen ground conditions on the Tibetan Plateau, 1981–2010, JGR Atmospheres, 118, 5216–5230, https://doi.org/10.1002/jgrd.50457, 2013.

Wang, W., Rinke, A., Moore, J. C., Ji, D., Cui, X., Peng, S., Lawrence, D. M., McGuire, A. D., Burke, E. J., Chen, X., Decharme, B., Koven, C., MacDougall, A., Saito, K., Zhang, W., Alkama, R., Bohn, T. J., Ciais, P., Delire, C., Gouttevin, I., Hajima, T., Krinner, G., Lettenmaier, D. P., Miller, P. A., Smith, B., Sueyoshi, T., and Sherstiukov, A. B.: Evaluation of air–soil temperature relationships simulated by land surface models during winter across the permafrost region, The Cryosphere, 10, 1721–1737, https://doi.org/10.5194/tc-10-1721-2016, 2016.

Damseaux, A.: Verbesserung der Permafrostdynamik in Landoberflächenmodellen: Erkenntnisse aus doppelten SensitivitätsexperimentenImproving permafrost dynamics in land surface models: insights from dual sensitivity experiments, Universität Potsdam, https://doi.org/10.25932/PUBLISHUP-63945, 2024.

Luo, Z., Risto, D., and Ahrens, B.: Assessing uncertainties in modeling the climate of the Siberian frozen soils by contrasting CMIP6 and LS3MIP, The Cryosphere, 19, 6547–6576, https://doi.org/10.5194/tc-19-6547-2025, 2025.

Response to Reviewer#3 Comments

Comment 1:

Differences in soil depth and vertical discretization can substantially affect heat and water transport in soils. Given that the three systems adopt different soil layering schemes, it remains unclear to what extent the simulated differences in soil thermal and hydrological processes arise from physical process representations versus structural differences in vertical discretization.

Response:

We thank the reviewer for this important and perceptive comment. We agree that differences in soil depth and vertical discretization can substantially influence the simulation of soil thermal and hydrological processes, and that this introduces inherent challenges in attributing inter-model differences to either physical process representations or structural configuration. Since JSBACH and CLM differ simultaneously in their vertical discretization, soil column depth, and physical parameterizations, separating the contributions of these factors in a fully controlled manner would require targeted sensitivity experiments with consistent model physics and varying vertical configurations. While such experiments would be scientifically valuable, they are beyond the scope of the present study.

This is a well-recognised challenge in the land surface modelling community. Koven et al. (2013) and Slater and Lawrence (2013) demonstrated a broad range of inter-model spread in simulated soil temperatures across CMIP5 models, attributing the uncertainty to a combination of snow insulation, land–atmosphere coupling, and vertical model resolution. González-Rouco et al. (2021) further demonstrated that increasing JSBACH's soil column depth from 9.83 m to over 1400 m substantially reduced deep soil warming by 0.5-1.5 K in scenario simulations, with energy storage 3-5 times larger in the deeper configuration, underscoring that vertical structure alone can introduce biases independent of process parameterizations.

The primary goal of this study is to evaluate how two widely used land surface models within ESMs, with fundamentally different configurations: JSBACH employing a comparatively coarse vertical discretization (5 soil layers to $\sim$9.83 m depth) and CLM5 a much finer soil column structure (20 soil layers extending to 8.6 m depth, plus 5 bedrock layers to $\sim$50 m), represent seasonally frozen ground characteristics. This comparison provides direct insight into their respective strengths and limitations for frozen ground applications.

Therefore, in response to the reviewer's comment, we agree that this distinction deserves explicit acknowledgement in the manuscript. **Accordingly, we have added the following paragraph in the revised manuscript at Line 345:**

In addition, differences in vertical soil discretization can affect heat and water transport in the soil column. Koven et al. (2013) and Slater and Lawrence (2013) identified vertical resolution, together with snow insulation and land surface physics, as key sources of uncertainty in simulating cold-region soil thermal states, while González-Rouco et al. (2021) demonstrated that soil column depth alone can substantially influence long-term subsurface temperature dynamics. Therefore, the inter-model differences reported in this study should be interpreted as reflecting the combined influence of structural configuration and physical process representation, rather than being attributable to either factor in isolation. Furthermore, since all simulations are driven by .... (Line 345-351)

We will also update Table 1 in the revised manuscript to include detailed information on the soil layer structures of both models as follows:

Table 1: Key structural and parametric differences in soil and snow process representations between JSBACH and CLM

Model features	**JSBACH**	**CLM**
Total soil column depth	∼9.8 m	∼49.5 m (bedrock below 8.6 m)
Number of soil layers (default)	5 soil layers	20 soil layers + 5 bedrock layers
Soil layer node depths (m)	0.0325, 0.192, 0.7755, 2.683, 6.984	0.01, 0.04, 0.09, 0.16, 0.26, 0.4, 0.58, 0.8, 1.06, 1.36, 1.7, 2.08, 2.5, 2.99, 3.58, 4.27, 5.06, 5.95, 6.94, 8.03, 9.795, 13.328, 19.483, 28.871, 41.998
Soil layer thickness (m)	0.065, 0.254, 0.913, 2.902, 5.7	0.02, 0.04, 0.06, 0.08, 0.12, 0.16, 0.2, 0.24, 0.28, 0.32, 0.36, 0.4, 0.44, 0.54, 0.64, 0.74, 0.84, 0.94, 1.04, 1.14, 2.39, 4.676, 7.635, 11.14, 15.115
Number of snow layers (default)	5 (hydrologically inactive)	10
Number of PFTs	11	15
Concept of supercooled water	Yes (Niu and Yang, 2006)	Yes (Niu and Yang, 2006)
Soil heat capacity scheme	(De Vries (1953))	(De Vries (1953))
Soil thermal conductivity	(Johansen (1977))	(Farouki (1981))
Bottom boundary condition	Zero flux	Zero flux
Snow density formulation	Temperature dependent	Multi-process: compaction, destructive and constructive, melt metamorphism (Anderson, 1976)
Fresh snow density parameterization	50 kg m^{-3} (constant)	Temperature dependent
Snow thermal conductivity	(Calonne et al. (2011))	(Jordan (1991))

Comment 2:

Soil ice plays an important role in regulating soil thermal properties and heat transfer. Although ERA5-Land does not explicitly represent soil ice, a more direct comparison between JSBACH and CLM in terms of soil ice processes could provide additional insight.

Response:

We thank the reviewer for this comment. The above mentioned aspect has already been addressed in the manuscript. While ERA5-Land does not explicitly represent soil ice, we provided a direct comparison of soil ice content between JSBACH and CLM in the Appendix (Figure B1), which offers additional insight into the role of soil ice processes. Furthermore, taking advantage of the availability of the ESA CCI surface soil moisture dataset, we have also included a small evaluation of simulated liquid soil moisture in the top layer against this reference data (Figure B2).
These figures together reveal important differences in soil ice and liquid water dynamics between the two models. In cold regions, CLM simulates a larger fraction of total soil moisture in the frozen/ice phase compared to JSBACH (Figure B1), indicative of stronger near-surface soil freezing in CLM. Evaluation of top-layer liquid soil moisture against ESA CCI observations (Figure B2) reveals that CLM tends to underestimate liquid soil moisture across boreal and Arctic regions. While summer thaw periods lead to increased liquid water content, autumn soil moisture in CLM remains systematically lower than observed, suggesting persistently drier near-surface conditions prior to freeze-up that may further promote and accelerate the onset of freezing in autumn.
The corresponding information is already included in the manuscript **(Lines 297-301)**.

Comment 3:

While JSBACH appears to perform better in simulating SFG, it lacks an explicit coupling between snow and soil water processes. This missing process may be important, particularly as the manuscript emphasizes the role of snow. Beyond albedo and insulation effects, snowmelt dynamics at high latitudes may also significantly influence soil thermal regimes. There is a concern that compensating errors among missing or simplified processes could lead to apparently improved performance. These limitations should be more explicitly acknowledged, potentially in the Conclusions.

Response:

We would like to clarify that in the manuscript and conclusion, we have only stated that JSBACH better captures frozen ground extent compared to the other two datasets. For other aspects, such as SFG and snow cover characteristics, JSBACH performs poorly due to different process representations (e.g., underestimation of snow depth and weaker insulation), just as CLM and ERA5L do for different reasons. We agree with the reviewer that, beyond albedo and insulation effects, snowmelt dynamics can play an important role in influencing soil thermal regimes, particularly in high-latitude regions. While our analysis discusses the role of snow cover timing (e.g., onset and end date), we acknowledge that the explicit role of snowmelt processes and their interaction with soil hydrology was not sufficiently emphasized. We also agree that this limitation should be more clearly acknowledged, particularly in the context of interpreting model performance.
Therefore, we revised the manuscript (1. Lines 510 and 2. Lines 522-524) to acknowledge this limitation and highlight the importance of snowmelt dynamics.
1. At most stations, the models exhibited premature soil freezing and delayed spring thaw, reflecting persistent challenges in capturing autumn land-atmosphere interactions and snow evolution processes **(including snowmelt dynamics)**.
2. "Future studies should focus **on site-level performance to investigate the combined effects of snow accumulation and snowmelt processes**, vegetation, soil properties, and soil moisture on soil freeze-thaw dynamics, in order to enhance land surface model performance."

Comment 4:

In Sect. 3.3.4, the authors provide a detailed analysis of controlling factors. However, a schematic diagram illustrating how these factors influence soil temperature and, consequently, SFG would greatly improve clarity, especially given the extensive use of abbreviations.

Response:

We thank the reviewer for this important suggestion. In response, we have decided to add a partial correlation table, as also suggested by Reviewer #2, together with a schematic diagram summarizing the statistically significant partial correlations between SFG characteristics, snow cover properties, and seasonal thermal drivers across observations and models (Fig. 1). The controlling variables used for each variable pair are listed in Table 2 and were selected based on physically motivated relationships to isolate direct from indirect pathways. To further visualise the relationships presented in Table 3, including both pearson correlation and partial correlation values calculated using the controlling parameters listed in Table 2, we constructed a schematic diagram (Fig. 1) for observations and each model. This diagram helps distinguish genuine physical drivers from spurious inter-correlations.

Variable definitions (for reference):

- SFG_Onset - onset date of seasonally frozen ground at 20 cm depth.
- SFG_End - end date of seasonally frozen ground at 20 cm depth
- SFGD - seasonally frozen ground duration (SFG_End - SFG_Onset)
- SC_Onset - snow cover onset date
- SC_End - snow cover end date
- SCD - snow cover duration (SC_End - SC_Onset)
- Autumn_AT - autumn air temperature
- Winter_AT - winter air temperature
- Spring_AT - spring air temperature
- Winter_SD - winter snow depth

Partial correlations (pc) were computed to remove the influence of shared drivers, particularly Autumn_AT, Spring_AT, and Winter_SD, which co-vary with multiple target variables simultaneously. This approach allows us to determine, for example, whether the relationship between SFG_Onset and SC_Onset is direct or primarily inherited from a common autumn temperature signal.

To illustrate the calculation, we provide an example of the partial correlation between parameters A and B, with C held as the control variable:

$$pc_{AB\cdot C} = \frac{R_{AB} - R_{AC}R_{BC}}{\sqrt{(1 - R_{AC}^2)\,(1 - R_{BC}^2)}} \tag{1}$$

Observation: In the observations, SC_Onset and Autumn_AT show a direct moderate positive relationship (R = 0.3), indicating that warmer autumn temperatures delay snow cover onset. SC_Onset and SFG_Onset exhibits virtually no direct relationship (partial r = 0.02 after controlling for Autumn_AT), indicating that frozen ground onset is not directly driven by snow arrival and is instead likely governed by local factors such as soil properties, vegetation cover, soil moisture, and microclimatic conditions. Similarly, after controlling for SC_Onset, the relationship between Autumn_AT and SFG_Onset is very weak (pc = 0.14), suggesting only limited direct influence of Autumn_AT on soil freezing onset. SC duration is primarily controlled by Winter_SD (pc = -0.52), Spring_AT (pc = -0.50), and Autumn_AT (pc = -0.33). SC_End shows strong dependence on Spring_AT (pc = -0.66) and Winter_SD (pc = +0.52), and a moderate relationship with SFG_End after controlling for spring temperature (pc = +0.37). In contrast, SFG_End is only weakly directly influenced by Spring_AT once the effect of SC_End is removed (pc = -0.20), highlighting that soil thaw timing is mediated through snow cover dynamics rather than temperature alone. In the observations, SFG duration shows a moderate negative relationship with both Autumn_AT (pc = -0.37) and Spring_AT (pc = -0.21) after controlling for SC duration, indicating that colder autumn and spring conditions contribute weakly to longer frozen ground duration independent of snow cover persistence.

ERA5L: In ERA5L, the schematic indicates no direct relationship between SFG_Onset and Autumn_AT. SC_End, however, shows relatively strong relationships with both SFG_End (pc = 0.85) and Spring_AT (pc = -0.69) when the influence of other variables is accounted for. In addition, the direct influence of Spring_AT on SFG_End (pc = -0.42) remains stronger than observed, suggesting that ERA5L overestimates the sensitivity of soil thaw to spring warming. A higher SFG duration-SC duration correlation (R= 0.53) in ERA5L than in observations (R = 0.38)) indicates that the model overestimates the coupling between snow cover duration and frozen ground duration. SFG duration maintains moderate negative relationships with both Autumn_AT (pc = -0.32) and Spring_AT (pc = -0.36) after controlling for SC duration, indicating a slightly stronger sensitivity of frozen ground duration to seasonal warming than observed.

CLM: CLM reproduces the correct directional sign across all major linkages, but SC_Onset's dependence on Autumn_AT is amplified (R = 0.73), suggesting snow onset timing is highly sensitive to autumn temperature conditions. Most critically, the SC_Onset-SFG_Onset relationship collapses to near zero, while the Autumn_AT-SFG_Onset relationship remains weak (pc = +0.09), which may suggest that the relatively high snow depth in CLM could partially thermally isolate the soil, reducing the influence of snow onset timing and autumn air temperature on soil freezing at 20 cm depth. SFG duration shows a stronger dependence on SC duration (R = 0.57) while SC duration is strongly controlled by Winter_SD (pc = +0.69), Spring_AT (pc = -0.52) and Autumn_AT (pc = -0.35), reflecting the dominant role of snow accumulation and melt-season temperature in regulating snow cover duration. Furthermore, the SC_End-SFG_End partial correlation is substantially amplified (pc = 0.84) after controlling for Spring_AT, indicating that snow disappearance strongly regulates soil thaw timing in CLM. Once SC duration and Spring_AT are controlled, Autumn_AT's independent effect on SFG duration weakens to pc = -0.20 in CLM (vs -0.37 in observations), while Spring_AT's independent effect (pc = -0.33) remains slightly stronger than observed (pc = -0.21), suggesting that

CLM further suppresses the direct autumn thermal pathway to the soil while retaining an elevated spring signal.

JSBACH: JSBACH exhibits a distinctly different relationship structure, likely reflecting its substantial snow depth underestimation. Unlike observations and CLM, SC_Onset shows virtually no dependence on Autumn_AT (R = 0.07), suggesting that snow accumulation timing in JSBACH is largely disconnected from autumn thermal conditions. The SFG_Onset and Autumn_AT partial correlation surges to pc = 0.71, the strongest signal in the entire network and far exceeding the observed value (pc = 0.14), while SC_Onset and SFG_Onset show no meaningful relationship (pc = -0.04), together confirming that in the absence of adequate snow cover, Autumn_AT drives soil freezing directly without the moderating influence of snow. SFG duration shows a stronger dependence on SC duration (R = +0.56) while SC duration is strongly controlled by Winter_SD (pr = +0.66), Spring_AT (pc = -0.35), and Autumn_AT (pc = -0.27). The SC_End and SFG_End coupling weakens substantially (pc = 0.18 vs 0.37 in observations), while Spring_AT's independent effect on SFG_End remains quite strong even after controlling for SC_End (pc = -0.68), indicating that without adequate snow insulation, spring air temperature forces soil thaw directly rather than through snow cover. In JSBACH, Autumn_AT's independent effect on SFG duration (pc = -0.52) and Spring_AT's independent effect (pc = -0.54) both remain considerably stronger than observed, confirming that seasonal air temperatures control frozen ground duration directly, bypassing snow insulation. In JSBACH, the sign change for SFG_Onset (strong negative dependency on SFGD, R = -0.77) together with the positive dependency of SFG_End on SFG duration (R = +0.84) suggests that both early onset and delayed end contribute to longer frozen ground duration, whereas in other datasets SFG duration shows stronger control by SFG_End.

Table 2: Relationships between seasonal air temperatures, snow depth, snow cover, and seasonally frozen ground characteristics with controlling factors and physical interpretation.

Parameters	**Controlling Factors**	**Remarks**
Autumn_AT × Winter_SD	–	Independent environmental drivers
Winter_AT × Winter_SD	–	Independent environmental drivers
SC_Onset × Autumn_AT	–	Direct autumn temperature control on snow onset
SFG_Onset × SC_Onset	Autumn_AT	Remove shared effect of autumn temperature
SFG_Onset × Autumn_AT	SC_Onset	Isolate direct thermal effect on soil freezing
SCD × Autumn_AT SCD × Spring_AT SCD × Winter_SD	Winter_AT, Spring_AT, Winter_SD Winter_AT, Autumn_AT, Winter_SD Winter_AT, Spring_AT, Autumn_AT	SCD is influenced by seasonal air temperatures and snow depth; controls remove indirect pathways and isolate direct effects.
SC_End × Spring_AT	Winter_AT, Winter_SD	Isolate spring warming effect on snow disappearance
SC_End × Winter_SD	Winter_AT, Spring_AT	Snow depth effect on melt date excluding temperature and thaw influences
SC_End × SCD SC_Onset × SCD	– –	Mathematically linked (duration = end - onset); no control variable required.
SC_End × SFG_End	Spring_AT	Remove shared effect of spring temperature
SFG_End × Spring_AT	SC_End	Direct spring temperature control on SFG onset.
SFGD × SCD	–	Includes the influence of all other factors to show total coupling effect.
SFGD × Autumn_AT	SCD, Spring_AT	Isolates the direct influence of autumn air temperature
SFGD × Spring_AT	SCD, Autumn_AT	Isolates the direct influence of spring air temperature
SFG_End × SFGD SFG_Onset × SFGD	– –	Mathematically linked (duration = end - onset): no control needed.

Table 3: Statistically significant (p value < 0.01) pearson and partial correlations between seasonal air temperatures, winter snow depth, snow cover, and seasonally frozen ground characteristics.

Parameters	**Pearson Correlation (R)**				**Partial Correlation (pc)**			
	Obs	**ERA5L**	**CLM**	**JSBACH**	**Obs**	**ERA5L**	**CLM**	**JSBACH**
Autumn_AT × Winter_SD	-0.59	-0.59	-0.51	-0.58	-	-	-	-
Winter_AT × Winter_SD	-0.34	-0.35	-0.16	-0.53	-	-	-	-
SC_Onset × Autumn_AT	0.3	0.71	0.73	0.07	-	-	-	-
SFG_Onset × SC_Onset	0.07	0.2	0.1	0.02	0.02	0.15	0.004	-0.04
SFG_Onset × Autumn_AT	0.15	0.13	0.13	0.71	0.14	-0.02	0.09	0.71
SCD × Autumn_AT	-0.7	-0.74	-0.68	-0.64	-0.33	-0.4	-0.35	-0.27
SCD × Spring_AT	-0.68	-0.7	-0.68	-0.52	-0.5	-0.5	-0.52	-0.35
SCD × Winter_SD	0.71	0.77	0.78	0.76	0.52	0.62	0.69	0.66
SC_End × Spring_AT	-0.74	-0.77	-0.72	-0.57	-0.66	-0.69	-0.65	-0.51
SC_End × Winter_SD	0.62	0.7	0.73	0.66	0.52	0.61	0.67	0.61
SC_End × SCD	0.89	0.91	0.92	0.84	-	-	-	-
SC_Onset × SCD	-0.34	-0.87	-0.87	0.02	-	-	-	-
SC_End × SFG_End	0.63	0.94	0.92	0.54	0.37	0.85	0.84	0.18
SFG_End × Spring_AT	-0.57	-0.82	-0.72	-0.78	-0.2	-0.42	-0.23	-0.68
SFGD × SCD	0.38	0.53	0.57	0.56	-	-	-	-
SFGD× Autumn_AT	-0.52	-0.58	-0.53	-0.73	-0.37	-0.32	-0.2	-0.52
SFGD × Spring_AT	-0.42	-0.59	-0.6	-0.71	-0.21	-0.36	-0.33	-0.54
SFG_End × SFGD	0.52	0.69	0.75	0.84	-	-	-	-
SFG_Onset × SFGD	0.26	0.31	0.14	-0.77	-	-	-	-

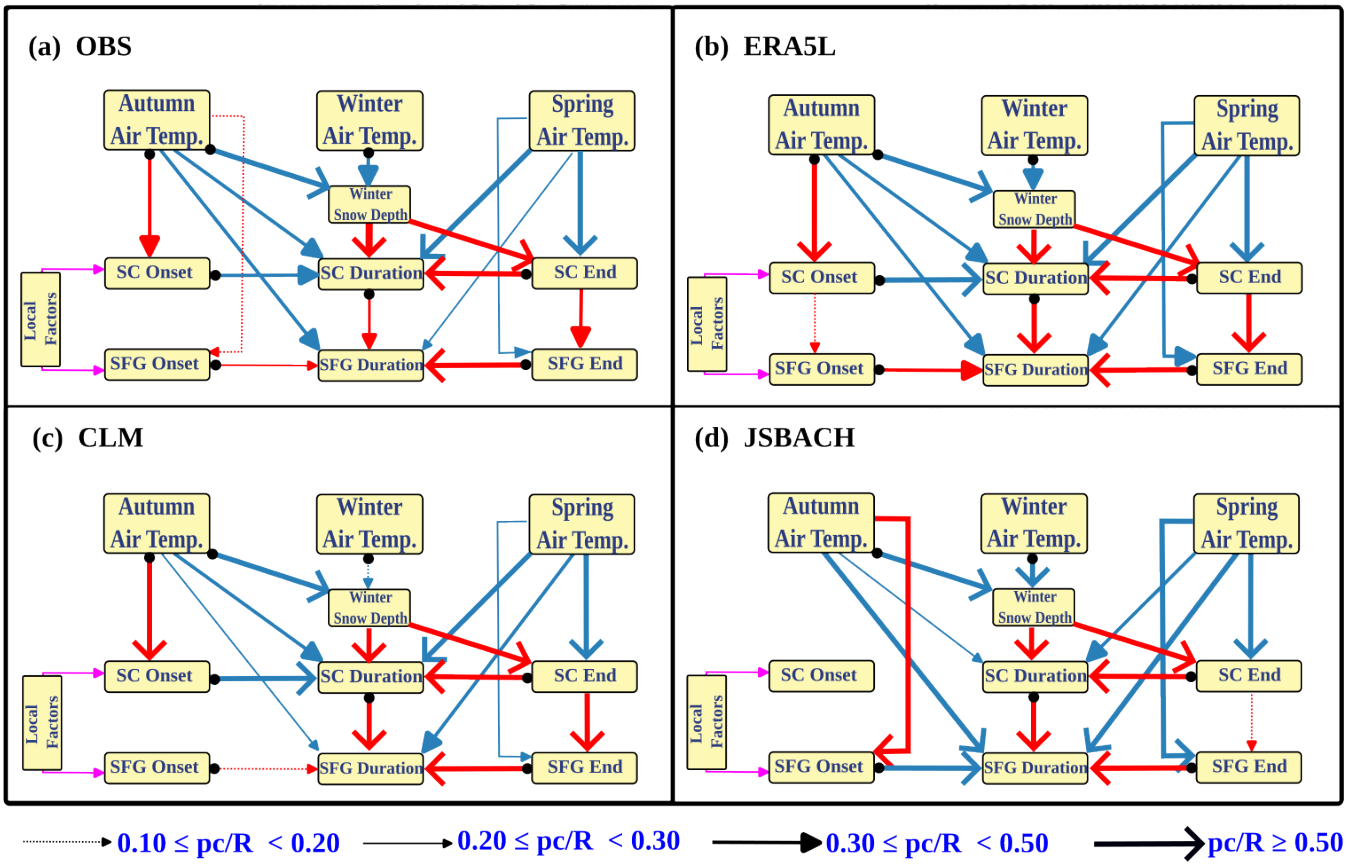


Figure 1: Schematic representation of the statistically significant pearson and partial correlations derived from Appendix Table 3, illustrating the interactions among seasonal air temperature, snow cover, snow depth, and seasonally frozen ground characteristics for (a) OBS, (b) ERA5L, (c) CLM, and (d) JSBACH. Red and blue lines denote positive and negative correlations, respectively, while the magenta-coloured line indicates possible local influences affecting the variables. Filled circles at the arrow origin indicate pearson correlation coefficients; all other values represent partial correlations.

Based on the information we have, we would like to correct our Sect 3.2.4 as follows in the revised manuscript:

Figure 10 shows statistically significant correlation matrices between SFG and snow cover characteristics, seasonal ATs, winter SD, as well as thermal indices for observations and models. To further distinguish direct relationships from shared seasonal influences, partial correlation (pc) analysis was additionally calculated (using `pingouin` package; `https://pingouin-stats.org/generated/pingouin.partial_corr.html`). The resulting partial correlation values are presented in Appendix Table 3, while the corresponding controlling factors used in each analysis are listed in Appendix Table 2. This approach removes the effect of co-varying seasonal controls (autumn AT, spring AT, winter AT and winter SD), allowing a clearer assessment of whether snow cover exerts an independent influence on soil freezing dynamics or whether observed linkages arise primarily through common atmospheric forcing. Figure 1 presents a schematic diagram based on Appendix Table 3, illustrating both pearson correlations (R) and partial correlations (pc) after accounting for shared drivers, thereby highlighting the relative strength and direction of the key interactions among variables.

In observations, SFG onset is not directly controlled by large-scale seasonal conditions. SC_Onset and autumn AT show a direct moderate positive relationship ($R = 0.30$), indicating that warmer autumn temperatures delay snow cover onset, consistent with a temperature-driven control, yet autumn AT exerts only limited direct influence on SFG onset (pc = 0.14), and the mutual partial correlation between SC_Onset and SFG_Onset is negligible (pc = 0.02 after controlling for autumn AT). SC duration is primarily controlled by winter SD (pc = −0.52), spring AT (pc = −0.50), and autumn AT (pc = −0.33). SC_End shows strong dependence on spring AT (pc = −0.66) and winter SD (pc = +0.52), and a moderate relationship with SFG_End after controlling for spring temperature (pc = +0.37). In contrast, SFG_End is correlated with the thawing index (R = -0.50) but is only weakly directly influenced by spring AT once the effect of SC_End is removed (pc = −0.20), highlighting that soil thaw timing is largely mediated through snow cover dynamics rather than temperature alone. SFG duration shows the strongest association with autumn AT (R = −0.52; pc = −0.37 after controlling for SC duration and spring AT), with spring AT retaining a moderate independent effect on SFG duration after controlling for SC duration (pc = −0.21). Altogether, the observations indicate that snow cover governs thaw timing and, together with the intensity of seasonal air temperature, determines the duration of seasonal ground freezing.

ERA5L and CLM broadly reproduce several of the observed dependencies but often amplify the strength of correlations, suggesting an excessive sensitivity of the simulated soil thermal regime to AT and snow controls. Moreover, both models do not account for the high heterogeneity in local conditions present at observation sites, as land surface properties are spatially smoothed, which likely contributes to the strong dependence of snow cover onset on seasonal variables, particularly autumn AT. In addition, precipitation phase partitioning in the models is governed by grid-scale temperature thresholds, making the onset of snow accumulation primarily temperature-driven. Consequently, local factors that modify the near-surface temperature regime, such as vegetation dynamics, soil thermal properties, microtopography, and surface energy exchanges, cannot be represented as specifically as in station observations. The SC_End-SFG_End coupling is particularly inflated ($R > 0.9$; pc = 0.85 in ERA5L and 0.84 in CLM after controlling for spring AT), far exceeding observations. In ERA5L, spring AT retains a stronger direct effect on SFG_End than observed (pc = −0.42), and the SFG duration-SC duration coupling is elevated (R = 0.53 vs. 0.38 in observations), with comparable sensitivity of SFG duration to both autumn AT (pc = −0.32) and spring AT (pc = −0.36). In CLM, SC_Onset is highly sensitive to autumn AT (R = 0.73), and the dominant role of snow accumulation in regulating SC duration is reflected in strong partial correlations with winter SD (pc = +0.69) and spring AT (pc = −0.52). CLM weakens the direct influence of autumn AT on soil freezing duration (pc = −0.20 compared to −0.37 in observations), while slightly strengthening the independent role of spring AT (pc = −0.33 compared to −0.21 in observations). Collectively, the amplified correlations indicate that excessive snow accumulation in ERA5L and CLM produces a tighter-than-observed coupling between snowpack evolution and subsurface freezing, thereby reducing the influence of other environmental controls on SFG variability.

JSBACH on the other hand, exhibits a distinctly different relationship structure, likely reflecting its substantial snow depth underestimation. SC_Onset shows no dependence on autumn AT (R = 0.07), while SFG_Onset is strongly tied to autumn AT (pc = 0.71) with no meaningful SC_Onset-SFG_Onset relationship (pc = -0.04). At the thaw end, the SC_End-SFG_End partial correlation weakens (pc = 0.18 vs. 0.37 in observations), while spring AT's direct effect on SFG_End remains very strong (pc = -0.68), indicating that spring warming forces soil thaw without mediation through snow cover. Consequently, both autumn AT (pc = -0.52) and spring AT (pc = -0.54) exert substantially stronger independent control on SFG duration than observed, bypassing the buffering role of snow insulation. A notable sign reversal in JSBACH where SFG_Onset shows a strong negative relationship with SFG duration (R = -0.77), alongside a strong positive SFG_End-SFG duration relationship (R = +0.84), indicates that both earlier onset and delayed end contribute to longer frozen ground duration, whereas in observations and the other models, SFG duration is primarily controlled through the end date.

The model's limited snow accumulation and relatively shallow snowpack fail to adequately decouple the soil from atmospheric temperature variations, resulting in a more direct soil thermal response to AT. Together, these results highlight that observed SFG characteristics arise from a balanced interaction between snow dynamics and seasonal thermal/freeze energy, whereas models diverge in how they distribute control between AT and snow insulation. ERA5L and CLM exaggerate the thermal decoupling effect of snow, while JSBACH underrepresents it, leading to contrasting biases in the simulated freeze-thaw behavior.

Minor comments

Comment 1: Lines 92-96: What quantities are aggregated to the grid scale (e.g., energy, hydrology, carbon)?

Response: As stated in Lines 91-92, JSBACH uses a tiling approach in which different land-cover types are simulated separately and then aggregated to the grid scale. Additive quantities, such as flux densities (energy, water, carbon) and albedo, are aggregated using area-weighted averaging based on tile fractions. However, non-additive quantities, such as surface temperature and roughness length, are treated differently: surface temperature is computed at the grid-box level to ensure energy conservation, while subsurface heat and moisture fluxes are calculated at the tile level and averaged at the end of each time step. We will update Lines 93-97 to clarify this.

Comment 2: Line 146: Why was JSBACH run without the two-phase spin-up?

Response: We thank the reviewer for this question. The spin-up procedures for both models are described in detail in our response to Reviewer #1 (Comment 1), where we clarify the differences in spin-up strategies between CLM and JSBACH. We also decided to add a brief paragraph in the revised manuscript to improve clarity.

Comment 3: Line 112: Please specify the version of CLM used.?

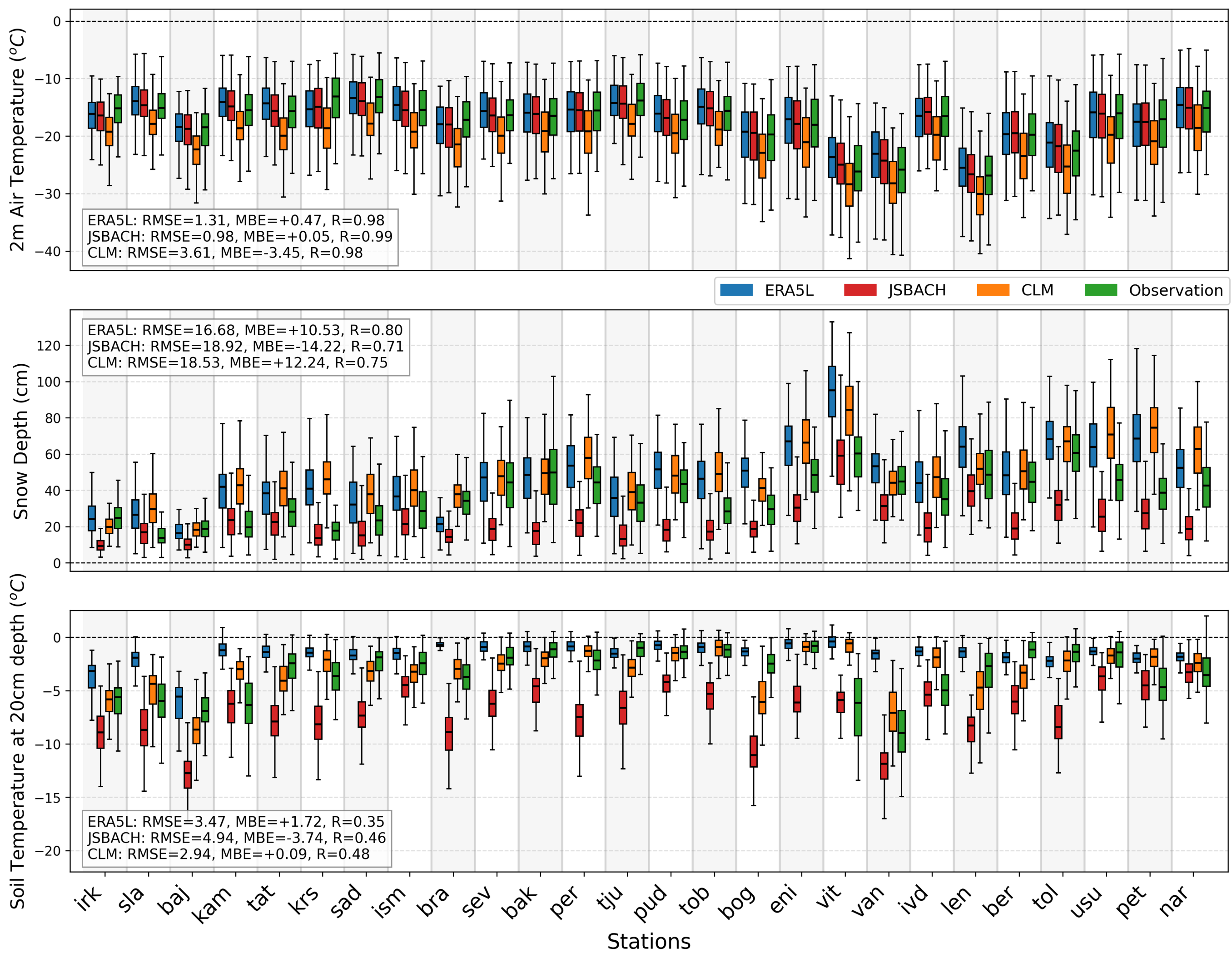
Response: The version of CLM used in this study has been specified in the manuscript (Line 120: CTSM5.1.dev128_cm3_v1 (`https://github.com/CMCC-Foundation/CTSM`; hereafter referred to as CLM)).

Comment 4: Section 3.1: Is this section better suited to the Methods rather than Results??

Response: We thank the reviewer for this helpful suggestion. We agree that this section is better suited to the Methods. Following this and similar suggestions from other reviewers, we have moved Section 3.1 to the end of the methodology section in the revised manuscript.

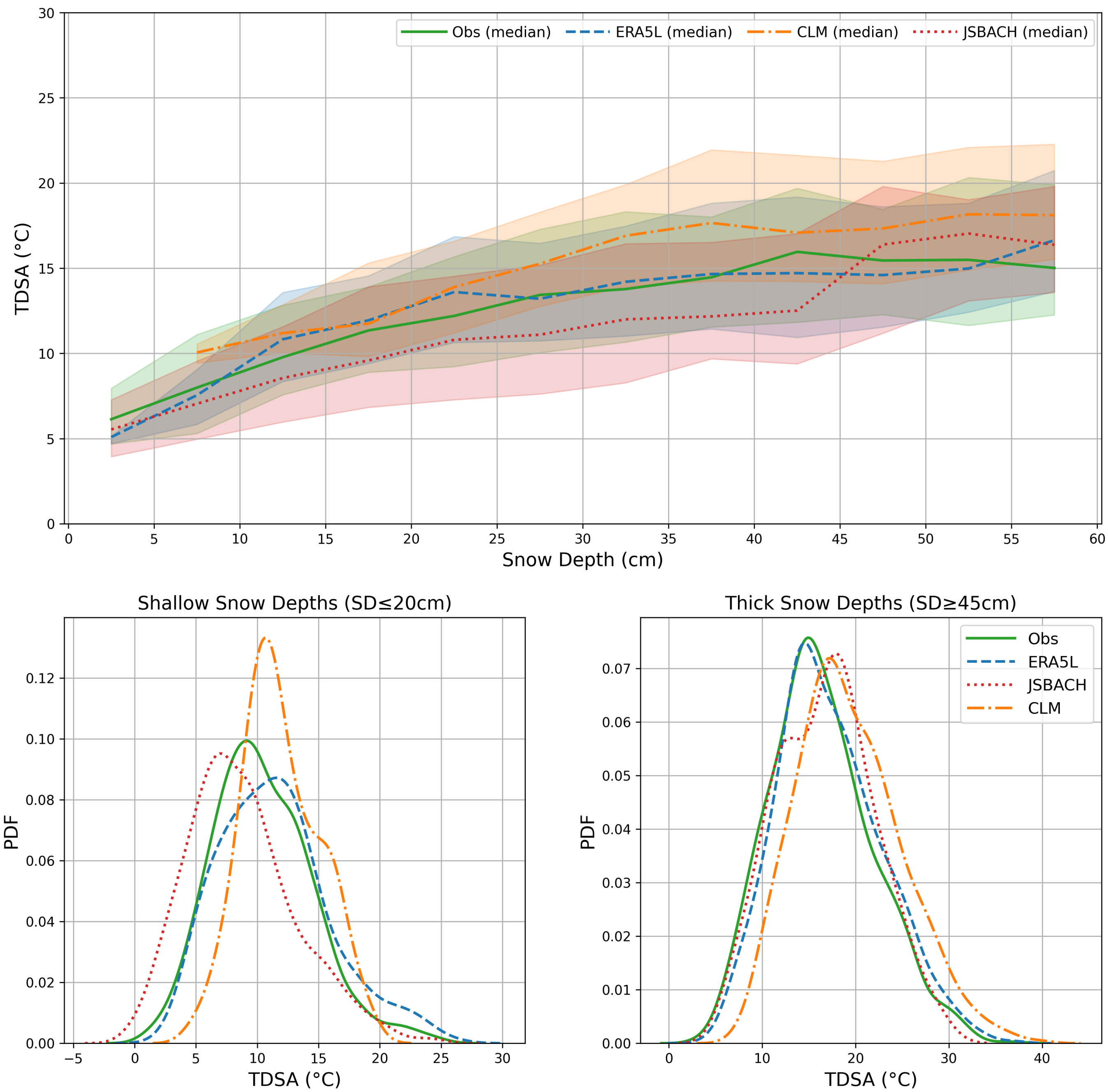
Comment 5:Figure 6: Showing results for all individual sites may reduce readability. Consider presenting site-averaged results or aggregating sites based on a classification scheme.

Response: We appreciate the concern regarding readability and acknowledge that presenting all individual sites in a single figure can be visually dense. However, we have chosen to retain this figure to allow the reader to directly assess site-specific variations and to better understand how underestimation or overestimation of snow depth influences model performance at individual locations. We believe this individual-site view provides valuable insight that would be lost through aggregation. In addition, we have provided aggregated performance metrics (RMSE, MBE, and correlation) across all sites, which offer a concise summary of overall model behaviour. To further improve readability, we plan to revise the figure layout by introducing clearer separation between sites (e.g., distinct vertical groupings), as illustrated in the figure below, which will be included in the revised manuscript.



Comment 6:Figure 7: It may improve clarity to show observations as solid black lines and model results as colored dashed lines.

Response: Following the reviewer's recommendation, we updated the figure by keeping the observational line as solid black and converting all model lines to different dashed styles while retaining the original color scheme. During this revision, we found that the shaded representation of snow-related effects reduced the visual clarity of overlapping curves, particularly against the solid black observational line. In addition, the use of different line styles improves accessibility and makes the distinction between datasets clearer, particularly for colorblind readers. This revision has been implemented in the updated figure shown below and will be included in the revised manuscript.



References

Koven, C. D., Riley, W. J., and Stern, A.: Analysis of Permafrost Thermal Dynamics and Response to Climate Change in the CMIP5 Earth System Models, Journal of Climate, 26, 1877-1900, https://doi.org/10.1175/JCLI-D-12-00228.1, 2013.

Slater, A. G. and Lawrence, D. M.: Diagnosing Present and Future Permafrost from Climate Models, Journal of Climate, 26, 5608-5623, https://doi.org/10.1175/JCLI-D-12-00341.1, 2013.

González-Rouco, J. F., Steinert, N. J., García-Bustamante, E., Hagemann, S., De Vrese, P., Jungclaus, J. H., Lorenz, S. J., Melo-Aguilar, C., García-Pereira, F., and Navarro, J.: Increasing the Depth of a Land Surface Model. Part I: Impacts on the Subsurface Thermal Regime and Energy Storage, Journal of Hydrometeorology, 22, 3211-3230, https://doi.org/10.1175/JHM-D-21-0024.1, 2021.