Response to Reviewer#2 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.

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 will clarify this point in the revised manuscript (Line 323).
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 moved 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.

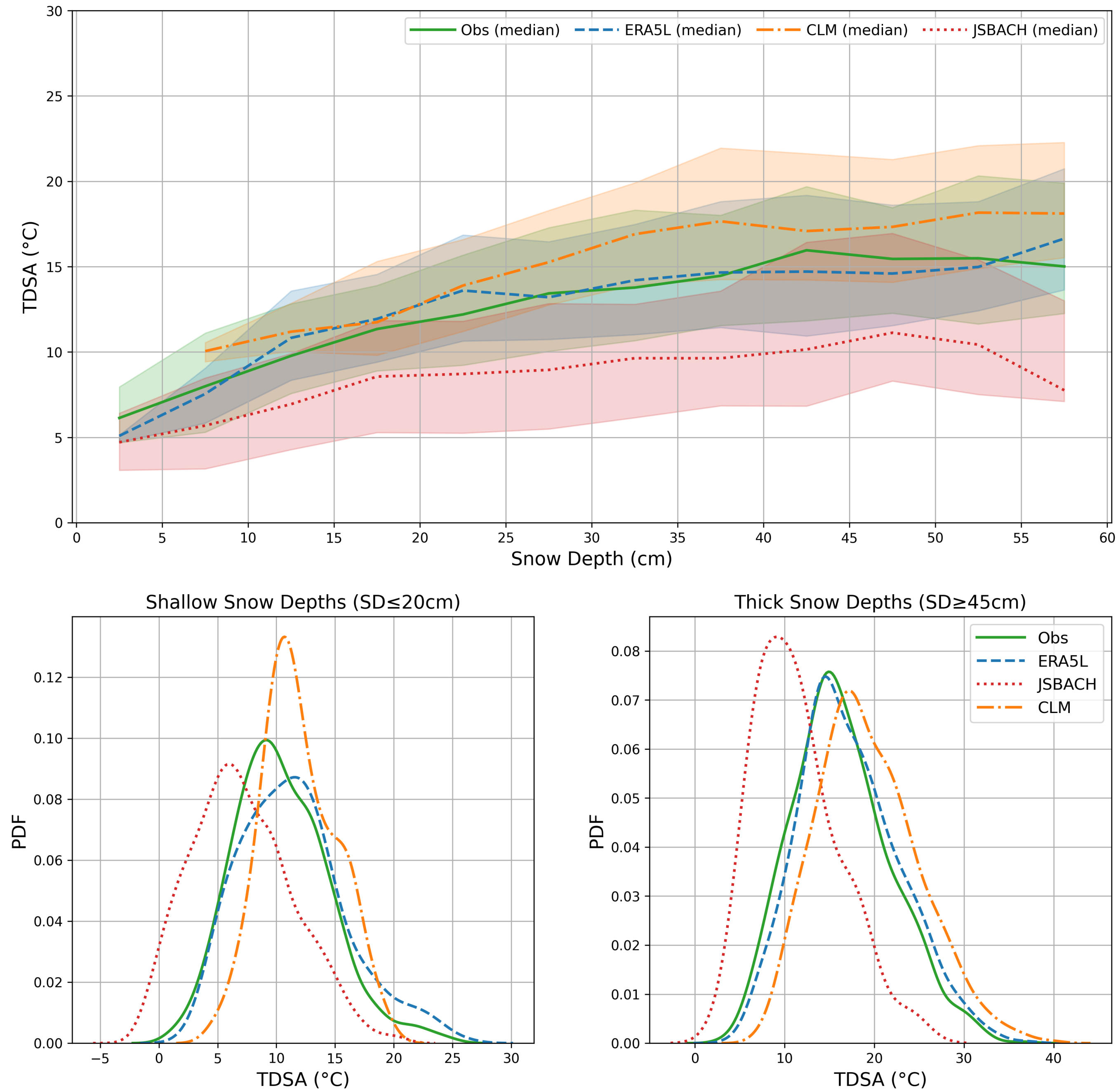


Figure 1: Relationship between wintertime monthly mean SD and soil-air temperature difference (TDSA) across observational and model datasets, shown as median TDSA (dotted lines, 5 cm SD bins) and IQR (shaded) for observations (Obs; green), ERA5L (blue), JSBACH (red), and CLM (orange) in upper panel. 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 327–330 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 influenced by seasonal AT and snow depth; control removes indirect pathways, isolating direct effect.
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 & thaw.
SC_End × SFG_End	Spring_AT	Remove shared effect of spring temperature.
SC_End × SCD SC_Onset × SCD	– –	Mathematically linked (duration = end - onset): no control needed.
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
SFD_End × SFGD	–	Mathematically linked (duration = end - onset): no control needed.
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	-	-	-	-
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	-	-	-	-

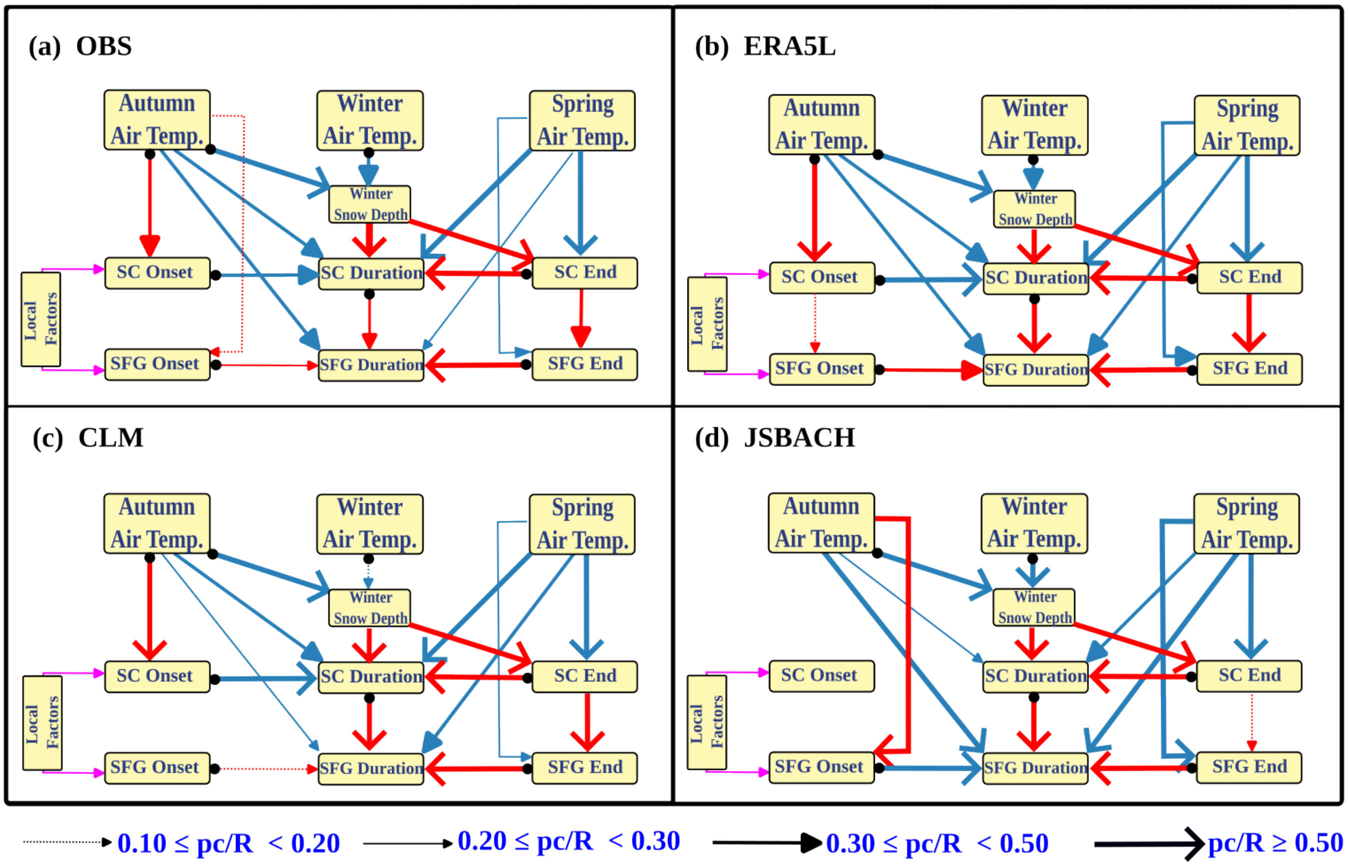


Figure 2: Schematic representation of the statistically significant pearson and partial correlations derived from Appendix Table 2, illustrating the interactions among seasonal air temperature, snow cover, snow depth, and seasonally frozen ground characteristics for (a) OBS, (b) ERA5L, (c) CLM, and (d) 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 3.3.4 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.)?

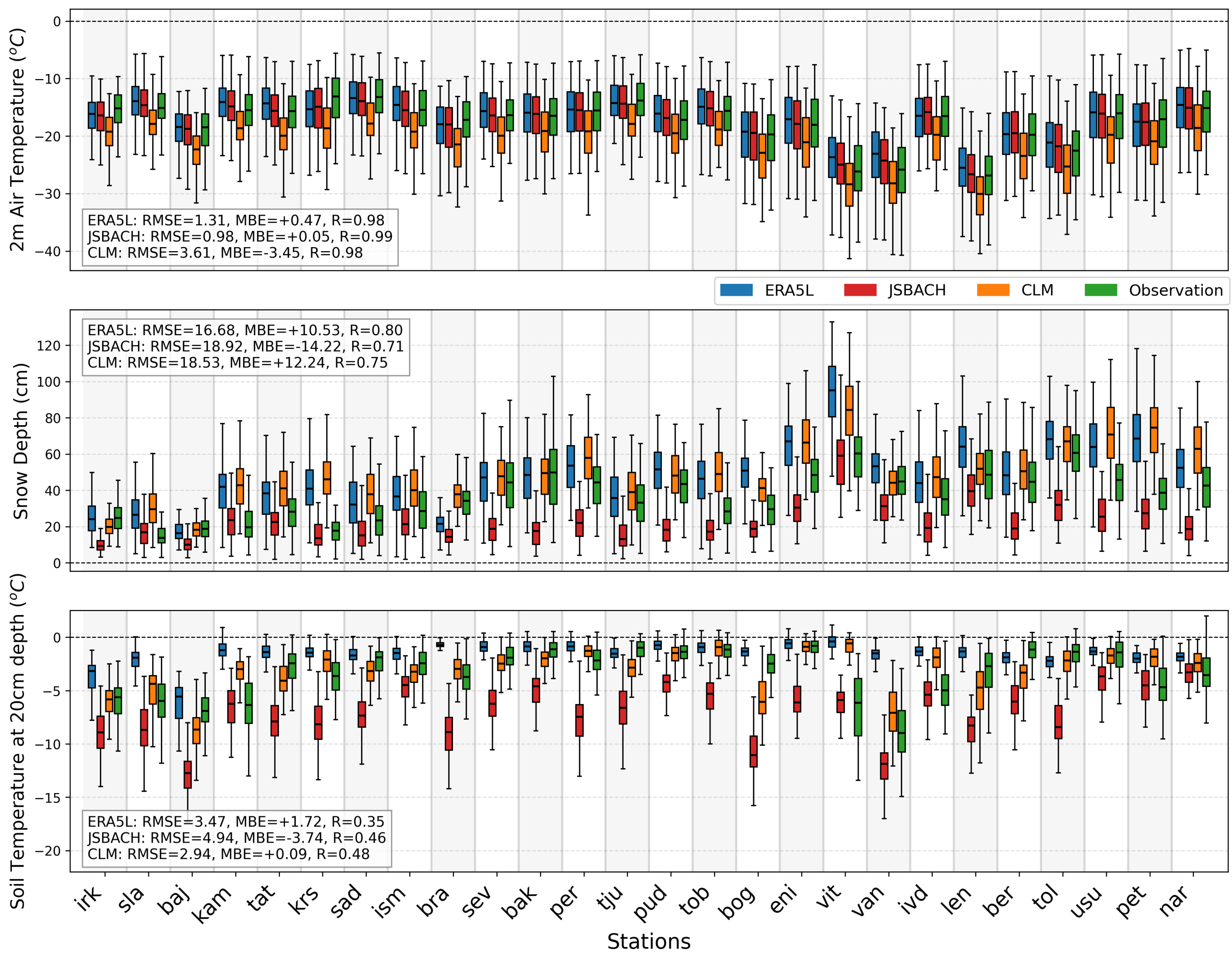
Response: We thank the reviewer for this comment. We apologize for the oversight and will revise the manuscript to define SFG at its first occurrence in the abstract.

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.