Assessing Uncertainties in Modeling the Climate of the Siberian Frozen Soils by Contrasting CMIP6 and LS3MIP

Zhicheng Luo[1], Danny Risto[1], and Bodo Ahrens[1]

[1]Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt a. M., Germany

Correspondence: Bodo Ahrens (Bodo.Ahrens@iau.uni-frankfurt.de)

Abstract. Climate models and their land components still ~~show pervasive~~ exhibit notable discrepancies in frozen soil simulations. Contrasting the historical runs of seven land-only models of the Land Surface, Snow, and Soil Moisture Model Intercomparison Project (LS3MIP) with their Coupled Model Intercomparison Project Phase 6 (CMIP6) counterparts allowed quantifying the contributions of the land surface parameterization scheme and the atmospheric forcing to the discrepancies. The simulation capabilities were assessed using observational data from 152 sites in Siberia and reanalysis data. In the winter months (December, January, and February), the LS3MIP ensemble bias in 0.2 m soil temperature was ~~higher~~ larger than the CMIP6 bias (~~-3.57~~-3.6 °C vs ~~-2.66~~-2.7 °C). The spread of winter 0.2 m soil temperatures was also larger in the LS3MIP ensemble (~~4.55~~4.6 °C) than in the CMIP6 ensemble (~~2.98~~3.0 °C). For permafrost sites, ~~the spatial correlation of the winter soil temperature simulations against observations was not better than~~ for all CMIP6 simulations, the correlations between winter soil temperatures with observations were below 0.6, and the correlations for spring/autumn ~~spatial~~ correlations of snow depth were ~~less than~~ below 0.8~~for all~~. In the CMIP6 ~~models. On average, the~~ simulations, the median 0.2 m soil ~~temperatures in the CMIP6 simulations were 0.34~~temperature was 0.3 °C warmer than in the observations when the simulated soil temperature dropped below -5 °C. However, the LS3MIP simulations were colder, with a ~~bias of -0.65~~cold bias in the median of 0.7 °C. The biases of 2 m temperature ~~had a different~~ in coupled simulations had an opposite sign and were amplified in magnitude compared to the biases of ~~the~~ their soil temperatures, especially below 0 °C. ~~Four of the climate models and their land components underestimated the snow insulation effect. We concluded~~ Our results indicate that land-only models have ~~difficulties in accurately simulating~~ limited capability in reproducing soil temperatures and snow depth under ~~low-temperature conditions . The CMIP6 models tended to compensate for errors in their land component with errors in the atmospheric model component. In~~ severe cold conditions (surface air temperature below -15 °C). Furthermore, four climate models and their land components underestimated the insulating role of snow. In cases with shallow snow depth (~~0 to 0.2~~0–0.2 m)~~cases, all models showed between 1 and 8~~, the models simulated air-soil temperature differences of up to 10 °C~~less air-soil temperature difference than in situ data. Therefore, a better representation of~~, whereas in situ measurements indicated even larger differences. The CMIP6 models tended to compensate for errors in their land component with errors in the atmospheric model component. Therefore, to improve frozen soil modeling in climate projections, a more accurate representation of the surface-soil insulation is essential~~for improvements in frozen soil land modeling~~.

1 Introduction

Under current climate conditions, an amplification effect up to 2–4 times is evident in warming trends within Arctic regions compared to global averages (Rantanen et al., 2022). Specifically, from 1979 to 2018, near-surface temperature growth rates in the land area surrounding the Arctic (0.51 °C per decade) were more than 1.5 times that of the Northern Hemisphere average (0.33 °C per decade), according to Climate Research Unit (CRU) TS4.02 (Wang et al., 2022). Climate simulations indicate that the Arctic experienced a warming rate of 0.66 ± 0.32 °C per decade from 1979 to 2014 (Cai et al., 2021). Additionally, under a moderate scenario, high-latitude regions could see temperature increases ranging from 1.2 to 5.3 °C between 2005 and 2100 (Koven et al., 2013). Higher temperatures drive the degradation of permafrost, especially in discontinuous and sporadic permafrost regions. The most significant changes occur in regions where the mean annual air temperatures are around 0 °C (Romanovsky et al., 2007; Åkerman and Johansson, 2008; You et al., 2021). The soil temperature at zero annual amplitude depth of continuous permafrost sites on a global scale warmed for 0.39 ± 0.15 °C during 2007–2016. This warming is two to three times larger than that observed at discontinuous permafrost sites (Biskaborn et al., 2019; Smith et al., 2022), raising concerns about the potential for more substantial permafrost degradation in the future.

Large amounts of soil carbon are stored in permafrost (Tarnocai et al., 2009; Fuchs et al., 2018), and when permafrost thaws, the soil carbon could be emitted as carbon dioxide or methane into the atmosphere at a faster rate (Schädel et al., 2018). In environments such as lakes and wetlands, the impact of thawed carbon on the climate is even more pronounced due to the low-oxygen conditions, which further ~~increases~~ increase the proportion of methane emitted alongside other greenhouse gases (Koven et al., 2015; Walter Anthony et al., 2018). Processes such as thermokarst result in sudden thaw events that greatly enhance the decomposition and release of frozen soil carbon, potentially increasing carbon emissions by up to 50 % (Abbott and Jones, 2015; Turetsky et al., 2019). The winter contribution to total Arctic methane emissions is predicted to reach 39 % (Rößger et al., 2022). Permafrost thaw alters hydrothermal conditions, which can alter surface vegetation depending on soil moisture. In lowland regions with ice-rich permafrost, abrupt thawing is often followed by vegetation recovery. Under stronger or prolonged changes, the system may reach a tipping point, beyond which widespread ecosystem disruption can occur (Heijmans et al., 2022).

Despite its critical role in climate feedback processes, frozen soil remains one of the ~~most~~ more uncertain components in Earth system models (Koven et al., 2013; Yokohata et al., 2020). In the land surface component of climate models, heat transfer through the soil is typically simulated as one-dimensional vertical transport. Models account for their specific soil layering schemes, where the thickness of soil layers generally increases with depth. By calculating the water and thermal balance at different depths, land surface models can derive the current state of soil moisture and temperature. Model parameters for hydrothermal transport are governed by soil texture, surface organic matter, moisture dynamics, and freeze-thaw conditions (Woo, 2012; Andresen et al., 2020; Yang et al., 2022). In permafrost regions, these factors determine the thermal offset (Kudryavtsev, V.A., 1977)—the temperature difference between the ground surface and the top of the permafrost—by altering soil hydrothermal properties. For example, the presence of solid and liquid water in frozen soil greatly affects the hydrothermal properties of the soil, which is described in various ways by different models (Niu and Yang, 2006; Li et al., 2010). Incorporating soil ice and water dynamics into land surface models improves simulations of active layer hydrothermal conditions by capturing seasonal freeze-thaw processes and moisture effects, such as summer cooling and winter warming of permafrost due to increased soil moisture (Swenson et al., 2012; Li et al., 2021; Du et al., 2023). Including soil ice dynamics in models allows for the simulation of ice-wedge degradation and associated ground subsidence, capturing rapid landscape changes such as thermokarst under strong warming scenarios (Liljedahl et al., 2016; Nitzbon et al., 2020; Cai et al., 2020). The parameterizations in land surface models are crucial for accurately representing permafrost dynamics and their associated climate feedback processes (Yokohata et al., 2020), and they also determine the inclusion of key physical and biogeochemical processes essential for modeling permafrost carbon emissions (Ekici et al., 2015; Matthes et al., 2017, 2025).

The timescales of major physical processes differ largely between the soil and the atmosphere. For instance, key variables such as temperature and humidity in the near-surface atmosphere can experience substantial fluctuations within hours or even minutes. In contrast, changes in water and thermal states within the soil become much slower as depth increases. At depths of several tens of meters beneath permafrost, soil temperatures may not exhibit any noticeable variation over decades. In this context, it is essential to recognize that the soil surface serves as a critical interface for atmospheric interactions (Beringer et al., 2001; Langer et al., 2011a, b). Snow cover plays an important role by insulating soils while affecting surface energy balance through changes to local albedo as well as other characteristics such as emissivity and roughness. The impact of snow on soil temperature exhibits spatial heterogeneity based on snow's own attributes—specifically, thickness, density, and duration (Zhang, 2005; Zhang et al., 2018). A thick snowpack provides stronger thermal insulation, which limits soil heat loss in winter and delays thawing in spring. Lower-density snow insulates more effectively due to its reduced thermal conductivity. The duration of snow cover determines the length of the insulated period, which affects the timing and amplitude of seasonal soil temperature changes. Research has shown that changes in snow conditions (snow depth, density, and duration) account for over 50 % of variations in soil temperatures observed in northeastern Siberia (Park et al., 2014, 2015). An accurate representation of snow cover is essential for climate models, as a recent study has shown significant discrepancies in snow representations across different seasonal forecasting systems over Siberia due to different snow parameterizations and initialization methods (Risto et al., 2022).

The Coupled Model Intercomparison Project Phase 6 (CMIP6), launched by the World Climate Research Program (WCRP), aims to explore various topics related to climate change (Eyring et al., 2016). It allows evaluation of the ability of the latest generation of climate models to simulate frozen soil by providing an ensemble of climate models at resolutions fine enough to distinguish different frozen soil regions. The extent and characteristics of frozen soil can vary abruptly over short distances, especially in complex terrain or transition zones between different types of permafrost. Research efforts have been conducted to improve the simulation capabilities of land surface models participating in CMIP, focusing on biological and physical processes in frozen soil areas (Ekici et al., 2014; Chadburn et al., 2015; Decharme et al., 2016; Brunke et al., 2016; Jafarov and Schaefer, 2016; Guimberteau et al., 2018; Cuntz and Haverd, 2018; Damseaux et al., 2025). The Land Surface, Snow, and Soil Moisture Model Intercomparison Project (LS3MIP) is designed to enhance our comprehension of land surface processes by assessing the effectiveness of various land-only models in simulating soil temperature and moisture, snow cover, and related hydrological dynamics. It also aims to generate valuable insights that can aid in refining land-only models (Van Den Hurk et al., 2016). In

LS3MIP, experiments are designed so that different land-only models use the same atmospheric forcing. Therefore, LS3MIP provides an opportunity to distinguish the impact of distinct climate variabilities produced by different atmospheric models in corresponding CMIP6 models.

In CMIP6 and LS3MIP, the setup of the land cover/land use scenario and the radiative forcing conditions follow the same protocol. However, the parameterization schemes of the climate models differ, and this is considered a main source of uncertainty in climate modeling (Deng et al., 2021; de Vrese et al., 2023; Kuma et al., 2023). In addition, the presence of internal climate variability can also lead to differences in uncertainty (Ye, 2021; Rashid, 2021; Schwarzwald and Lenssen, 2022; Jain et al., 2023). Understanding and isolating these uncertainties is essential for improving model reliability.

~~Here, we focus on the Siberian region with frozen soils, which constitutes a significant portion of the Eurasian continent's frozen terrain~~Northern Eurasia contains more than two-thirds of the Earth's permafrost area (Groisman and Bartalev, 2007), with the majority located in Siberia. The potential degradation of permafrost in Siberia could have far-reaching consequences for climate and ecosystems throughout Eurasia and globally (Schuur et al., 2015; Streletskiy et al., 2025). Within this region, ~~the observational dataset~~ we used the soil temperature observational dataset at standardized depths, as provided by the All-Russian Scientific Research Institute of Hydrometeorological Information-World Data Center ~~(RIHMI-WDC) (Frauenfeld and Zhang, 2011; Shersti~~ ~~can be used. This dataset provides consistent soil temperature measurements at standardized depths and can thus be used as a reference in climate model evaluation.~~ (RIHMI-WDC; Frauenfeld and Zhang, 2011; Sherstiukov, 2012a; Zhang et al., 2018).

The characteristics of frozen soil surface dynamics are assessed by comparing model outputs with references, including reanalysis and observational data. This research focuses on the shallow soil temperature response to atmospheric forcing, explicitly targeting a depth of 0.2 m. We will analyze the discrepancies between the climate models in CMIP6 and their land-only models in LS3MIP to quantify the bias and uncertainty present in frozen soil regions, attributing them to land-only models versus those resulting from atmospheric forcing. CMIP6 models include fully coupled components, such as the atmosphere, ocean, land, and sea ice, whereas LS3MIP models only include the land surface model and prescribe atmospheric forcing from reanalysis or historical simulations. ~~Consequently, the differences between the two can be used to attribute model biases to either the land surface model structure and parameterizations or coupled atmosphere-land interactions. Under identical,~~ As land surface models are designed to simulate the response of surface and soil parameters to prescribed atmospheric forcing, they are better constrained as driven by observation-based atmospheric conditions~~, the~~ . Therefore, under identical observation-based atmospheric forcing, LS3MIP models are ~~expected to simulate soil temperature~~ generally expected to reproduce soil conditions more accurately than their CMIP6 counterparts. If not, discrepancies may indicate limitations within the land surface models themselves, such as deficiencies in parameterization or missing processes that impair their ability to respond appropriately to atmospheric forcing. Conversely, errors found in coupled CMIP6 simulations may result from biases in atmospheric forcing, such as misrepresentation of near-surface air temperature, precipitation, or surface radiation. ~~This experimental design allows us to distinguish the sources of uncertainty between land-only and coupled simulations . Additionally~~The experimental design of LS3MIP provides a framework that enables an evaluation of the respective contributions of these models to biases and uncertainties in soil temperature simulations within climate models. Specifically, we will explore inter-model variations within LS3MIP and assess how specific structural features, such as bottom boundary conditions and snow thermal conductivity parameterizations, relate to model performance in frozen soil regions.

2 Data and Methods

We used the data from climate models, reanalysis, observations, and processing methods of target variables for our analysis. We only included data from 1985 to 2014 in this research, as this period offers the best collection of observation records, and CMIP6 historical experiments are limited to 2014.

2.1 CMIP6 and LS3MIP Simulations

The CMIP6 multi-model ensemble provides historical climate simulations based on the same external forcing (solar radiation, greenhouse gases, aerosols, etc.) (Eyring et al., 2016). Our study used the *historical* simulations with predefined CO_2 concentrations, which contain distinctive combinations of atmospheric and land models. The *Land-Hist* experiments from LS3MIP are offline, land-only simulations with no feedback to the atmosphere and no dynamic forcing from atmospheric models (Van Den Hurk et al., 2016). All the LS3MIP simulations employed the same atmospheric forcing derived from the Global Soil Wetness Project Phase 3 (GSWP3) and the same land surface setup as in the CMIP6 experiments (Van Den Hurk et al., 2016). GSWP3 is a global land surface modeling project that provides long-term meteorological gridded forcing data based on the 20th Century Reanalysis (20CR), bias-corrected with observational datasets. This setup allowed us to directly compare CMIP6 and LS3MIP results and disentangle the relative contributions of coupling-related errors and land model deficiencies to biases in frozen soil regions.

We chose seven climate models involved in both projects, incorporating six different land models. The selected models are listed in Table 1. Other climate models, which also participated in both projects, could not be included in this study as they either turned off the freeze option in frozen soil in the CMIP6 version or did not provide data for all our target variables. Hereafter, we refer to the CMIP6 as Group C and the LS3MIP as Group L in plots and analysis. Four variables are collected for this study: 2 m air temperature (tas), soil temperature in 0.2 m depth (tsl), snow depth (snd), and precipitation (pr). Both CMIP6 and LS3MIP data can be accessed at https://aims2.llnl.gov/search/cmip6.

Table 1. Selected CMIP6/LS3MIP experiment pairs, the layering and resolution. For other features and references, see Table 2.

Model Name	**Land Surface Model**	**Total Soil Layers** (max. node depth (m))	**Soil Layers** in Top 3 m	**max. Snow Layers**	**Resolution** (lat×lon)
CESM2	CLM5.0	25 (42.0)	14	12	$0.9^\circ \times 1.25^\circ$
CNRM-CM6.1	Surfex 8.0c	14 (10.0)	11	10	$1.4^\circ \times 1.4^\circ$
CNRM-ESM2.1	Surfex 8.0c	14 (10.0)	11	10	$1.4^\circ \times 1.4^\circ$
IPSL-CM6A-LR	ORCHIDEE v2.0	18 (65.56)	12	3	$1.25^\circ \times 1.875^\circ$
HadGEM3-GC31-LL	JULES-HadGEM3-GL7.1	4 (2.0)	4	3	$1.25^\circ \times 1.875^\circ$
UKESM1.0-LL	JULES-ES-1.0	4 (2.0)	4	3	$1.25^\circ \times 1.875^\circ$
MIROC6	MATSIRO6.0	6 (9.0)	5	3	$1.4^\circ \times 1.4^\circ$

Table 2. Features of the land surface models. MC indicates mechanical compaction.

Land Model	**Snow Density**	**Snow Thermal Conductivity**	**Bottom Boundary Condition**	**References**
CLM5.0	MC	Density-dependent Jordan (1991)	Zero-flux	Van Kampenhout et al. (2017) Lawrence et al. (2019)
Surfex 8.0c	MC	Density-dependent Yen (1981)	Fixed temperature	Vionnet et al. (2012) Decharme et al. (2019)
ORCHIDEE v2.0	MC	Depends on density, temperature, and pressure	Fixed flux	Wang et al. (2013) Bowring et al. (2019)
JULES-HadGEM3-GL7.1	MC	Density-dependent Calonne et al. (2011)	Zero-flux	Clark et al. (2011) Walters et al. (2019) Wiltshire et al. (2020c)
JULES-ES-1.0	MC	Density-dependent Calonne et al. (2011)	Zero-flux	Sellar et al. (2019) McNeall et al. (2024)
MATSIRO6.0	Fixed ($300\,\mathrm{kg\,m^{-3}}$)	Fixed ($0.3\,\mathrm{W\,m^{-1}\,K^{-1}}$)	Fixed temperature	Takata et al. (2003) Guo et al. (2021)

Table 2 highlights several attributes of each land model, focusing on their key characteristics related to the surface energy balance and associated processes. For a more general comparison of land-only models and their features regarding snow parameterization, see Menard et al. (2021). The land models differ in their representation of critical processes related to surface energy balance, snow physics, boundary conditions, and surface organic matter. The thermal conductivity of snow is modeled using density-dependent formulations as a power function (Yen, 1981) or a quadratic function (Jordan, 1991; Calonne et al., 2011; Wang et al., 2013), or using fixed values. Snow density is treated either dynamically through mechanical compaction (MC) or as a constant. The soil bottom boundary is handled with zero-flux assumption, fixed temperature, or fixed temperature gradients.

2.2 ERA5-Land Reanalysis

We utilized monthly averaged ERA5-Land reanalysis data from the European Centre for Medium-Range Weather Forecasts. ERA5-Land is a numerical land surface model product forced by atmospheric variables of ERA5, featuring a high horizontal resolution of $0.1° \times 0.1°$(Muñoz-Sabater et al., 2021; Copernicus Climate Change Service, 2019). Monthly data provided by ERA5-Land include 2 m temperature, snow depth, and soil temperature at four depths (0 to 0.07 m, 0.07 to 0.28 m, 0.28 to 1.0 m, 1.0 to 2.89 m). Soil temperature was linearly interpolated to a depth of 0.2 m to directly compare with observations. We also remapped ERA5-Land to a coarser horizontal resolution of 100 km (the finest resolution of the selected CMIP6 models). This remapped dataset is referred to as E5LC. Comparing ERA5-Land with its coarsened version allowed us to assess the impact of resolution on simulation outcomes. This also provided an opportunity to evaluate the reliability of ERA5-Land in representing conditions within permafrost-affected areas.

2.3 Observational Data

Observational daily data were gathered from 236 meteorology sites by RIHMI-WDC ~~(Sherstiukov, 2012a; Bulygina et al., 2014a, b; Zhang~~ (Sherstiukov, 2012a; Bulygina et al., 2014a, b; Zhang et al., 2018). We filtered the data from 1985 to 2014 based on the quality flag (Sherstiukov, 2012b) provided in the dataset and employed only sites with a minimum of 330 valid days per year for all four target variables and at least 15 years of valid data. We only used the highest quality data, flagged as 0. Stations west of 60° E, east of 120° E, and south of 45° N were excluded to eliminate most stations with warmer climates. Moreover, sites were classified based on their winter 2 m air temperature. These criteria were put in place to ensure the accuracy and reliability of the data analyzed. A total of 152 stations were selected.

2.4 Data Preprocessing

We interpolated tas, tsl, pr, and snd for all model simulations at the station sites' coordinates by choosing the nearest-neighbor grid cell values. This introduced additional, but small uncertainty when comparing modeled data with observed data ~~. Notably, the CMIP6 historical runs can differ in phase from the actual climate phase due to internal climate variability. To minimize this uncertainty source,~~ as will be shown below. In time dimension, we considered only 30-year monthly averaged data for evaluations aiming to minimize uncertainties due to climate variability and phase shifts in CMIP6 simulations.

2.5 Evaluation Metrics

To evaluate the models, we quantified their ability to simulate a reasonable climate mean state and internal climate variability.

The variability of target variables was quantified using the interquartile range (~~IQR~~IQR), which measures the spread of the simulations. If the model over- or underestimated the observed IQR_o, the model over- or underestimated climate variability, respectively. Thus, we used the relative spread

$$RS_{m,i} = \frac{\mathrm{IQR}_{m,i}}{\mathrm{IQR}_{o,i}} \tag{1}$$

with m indicating the model, o the observation, and i the target variable.

The central climate state of the models and the observations were quantified by the median (med). We used the standardized model medians $\mathrm{med}_{m,i}$, naming it relative bias RB,

$$RB_{m,i} = \frac{\mathrm{med}_{m,i} - \mathrm{med}_{o,i}}{\mathrm{IQR}_{o,i}} \tag{2}$$

as a measure of systematic model errors.

RS assesses a model's capability to reproduce the observed variability. This is essential for determining a model's reliability in simulating dynamic climate systems. RB, on the other hand, addresses systematic deviations using the median data values. It was standardized using observed variability (~~IQR_o~~IQR_o), which facilitates comparisons of biases across different variables. RB emphasizes whether systematic errors are pronounced relative to natural variability (informs if the error is smaller ($RB_{m,i} < 1$) or larger (> 1) compared to the observed climate variability), helping prioritize improvements in model development.

For the qualification of the error heterogeneity of the 2 multi-model ensembles at the sites' locations, we defined the ensemble means of the models' 30-year median biases as follows

$$EB_{i,s} = \sum_{m=1}^{M} \frac{1}{FN_m} (\mathrm{med}_{m,i,s} - \mathrm{med}_{o,i,s}) \tag{3}$$

where s indicates the seasons, $F = 5$ represents the number of "model families" and N_m, which represents the number of "family members" (either 1 or 2 in this study), was used to properly weight similar models, as described in Kuma et al. (2023). We then calculated the ensemble spreads of median biases $ES_{i,s}$, which were defined as the standard deviations.

The ensemble mean biases $EB_{i,s}$ and spreads were calculated for both the CMIP6 ~~(Group C)~~ and LS3MIP ~~(Group L)~~ model ensembles with $M = 7$ members each. The analyses were done in different seasons to distinguish the impact of different freeze/thaw periods on ensemble performance. In this study, seasons were defined by months. Winter was defined as December, January, and February (DJF); spring as March, April, and May (MAM); summer as June, July, and August (JJA); and autumn as September, October, and November (SON).

2.6 ~~Permafrost Stations~~

The simulation performance at specified sites was quantified using Taylor diagrams (Taylor, 2001). The data were compared with observations across all seasons in the permafrost region. The correlation coefficient was computed from two-dimensional data (19 stations × 3 months), thereby reflecting the agreement between the model and observations in both spatial distribution and sub-seasonal variation.

~~To~~ We applied additional statistical analysis to the temperature variables of CMIP6 and LS3MIP ensembles, focusing on temperature states, including interquartiles, medians, means, standard deviations, and kurtosis. The kurtosis characterizes the shape of a distribution, where a value of 0 indicates a normal distribution. When the standard deviations are comparable, a higher kurtosis indicates a distribution with heavier tails, i.e., more extreme deviations from the mean.

2.6 Categorization of Stations

We aimed to assess the models' performance under different climate conditions to determine whether simulation uncertainties increase at lower temperatures or remain similar. To distinguish between different climate regimes, one possible approach is to categorize stations based on their average DJF 2 m air temperature (Wang et al., 2016).

We expanded the scope by focusing on air temperatures when snow is present, not just during the DJF period, and linked the results to the insulating effect of snow. To investigate the climatology in the permafrost area, sites were ~~identified as~~ further identified as valid permafrost sites using the definition of ~~Lawrence and Slater (2005)~~Lawrence and Slater (2005). If a station had at least 300 days of valid data every year ~~,~~ and 24 consecutive months of lower than 0 °C in any layer below 0.2 m depth, it was defined as a valid permafrost station. Using this method, we selected 19 valid permafrost stations among all observation sites.

3 Results

3.1 Winter 2 m Temperature in Target Area

Figure 1 shows the map of average winter (~~DJF~~ DJF from 1985 to 2014) 2 m air temperatures (tas) from the CMIP6 multi-model ensemble, and the symbols correspond to the observation data from RIHMI-WDC. As shown in the map, winter-time tas in the target area is colder in the northeast and warmer in the southwest. Within the area 50° E to 185° E, north of 45° N, the average ~~DJF~~ DJF tas is generally below 0 °C. The region ~~with less~~ colder than -25 °C had a large overlap with the continuous permafrost region (Brown et al., 1997; Obu et al., 2019).

The temperature categories were listed in the legend of Fig. 1. There were 3 sites with average tas warmer than -5 °C, 25 sites between -15 and -5 °C, 76 sites between -25 and -15 °C, and 29 stations with tas below -25 °C. Besides, 19 sites were identified as valid permafrost ('~~perma~~perma') sites using the method introduced by Lawrence and Slater (2005) (soil layer temperatures continuously below 0 °C for at least two years) and labeled by stars. All '~~perma~~perma' sites had average winter tas values ~~less than~~ below -23 °C.

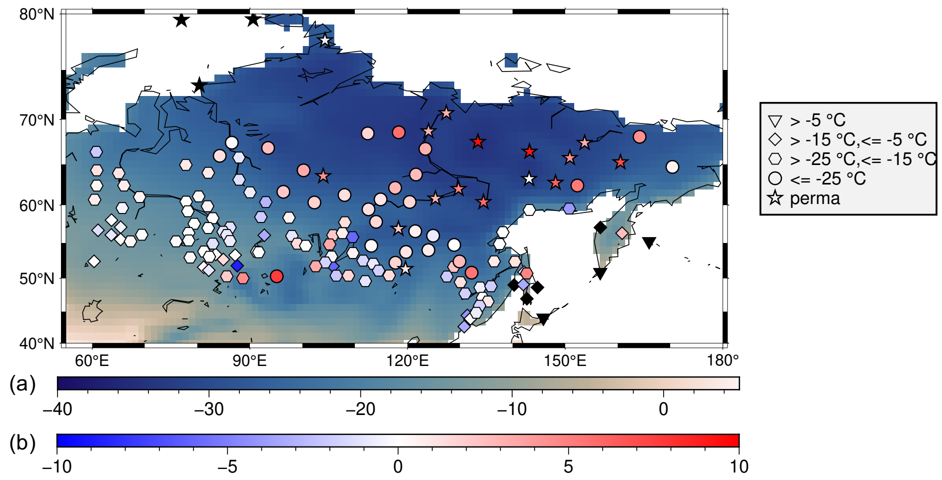


Figure 1. Mean near-surface air temperature (a) as given by the weighted multi-model ensemble from CMIP6, and the difference between the ensemble mean and the observational data (b) at sites for winter (DJF) in 1985–2014 in °C. Symbols indicate the climate state at the observational sites (see legend), and symbol colors show the difference between the ensemble mean and the observational data.

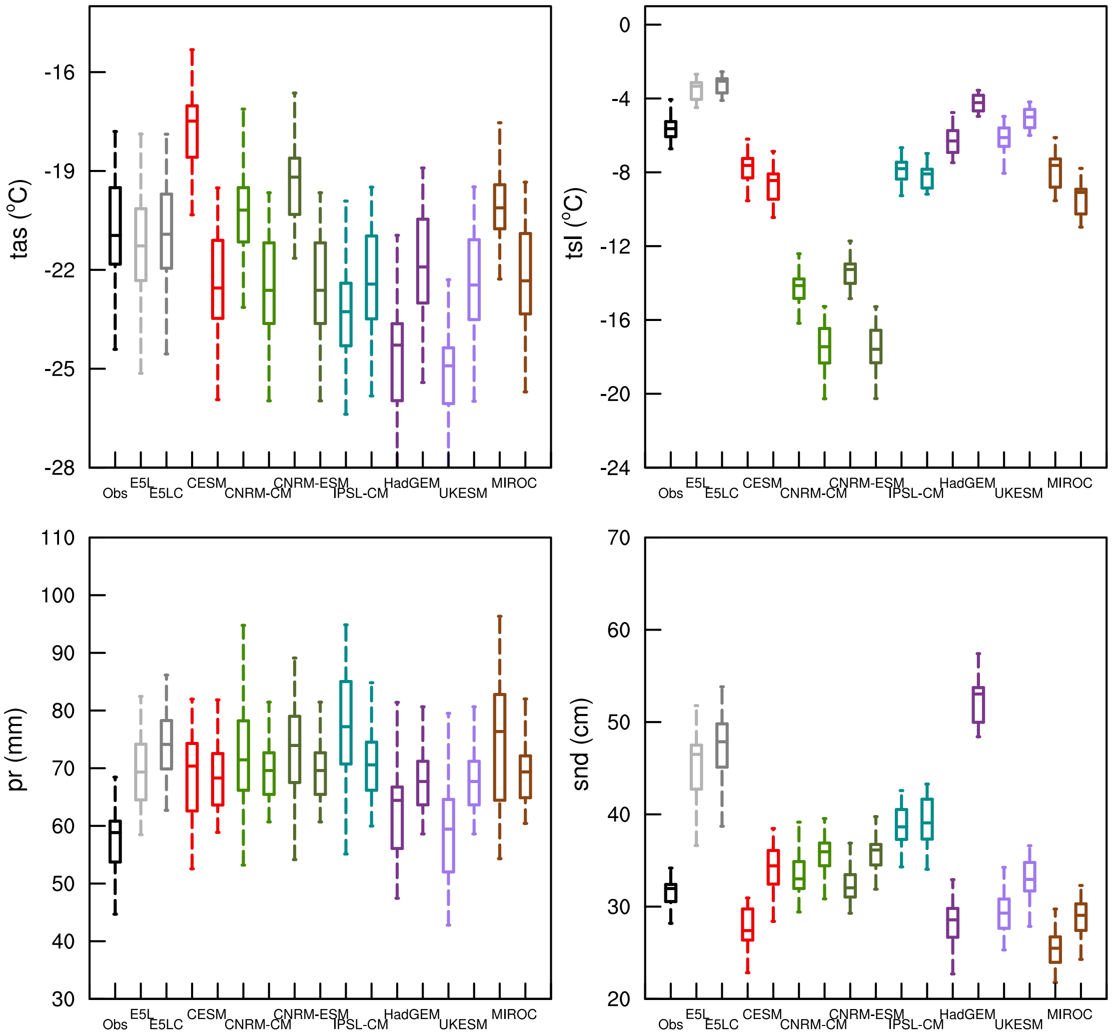


Figure 2. Sites' averaged ~~DJF-climates~~ DJF-climates (1985–2014) of hydrothermal variables as observed and simulated. Names on the x-axis and the colors indicate the different data sources. In addition to observations, ERA5-Land (E5L) and its coarsened variant, E5LC, are included to enable comparison at the same horizontal resolution as CMIP6 and LS3MIP. Model names indicate CMIP6 model output (left), and LS3MIP (right; see Table 1). Each CMIP6-LS3MIP pair shares the same color. The boxes represent the medians, first and third quartiles; the $\pm$ 1.5$\times$~~IQR~~ IQR or the maximum and minimum values, if within the former range, are taken as the whiskers' length.

3.2 Model Climatologies

The sites' averaged winter climatologies (DJF) of the four target variables for the different datasets from 1985 to 2014 are shown in Fig. 2. All model variables were interpolated into the site's locations using the nearest neighbor method.

ERA5-Land's tas and tsl climatologies are closer to the observed climatologies than those of most models, though this is not the case for pr and snd. Both ERA5-Land and E5LC were interpolated into the sites' locations using the nearest neighbor method. ~~The coarser resolution resulted in~~ Although the coarser resolution yielded slightly higher values ~~of the four variables we analyzed , all less~~ for the four analyzed variables, the median differences between ERA5-Land and E5LC were all smaller than one interquartile range of ~~observations.~~ the observations.

In LS3MIP, tas and pr were derived from the same forcing data. However, Fig. 2 shows slight discrepancies of tas (less than 1 °C) and pr (less than 3 mm) among land-only models. The LS3MIPs' tas climatologies were systematically colder by more than 1 °C than observations and E5LC, and their pr climatologies aligned better with E5LC than with observations (which have on average about 15 % smaller values).

The CMIP6 models' tas climatologies scattered substantially. CESM2's tas median was shifted by more than +3 °C against observations, while tas medians of HadGEM3-GC31-LL and UKESM1.0-LL were shifted by more than -3 °C.

Most models' tsl climatologies were colder than observed. The CNRM simulations exhibited the lowest soil temperatures overall (Fig. 2) with the climatologies of being more than 8 °C lower in both CMIP6 and LS3MIP. The strong cold bias of CNRM-CM6.1 and CNRM-ESM2.1 is also shown during spring and autumn (Fig. ~~A2~~A3). As expected, the temporal variability, quantified by the box plots' IQR, was smaller in tsl than in tas. Substantial inter-model variability was observed in soil temperature simulations. The differences between a model's CMIP6 and LS3MIP tsl were much smaller than their differences from observations. Land-only models that belong to the same family (the land components of two CNRM models and the models HadGEM and UKESM, see Table 1) exhibited similarities in tsl medians and IQRs.

In the other seasons, the spread of tas and tsl climatologies ~~(Fig. A)~~ was respectively smaller than in ~~DJF. During JJA~~DJF. During JJA (Fig. A1), model differences in tas variability were clearly reflected in tsl, particularly as snow was absent at most sites during summer (median value less than 0.3 cm).

In both CMIP6 and LS3MIP runs, the DJF pr values were typically $10\,\mathrm{mm}$ higher than in the observations. UKESM1.0-LL simulated relatively good pr in CMIP6, with less than $3\,\mathrm{mm}$ overestimation in DJF average. The observed DJF snd was approximately 30 cm. It was overestimated by 70 % in HadGEM3-GC31-LL (LS3MIP run) and by 50 % in E5LC. ~~The~~ Overall, the models were diverse in simulating DJF medians snd, with high temporal variability.

3.2.1 Relative Spread and Relative Bias

This subsection compares the relative spread (RS) and relative biases (RB) of E5LC and the different models compared to observations for all four target variables and all seasons. The RS assesses the sites-averaged temporal climate variability, and RB is the shift of the climatologies.

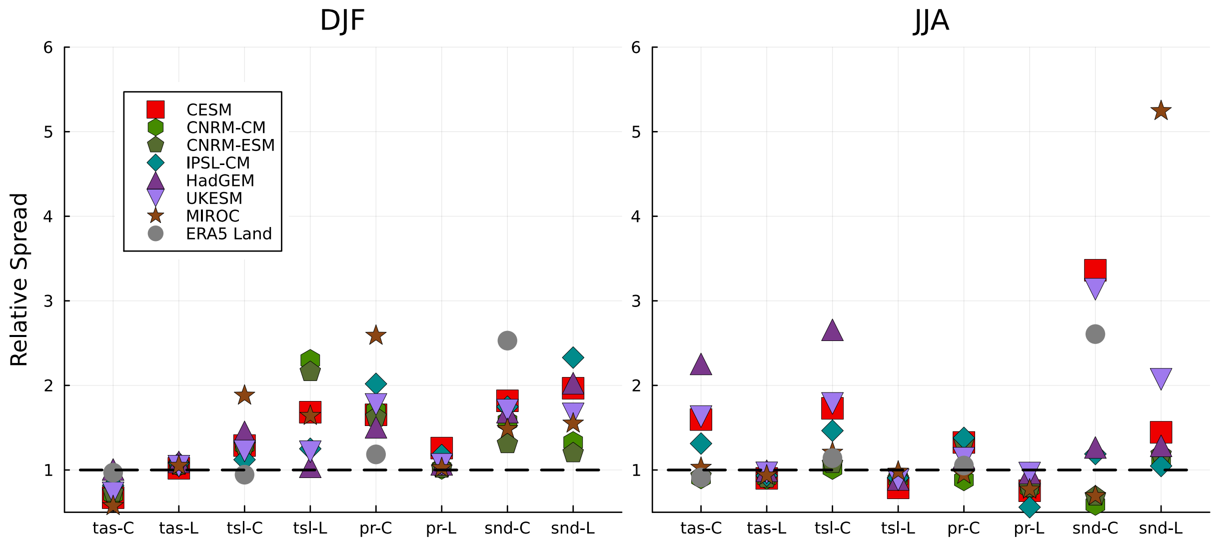


Figure 3. Relative spread (RS) of the sites-averaged climates (~~1985–2014~~1985–2014) of the four variables in both ensembles and in E5LC with reference observations for all seasons. The colors indicate the models, and the x-axes show the variables and ensembles (C indicates CMIP6, and L indicates LS3MIP runs).

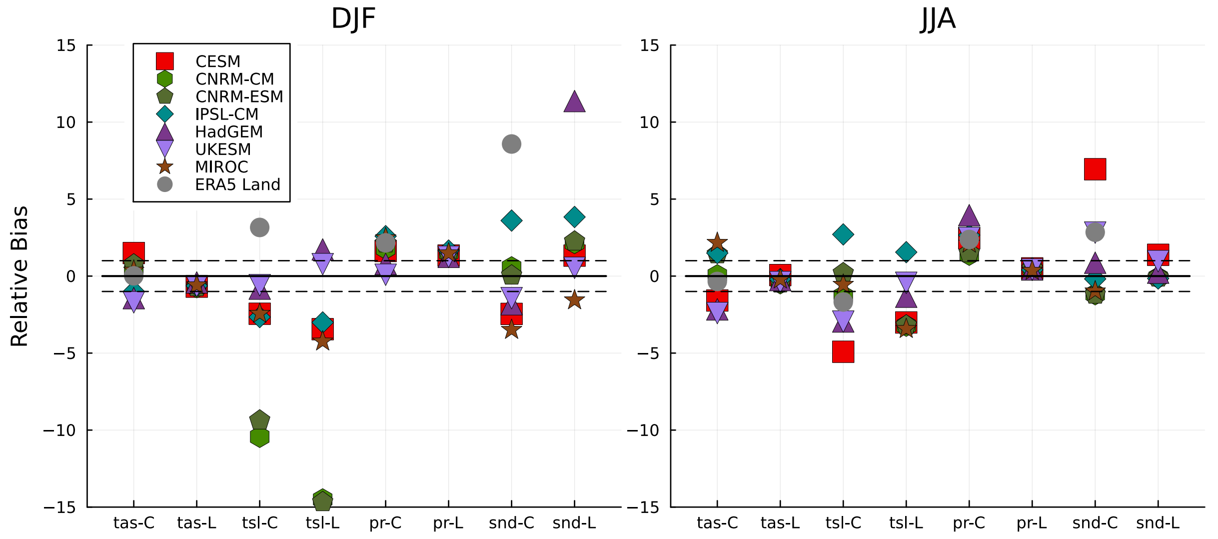


Figure 4. Same as in Fig. 3, but for relative biases (RB). The dashed lines ~~(from -1 to 1)~~ indicate the range of absolute median differences smaller than the observation's ~~IQR~~IQR.

Most CMIP6 models underestimated tas ~~DJF~~ DJF climate variability and ~~models~~ overestimated tsl climate variability (Fig. 3). The ~~CMIP6 models showed a larger diversity in tas, and all values were below 1. However, the LS3MIP models showed a large diversity for tsl and pr in~~ LS3MIP models showed an even larger tsl diversity in winter, ranging from 1 to almost 2.7~~. The LS3MIP models performed well in simulating JJA~~ , but performed better in simulating JJA variability of tsl, though ~~there was~~ with a general underestimation. In DJF and SON, the tas RS for all CMIP6 models were within the range of 0.5 to 1. In MAM, they were within the range of 0.7 to 1.7. CESM2, HadGEM3-GC31-LL, and UKESM1.0-LL exceeded 1.5 RS of tas in JJA. Larger variability differences in simulated tsl were observed for LS3MIP simulations in DJF and SON, even though their tas share almost identical RS.

The DJF pr relative spreads of ~~Group C~~ CMIP6 models were all higher than 1.3. In spring and autumn, most climate models simulated good interannual variability of pr, with less than a 20 % difference to observation (Fig. ~~A1~~A2). Both groups and E5LC overestimated the snd spread in DJF ~~snd, having high discrepanciesamong models~~with large inter-model discrepancies. The RS of snd scatter between 0.8 and 1.8 in SON, and 1 and 1.6 in MAM for most models (except for E5LC, CNRM-CM and MIROC in CMIP6, and HadGEM in LS3MIP).

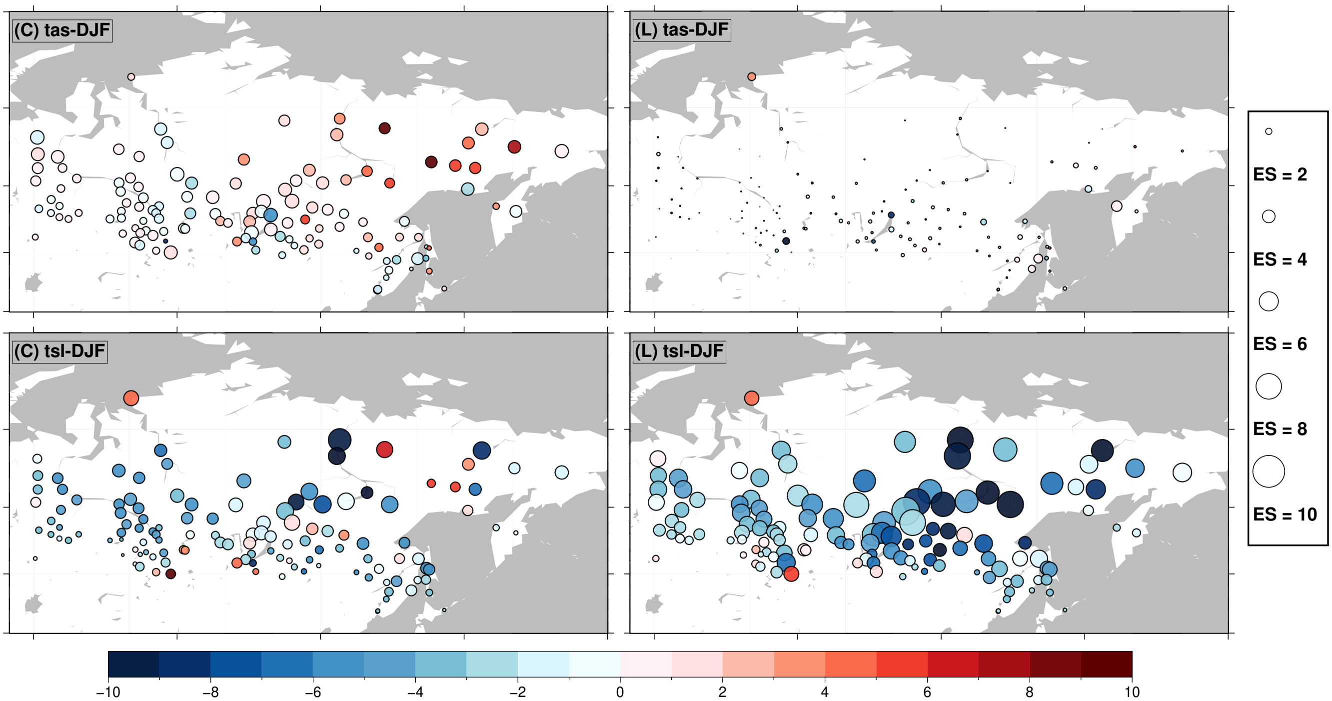


Figure 5. Weighted DJF multi-model mean biases (EB) and their standard deviations (ES) of tas and tsl at observational sites in °C. ~~Panels A–H and I–P are for tas and tsl, respectively.~~ The ~~first and third columns are the results~~ top-left subtitles of each subplot indicate the presented model ensemble (C for CMIP6 ~~models,~~ and ~~the second and fourth columns are the results of~~ L for LS3MIP~~models. The first to fourth rows show~~), the ~~DJF~~variable, ~~MAM, JJA,~~ and ~~SON seasons, respectively~~the season. Colors indicate the bias value (in °C), while the circle radii represent biases' standard deviations (in °C), i.e., the ensemble spread.

The relative bias RB of all variables calculated in all seasons is shown in Fig. 4. ~~The zone within the dashed lines indicates a one-time IQR of the observed value.~~ If a model's absolute RB is less than one, its performance is considered adequate. For example, the winter RB of all LS3MIP models was within the dashed line zone in Fig. 4 for tas, and ~~was~~ at or slightly above 1 for the DJF pr ~~of all LS3MIP models~~ as they are derived from the same atmospheric forcing data. ~~Nearly all~~ All CMIP6 and LS3MIP models exhibited a positive relative pr bias. The relative snd bias was much more diverse; models only showed a consistent positive snd bias in MAM (Fig. ~~A2~~A3). The CMIP6 runs of HadGEM3-GC31-LL and UKESM1.0-LL underestimated the values of tas and tsl in all seasons. The tsl RB in DJF for CNRM simulations exceeded -15 to -9 times the observed IQR, while other models were all within a range of ±5 times in both groups. The RB of CMIP6 and LS3MIP in tsl were all negative in the transition seasons (Fig. A3), and also high discrepancies were shown in snd RB values.

3.2.2 Spatial Heterogeneity

In ~~DJF,~~ DJF, the multi-model ensembles ~~had~~ exhibited the largest EB and ES (Fig. 5). ~~In DJF~~On average, the CMIP6 models overestimated tas ~~on average~~ by more than 0.5 °C, ~~but~~ with a large ES of ~~2.96~~3.0 °C~~. The DJF average EB and ES for~~, while LS3MIP ~~were -0.51~~showed a mean EB of -0.5 °C and ~~0.69~~a smaller ES of 0.7 °C~~, respectively~~. CMIP6 models tended to produce warm tas biases, whereas LS3MIP land models exhibited cold tsl biases at the same locations, particularly in Siberia.

In the other seasons (Fig. A4), EB and ES were distinctly smaller, with most CMIP6 runs' tas slightly too cold in spring at most stations and slightly too warm in autumn in northeastern Siberia. The tas EB of the forced models was small at most sites (exceptions were near water bodies, e.g., Lake Baikal and sea coasts, which indicated interpolation artifacts due to the selected grids).

The tsl EB and ES were larger in magnitude than for tas, especially in winter, with spatial averaged EB of ~~-2.66~~-2.7 °C and ~~-3.57~~-3.6 °C for CMIP6 and LS3MIP runs, respectively. Additionally, the LS3MIP runs had a larger bias spread than the CMIP6 runs in all seasons except summer. In winter, the ES were ~~2.98~~3.0 °C for the CMIP6 and as large as ~~4.55~~4.6 °C for the LS3MIP runs.

3.3 Permafrost Region

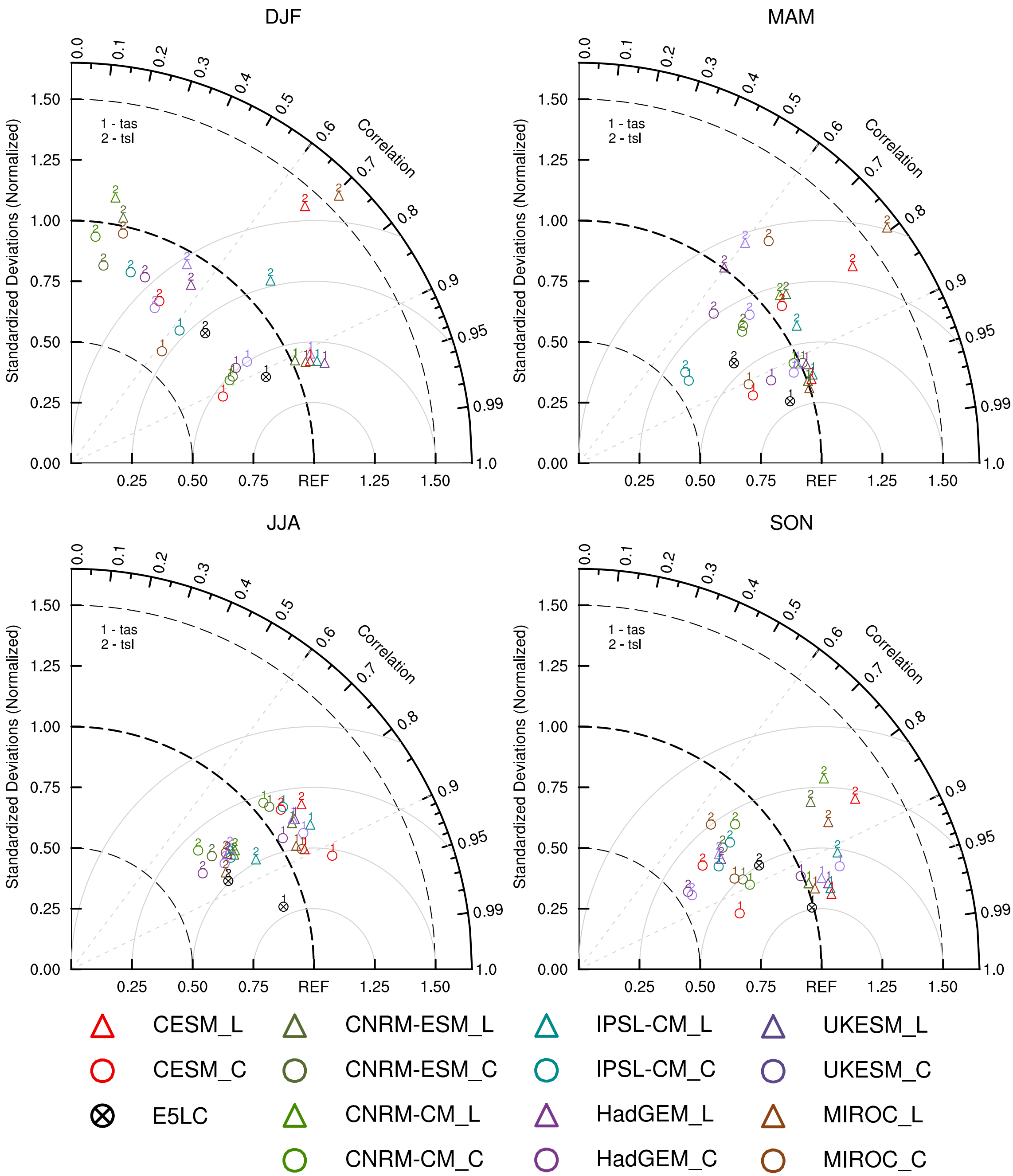


Figure 6. Taylor diagrams indicating the ~~spatial~~ correlation, normalized standard deviation, and unbiased root-mean-square (RMS) difference (gray circles) for the four seasons of the simulated variables tas, tsl (labeled by numbers) against observations (REF) at valid permafrost sites (here: sites with mean winter $tas < -25\,^{\circ}\mathrm{C}$). Normalization is applied using the standard deviations of the observations. The correlation used in this figure is the Pearson correlation coefficient. Colors indicate different models, and black crossed circles indicate E5LC. Triangles show the results of the LS3MIP runs, and solid circles of the CMIP6 runs. The REF indicates the corresponding observation.

Using the classic definition of permafrost, which requires more than two consecutive years of temperatures below 0 °C, we categorized 19 ~~permafrost sites were selected~~valid permafrost sites (see section 2.6). Notice that in the permafrost region, the largest EB and ES are shown (see Fig. 5). ~~The simulation performance at the permafrost sites was quantified using Taylor diagrams (Taylor, 2001). The 30-year average of simulated data was compared to the observations in all seasons at all sites in the permafrost region. The REF in Fig. 6 is the corresponding observation.~~

In JJA, both ensembles showed a high correlation with the observations (higher than 0.7). Both ensembles simulated JJA tas within a normalized standard deviation ranging from ~~0.95 to 1.1~~1.00 to 1.20, while E5LC's normalized standard deviation was lower than ~~0.75~~0.95. In the other seasons, the tas simulations had correlations higher than 0.6; however, ~~they all underestimated the~~ most CMIP6 models underestimated the normalized standard deviation of tas. In MAM and SON, tas correlations were ~~almost above 0.9~~all above 0.85, which was slightly higher than tas correlations in JJA.

Almost all CMIP6 and LS3MIP models struggled to simulate tsl ~~in~~ at valid permafrost sites, exhibiting low correlations and large RMSE, especially during DJF. The correlations were generally higher than ~~0.5 for all~~ 0.6 for both temperature variables in the other seasons. E5LC had similar DJF tas simulation performance ~~to~~ as the CMIP6 models, but it simulated DJF tsl with the smallest RMSE ~~compared to~~ of all models. ~~When considering the same model across~~ For model pairs from CMIP6 and LS3MIP, the land-only simulations consistently ~~outperformed others in~~ showed higher DJF tsl ~~. This is because tas and snd were superior in LS3MIP (Fig. A3).~~ correlations than the corresponding coupled simulations.

From spring to autumn, the CMIP6 models tended to overestimate precipitation variability (see Fig. ~~A3~~A5). Overall, the ~~Group L~~ LS3MIP models and E5LC ~~performed better than the Group C~~ showed higher correlations than the CMIP6 models in the MAM and SON snd simulations. In MAM, the normalized standard deviation of LS3MIP snd was considerably larger than in DJF and SON, generally exceeding 1.25 ~~for Group C models and 1.5 for Group L models~~times the reference value. Using the same forcing, the snd correlations improved ~~in SON~~from autumn to spring, with all values above 0.8 except for UKESM1.0-LL, which had a correlation ~~of~~ around 0.7 in SON and MAM. The correlations of CMIP6 pr were similarly ~~low in all seasons. Despite~~ lower than 0.9 from autumn to spring. LS3MIP ~~showing~~ showed a high correlation (higher than ~~0.8) and good~~ 0.85) and a stable normalized standard deviation (~~0.9 to 1.1) of~~ 1.25 to 1.50) for pr in ~~DJF and MAM, CMIP6 and~~ SON, DJF, and MAM. Most LS3MIP ~~were unable~~ models were able to simulate snd correlations higher than 0.8 in these ~~two~~ seasons.

3.4 Climate Dependency of Modeled Temperatures

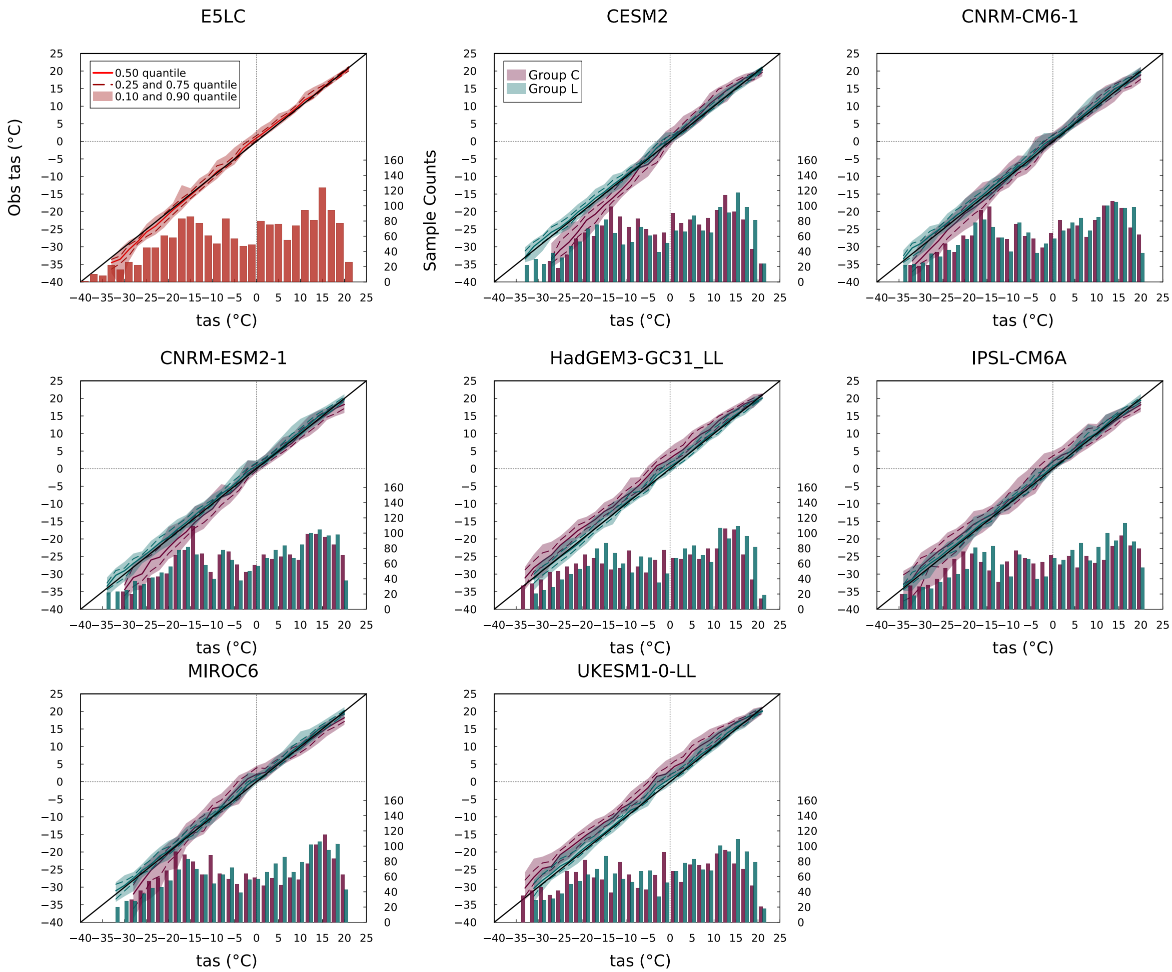


Figure 7. Quantile-Quantile (~~QQ~~Q-Q) and histogram plots (Wilks, 2019) for tas~~(Wilks, 2019)~~. The plots show 30-year mean monthly tas data of models at all sites and their observational values. The model ~~simulation outputs~~ simulations are binned ~~at~~ into 2 °C intervals. Each bar (or pair of bars for CMIP6 and LS3MIP) is plotted at the center of its interval (e.g., 1 °C for 0 to 2 °C). The ~~colored~~ solid and ~~dash~~ dashed colored curves ~~are~~ represent the median and 1st/3rd quartiles, respectively, of the corresponding observations for all data points ~~in~~ within the temperature interval~~, and the~~. The shaded area ~~is~~ indicates the ~~inter-decide~~ interdecile range. The histograms represent the sample size within each temperature interval, and temperature intervals with sample sizes smaller than 20 are excluded from the Q-Q plots. The further away the data is from the diagonal line, the larger the model's simulation bias is at certain temperature states. A higher vertical quantile range indicates more inconsistency with observation under identical temperature conditions.

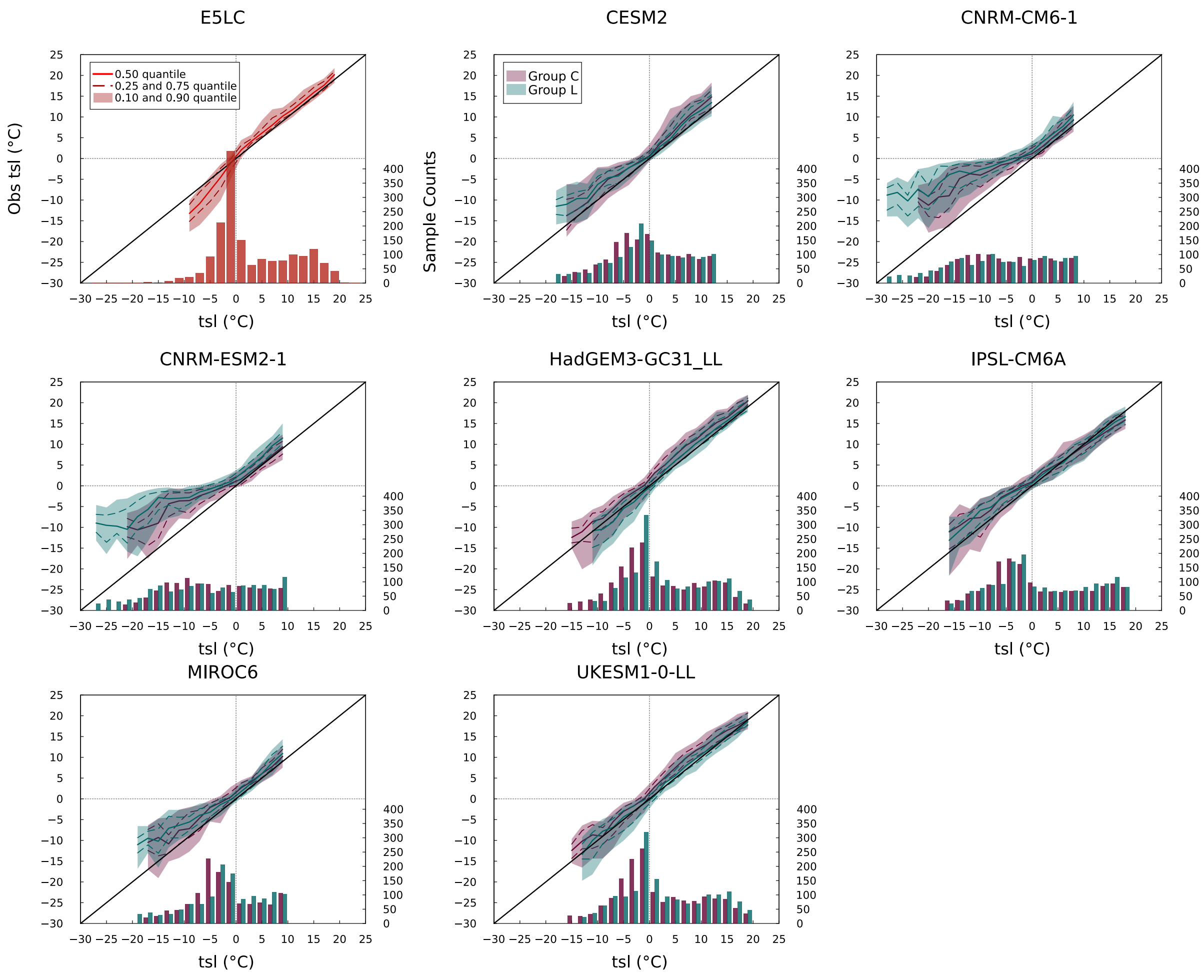


Figure 8. Same as Fig. 7, but for ~~tsl~~tsl.

Larger model biases and discrepancies were found in permafrost regions. To evaluate how model errors depend on the temperature conditions, we categorized all the model outputs of tas and tsl based on their values and compared these outputs to the corresponding observations. For each model, the monthly data from every site was calculated as an average over 30 years. We then examined the corresponding values for the same site and month with valid observations, calculating the median values, quartiles, and deciles for each interval.

As shown in Fig. 7, ~~Group L~~ LS3MIP models generally reproduced the observed sample distribution of tas and maintained narrow quantile spreads across the full temperature range. In contrast, ~~Group C~~ CMIP6 models exhibited larger deviations. For example, CESM2, CNRM-CM6.1, and CNRM-ESM2.1 tended to simulate tas higher than observed when the observed tas was below 0,°C, with median differences exceeding +3 °C. HadGEM3-GC31-LL and UKESM1.0-LL, by contrast, consistently produced tas values approximately 5 °C lower than observed across a broad range of frozen conditions. IPSL-CM6A-LR showed persistent cold biases of over -5 °C in most subzero-temperature cases. MIROC6 produced tas values that were lower than the observed values between -10 and 2 °C, but the simulated tas values were increasingly higher than the observed values when simulated tas dropped below -20 °C. A common feature of ~~Group C~~ CMIP6 models was a wider ~~vertical~~ spread of quantiles in the below-freezing temperature range, especially when tas dropped below -15 °C.

Compared to tas, quantile spreads were generally larger in the tsl simulations. For instance, the E5LC model exhibited a bimodal tas sample distribution (Fig. 7), with one peak near 15 °C and another around -15 °C, whereas its tsl sample distribution (Fig. 8) showed a single dominant peak between -5 and 2 °C. E5LC also displayed smaller and more centered tsl quantiles above 0 °C, but it showed an increasing warm bias and quantile spread as tsl dropped below freezing. Similar near-surface soil warming was observed in the LS3MIP run of HadGEM3-GC31-LL, which was associated with excessive snow depth. CNRM-CM6.1 and CNRM-ESM2.1 were the only models that did not show a cluster of tsl values near 0 °C. Instead, their tsl sample distributions were shifted toward much lower values (with extreme values lower than -15 °C), showing median differences from observations reaching -17 °C. CESM2 and MIROC6 also displayed systematic cold biases, with median tsl differences of -5 and -6 °C, respectively, when the observed tsl was around -10 °C. ~~Group L~~ LS3MIP models varied more in magnitude but generally showed positive tsl biases below 0 °C.

The two groups also showed distinct differences in simulating cold temperature extremes, particularly in the minimum tas. For example, CESM2's minimum tas in the coupled simulation was 6 °C higher than in its land-only simulation, indicating a notable warm bias in the coupled system at the lower extreme. In contrast, HadGEM3-GC31-LL and UKESM1.0-LL simulated minimum tas values that were 2 °C lower in the coupled runs than in their land-only counterparts. These discrepancies in tas at the cold end of the sample distribution were also reflected in the tsl simulations. In particular, the minimum tsl values differed by up to $\pm$6 °C between models in the two groups.

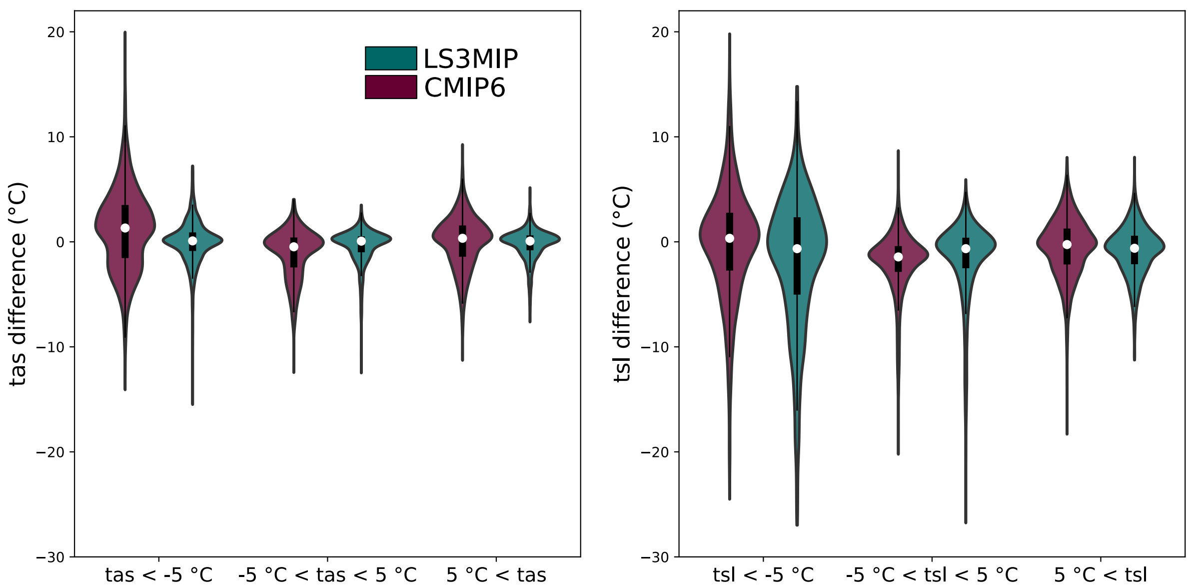


Figure 9. Difference between model tas (left), tsl (right) and corresponding observations. The x-axis represents different temperature intervals, and the y-axis is the 30-year average temperature difference ($T_{model} - T_{obs}$). Differences are categorized into three sets according to the 30-year average temperatures of every month at the observation sites: below -5 $^{\circ}$C (Set Frozen), -5 $^{\circ}$C to 5 $^{\circ}$C (Set Intermediate), and above 5 $^{\circ}$C (Set Warm). The violin plots show the distribution of the data. The width represents the density of the data points. The white dots show the median values, and the thick vertical black lines show the interquartile range.

Table 3. Statistics of the differences between the weighted multi-model ensembles and the observations of tas and tsl, as illustrated in Fig. 9. Here, kurtosis refers to excess kurtosis, which is kurtosis minus three.

Category	Sample Size	Ensemble	Q_1	Median	Q_3	Mean	Std. Dev.	Kurtosis
tas Set Warm	3475	CMIP6	~~-1.42~~ -1.4	~~0.34~~ 0.3	~~1.56~~ 1.6	~~0.01~~ 0.0	~~2.46~~ 2.5	~~0.76~~ 0.8
		LS3MIP	~~-0.81~~ -0.8	~~0.07~~ 0.1	~~0.63~~ 0.6	~~-0.22~~ -0.2	~~1.43~~ 1.4	~~2.93~~ 2.9
tas Set Intermediate	1480	CMIP6	~~-2.44~~ -2.4	~~-0.47~~ -0.5	~~0.41~~ 0.4	~~-1.12~~ -1.1	~~2.28~~ 2.3	~~0.77~~ 0.8
		LS3MIP	~~-0.99~~ -1.0	~~0.06~~ 0.1	~~0.52~~ 0.5	~~-0.38~~ -0.4	~~1.61~~ 1.6	~~7.79~~ 7.8
tas Set Frozen	3805	CMIP6	~~-1.55~~ -1.6	~~1.32~~ 1.3	~~3.51~~ 3.5	~~1.16~~ 1.2	~~3.85~~ 3.9	~~0.59~~ 0.6
		LS3MIP	~~-0.87~~ -0.9	~~0.09~~ 0.1	~~0.90~~ 0.9	~~-0.04~~ -0.0	~~2.06~~ 2.0	~~6.64~~ 6.6
tsl Set Warm	3295	CMIP6	~~-2.15~~ -2.2	~~-0.26~~ -0.3	~~1.26~~ 1.3	~~-0.51~~ -0.5	~~2.81~~ 2.8	~~1.60~~ 1.6
		LS3MIP	~~-2.13~~ -2.1	~~-0.61~~ -0.6	~~0.58~~ 0.6	~~-0.86~~ -0.9	~~2.35~~ 2.4	~~1.21~~ 1.2
tsl Set Intermediate	3655	CMIP6	~~-2.86~~ -2.9	~~-1.44~~ -1.4	~~-0.41~~ -0.4	~~-1.95~~ -2.0	~~2.87~~ 2.9	~~4.33~~ 4.3
		LS3MIP	~~-2.51~~ -2.5	~~-0.65~~ -0.7	~~0.40~~ 0.4	~~-1.70~~ -1.7	~~3.81~~ 3.8	~~6.41~~ 6.4
tsl Set Frozen	1445	CMIP6	~~-2.74~~ -2.7	~~0.34~~ 0.3	~~2.77~~ 2.8	~~-0.12~~ -0.1	~~5.26~~ 5.3	~~1.68~~ 1.7
		LS3MIP	~~-5.03~~ -5.0	~~-0.65~~ -0.7	~~2.34~~ 2.3	~~-1.81~~ -1.8	~~6.43~~ 6.4	~~1.24~~ 1.2

We subtracted the 30-year average of monthly observational data from all stations with the corresponding simulated values and then sorted them into observed temperature intervals, as shown in Fig. 9. The boundary of -5 °C and 5 °C was chosen to make sure the soil is completely frozen/thawed in the Set ~~Froze~~Frozen/Set Warm.

Statistical values of Fig. 9 were listed in Table 3. Set Frozen and Set Warm tas data sample sizes were more than twice as large as in Set Intermediate. Set Frozen had the largest standard deviations, ~~3.85~~3.9 °C in CMIP6 runs, and ~~2.06~~2.1 °C in LS3MIP runs. The mean tas values of the CMIP6 runs ~~were divided, with -1.12~~ranged from -1.1 °C and ~~0.01~~1.2 °C~~in~~. CMIP6 shows tas kurtosis from 0.6 to 0.8 in all three Sets. LS3MIP shows higher kurtosis, especially in Set Intermediate and Set ~~Warm, respectively~~Frozen (7.8 and 6.6, respectively).

The tsl samples were mainly concentrated in Set Intermediate and Set Warm. In contrast, there were also higher standard deviations in Set Frozen, ~~5.26~~5.3 °C and ~~6.43~~6.4 °C for CMIP6 and LS3MIP runs, respectively. The standard deviations of Set Frozen and Set Intermediate in LS3MIP runs were higher than those in CMIP6 runs by ~~1.18~~1.2 °C and ~~0.94~~0.9 °C, respectively. The mean and minimum ~~value~~ values of tsl bias ~~was~~ were much lower than ~~that~~ those of tas bias, and this negative bias was shown in all corresponding sets of both groups in Table 3. And tsl below -5 °C showed higher overall variability in ~~Group L than in Group C~~LS3MIP than in CMIP6, with the highest IQR of ~~7.37~~7.4 °C, and the highest standard deviation of ~~6.43~~6.4 °C among all categories. The kurtosis values of the two ensembles are more similar in tsl than in tas.

3.5 Snow Insulation

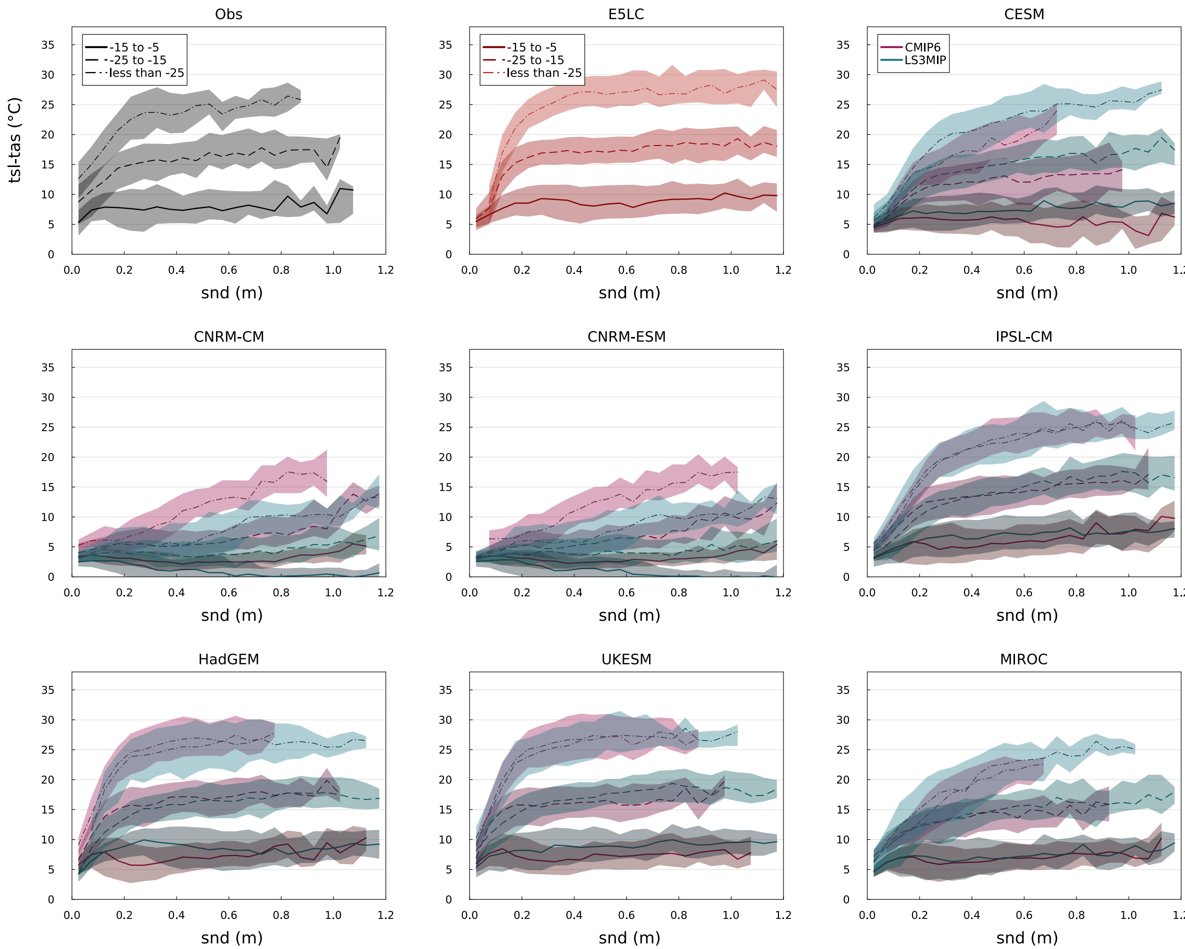


Figure 10. Snow depth and soil-surface temperature differences. The temperature categorization method follows ~~Wang et al. (2016)~~ Wang et al. (2016) but includes data from all seasons. The plots use different line styles and color schemes to distinguish between different temperature (tas) categories and ensembles (CMIP6 in red and LS3MIP in blue). Our sampling technique involves binning data at 0.05 m intervals, with coverage of ~~0–1.2~~0–1.2 m snd, and each interval contains a minimum of ten samples for accuracy. The lines represent the median values, while the shaded areas indicate the interquartile range (25th–75th percentile).

To better understand the insulating effect of snow in the models, we applied a method similar to that used in previous model evaluations (Wang et al., 2016; Burke et al., 2020). We collected the monthly observational data samples containing tas, tsl and snd and characterized them by three different tas intervals (-15 to -5 °C, -25 to -15 °C, and less than -25 °C). Analyzing the relationship between ~~temperature difference~~ the temperature difference $\Delta T = tsl - tas$ and snow depth allowed us to understand each model's snow insulation effect under different tas and snd conditions.

According to the observations shown in Fig. 10, even when snd ~~is very thin , the insulating effect of the soil surface layer was still large . Its intensity could range~~ was very thin there was a large insulating effect. The observed insulating effect ranged from 3 to 16 °~~C to 15~~C ~~according to~~ depending on tas. The impact of snow ~~on ΔT ($tsl - tas$) showed the greatest variability when the snow depth is shallow~~ varied most with shallow snow depths. However, once the snow depth exceeded ~~about 0.2 m~~some threshold, tsl stabilized near 0 °C or ~~lower~~slightly below, becoming nearly constant and minimally affected by tas. For ~~temperatures ranging from~~ tas in -15 to -5 °C, this ~~inflection point was reached when~~ threshold was reached with snd ~~was~~ about 0.1 m~~high~~. For colder temperatures, it was reached at depths around 0.2 ~~m~~ to 0.3 m. Thicker snow conditions limited the cooling of the underlying soils, as indicated by the strong relationship observed between ΔT and tas when snd exceeded ~~0.3 m. The median value of maximum~~ 0.3 m. Beyond the stabilization threshold, ΔT for the -5 ~~C~~ to -15 °C category was approximately 7.5 °C, ~~while~~ for the -15 ~~C~~ to -25 °C category ~~,~~ it was 15 °C, and it reached 24 °C when tas was below -25 °C. ~~When snd exceeded 0.3 m~~Thus, tsl ~~drops by 1~~dropped by ~ 1 °C for every 4 ~~C~~ to 5 °C ~~decreased~~ decrease in tas~~, considering the range of interquartile variations~~. E5LC showed similar insulation ~~in the two warmer temperature categories. However, it showed lower insulation when there was almost no snowand generally larger insulation in the coldest category~~for thick snow, but less and tas-independent insulation for very thin snow.

Only the land-only models of HadGEM3-GC31-LL ~~, IPSL-CM6A-LR,~~ and UKESM1.0-LL exhibited ΔT curves ~~that were similar to observations. Despite the generally good performance, HadGEM3-GC31-LL and UKESM1.0-LL had deficiencies in simulating soil insulation below -15 C, with ΔT 1~~ similar to ~~2 C higher than observation when snd was higher than 0.3 m~~the observations', with comparable insulation for thin snow depths and with stabilization beyond comparable thresholds. IPSL-CM6A-LR ~~also underestimated the insulating effect when there was almost no snow. However, it accurately reproduced the inflection point of the snd-T curve. After reaching this point, the insulation effect still continued to increase in IPSL-CM6A-LR. In contrast, HadGEM3-GC31-LL~~ and ~~UKESM1.0-LL showed a slower rate of increasing ΔT that was more comparable with the observations. The other models tended to underestimate the snow insulation , with the LS3MIP simulations from CESM showing results closer to observations, while the CNRM simulations showed the largest underestimation~~MIROC6's and CESM2's land models identified the thresholds, but did not stabilize. The CNRM models did not identify the threshold and largely underestimated the insulation with even negative ΔT in the warmest tas category.

Comparing the CMIP6 runs with the corresponding LS3MIP runs, HadGEM3-GC31-LL, UKESM1.0-LL, IPSL-CM6A-LR, and MIROC6 showed similar snow insulation effects in both ensembles ~~(CMIP6 ensemble and LS3MIP ensemble). HadGEM3-GC31-LL also showed comparable characteristics, although it differed under warmer conditions with lower snow depths.~~ with some underestimation in the warm tas category. The coupled CESM2 ~~'s LS3MIP runs showed a snow insulation effect closer to observations, but CESM2's CMIP6 runs simulated a much lower ΔT curve. In the LS3MIP CESM2 runs, under~~

~~the colder conditions (i.e., below -15 C), the snow insulation effect increased consistently with snow depth over all depths. Similar results were observed in MIROC6 as well. Both models showed realistic snow insulation effect values when~~ ~~snd~~ ~~is above 0.4 m. However, the models' initial insulation started low and increases more slowly in~~ ~~ΔT~~ ~~when~~ ~~snd~~ ~~was below 0.2 m. This was even more pronounced in~~ underestimated insulation in all tas categories, the CNRM-CM6.1 and CNRM-ESM2.1 ~~, which had a low~~ ~~ΔT-snd~~ ~~curve and a negative~~ ~~ΔT~~ ~~value for~~ ~~snd~~ ~~above 0.2 m under the -5 C to -15 C category of~~ overestimated insulation in all categories compared to its LS3MIP runs.

4 Discussion

4.1 ~~Winter 2 m Temperature~~ Atmospheric Forcing in ~~Target Area~~LS3MIP

~~We aimed to assess the models' performance under varying climate conditions to determine whether simulation uncertainties increase at lower temperatures or remain similar. To distinguish different climate regimes, a practical approach to categorizing the stations is to use their average DJF 2 m air temperature, following Wang et al. (2016). By focusing on winter temperatures, we could further link the results to the insulating effect of snow.~~

4.2 ~~Model Climatologies~~

~~Horizontal resolutiondid not have a major contribution to bias, concerning differences between~~ In this study, one focus is the LS3MIP models' ability to simulate snd and tsl when driven by the same prescribed atmospheric forcing. The similar underestimation of tas and overestimation of pr in LS3MIP and E5LC of about the observations' IQR, suggest similar biases in forcing and reanalysis. This reflects (a) the scale mismatch between the site observations and the gridded data that are aggregated data constrained by grid resolution, and (b) that the forcing is based on uncertain reanalysis data. The representation on a coarser horizontal grid has a minimal impact on the bias, as demonstrated by the differences between the ERA5-Land and E5LC models shown in Fig. 2. ~~Even though the LS3MIP models were forced with identical reanalysis data, their~~ ~~tas~~ ~~and~~ ~~pr~~ ~~medians differed slightly, as they have different grid resolutions and methods for calculating and regridding~~ ~~tas~~ ~~and~~ ~~pr~~~~. Also, the LS3MIP~~ ~~tas~~ ~~and~~ ~~pr~~ ~~differed from observations , highlighting the fact that a site's climate does not necessarily represent the climate of the corresponding grid. This emphasizes the importance of recognizing that gridded model outputs are spatial averages constrained by grid cell resolutions. Station observations, on the other hand,~~ But station observations are influenced by local conditions and may not represent the broader grid-scale climate. Therefore, ~~higher consistency between the two~~ high consistency between station observations and modeled climate does not necessarily imply ~~better model performance . Furthermore, discrepancies in temporal variability between modeled and observed climate can complicate interpretation when comparing with CMIP6 outputs. In this study, we focused on evaluating the models' ability to simulate~~ ~~snd~~ good model performance and ~~tsl~~ ~~with similar atmospheric forcing(~~vice versa without further representativity investigation and without averaging across an ensemble of observation sites.

Even though the LS3MIP models were driven with identical atmospheric forcing, their tas and pr ~~).The cold bias in LS3MIP tas climatologies and their closer pr agreement with ERA5-Land than with observations suggested shared biases from reanalysis and model forcing, which was also possibly~~ differed slightly (Figs. 2 and A1). These variations result from differences in how each model pre-processes the atmospheric forcing (e.g., elevation-dependent temperature adjustments based on the lapse rate due to differences ~~between grid data and site-level data.~~ in elevation between the model and the forcing data).

4.2 General Performance

The large inter-model spread in CMIP6 tas climatologies ~~indicated~~ highlights that even under the same historical radiative forcing, models simulate divergent near-surface climates in regions with frozen soils. This spread suggests substantial differences in model ~~configuration or parameterization. The family resemblance in tsl biases (e.g., CNRM,~~ configurations and parameterizations. However, there are four models which belong to two model families (the two CNRM models, and the HadGEM/UKESM ~~) points to structural similarities within model families. The reduced seasonal spread in JJA suggests that model performance was more consistent during snow-free periods, while snow dynamics played a role in winter divergence.~~ models) and show similar biases in tsl within their model family. Therefore, as suggested by Kuma et al. (2023), weighting by family-size should be considered when evaluating multi-model ensembles.

~~The excessive spread in JJA tas and tsl for specific CMIP6 models (CESM2, HadGEM3-GC31-LL, UKESM1.0-LL) points to seasonal over-responsiveness to summer variability. The excessive DJF RS and generally positive RB of pr were shown in Group C, in contrast to the well-constrained RB and RS of pr shown in Group L. However, both groups exhibited similarly large spreads and inconsistent biases in snd, suggesting that the snow-rain criterion remains a~~

Besides the snow model itself, the criterion for snowfall can be a potential source of uncertainty for the ~~frozen soil simulation. Each land-surface model's snow scheme—and its sensitivity to air and soil temperature—limits how much of that precipitation is accounted for as snow~~snd simulation. The land-only models use different criteria to determine whether precipitation falls as rain or snow, as well as the density of freshly fallen snow. The positive winter ~~relative pr bias and smaller, less systematic relative snd bias~~ bias of pr, along with the larger bias of snd across most models~~points to possible compensation mechanisms or~~, suggests a possible misalignment between precipitation input and snow accumulation. ~~With better RS and RB of tas and pr in Group L, models still had diverse abilities simulating tsl and~~ Both ensembles exhibited similarly large spreads and inconsistent biases in snd~~. This indicates challenges in accurately capturing the variability and relative bias for these specific variables across different models and seasons.~~, even under the identical precipitation conditions in LS3MIP.

The land surface models themselves were the major source of error in frozen soil climate simulations, even more influential than the atmospheric models. The ~~systematic over- and underestimations highlight the importance of further refining model approaches to improve performance in simulating snow and soil temperature dynamics. Snow in DJF, MAM, and SON (Fig. A1) and the prevailing phase change processes in the soil were likely the main reason for higher tsl interannual variability in these seasons. In JJA, all~~ LS3MIP ~~models showed low interannual variability, further indicating the role of snow in stable~~ simulations' tsl ~~simulation.~~

~~Given the smaller tas~~ mean ensemble bias (EB~~and the larger tsl EB in the LS3MIP than in the~~) magnitudes were substantially larger than their tas EB, exceeding the magnitude of CMIP6 ~~simulations, the bias spread introduced by the land-only models was large and is partially compensated for by the atmospheric component in the CMIP6~~ simulations~~. This explains why CMIP6 models show better~~' tsl ~~results. As shown in the subplots of~~ and even tas EBs (e.g., Fig. 5~~((C)tas-DJF) and (L) tsl-DJF), atmospheric models in CMIP6 tended to show warm biases in air temperature (tas), while LS3MIP land models exhibited cold biases in soil temperature (tsl) at the same locations, particularly in Siberia. When driven by the same forcing, however, land models revealed their intrinsic tendencies, which also contributed to differences in snow accumulation despite similar precipitation inputs. The notable differences in ES and EB across seasons further emphasize the importance of understanding diverse land-surface interactions in different seasons.Moreover, the seasonal consistency of Group L's tas ES suggests that the bias primarily arises from geographical features at station locations and how these were further represented at different model resolutions and surface configurations. In DJF,~~). This indicates that there were error compensation mechanisms between the land and atmospheric components in the ~~multi-model EB of~~ CMIP6 ~~and LS3MIP~~ ensemble. Especially, in frozen conditions, the spread in tsl ~~was negative, but it's positive for CMIP6 tas. And the ES of~~ bias is large in both ensembles (Fig. 9 and Table 3), even larger in the LS3MIP ~~tsl was larger than ES of~~ than in the CMIP6 ~~tas, although LS3MIP models were forced by the same atmospheric data~~ensemble. This indicates ~~larger variability caused by the land surface model than the atmospheric model~~a large spread in soil-snow insulation parameterization in frozen conditions.

4.3 ~~Permafrost Region~~

~~The large RMSE~~ The large RMSEs and low correlations of tsl in DJF (Fig. 6), compared to ~~better results in~~ the other seasons, highlight ~~a~~ seasonal dependency in simulation skill ~~, possibly tied to limitations in frozen-process representations. The consistency of~~ and reflect model deficiencies in the representation of snow insulation. The cold biases in tsl ~~accuracy with tas in JJA suggests that under unfrozen conditions, the models can simulate soil–atmosphere coupling more reliably.Moreover, better snow simulation ability improved soil temperature simulation performance. The lower tas correlation in JJA than in MAM and SON indicates possible deficiencies in representing surface energy exchange during summer. The~~ of some models (Fig. 4) were primarily due to insufficient surface insulation in these models, which will be discussed in more detail in the following subsection.

The large difference in the correlation performance of SON snd between ~~Group L and Group C highlights~~ LS3MIP and CMIP6 (Fig. A5) underlines the importance of a ~~good simulation~~ reliable pr simulation during the snow accumulation period. ~~Nevertheless,~~ Additionally, the low correlations and high normalized standard deviations of ~~Group C~~ the MAM and SON pr ~~from spring to autumn suggest that improvements are still needed~~ of CMIP6 models indicate the need for improvements in simulating precipitation~~seasonality~~.

4.3 ~~Climate Dependency of Modeled Temperatures~~

~~Cao et al. (2022) discussed an unreasonable warm tsl bias that is possibly due to the overestimation of permafrost snow depth in ERA5-Land. Similar warming of tsl due to excessive snd occurred in the LS3MIP simulation of HadGEM3-GC31-LL.~~

~~Similar to findings from CLM5 simulations Dutch et al. (2022), several~~ In JJA, three CMIP6 models (CESM2, ~~CNRM-CM6.1, CNRM-ESM2.1, and MIROC6) exhibited cold tsl biases despite warm~~ HadGEM3-GC31-LL, and UKESM1.0-LL) exhibited an excessive spread in tas, ~~likely due to underestimated snow and soil insulation. This resulted in excessive energy loss from soil to atmosphere when~~ pointing to an overestimation of summer variability by the atmospheric component. All CMIP6 models simulated lower tas ~~< -5 C. There was an excessively low tsl shown in Fig. 8. At 0.2 m depth, this cold bias mainly occurred when soil temperatures were below 0 C. As shown in Fig. A4 and~~ correlations with observations in JJA than in MAM and SON at the permafrost sites (Fig. ~~A5, at deeper soil layers (~~6). This reduced skill reflects deficiencies in representing summer surface energy exchange, when the absence of snow cover enhances land-atmosphere coupling and its complexity (more spatial variability in surface albedo, higher solar radiation, more intense soil-atmosphere energy flux, etc.). Nevertheless, tsl simulations performed best in the summer season, suggesting that, under unfrozen conditions, the representation of soil processes is comparably reliable.

In deeper soil layers (0.8 m and 1.6 m) ~~, the cold bias was also present when soil temperatures were above 0~~ tsl cold biases were present in all thermal regimes in both ensembles (Figs. ~~C. This phenomenon exhibited strong model dependence, suggesting that the cause was not only related to insufficient representation of surface and snow insulation but may also stem from factors such as overly strong thermal resistance in~~ A and A7). This suggests deficiencies in the ~~land-only model parameterizations or low soil moisture content.~~ representation of soil heat conductivity and capacity. One possible cause is an underestimation of soil moisture, which reduces the effective thermal conductivity and heat capacity while weakening the buffering effect of phase-change latent heat.

~~In addition, we categorized the model output by the freeze/thaw state of observation.This enabled us to compare the performance of two groups in different temperature states. To avoid assessment errors caused by different sample sizes, we discussed the overall uncertainty exhibited by the model in the thawed state, the freeze-thaw transition state, and the frozen state, using the boundaries of -5 C and 5 C, as the phase change process occurred most frequently between the boundaries. Low kurtosis of CMIP6 tas difference shows that the CMIP6 ensemble fails to reproduce the realistic spatio-temporal variation of permafrost tas. The larger standard deviations in LS3MIP tsl results compared to CMIP6 indicate that the land-only models produce higher variability in Set Intermediate and Set Frozen. The consistently negative tsl bias implies that land models systematically simulate colder than observed soil temperatures. Notably, only in Set Intermediate LS3MIP and CMIP6 simulated high kurtosis.~~

4.3 Model Features

The ~~close-to-zero median and high kurtosis in Group L tas did not translate into similarly good results of tsl. This further demonstrates that land surface models exhibit distinct simulation tendencies in frozen soil. When driven by identical atmospheric forcing, the differences in LS3MIP tsl among models are substantially greater than those observed in their corresponding CMIP6 simulations. Furthermore, the elevated IQR and standard deviation of the tsl difference in Group L when tsl was below -5 C suggest that discrepancies in snow insulation representation, soil layering schemes, or freeze-thaw parameterizations contribute to greater simulation inconsistency within the model ensemble.~~

4.4 ~~Snow Insulation~~

~~Two key phenomena should be considered regarding the snow insulation effect . Firstly, the insulating effect of the soil surface layer itself remained influential even when snd is very thin. Secondly, the~~ impact of snow on ΔT ($tsl-tas$) showed the greatest variability when the snow depth was shallow ~~, stabilizing~~ and tsl stabilized near 0 °C or lower once the snow depth exceeded about 0.2 m. This stabilization was minimally affected by tas, with ~~inflection points~~ snd threshold reached at different depths depending on temperature~~ranges~~. The insulating effect of snow is due to its low thermal conductivity~~(about), which~~ , which ranges from about $0.05\,\mathrm{W\,m^{-1}\,K^{-1}}$ for fresh snow to about $0.7\,\mathrm{W\,m^{-1}\,K^{-1}}$ for dense snow (e.g., Sturm et al., 1997; Calonne et al., 2011) . This low thermal conductivity reduces heat loss from the soil ~~. This effect became~~ and depends mainly on snow density. The insulation effect becomes more pronounced at lower tas as the temperature difference (ΔT) increases.

~~The results from Fig. 10 suggest that factors beyond tas and snd impact the ability of the snow layer to impede energy transfer. Evaluating the snow insulation effect based on LS3MIP or CMIP6 runs alone may lead to different conclusions. For instance, the CESM2 LS3MIP run showed a snow insulation effect closer to observations, whereas the CESM2 CMIP6 run simulated a much lower ΔT curve. The issue with their snow insulation dynamics was the low initial insulation and slow increase in ΔT when snd was below 0.2 m.~~

Four land models mentioned in this study ~~are~~ were newer versions of the land models studied by ~~Wang et al. (2016)~~ Wang et al. (2016). Comparing Table A1 with ~~findings of Wang et al. (2016) revealed notable advancements in newer models.~~ the findings of Wang et al. (2016) revealed advancements in the newer model versions. The newer versions of JULES, used in HadGEM3-GC31-LL and UKESM1.0-LL, provided more realistic initial insulation at low snd values and did not overestimate the insulation effect at higher snd values (lower RMSEs in all 3 categories). In particular, they updated the snow thermal conductivity formulation from that described in Yen (1981) to that described in Calonne et al. (2011). Compared to their previous versions, ORCHIDEE (IPSL-CM6A-LR) and MIROC6 exhibited a larger, closer to observations snow insulation effect (lower RMSEs in all 3 categories).

CLM5.0 (indicated as CESM LS3MIP in Fig. 10) generated ΔT profiles that more closely resemble observed values within the range of -5 ~~C~~ to -15 °C with an RMSE of ~~0.79~~0.8 °C (Table A1) compared to CLM4.5 (RMSE of ~~1.46~~1.5 °C in ~~Wang et al. (2016)~~ Wang et al. (2016)). However, in the colder categories, it did not ~~perform better. The newer versions of JULES, used in HadGEM3-GC31-LL and UKESM1~~show improvements. Dutch et al. (2022) and Damseaux et al. (2025) used an alternative parameterization of snow thermal conductivity (Sturm et al., 1997) in CLM5.0 ~~-LL, provided better initial insulation at low snd values and did not overestimate the insulation effect at higher snd values (lower RMSE in all 3 categories). Compared to its previous version, ORCHIDEE (IPSL-CM6A-LR) exhibited a larger snow insulation effect that was closer to observed values (lower RMSE in all 3 categories). MIROC6 also showed lower RMSE compared to its older version in all three temperature categories~~simulations in the Arctic tundra, which reduced the cold-soil temperature bias. Furthermore, the study by ~~Burke et al. (2020) also~~ Burke et al. (2020) compared the insulation effect of various CMIP6 and CMIP5 models, but only within the -15 ~~C~~ to -25 °C category~~and among coupled climate models from CMIP6 and CMIP5, not land-only models~~. Surprisingly, their results showed a degradation from CESM1 to CESM2 (CLM4.5 to CLM5.0) when it comes to representing snow insulation. Our results (CESM CMIP6 in Fig. 10) confirmed ~~this finding, though~~ their finding, but with a better performance of the land-only ~~simulations~~ simulation (CESM LS3MIP)~~performed better~~.

4.4 ~~Impact of Land Model Features on Performance~~

In CESM2, the coupled run generally exhibits weaker insulating effects than the land-only run across all temperatures. If the biases in snow insulation were solely due to snow depth, the two ensembles would exhibit converging temperature difference curves in Fig. 10. Thus, additional processes in the coupling influence and potentially diminish the insulating effect of snow. Evaluating the snow insulation effect based on LS3MIP or CMIP6 runs separately would lead to different conclusions.

~~The strong cold bias of CNRM models may be linked to how the models handle snow cover fraction.These models allow~~ The CNRM models simulated a winter cold tsl bias and low snow insulation compared to observations (see Figs. 4 and 10). Their land model allows for a snow-free fraction within vegetated areas, ~~thereby reducing the thermal insulation effect and resulting in colder simulated soil temperatures. The weaker snow thermal conductivity~~ even at higher snow depths, depending on the vegetation fraction, roughness length, and snow depth (Decharme et al., 2019). This approach reduces the grid-scale thermal insulating effect of snow in boreal forests, where only 20–50 % of the grid cell is considered to be snow-covered (Wang et al., 2016). The low snow insulation in the CNRM ~~models was also confirmed by the results in Fig. 10. As highlighted in Decharme et al. (2019) and Wang et al. (2016), this issue contrasts with observations, which are mostly free of intercepting vegetation. This exaggerates the snow 's insulation effect in observations in the region. However~~simulations cannot be attributed to its snow thermal conductivity formulation, as it uses the same formulation as ERA5-Land. Here, we use the observation as the reference, but it should be noted that observational sites are typically located in areas with low vegetation, which might yield large snow insulation effects at site locations. Furthermore, land surface temperatures in boreal forests are typically warmer in winter ~~, mainly due to higher albedo~~ compared to openly snow-covered areas (Li et al., 2015). ~~Regardless of differences in parameterizations or whether the simulations were coupled or uncoupled, all models analyzed underestimated soil temperatures in autumn~~Snow-vegetation interaction is a key source of uncertainty in soil temperature that must be considered when evaluating model performance against site observations.

~~Better snow and soil temperature (under snow)simulations are related not only to snow parameterization but also to other land model processes, such as soil~~ MIROC6 relies on fixed-value parameters in snow density, snow thermal conductivity, and bottom boundary condition (Table 2). However, the performance in the land-only simulations was average compared to the other models. Still, the snow thermal insulation effect was underestimated (Fig. 10). Also, the tsl of coupled MIROC6 shows the highest variability (about two times the observed variability) among CMIP6 models (see Fig. 3). Thus, this model could benefit from using a more dynamic snow parameterization.

Another important factor for frozen soil modeling performance is the soil bottom boundary conditions. HadGEM3-GC31-LL and UKESM1.0-LL had good performance in simulating tsl with their shallow soil column. ~~Moreover~~Specifically, the zero-flux assumption is ~~possibly~~ likely more influential on the soil surface when ~~used in~~ applied to a shallower soil column~~;~~ , as it constrains the RS of soil temperature in winter. In general, defining a bottom flux showed better results than providing fixed values when considering the impact of bottom boundary conditions on soil temperature ~~.~~ (see Table 2 and Figs. A and A7).

~~None of the models could accurately depict soil insulation.~~ When snd was below 0.05 m, all models simulated too small ΔT ~~(Fig. 10) that was lower than the observed value. This shortage could lead to an overestimation of the total ΔT if the model soil surface insulation is increased.~~ values. The shortage could be due to an insufficient representation of the ~~thermal offset~~soil insulation, which is controlled by factors such as soil texture, soil moisture, and surface organic matter. Organic ~~layer can accumulate~~ matter accumulates up to 15 cm on top of the surface of frozen soil with porosity greater than 0.95 (Boike et al., 2013)~~. It~~, and it is an important factor in the thermal dynamics of the soil surface (Zhu et al., 2019). ~~Incorporating~~ Therefore, incorporating the impact of the surface organic layer improved the soil temperature simulation in land surface models (Ekici et al., 2014; Chadburn et al., 2015). ~~It was especially true when some models adequately reproduce the insulation magnitude at high snd.~~ The HadGEM3-GC31-LL, UKESM1.0-LL, and IPSL-CM6A-LR models exhibited the best surface-soil insulation (cf. RS and RB of summer tsl). Under low snd (lower than 0.05 m), the UKESM1.0-LL simulated large insulation effects of 4 to 12 °C~~under low snd (lower than 0.05 m)~~, which was closest to the observation and which might guide ~~to~~ further refinements. However, the contribution of surface-soil insulation remained insufficiently represented in the other models. Soil moisture critically governs permafrost thermal behavior: high water content lowers frozen-soil thermal conductivity and heat capacity (Langer et al., 2011a; Jafarov et al., 2020). Low soil moisture content and the absence of an explicit phase change process can lead to a cold bias in model simulations, making soil temperatures more sensitive to atmospheric forcing. For example, models such as IPSL-CM6A-LR exhibit rapid summer thawing, likely due to insufficient latent heat buffering from soil moisture (Burke et al., 2020).

~~The HadGEM3-GC31-LL, UKESM1.0-LL, and IPSL-CM6A-LR models exhibited the best soil surface insulation when considering the RS and RB of summer tsl. However, the significance of surface soil insulation was still underestimated.~~

~~It is challenging to identify how model features influence the vertical energy transportation process without conducting sensitivity experiments. Different physical processes represented in land surface models may interact in complex ways, either synergistically or oppositely, affecting the model's simulation capability. Without the ability to isolate the effects of these various processes, it is difficult to determine whether simulation errors resulted from one specific scheme or multiple overlapping processes.~~

~~However, it should be noted that the models with a better representation of snow insulation (Fig. 10), namely HadGEM3-GC31-LL, UKESM1.0-LL, and IPSL-CM6A-LR, use more recent formulations for snow thermal conductivity (Calonne et al., 2011; Wang et al., 2013). In particular, the updated snow thermal conductivity formulation in JULES (HadGEM and UKESM) from Yen (1981) to Calonne et al. (2011) improved the snow insulation effect, as demonstrated by a comparison of our results with Wang et al. (2013). MIROC6 was not the worst in terms of snow insulation, even though it assumes snow density and thermal conductivity to be constant. Furthermore, the low snow insulation in the CNRM simulations could not be attributed to its snow thermal conductivity formulation because it uses the same formulation as ERA5-Land. Their total insulation also depends on snow density and snow cover fraction.~~

5 Conclusions

This research investigated coupled CMIP6 and land-only LS3MIP historical climate simulations in frozen soil areas. Errors caused by the land surface models versus the errors caused by atmospheric forcing or coupled models were quantified and discussed.

~~Except in summer months, inaccurate interannual variability in the simulation of soil temperature by~~ In winter and in the transition seasons, relative biases (down to -15) and overestimated relative variability (up to 2.5) in the CMIP6 ~~models is~~ and LS3MIP simulations' soil temperature were mainly caused by deficiencies in the land surface models~~and is less inherited from atmospheric components. Biases in~~. Better soil temperature and snow performance in CMIP6 than in LS3MIP simulations do not indicate that the land surface ~~models even partially compensate for the influence of air temperature biases~~component is responding realistically to atmospheric forcing, but that the atmospheric component compensated for errors in the land component. Similarly, improved precipitation simulation does not necessarily lead to ~~better~~ improved snow depth results in winter and spring~~. However, in autumn, better *snd* was simulated with better *pr* in~~, but in autumn (as shown by comparison with the realistically forced LS3MIP ~~. Land surface models performed better when coupled to an atmospheric model (CMIP6~~simulations). Balancing ~~tuning in~~ and tuning of coupled climate models ~~with achieving physically accurate~~ against forced land-only simulations is a ~~key consideration for future model development. Good soil temperature and snow performance in the coupled model do not necessarily indicate that the land surface component is responding realistically to atmospheric forcing.~~ path to more robust future CMIP simulations.

~~Soil temperature biases and their spread between models are much more evident in winter than in the other seasons. Spatially, the models exhibit larger disagreements in reproducing the soil temperature of permafrost sites. The largest bias standard deviation disagreements of *tas* and *tsl* are observed in Set Frozen (temperature lower than -5 C) for both ensembles (see Table 3). These indicate a limitation of models reproducing the *tsl* relationship with *tas* in freezing conditions. Land surface models need to incorporate or improve processes related to frozen soil and soil hydrothermal dynamics in frozen conditions. This includes enhancing the simulation of soil moisture content, refining soil thermal and hydraulic parameterizations in frozen states, and representing key features of frozen soil, such as excess ground ice and surface organic matter.~~ While improved parameterization schemes can enhance the realism of permafrost simulations, they often increase model complexity and may introduce additional sources of uncertainty. For example, increasing the number of soil layers and depth in land surface models can improve the physical representation, but this comes at the cost of larger parameter uncertainty. Achieving a balance between physical realism and model performance remains a necessary step in advancing climate modeling in permafrost.

~~The deficiency of land surface models is reflected in the ability to simulate snow depth and/or to represent the effect of thermal insulation~~ The models showed limitations in reproducing the thermal insulation effect in freezing conditions. Snow insulation plays a critical role in modulating soil temperatures. Updating the parameterization of snow thermal conductivity, as ~~demonstrated~~ done in recent models such as HadGEM3-GC31-LL and UKESM1.0-LL, ~~can enhance the insulation effect of snow~~ enhanced the representation of the snow insulation effect. Similarly, Damseaux et al. (2025) showed that using a snow thermal conductivity parameterization better suited to permafrost regions in CLM5.0 improved the thermal insulation effect. However, the insulation effect is not solely determined by thermal conductivity parameterization. Accurate parameterizations of snow density, snow depth, and snow cover fraction are also important factors. ~~In particular, snow cover controls the spatial continuity of insulation. Even when snow depth is accurately simulated, inadequate thermal insulation can result from insufficient snow cover because exposed ground patches allow greater heat loss. Addressing these factors in future model development is essential for improving the representation of snow insulation and, consequently, soil temperatures.~~ Furthermore, land surface models need to incorporate or improve processes related to frozen soil and soil hydrothermal dynamics in frozen conditions. This includes enhancing the simulation of soil moisture content, refining soil thermal and hydraulic parameterizations in frozen states, and representing key features of frozen soil, such as excess ground ice and surface organic matter.

It is challenging to identify specific model features that influence the surface energy balance without conducting sensitivity experiments. As discussed, different representations of physical processes in land surface models interact with each other in complex ways. These interactions can either amplify or dampen climate variability, and thus affect the models' simulation capability. Future studies should focus on isolating the effects of various processes to determine if simulation errors resulted from one specific parameterization or from multiple interacting parameterizations.

Note that the main scope of this study ~~is~~ was limited to soil depths down to 0.2 m and that the thermal state of frozen soils is not determined solely by temperature (Groenke et al., 2023). ~~It is essential to consider various~~ Various hydrothermal processes within the deeper soil, including thermal offset, permafrost active layers, ~~seasonally frozen soil~~ seasonal freezing depth, and transport of heat and water~~transport. These factors play a critical role~~, are additional critical factors in capturing frozen soil dynamics and must be investigated further.

Code availability. The scripts for the data processing and analysis can be found at Zenodo: https://doi.org/10.5281/zenodo.17157176

Author contributions. ZL and BA determined the research outline and methodology, and wrote the initial manuscript. ZL and DR analysed the results and edited the manuscript. BA provided guidance on data analysis. ZL collected the data, generated the figures, and was responsible for the code calculations. Results were discussed by all authors.

Data availability. CMIP6 and LS3MIP multi-model ensemble data (doi: 10.22033/ESGF/CMIP6.4066, Voldoire (2018), doi: 10.22033/ESGF/CMIP6.4095, Voldoire (2019b), doi: 10.22033/ESGF/CMIP6.4068, Seferian (2018), doi: 10.22033/ESGF/CMIP6.9599, Voldoire (2019a), doi: 10.22033/ESGF/CMIP6.5195, Boucher et al. (2018), doi: 10.22033/ESGF/CMIP6.5205, Boucher et al. (2019), doi: 10.22033/ESGF/CMIP6.5603, Tatebe and Watanabe (2018), doi: 10.22033/ESGF/CMIP6.5622, Onuma and Kim (2020), doi: 10.22033/ESGF/CMIP6.6109, Ridley et al. (2019), doi: 10.22033/ESGF/CMIP6.14460, Wiltshire et al. (2020b), doi: 10.22033/ESGF/CMIP6.6113, Tang et al. (2019), doi: 10.22033/ESGF/CMIP6.14462, Wiltshire et al. (2020a), doi: 10.22033/ESGF/CMIP6.7627, Danabasoglu (2019b), doi: 10.22033/ESGF/CMIP6.7650, Danabasoglu (2019a)) were downloaded from esgf-node.llnl.gov/projects/cmip6 on 2023-05-22.

ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi: 10.24381/cds.68d2bb30 were accessed on 2024-08-08.

The daily observational data from RIHMI-WDC can be collected from aisori-m.meteo.ru.

Competing interests. The contact author has declared that none of the authors has any competing interests

Acknowledgements. Zhicheng Luo gratefully acknowledges the China Scholarship Council (CSC) sponsorship for Z.L. (No.202006040064). BA acknowledges support by DWD IDEA S4S ~~–~~- project FS-SF (4823IDEAP2). This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bb1064 and of Goethe-HLR. The authors thank Professor Duoying Ji from Beijing Normal University for providing valuable insights and support for this research. Thanks to Mittal Parmar for her ~~contribution to this research~~comments and discussions regarding this paper. We express our gratitude to the World Climate Research Programme for its coordination and support of CMIP6 through the efforts of its Working Group on Coupled Modelling. We appreciate the climate modeling groups for producing and providing their model output, the Earth System Grid Federation (ESGF) for archiving the data and ensuring access, and the multiple funding agencies that support CMIP6 and ESGF.

Appendix A: Additional Figures

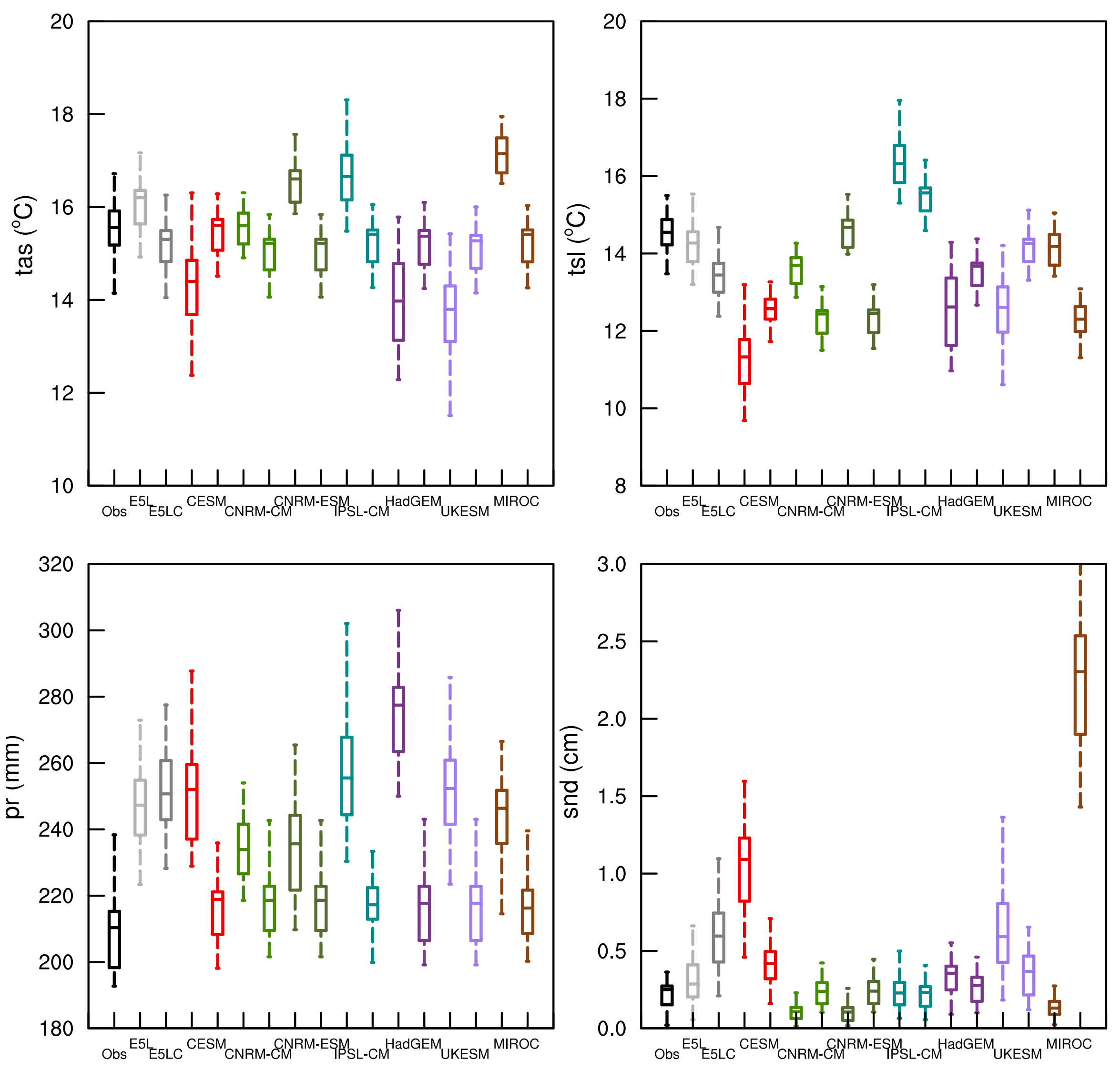


Figure A1. Sites' locations averaged JJA-climates (1985–2014) of hydrothermal variables as observed and simulated. Names on the x-axis and the colors indicate the different data sources. Model names indicate CMIP6 model output (left), and LS3MIP (right; see Table 1). Each CMIP6-LS3MIP pair shares the same color. The boxes represent the medians, first and third quartiles; the $\pm$ 1.5×IQR or the maximum and minimum values, if within the former range, are taken as the whiskers' length.

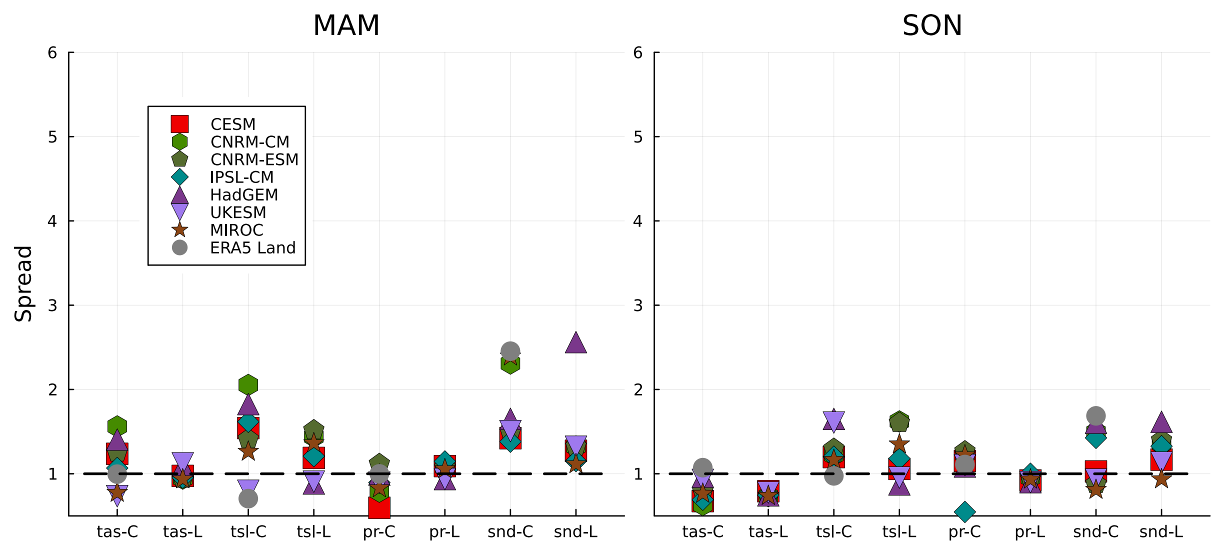


Figure A2. Relative spread (RS) of the sites-averaged climates (~~1985–2014~~1985–2014) of the four variables in both ensembles and in E5LC with reference observations for all seasons. The colors indicate the models, and the x-axes show the variables and ensembles (C indicates CMIP6, and L indicates LS3MIP runs).

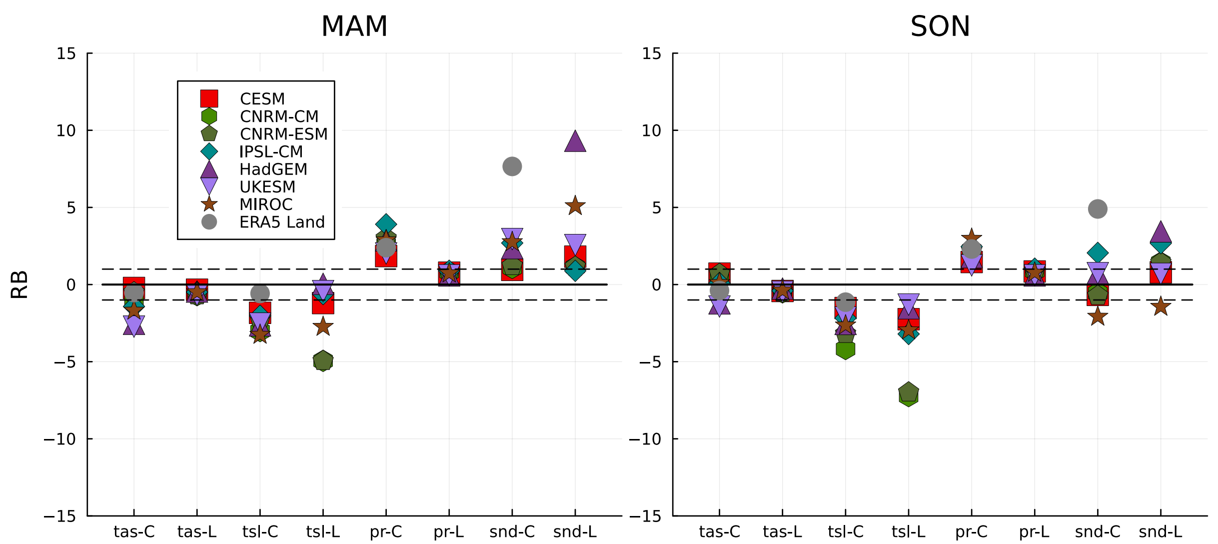


Figure A3. Same as in Fig. ~~A1~~A2, but for relative biases (RB). The dashed lines (from -1 to 1) indicate the range of absolute median differences smaller than the observation's ~~IQR~~IQR.

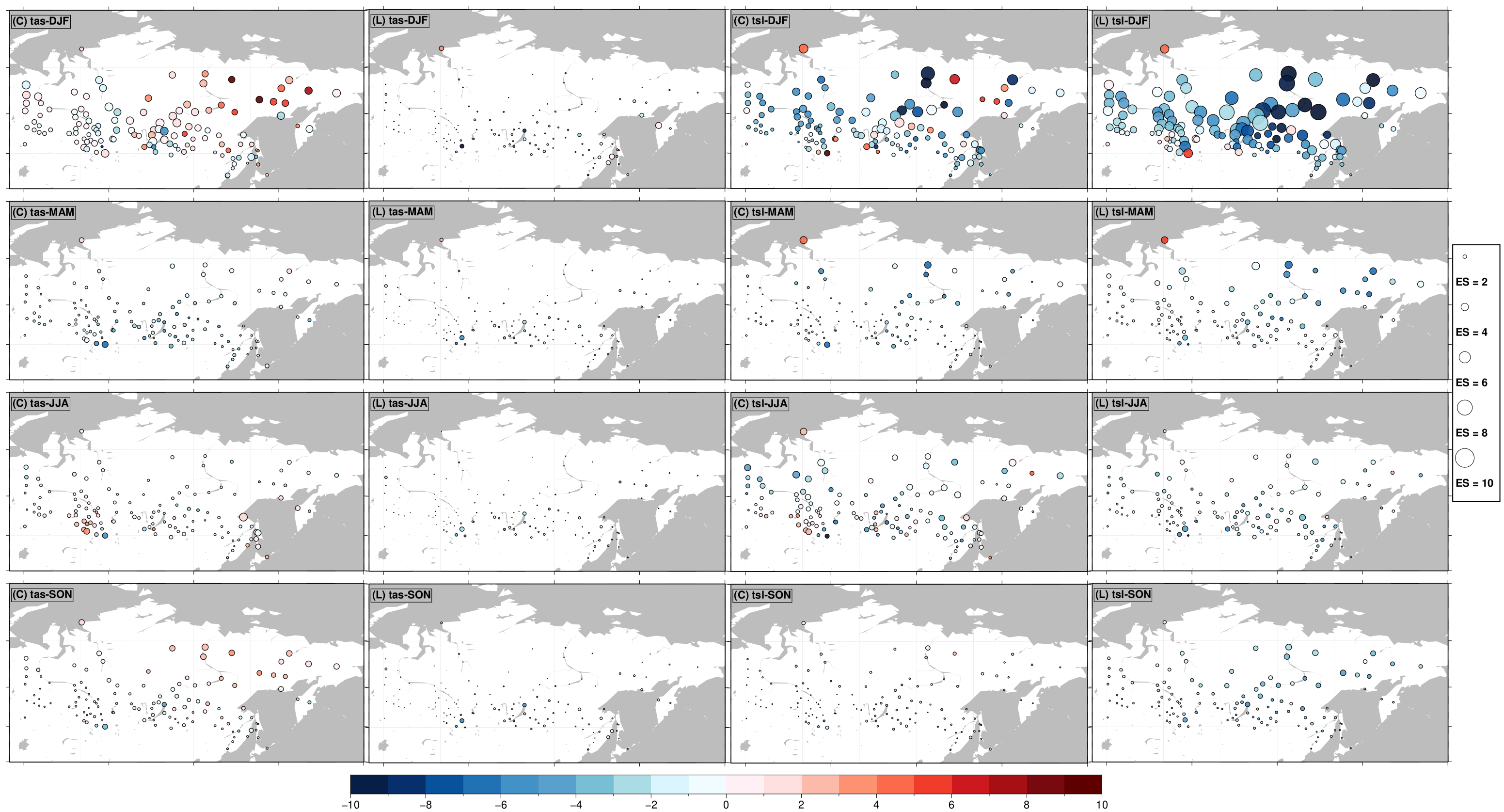


Figure A4. Weighted multi-model mean biases (EB) and their standard deviations (ES) of tas and tsl at observational sites in °C. The first and third columns are the results of CMIP6 models, and the second and fourth columns are the results of LS3MIP models. The first to fourth rows show the DJF, MAM, JJA, and SON seasons, respectively. Colors indicate the bias value, while the circle radii represent biases' standard deviations, i.e., the ensemble spread.

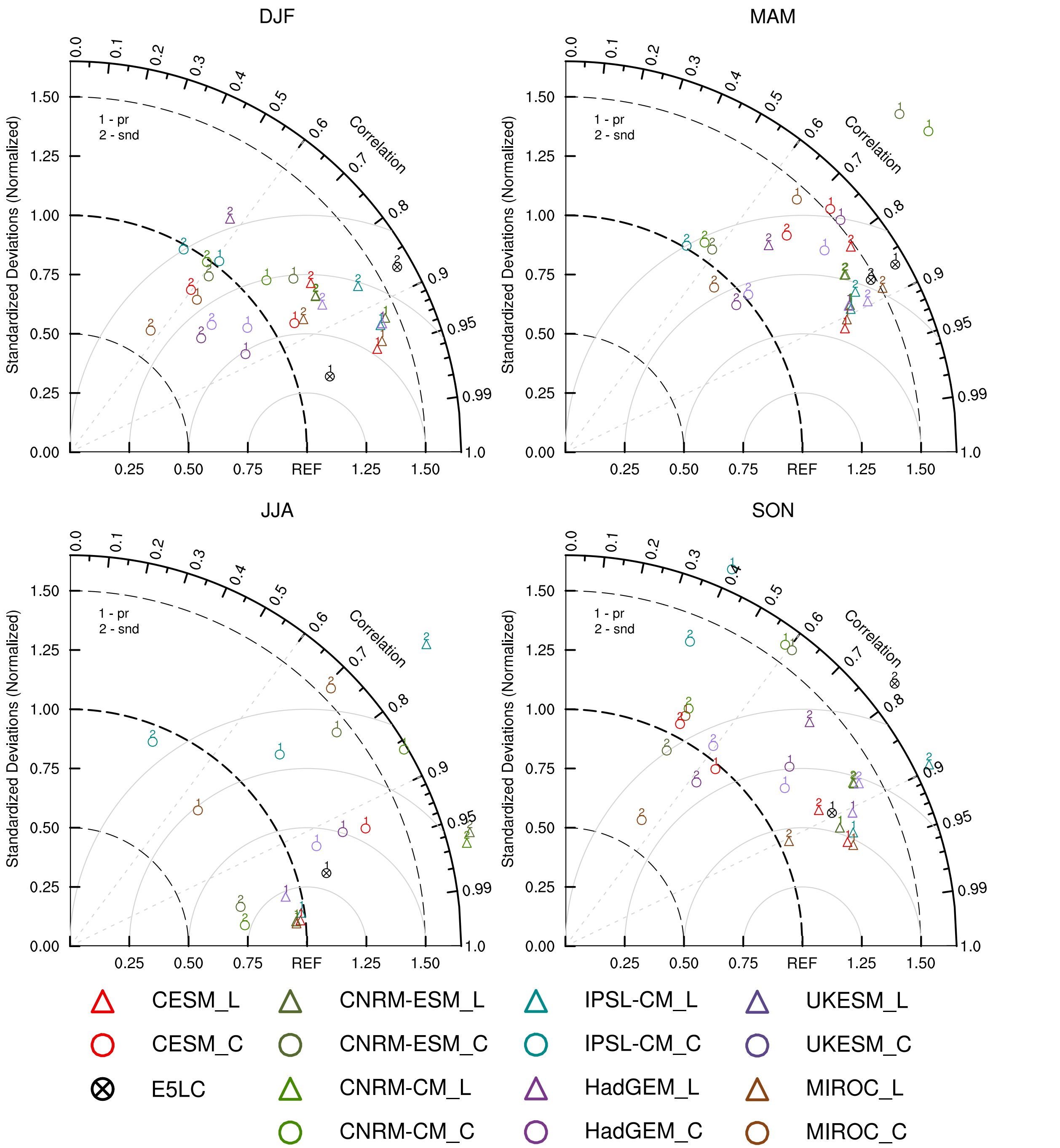
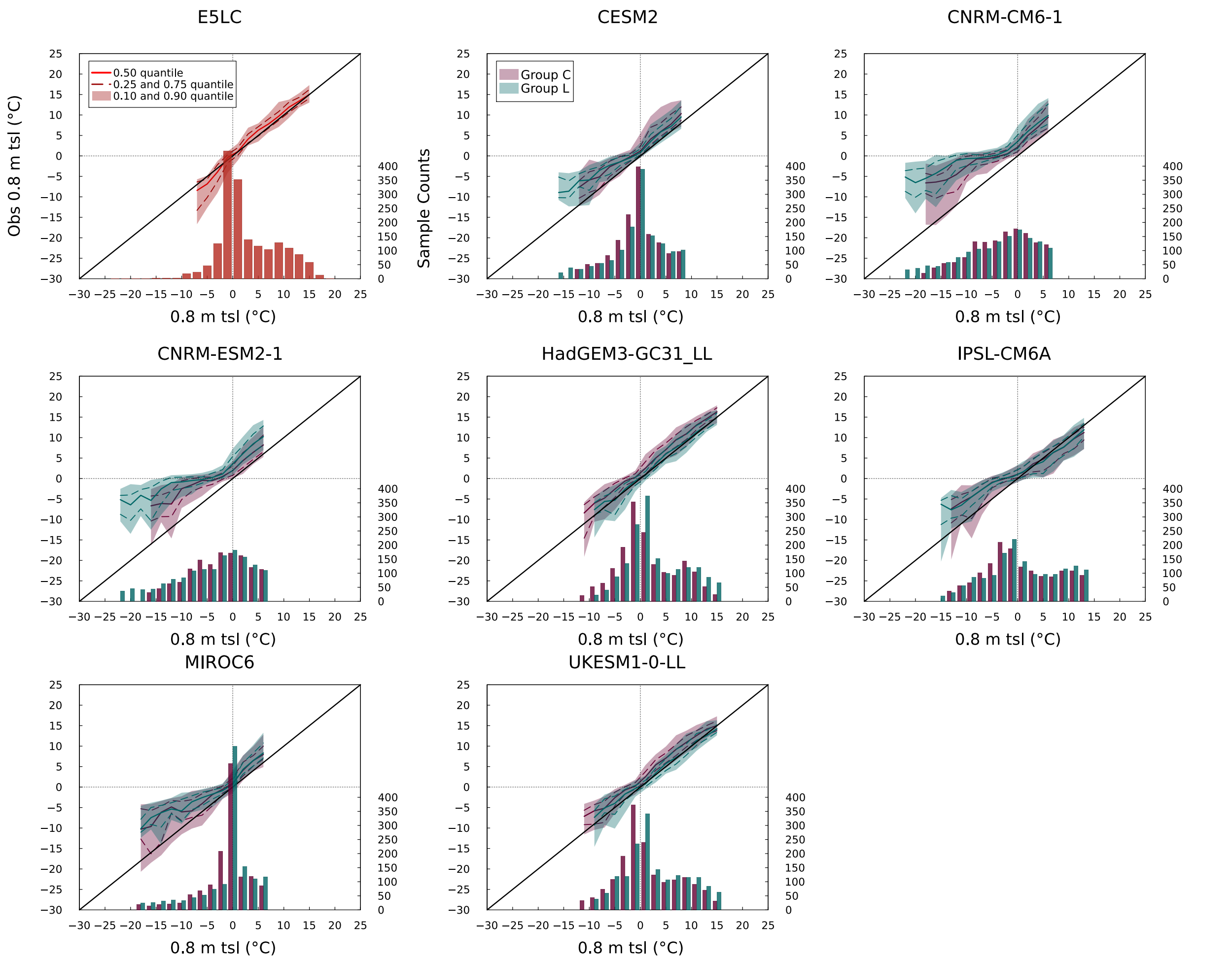


Figure A5. Taylor diagrams indicating the spatial correlation, standard deviation, and unbiased root-mean-square (RMS) difference (gray circles) for the four seasons of the simulated variables pr, snd (labeled by numbers) against observations at valid permafrost sites (here: sites with mean winter $tas < -25\,^{\circ}\mathrm{C}$). Normalization is applied using the standard deviations of the observations. Colors indicate different models, and black crossed circles indicate E5LC. Triangles show the results of the LS3MIP runs, and solid circles of the CMIP6 runs.



~~Same method as Fig. A4, but for soil temperature in depth 1.6 .~~

~~Sites' locations averaged JJA-climates (1985–2014) of hydrothermal variables as observed and simulated. Names on~~ The further away the ~~x-axis and the colors indicate the different~~ data ~~sources. Model names indicate CMIP6 model output (left), and LS3MIP (right; see Table 1). Each CMIP6-LS3MIP pair shares~~ is from the ~~same color. The boxes represent the medians~~diagonal line, ~~first and third quartiles;~~ the ~~± 1.5×IQR or~~ larger the ~~maximum and minimum values, if within the former range, are taken as the whiskers~~model'~~length~~s simulation bias is at certain temperature states. A higher vertical quantile range indicates more inconsistency with observation under identical temperature conditions.

~~Same method as Fig. A4, but for soil temperature in depth 1.6 .~~

~~Sites' locations averaged JJA-climates (1985–2014) of hydrothermal variables as observed and simulated. Names on~~ The further away the ~~x-axis and the colors indicate the different~~ data ~~sources. Model names indicate CMIP6 model output (left), and LS3MIP (right; see Table 1). Each CMIP6-LS3MIP pair shares~~ is from the ~~same color. The boxes represent the medians~~diagonal line, ~~first and third quartiles;~~ the ~~± 1.5×IQR or~~ larger the ~~maximum and minimum values, if within the former range, are taken as the whiskers~~model'~~length~~s simulation bias is at certain temperature states. A higher vertical quantile range indicates more inconsistency with observation under identical temperature conditions.

Figure A6. Quantile-Quantile (~~QQ~~Q-Q) and histogram plots for soil temperature in depth 0.8 m. The plots show 30-year mean monthly 0.8 m *tsl* data of models at all sites and their observational values. The model ~~simulation outputs~~ simulations are binned ~~at~~ into 2 °C intervals. Each bar (or pair of bars for CMIP6 and LS3MIP) is plotted at the center of its interval (e.g., 1 °C for 0 to 2 °C). The ~~colored~~ solid and ~~dash~~ dashed

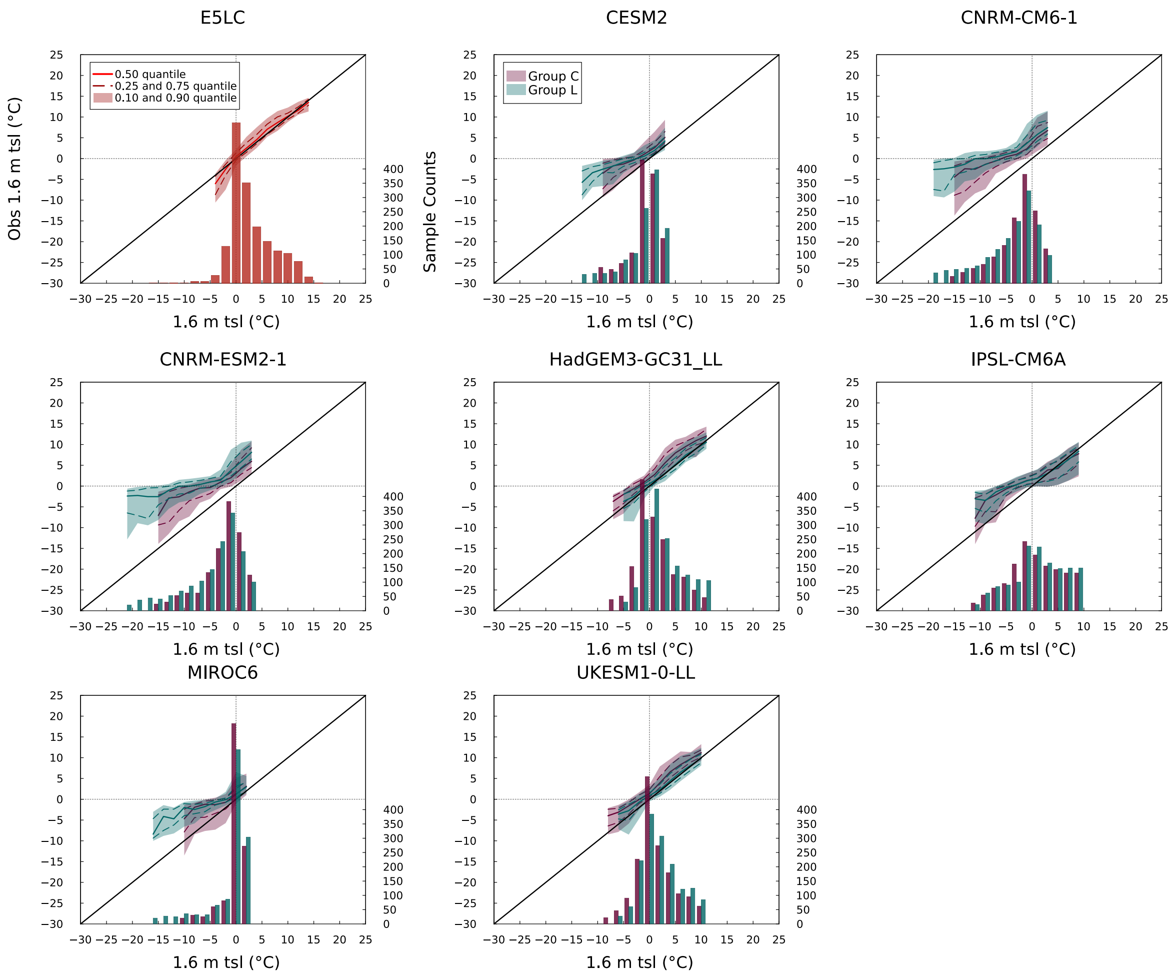


Figure A7. Same method as Fig. A, but for soil temperature at a depth of 1.6 m.

Table A1. Root-mean-square error (RMSE) between modelled and observed snow insulation effect (relationship between snow depth and soil-air temperature differences) across three tas categories, within the 0–0.8 m snow depth range (excluding fill-values) in °C, as illustrated in Fig. 10.

Model	-15 °C to -5 °C	-25 °C to -15 °C	less than -25 °C
E5LC	~~1.05~~ 1.1	~~1.52~~ 1.5	~~2.73~~ 2.7
CESM (CMIP6)	~~2.09~~ 2.1	~~3.97~~ 4.0	~~6.22~~ 6.2
CESM (LS3MIP)	**~~0.79~~0.8**	~~2.18~~ 2.2	~~4.09~~ 4.1
CNRM-CM (CMIP6)	~~4.83~~ 4.8	~~9.58~~ 9.6	~~12.39~~ 12.4
CNRM-CM (LS3MIP)	~~6.30~~ 6.3	~~11.47~~ 11.5	~~16.01~~ 16.0
CNRM-ESM (CMIP6)	~~4.75~~ 4.8	~~9.67~~ 9.7	~~12.15~~ 12.2
CNRM-ESM (LS3MIP)	~~6.20~~ 6.2	~~11.41~~ 11.4	~~15.75~~ 15.8
IPSL-CM (CMIP6)	~~2.26~~ 2.3	~~2.79~~ 2.8	~~4.34~~ 4.3
IPSL-CM (LS3MIP)	~~1.56~~ 1.6	~~3.25~~ 3.3	~~3.94~~ 3.9
HadGEM (CMIP6)	~~1.09~~ 1.1	**~~1.05~~1.0**	**~~2.33~~2.3**
HadGEM (LS3MIP)	~~1.26~~ 1.3	~~1.43~~ 1.4	**~~1.57~~1.6**
UKESM (CMIP6)	**~~0.71~~0.7**	**~~1.12~~1.1**	~~2.56~~ 2.6
UKESM (LS3MIP)	~~1.20~~ 1.2	**~~1.16~~1.2**	**~~2.15~~2.2**
MIROC (CMIP6)	~~1.11~~ 1.1	~~2.14~~ 2.1	~~5.41~~ 5.4
MIROC (LS3MIP)	**~~0.79~~0.8**	~~1.72~~ 1.7	~~4.32~~ 4.3

References

Abbott, B. W. and Jones, J. B.: Permafrost collapse alters soil carbon stocks, respiration, CH$_4$, and N_2 O in upland tundra, Global Change Biology, 21, 4570–4587, https://doi.org/10.1111/gcb.13069, 2015.

Andresen, C. G., Lawrence, D. M., Wilson, C. J., McGuire, A. D., Koven, C., Schaefer, K., Jafarov, E., Peng, S., Chen, X., Gouttevin, I., Burke, E., Chadburn, S., Ji, D., Chen, G., Hayes, D., and Zhang, W.: Soil moisture and hydrology projections of the permafrost region – a model intercomparison, The Cryosphere, 14, 445–459, https://doi.org/10.5194/tc-14-445-2020, 2020.

Beringer, J., Lynch, A. H., Chapin, F. S., Mack, M., and Bonan, G. B.: The Representation of Arctic Soils in the Land Surface Model: The Importance of Mosses, Journal of Climate, 14, 3324–3335, https://doi.org/10.1175/1520-0442(2001)014<3324:TROASI>2.0.CO;2, 2001.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A. B. K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., and Lantuit, H.: Permafrost is warming at a global scale, Nature Communications, 10, 264, https://doi.org/10.1038/s41467-018-08240-4, 2019.

Boike, J., Kattenstroth, B., Abramova, K., Bornemann, N., Chetverova, A., Fedorova, I., Fröb, K., Grigoriev, M., Grüber, M., Kutzbach, L., Langer, M., Minke, M., Muster, S., Piel, K., Pfeiffer, E.-M., Stoof, G., Westermann, S., Wischnewski, K., Wille, C., and Hubberten, H.-W.: Baseline characteristics of climate, permafrost and land cover from a new permafrost observatory in the Lena River Delta, Siberia (1998–2011), Biogeosciences, 10, 2105–2128, https://doi.org/10.5194/bg-10-2105-2013, 2013.

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., Meurdesoif, Y., Cadule, P., Devilliers, M., Ghattas, J., Lebas, N., Lurton, T., Mellul, L., Musat, I., Mignot, J., and Cheruy, F.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5195, 2018.

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., Meurdesoif, Y., Ghattas, J., Cadule, P., Ducharne, A., Vuichard, N., and Cheruy, F.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 LS3MIP land-hist, https://doi.org/10.22033/ESGF/CMIP6.5205, 2019.

Bowring, S. P. K., Lauerwald, R., Guenet, B., Zhu, D., Guimberteau, M., Tootchi, A., Ducharne, A., and Ciais, P.: ORCHIDEE MICT-LEAK (r5459), a global model for the production, transport, and transformation of dissolved organic carbon from Arctic permafrost regions – Part 1: Rationale, model description, and simulation protocol, Geoscientific Model Development, 12, 3503–3521, https://doi.org/10.5194/gmd-12-3503-2019, 2019.

Brown, J., Sidlauskas, F. J., and Delinski, G.: International Permafrost Association circum-Arctic map of permafrost and ground ice conditions, The Survey ; For sale by Information Services, Reston, Va., Denver, Colo., ISBN 978-0-607-88745-7, oCLC: 38148545, 1997.

Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., Leung, L. R., Niu, G.-Y., Troch, P. A., and Zeng, X.: Implementing and Evaluating Variable Soil Thickness in the Community Land Model, Version 4.5 (CLM4.5), Journal of Climate, 29, 3441–3461, https://doi.org/10.1175/JCLI-D-15-0307.1, 2016.

Bulygina, O., Razuvaev, V., and Aleksandrova, T.: Description of the dataset of snow characteristics at meteorological stations in Russia and the former USSR, http://aisori-m.meteo.ru/waisori/index.xhtml?idata=7, 2014a.

Bulygina, O., Razuvaev, V., and Alexandrova, T.: Description of the data array of daily air temperature and precipitation at meteorological stations in Russia and the former USSR (TTTR), http://aisori-m.meteo.ru/waisori/index.xhtml?idata=5, 2014b.

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.

Cai, L., Lee, H., Aas, K. S., and Westermann, S.: Projecting circum-Arctic excess-ground-ice melt with a sub-grid representation in the Community Land Model, The Cryosphere, 14, 4611–4626, https://doi.org/10.5194/tc-14-4611-2020, 2020.

Cai, Z., You, Q., Wu, F., Chen, H. W., Chen, D., and Cohen, J.: Arctic Warming Revealed by Multiple CMIP6 Models: Evaluation of Historical Simulations and Quantification of Future Projection Uncertainties, Journal of Climate, 34, 4871–4892, https://doi.org/10.1175/JCLI-D-20-0791.1, 2021.

Calonne, N., Flin, F., Morin, S., Lesaffre, B., Du Roscoat, S. R., and Geindreau, C.: Numerical and experimental investigations of the effective thermal conductivity of snow, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL049234, 2011.

Cao, B., Arduini, G., and Zsoter, E.: Brief communication: Improving ERA5-Land soil temperature in permafrost regions using an optimized multi-layer snow scheme, The Cryosphere, 16, 2701–2708, https://doi.org/10.5194/tc-16-2701-2022, 2022.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, Geoscientific Model Development, 8, 1493–1508, https://doi.org/10.5194/gmd-8-1493-2015, 2015.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701–722, https://doi.org/10.5194/gmd-4-701-2011, 2011.

Copernicus Climate Change Service: ERA5-Land monthly averaged data from 1950 to present, https://doi.org/10.24381/CDS.68D2BB30, 2019.

Cuntz, M. and Haverd, V.: Physically Accurate Soil Freeze-Thaw Processes in a Global Land Surface Scheme, Journal of Advances in Modeling Earth Systems, 10, 54–77, https://doi.org/10.1002/2017MS001100, 2018.

Damseaux, A., Matthes, H., Dutch, V. R., Wake, L., and Rutter, N.: Impact of snow thermal conductivity schemes on pan-Arctic permafrost dynamics in the Community Land Model version 5.0, The Cryosphere, 19, 1539–1558, https://doi.org/10.5194/tc-19-1539-2025, 2025.

Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 LS3MIP land-hist, https://doi.org/10.22033/ESGF/CMIP6.7650, 2019a.

Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.7627, 2019b.

de Vrese, P., Georgievski, G., Gonzalez Rouco, J. F., Notz, D., Stacke, T., Steinert, N. J., Wilkenskjeld, S., and Brovkin, V.: Representation of soil hydrology in permafrost regions may explain large part of inter-model spread in simulated Arctic and subarctic climate, The Cryosphere, 17, 2095–2118, https://doi.org/10.5194/tc-17-2095-2023, 2023.

Decharme, B., Brun, E., Boone, A., Christine Delire, Le Moigne, P., and Morin, S.: Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model, The Cryosphere, 10, 853–877, https://doi.org/10.5194/tc-10-853-2016, 2016.

Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S., and Voldoire, A.: Recent Changes in the ISBA-CTRIP Land Surface System for Use in the CNRM-CM6 Climate Model and in Global Off-Line Hydrological Applications, Journal of Advances in Modeling Earth Systems, 11, 1207–1252, https://doi.org/10.1029/2018MS001545, 2019.

Deng, M., Meng, X., Lu, Y., Li, Z., Zhao, L., Hu, Z., Chen, H., Shang, L., Wang, S., and Li, Q.: Impact and Sensitivity Analysis of Soil Water and Heat Transfer Parameterizations in Community Land Surface Model on the Tibetan Plateau, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 670, https://doi.org/10.1029/2021MS002670, 2021.

Du, R., Peng, X., Frauenfeld, O. W., Jin, H., Wang, K., Zhao, Y., Luo, D., and Mu, C.: Quantitative Impact of Organic Matter and Soil Moisture on Permafrost, Journal of Geophysical Research: Atmospheres, 128, e2022JD037 686, https://doi.org/10.1029/2022JD037686, 2023.

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

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.

Ekici, A., Chadburn, S., Chaudhary, N., Hajdu, L. H., Marmy, A., Peng, S., Boike, J., Burke, E., Friend, A. D., Hauck, C., Krinner, G., Langer, M., Miller, P. A., and Beer, C.: Site-level model intercomparison of high latitude and high altitude soil thermal dynamics in tundra and barren landscapes, The Cryosphere, 9, 1343–1361, https://doi.org/10.5194/tc-9-1343-2015, 2015.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Frauenfeld, O. W. and Zhang, T.: An observational 71-year history of seasonally frozen ground changes in the Eurasian high latitudes, Environmental Research Letters, 6, 044 024, https://doi.org/10.1088/1748-9326/6/4/044024, 2011.

Fuchs, M., Grosse, G., Strauss, J., Günther, F., Grigoriev, M., Maximov, G. M., and Hugelius, G.: Carbon and nitrogen pools in thermokarst-affected permafrost landscapes in Arctic Siberia, Biogeosciences, 15, 953–971, https://doi.org/10.5194/bg-15-953-2018, 2018.

Groenke, B., Langer, M., Nitzbon, J., Westermann, S., Gallego, G., and Boike, J.: Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer, The Cryosphere, 17, 3505–3533, https://doi.org/10.5194/tc-17-3505-2023, 2023.

Groisman, P. Y. and Bartalev, S. A.: Northern Eurasia earth science partnership initiative (NEESPI), science plan overview, Global and Planetary Change, 56, 215–234, https://doi.org/10.1016/j.gloplacha.2006.07.027, 2007.

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., Ottlé, C., Jornet-Puig, A., Bastos, A., Laurent, P., Goll, D., Bowring, S., Chang, J., Guenet, B., Tifafi, M., Peng, S., Krinner, G., Ducharne, A., Wang, F., Wang, T., Wang, X., Wang, Y., Yin, Z., Lauerwald, R., Joetzjer, E., Qiu, C., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, Geoscientific Model Development, 11, 121–163, https://doi.org/10.5194/gmd-11-121-2018, 2018.

Guo, Q., Kino, K., Li, S., Nitta, T., Takeshima, A., Nitta, T., Onuma, Y., Satoh, Y., Suzuki, T., Takata, K., Yoshida, N., and Yoshimura, K.: Description of MATSIRO6, https://doi.org/10.15083/0002000181, 2021.

Heijmans, M. M. P. D., Magnússon, R. , Lara, M. J., Frost, G. V., Myers-Smith, I. H., Van Huissteden, J., Jorgenson, M. T., Fedorov, A. N., Epstein, H. E., Lawrence, D. M., and Limpens, J.: Tundra vegetation change and impacts on permafrost, Nature Reviews Earth & Environment, 3, 68–84, https://doi.org/10.1038/s43017-021-00233-0, 2022.

Jafarov, E. and Schaefer, K.: The importance of a surface organic layer in simulating permafrost thermal and carbon dynamics, The Cryosphere, 10, 465–475, https://doi.org/10.5194/tc-10-465-2016, 2016.

Jafarov, E. E., Harp, D. R., Coon, E. T., Dafflon, B., Tran, A. P., Atchley, A. L., Lin, Y., and Wilson, C. J.: Estimation of subsurface porosities and thermal conductivities of polygonal tundra by coupled inversion of electrical resistivity, temperature, and moisture content data, The Cryosphere, 14, 77–91, https://doi.org/10.5194/tc-14-77-2020, 2020.

Jain, S., Scaife, A. A., Shepherd, T. G., Deser, C., Dunstone, N., Schmidt, G. A., Trenberth, K. E., and Turkington, T.: Importance of internal variability for climate model assessment, npj Climate and Atmospheric Science, 6, 68, https://doi.org/10.1038/s41612-023-00389-0, 2023.

Jordan, R. E.: A One-dimensional temperature model for a snow cover : technical documentation for SNTHERM.89, https://hdl.handle.net/11681/11677, 1991.

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.

Koven, C. D., Lawrence, D. M., and Riley, W. J.: Permafrost carbon-climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics, Proceedings of the National Academy of Sciences, 112, 3752–3757, https://doi.org/10.1073/pnas.1415123112, 2015.

Kudryavtsev, V.A.: Fundamentals of frost forecasting in geological engineering investigations, vol. 606, Cold Regions Research and Engineering Laboratory, 1977.

Kuma, P., Bender, F. A., and Jönsson, A. R.: Climate Model Code Genealogy and Its Relation to Climate Feedbacks and Sensitivity, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 588, https://doi.org/10.1029/2022MS003588, 2023.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia – Part 2: Winter, The Cryosphere, 5, 509–524, https://doi.org/10.5194/tc-5-509-2011, 2011a.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The surface energy balance of a polygonal tundra site in northern Siberia – Part 1: Spring to fall, The Cryosphere, 5, 151–171, https://doi.org/10.5194/tc-5-151-2011, 2011b.

Lawrence, D. M. and Slater, A. G.: A projection of severe near-surface permafrost degradation during the 21st century, Geophysical Research Letters, 32, L24 401, https://doi.org/10.1029/2005GL025080, 2005.

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.

Li, Q., Sun, S., and Xue, Y.: Analyses and development of a hierarchy of frozen soil models for cold region study, Journal of Geophysical Research: Atmospheres, 115, 2009JD012 530, https://doi.org/10.1029/2009JD012530, 2010.

Li, X., Wu, T., Wu, X., Chen, J., Zhu, X., Hu, G., Li, R., Qiao, Y., Yang, C., Hao, J., Ni, J., and Ma, W.: Assessing the simulated soil hydrothermal regime of the active layer from the Noah-MP land surface model (v1.1) in the permafrost regions of the Qinghai–Tibet Plateau, Geoscientific Model Development, 14, 1753–1771, https://doi.org/10.5194/gmd-14-1753-2021, 2021.

Li, Y., Zhao, M., Motesharrei, S., Mu, Q., Kalnay, E., and Li, S.: Local cooling and warming effects of forests based on satellite observations, Nature Communications, 6, 6603, https://doi.org/10.1038/ncomms7603, 2015.

Liljedahl, A. K., Boike, J., Daanen, R. P., Fedorov, A. N., Frost, G. V., Grosse, G., Hinzman, L. D., Iijma, Y., Jorgenson, J. C., Matveyeva, N., Necsoiu, M., Raynolds, M. K., Romanovsky, V. E., Schulla, J., Tape, K. D., Walker, D. A., Wilson, C. J., Yabuki, H., and Zona, D.: Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology, Nature Geoscience, https://doi.org/10.1038/ngeo2674, 2016.

Matthes, H., Rinke, A., Zhou, X., and Dethloff, K.: Uncertainties in coupled regional Arctic climate simulations associated with the used land surface model, Journal of Geophysical Research: Atmospheres, 122, 7755–7771, https://doi.org/10.1002/2016JD026213, 2017.

Matthes, H., Damseaux, A., Westermann, S., Beer, C., Boone, A., Burke, E., Decharme, B., Genet, H., Jafarov, E., Langer, M., Parmentier, F., Porada, P., Gagne-Landmann, A., Huntzinger, D., Rogers, B., Schädel, C., Stacke, T., Wells, J., and Wieder, W.: Advances in Permafrost Representation: Biophysical Processes in Earth System Models and the Role of Offline Models, Permafrost and Periglacial Processes, 36, 302–318, https://doi.org/10.1002/ppp.2269, 2025.

McNeall, D., Robertson, E., and Wiltshire, A.: Constraining the carbon cycle in JULES-ES-1.0, Geoscientific Model Development, 17, 1059–1089, https://doi.org/10.5194/gmd-17-1059-2024, 2024.

Menard, C. B., Essery, R., Krinner, G., Arduini, G., Bartlett, P., Boone, A., Brutel-Vuilmet, C., Burke, E., Cuntz, M., Dai, Y., Decharme, B., Dutra, E., Fang, X., Fierz, C., Gusev, Y., Hagemann, S., Haverd, V., Kim, H., Lafaysse, M., Marke, T., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Schädler, G., Semenov, V. A., Smirnova, T., Strasser, U., Swenson, S., Turkov, D., Wever, N., and Yuan, H.: Scientific and Human Errors in a Snow Model Intercomparison, Bulletin of the American Meteorological Society, 102, E61–E79, https://doi.org/10.1175/BAMS-D-19-0329.1, 2021.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Nitzbon, J., Westermann, S., Langer, M., Martin, L. C. P., Strauss, J., Laboor, S., and Boike, J.: Fast response of cold ice-rich permafrost in northeast Siberia to a warming climate, Nature Communications, 11, 2201, https://doi.org/10.1038/s41467-020-15725-8, 2020.

Niu, G.-Y. and Yang, Z.-L.: Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale, Journal of Hydrometeorology, 7, 937–952, https://doi.org/10.1175/JHM538.1, 2006.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J., and Zou, D.: Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale, Earth-Science Reviews, 193, 299–316, https://doi.org/10.1016/j.earscirev.2019.04.023, 2019.

Onuma, Y. and Kim, H.: MIROC MIROC6 model output prepared for CMIP6 LS3MIP land-hist, https://doi.org/10.22033/ESGF/CMIP6.5622, 2020.

Park, H., Sherstiukov, A. B., Fedorov, A. N., Polyakov, I. V., and Walsh, J. E.: An observation-based assessment of the influences of air temperature and snow depth on soil temperature in Russia, Environmental Research Letters, 9, 064 026, https://doi.org/10.1088/1748-9326/9/6/064026, 2014.

Park, H., Fedorov, A. N., Zheleznyak, M. N., Konstantinov, P. Y., and Walsh, J. E.: Effect of snow cover on pan-Arctic permafrost thermal regimes, Climate Dynamics, 44, 2873–2895, https://doi.org/10.1007/s00382-014-2356-5, 2015.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Communications Earth & Environment, 3, 168, https://doi.org/10.1038/s43247-022-00498-3, 2022.

Rashid, H. A.: Diverse Responses of Global-Mean Surface Temperature to External Forcings and Internal Climate Variability in Observations and CMIP6 Models, Geophysical Research Letters, 48, e2021GL093 194, https://doi.org/10.1029/2021GL093194, 2021.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6109, 2019.

Risto, D., Fröhlich, K., and Ahrens, B.: Snow Representation over Siberia in Operational Seasonal Forecasting Systems, Atmosphere, 13, 1002, https://doi.org/10.3390/atmos13071002, 2022.

Romanovsky, V., Sazonova, T., Balobaev, V., Shender, N., and Sergueev, D.: Past and recent changes in air and permafrost temperatures in eastern Siberia, Global and Planetary Change, 56, 399–413, https://doi.org/10.1016/j.gloplacha.2006.07.022, 2007.

Rößger, N., Sachs, T., Wille, C., Boike, J., and Kutzbach, L.: Seasonal increase of methane emissions linked to warming in Siberian tundra, Nature Climate Change, 12, 1031–1036, https://doi.org/10.1038/s41558-022-01512-4, 2022.

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E.: Climate change and the permafrost carbon feedback, Nature, 520, 171–179, https://doi.org/10.1038/nature14338, 2015.

Schwarzwald, K. and Lenssen, N.: The importance of internal climate variability in climate impact projections, Proceedings of the National Academy of Sciences, 119, e2208095 119, https://doi.org/10.1073/pnas.2208095119, 2022.

Schädel, C., Koven, C. D., Lawrence, D. M., Celis, G., Garnello, A. J., Hutchings, J., Mauritz, M., Natali, S. M., Pegoraro, E., Rodenhizer, H., Salmon, V. G., Taylor, M. A., Webb, E. E., Wieder, W. R., and Schuur, E. A.: Divergent patterns of experimental and model-derived permafrost ecosystem carbon dynamics in response to Arctic warming, Environmental Research Letters, 13, 105 002, https://doi.org/10.1088/1748-9326/aae0ff, 2018.

Seferian, R.: CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.4068, 2018.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., De Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, https://doi.org/10.1029/2019MS001739, 2019.

Sherstiukov, A.: Dataset of daily soil temperature up to 320 cm depth based on meteorological stations of Russian Federation, RIHMI-WDC, 176, 224–232, 2012a.

Sherstiukov, A.: Statistical quality control of a daily soil temperature dataset, 2012b.

Smith, S. L., O'Neill, H. B., Isaksen, K., Noetzli, J., and Romanovsky, V. E.: The changing thermal state of permafrost, Nature Reviews Earth & Environment, 3, 10–23, https://doi.org/10.1038/s43017-021-00240-1, 2022.

Streletskiy, D. A., Maslakov, A., Grosse, G., Shiklomanov, N. I., Farquharson, L., Zwieback, S., Iwahana, G., Bartsch, A., Liu, L., Strozzi, T., Lee, H., and Debolskiy, M. V.: Thawing permafrost is subsiding in the Northern Hemisphere—review and perspectives, Environmental Research Letters, 20, 013 006, https://doi.org/10.1088/1748-9326/ada2ff, 2025.

Sturm, M., Holmgren, J., König, M., and Morris, K.: The thermal conductivity of seasonal snow, Journal of Glaciology, 43, 26–41, https://doi.org/10.3189/S0022143000002781, 1997.

Swenson, S. C., Lawrence, D. M., and Lee, H.: Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model, Journal of Advances in Modeling Earth Systems, 4, 2012MS000 165, https://doi.org/10.1029/2012MS000165, 2012.

Takata, K., Emori, S., and Watanabe, T.: Development of the minimal advanced treatments of surface interaction and runoff, Global and Planetary Change, 38, 209–222, https://doi.org/10.1016/S0921-8181(03)00030-4, 2003.

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., Walton, J., and Jones, C.: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6113, 2019.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Global Biogeochemical Cycles, 23, 2008GB003 327, https://doi.org/10.1029/2008GB003327, 2009.

Tatebe, H. and Watanabe, M.: MIROC MIROC6 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5603, 2018.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Turetsky, M. R., Abbott, B. W., Jones, M. C., Walter Anthony, K., Olefeldt, D., Schuur, E. A. G., Koven, C., McGuire, A. D., Grosse, G., Kuhry, P., Hugelius, G., Lawrence, D. M., Gibson, C., and Sannel, A. B. K.: Permafrost collapse is accelerating carbon release, Nature, 569, 32–34, https://doi.org/10.1038/d41586-019-01313-4, 2019.

Van Den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soilmoisture Model Intercomparison Project – aims, setup and expected outcome, Geoscientific Model Development, 9, 2809–2832, https://doi.org/10.5194/gmd-9-2809-2016, 2016.

Van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G., and Van Den Broeke, M. R.: Improving the Representation of Polar Snow and Firn in the Community Earth System Model, Journal of Advances in Modeling Earth Systems, 9, 2583–2600, https://doi.org/10.1002/2017MS000988, 2017.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geoscientific Model Development, 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, 2012.

Voldoire, A.: CMIP6 simulations of the CNRM-CERFACS based on CNRM-CM6-1 model for CMIP experiment historical, https://doi.org/10.22033/ESGF/CMIP6.4066, 2018.

Voldoire, A.: CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 LS3MIP land-hist, https://doi.org/10.22033/ESGF/CMIP6.9599, 2019a.

Voldoire, A.: CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 LS3MIP land-hist, https://doi.org/10.22033/ESGF/CMIP6.4095, 2019b.

Walter Anthony, K., Schneider Von Deimling, T., Nitze, I., Frolking, S., Emond, A., Daanen, R., Anthony, P., Lindgren, P., Jones, B., and Grosse, G.: 21st-century modeled permafrost carbon emissions accelerated by abrupt thaw beneath lakes, Nature Communications, 9, 3262, https://doi.org/10.1038/s41467-018-05738-9, 2018.

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geoscientific Model Development, 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

Wang, T., Ottlé, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S., and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, Journal of Geophysical Research: Atmospheres, 118, 6064–6079, https://doi.org/10.1002/jgrd.50395, 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 surfacemodels during winter across the permafrost region, The Cryosphere, 10, 1721–1737, https://doi.org/10.5194/tc-10-1721-2016, 2016.

Wang, X., Ran, Y., Pang, G., Chen, D., Su, B., Chen, R., Li, X., Chen, H. W., Yang, M., Gou, X., Jorgenson, M. T., Aalto, J., Li, R., Peng, X., Wu, T., Clow, G. D., Wan, G., Wu, X., and Luo, D.: Contrasting characteristics, changes, and linkages of permafrost between the Arctic and the Third Pole, Earth-Science Reviews, 230, 104 042, https://doi.org/10.1016/j.earscirev.2022.104042, 2022.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Elsevier, Amsterdam, Netherlands ; Cambridge, MA, fourth edition edn., ISBN 978-0-12-815823-4, 2019.

Wiltshire, A., Robertson, E., Burke, E., and Liddicoat, S.: MOHC UKESM1.0-LL model output prepared for CMIP6 LS3MIP, https://doi.org/10.22033/ESGF/CMIP6.14462, 2020a.

Wiltshire, A., Robertson, E., Burke, E., and Liddicoat, S.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 LS3MIP, https://doi.org/10.22033/ESGF/CMIP6.14460, 2020b.

Wiltshire, A. J., Duran Rojas, M. C., Edwards, J. M., Gedney, N., Harper, A. B., Hartley, A. J., Hendry, M. A., Robertson, E., and Smout-Day, K.: JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2, Geoscientific Model Development, 13, 483–505, https://doi.org/10.5194/gmd-13-483-2020, 2020c.

Woo, M.-k.: Permafrost Hydrology, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 9783642234613 9783642234620, https://doi.org/10.1007/978-3-642-23462-0, 2012.

Yang, S., Li, R., Zhao, L., Wu, T., Wu, X., Zhang, Y., Shi, J., and Qiao, Y.: Evaluation of the Performance of CLM5.0 in Soil Hydrothermal Dynamics in Permafrost Regions on the Qinghai–Tibet Plateau, Remote Sensing, 14, 6228, https://doi.org/10.3390/rs14246228, 2022.

Ye, K.: Inter-model spread in the wintertime Arctic amplification in the CMIP6 models and the important role of internal climate variability, Global and Planetary Change, 2021.

Yen, Y.-C.: Review of thermal properties of snow, ice, and sea ice, vol. 81, US Army, Corps of Engineers, Cold Regions Research and Engineering Laboratory, 1981.

Yokohata, T., Saito, K., Takata, K., Nitta, T., Satoh, Y., Hajima, T., Sueyoshi, T., and Iwahana, G.: Model improvement and future projection of permafrost processes in a global land surface model, Progress in Earth and Planetary Science, 7, 69, https://doi.org/10.1186/s40645-020-00380-w, 2020.

You, Q., Cai, Z., Pepin, N., Chen, D., Ahrens, B., Jiang, Z., Wu, F., Kang, S., Zhang, R., Wu, T., Wang, P., Li, M., Zuo, Z., Gao, Y., Zhai, P., and Zhang, Y.: Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences, Earth-Science Reviews, 217, 103 625, https://doi.org/10.1016/j.earscirev.2021.103625, 2021.

Zhang, T.: Influence of the seasonal snow cover on the ground thermal regime: An overview, Reviews of Geophysics, 43, 2004RG000 157, https://doi.org/10.1029/2004RG000157, 2005.

Zhang, Y., Sherstiukov, A. B., Qian, B., Kokelj, S. V., and Lantz, T. C.: Impacts of snow on soil temperature observed across the circumpolar north, Environmental Research Letters, 13, 044 012, https://doi.org/10.1088/1748-9326/aab1e7, 2018.

Zhu, D., Ciais, P., Krinner, G., Maignan, F., Jornet Puig, A., and Hugelius, G.: Controls of soil organic matter on soil thermal dynamics in the northern high latitudes, Nature Communications, 10, 3172, https://doi.org/10.1038/s41467-019-11103-1, 2019.

Åkerman, H. J. and Johansson, M.: Thawing permafrost and thicker active layers in sub-arctic Sweden, Permafrost and Periglacial Processes, 19, 279–292, https://doi.org/10.1002/ppp.626, 2008.