Assessing ~~Climate Modeling~~ Uncertainties in Modeling the Climate of the Siberian Frozen ~~Soil Regions~~ 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 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 ~~allows~~ allowed quantifying the contributions of the land surface parameterization scheme and the atmospheric forcing to the discrepancies. The simulation capabilities ~~are~~ were assessed using observational data from 152 sites in Siberia and reanalysis data. ~~On average, the 0.2-m soil temperatures in the CMIP6 simulations are 5.4 °C colder than the observations if the simulated soil temperature drops below -5 °C. The LS3MIP simulations are even colder with a bias of -6.7 °C.~~ In the winter months (December, January, and February), the LS3MIP ensemble ~~diversity in 2-m temperature is less~~ bias in 0.2 m soil temperature was higher than the CMIP6 ~~diversity (0.8 °C vs 3.2 °~~bias (-3.57 °C vs -2.66 °C). ~~In contrast, the diversity~~ The spread of winter 0.2~~-m soil temperatures is~~ m soil temperatures was also larger in the LS3MIP ensemble (~~5.7 °~~4.55 °C) than in the CMIP6 ensemble (~~3.6 °~~2.98 °C). For permafrost sites, the spatial correlation of the ~~simulations of~~ winter soil temperature ~~against observation is~~ simulations against observations was not better than ~~0.7, and~~ 0.6, and the spring/autumn spatial correlations of snow depth ~~are less than 0.75 for the~~ were less than 0.8 for all CMIP6 models. On average, the 0.2 m soil temperatures in the CMIP6 simulations were 0.34 °C warmer than the observations when the simulated soil temperature dropped below -5 °C. However, the LS3MIP simulations were colder, with a bias of -0.65 °C. The biases of ~~2-m temperature have~~ 2 m temperature had a different sign and ~~are~~ were amplified in magnitude compared to the biases of the soil temperatures, especially below 0~~°C.~~ °C. Four of the climate models and their land components ~~underestimate~~ underestimated the snow insulation effect. We ~~conclude that land surface models struggle to well simulate~~ concluded that land-only models have difficulties in accurately simulating soil temperatures and snow depth under low-temperature conditions. The CMIP6 models ~~tend~~ tended to compensate for errors in their land component ~~by errors through errors~~ with errors in the atmospheric model component. In shallow snow depth (0 to 0.2~~m~~ m) cases, all models ~~show~~ showed between 1 ~~to~~ and 8~~°C~~ °C less air-soil temperature difference than in situ data. Therefore, a better representation of surface-soil insulation is essential for improvements in frozen soil land modeling.

1 Introduction

~~The recent rise in~~ 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 ~~air temperature is almost twice the global average in the Arctic~~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 ~~show~~ indicate that the Arctic ~~could warm by as much as 7.5°C in 1900-2100 (Lawrence and Slater, 2005; Lawrence et al., 2008)~~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 ~~greatest~~ most significant changes occur in regions where the mean annual air temperatures are around 0~~°C~~ °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, which~~ °C during 2007–2016. This warming is two to three times ~~the warming~~ larger than that observed at discontinuous permafrost sites (Biskaborn et al., 2019; Smith et al., 2022), ~~warning~~ raising concerns about the potential ~~of~~ for more substantial permafrost degradation in the future.

Large amounts of soil carbon are ~~emitted~~ 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 ~~during the degradation of permafrost. Permafrost's abrupt thawing results in 1.5 times more carbon emissions than normal degradation (Turetsky et al., 2019; Miner et al., 2022). The~~ 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 ~~environment~~ conditions, which further increases the ~~share of methane in such carbon emissions (Koven et al., 2015)~~proportion of methane emitted alongside other greenhouse gases (Koven et al., 2015; Walter A. 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). ~~Changes in hydrothermal conditionsthat occur during permafrostthawing can lead to changesin surface vegetation types based on different soil-frozen water contents (Heijmans et al., 2022). The model parameters used to calculate~~ 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 uncertain components in Earth system models. 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 ~~controlled~~ governed by soil texture, surface organic matter, ~~frequently varying soil water content~~moisture dynamics, and freeze-thaw conditions ~~, which~~ (Woo, 2012; Andresen et al., 2020; Yang et al., 2022). In permafrost regions, these factors determine the thermal offset ~~or soil insulation effect. There are differences in the time scales of major physical processes between the soil and the atmosphere, and the soil surface is the window through which the atmosphere interacts with the soil (Beringer et al., 2001; Langer et al., 2011a, b). The~~ (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 ~~manners~~ ways by different models (Niu and Yang, 2006; Li et al., 2010). ~~Snow cover insulates the soil and affects the surface energy balance by changing local albedo and other conditions~~ 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. ~~Its effect~~ The impact of snow on soil temperature ~~is spatially heterogeneous, depending on its characteristics, including~~ exhibits spatial heterogeneity based on snow's own attributes—specifically, thickness, density, and duration (Zhang, 2005; Zhang et al., 2018). ~~In northeastern Siberia, the change~~ 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 ~~accounts for more than~~ (snow depth, density, and duration) account for over 50~~% of soil temperature changes (Park et al., 2014, 2015).~~ % 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 ~~is currently the most suitable dataset for~~

~~evaluating the~~ allows evaluation of the ability of the latest generation of climate models to simulate frozen soil ~~, as it provides~~ by providing an ensemble of climate models at resolutions fine enough ~~for distinguishing~~ to distinguish different frozen soil regions. The ~~models have been further developed relative to CMIP5 to varying degrees, including~~ 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; . The Land Surface, Snow, and Soil Moisture Model Intercomparison Project (LS3MIP~~provides the opportunity to exclude~~) 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 ~~variability~~ variabilities produced by different atmospheric models ~~(Van Den Hurk et al., 2016)~~in corresponding CMIP6 models.

In CMIP6 and LS3MIP, the ~~set-up~~ setup of the land cover/land use scenario and the radiative forcing conditions ~~are based on~~ 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 ~~constitute~~ constitutes a significant portion of the Eurasian continent's frozen terrain. 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 ~~observation~~ observational dataset provided by the All-Russian Scientific Research Institute of Hydrometeorological Information-World Data Center (RIHMI-WDC) (Frauenfeld and Zhang, 2011; Sherstiukov, 2012a; Zhang et al., 2018) can be ~~applied, which~~ used. This dataset provides consistent soil temperature measurements at standardized depths and ~~thus can~~ can thus be used as a reference in climate model evaluation.

The characteristics of frozen soil surface dynamics are assessed by comparing model outputs with ~~benchmarks~~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 ~~same model~~ 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 surface~~ land-only models versus those resulting from atmospheric ~~forcings. With identical and more realistic atmospheric conditions, we anticipate that~~ 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, observation-based atmospheric conditions, the LS3MIP models ~~will more accurately simulate soil conditions. If these models~~

~~fail to produce soil variable outputs that align better with observed data than the~~ are expected to simulate soil temperature 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 ~~, it is regarded as an error in the land surface models. We will discuss the variations among different models of~~ 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, we will explore inter-model variations within LS3MIP and ~~try to establish a connection between model performance and their specific features~~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 ~~uses~~ 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 surface~~ , 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 ~~. This enables~~ (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 ~~and assess the impact of errors due to coupling versus deficiencies in the land model on the simulation results in the frozen soil region~~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 ~~/earth system models~~ involved in both projects, incorporating six different land models. The selected models are listed in Table~~1.~~ 1. Other climate models, which also participated in both projects, ~~can not be considered as they~~ 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 ~~runs from CMIP6 historical simulations~~ as Group C and the LS3MIP ~~runs from LS3MIP~~ as Group L in plots and analysis. Four variables are collected for this study: ~~2-meter~~ 2 m air temperature (tas), soil temperature in 0.2~~-meter~~ m depth (tsl), snow depth (snd), and precipitation (pr).

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

Model Name	**Land Surface Model**	**Total Soil La** (max. node depth
CESM2	CLM5.0	25 (42.0)
CNRM-CM6.1	Surfex 8.0c	14 (10.0)
CNRM-ESM2.1	Surfex 8.0c	14 (10.0)
~~HadGEM3-GC31-LL JULES-HadGEM3-GL7.14 (2.0)431.25×1.875~~IPSL-CM6A-LR	ORCHIDEE v2.0	18 (65.56)
~~MIROC6~~ HadGEM3-GC31-LL	~~MATSIRO6.0~~ JULES-HadGEM3-GL7.1	~~6 (9.0~~4 (2.0
UKESM1.0-LL	JULES-ES-1.0	4 (2.0)
MIROC6	MATSIRO6.0	6 (9.0)

160 Both CMIP6 and LS3MIP data can be accessed at ~~https://aims2. llnl.gov/search/cmip6/.~~ https://aims2.llnl.gov/search/cmip6.

Table 2. Features of the land surface models. ~~SA indicates spectrally averaged, and~~ MC indicates mechanical compaction.~~Surface organic matter indicates how models consider the function of surface organic matter.~~

Land Model	**~~Albedo~~ Snow**	**Snow**	**Bottom**
	~~Conductivity~~ Density	**Thermal Conductivity**	**Boundary Condition**
CLM5.0	~~Spectral Density-Dependent~~MC	~~Zero Flux~~Density-dependent	~~Hydro-thermodynamic~~Zero-fl
		Jordan (1991)	
Surfex 8.0c	~~Spectral~~ MC	~~Power Function~~ Density-dependent	Fixed temperature
	~~Fixed~~	~~Hydro-thermodynamic~~Yen (1981)	~~Vionnet et al. (2012)~~
ORCHIDEE v2.0	MC	Depends on density,	Fixed flux
		~~Decharme et al. (2019)~~temperature, and pressure	
JULES-HadGEM3-GL7.1	~~Spectral Power Function~~ MC	Density-dependent	Zero-flux
		Calonne et al. (2011)	
~~ORCHIDEE v2~~JULES-ES-1.0	~~SA Quadratic Equation~~MC	~~Fixed Flux~~Density-dependent	~~Water/Carbon Cycle~~Zero-flu
		Calonne et al. (2011)	
MATSIRO6.0	~~Spectral~~ Fixed	Fixed	Fixed temperature
~~JULES-ES-1.0~~	~~Spectral~~ (300 $\mathrm{kg\,m^{-3}}$)	~~Power Function~~ (0.3 $\mathrm{W\,m^{-1}\,K^{-1}}$)	~~MC~~

~~Table 2~~

Table 2 highlights several attributes of each land model, focusing on their key characteristics related to the surface energy balance and associated processes~~(Menard et al., 2021). These~~. 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, ~~and soil-vegetation-atmosphere interactions. Albedo is treated either spectrally, accounting for wavelength-dependent variations, or as a time-dependent parameter. Snow conductivity~~ boundary conditions, and surface organic matter. The thermal conductivity of snow is modeled using density-dependent formulations ~~, empirical power functions, or~~ as a power function (Yen, 1981) or a quadratic function (Jordan, 1991; Calonne et al., 2011; Wang et al., 201 , or using fixed values. ~~Similarly, snow~~ 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 ~~values~~temperature, or fixed temperature gradients. ~~Models incorporate surface organic matter to varying degrees of complexity. Some focus on its hydrological and thermal impacts, while others emphasize its role in modifying surface albedo and roughness or as a component of the carbon cycle. For instance, models with a "Carbon Cycle Focus" simulate organic matter decomposition and carbon fluxes, whereas those with a "Hydrological Focus" prioritize water retention and flow.~~

2.2 ERA5-Land Reanalysis

We utilized monthly averaged ERA5-Land reanalysis data from the European Centre for Medium-Range Weather Forecasts~~, as it can assist our assessment in determining the impact of comparing gridded data to site data~~. ERA5-Land is a numerical land surface model product forced by atmospheric variables of ERA5, featuring a high horizontal resolution of 0.1~~° ×~~° ×0.1~~° (Muñoz-Sabater et al., 2021)~~°(Muñoz-Sabater et al., 2021; Copernicus Climate Change Service, 2019). Monthly data provided by ERA5-Land include ~~2-meter~~ 2 m temperature, snow depth, and soil temperature ~~and moisture~~ at four depths (0 to 0.07~~m~~ m, 0.07 to 0.28~~m~~ m, 0.28 to 1.0~~m~~ m, 1.0 to 2.89~~m)are provided by~~ m). Soil temperature was linearly interpolated to a depth of 0.2 m to directly compare with observations. We also remapped ERA5-Land ~~. The dataset can be obtained from https://cds. climate. copernicus. eu/.~~ 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 ~~are~~ were gathered from 236 meteorology sites by RIHMI-WDC (Sherstiukov, 2012a; Bulygina et al., 2014a, b; Zh . 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~~° E, east of 120~~°E~~° E, and south of 45~~°N are~~ ° N were excluded to eliminate most stations with warmer climates. ~~These criteria are~~ 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 ~~have been selected. The data is collected from http://meteo.ru/.~~ were selected.

2.4 Data Preprocessing

We ~~interpolate~~ interpolated tas, tsl, pr, and snd for all model simulations at the station sites' coordinates by choosing the nearest-neighbor grid cell. This ~~introduces~~ introduced additional 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, we ~~consider~~ considered only 30-year averaged data for evaluations.

2.5 Evaluation Metrics

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

The variability of target variables ~~is~~ was quantified using the ~~Inter-Quartile Range~~ interquartile range (IQR), which measures the spread of the simulations. If the model over- or ~~underestimates~~ underestimated the observed IQR_o, the model over- or

~~underestimates~~ underestimated climate variability, respectively. Thus, we ~~use~~ used the relative spread

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

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

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

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

as a measure ~~for~~ 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 ~~in the central tendency of the data . It is~~ using the median data values. It was standardized using observed variability (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 ~~model~~ 2 multi-model ensembles at the sites' locations, we ~~define~~ defined the ensemble means of the models' 30-year median biases as follows

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

~~and the ensemble-spreads of~~ 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 ~~are calculated as~~ were defined as the standard deviations.

The ensemble mean biases $EB_{i,s}$ and spreads ~~are~~ were calculated for both the CMIP6 (Group C) and LS3MIP (Group L) model ensembles with $M = 7$ members each. The analyses ~~are~~ 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).

3 ~~Results and Discussions~~

2.1 ~~Winter 2-m Temperature in Target Area~~Permafrost Stations

~~Fig.~~

To investigate the climatology in the permafrost area, sites were identified as permafrost sites using the definition of 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 ~~1 is based on~~ °C in any layer below 0.2 m depth, it was defined as a permafrost station. Using this method, we selected 19 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-meter~~ 2 m air temperatures (tas) from ~~ERA5-Land~~the CMIP6 multi-model ensemble, and the symbols correspond to the ~~matching~~ 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~~° E to 185~~°E~~° E, north of 45~~°N~~° N, the average DJF tas is generally below 0~~° C~~°C. The region with less than -25~~° C has~~ °C had a large overlap with the continuous permafrost region (Brown et al., 1997; Obu et al., 2019).

~~We aim 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 categorize the stations is by using their average DJF 2-m air temperature following Wang et al. (2016). By focusing on winter temperatures, we can further link the outcomes to the insulating effect of snow.~~ The temperature categories ~~are~~ were listed in the legend of Fig. ~~1. There are~~ 1. There were 3 sites with average tas warmer than -5~~° C~~ °C, 25 sites between -15 and -5~~° C, 79~~ °C, 76 sites between -25 and -15~~° C, and 40~~ °C, and 29 stations with tas below -25~~° C~~ °C. Besides, ~~five sites are~~ 19 sites were identified as permafrost (’perma’) sites using the method introduced by Lawrence and Slater (2005) (soil layer temperatures continuously below 0~~° C~~ °C for at least two years) and labeled by stars. All ’perma’ sites ~~have~~ had average winter tas values less than ~~-30 ° C~~-23 °C.

~~The spatial distribution of average winter tas conditions in ERA5-Land exhibits high consistency with the observations. For example, the site observations colder than -25 °C are also colder than -25 °C in ERA5-Land, including observations in isolated locations south of 55 °N. This proves that ERA5-Land can be a solid benchmark that supports observation as gridded data.~~

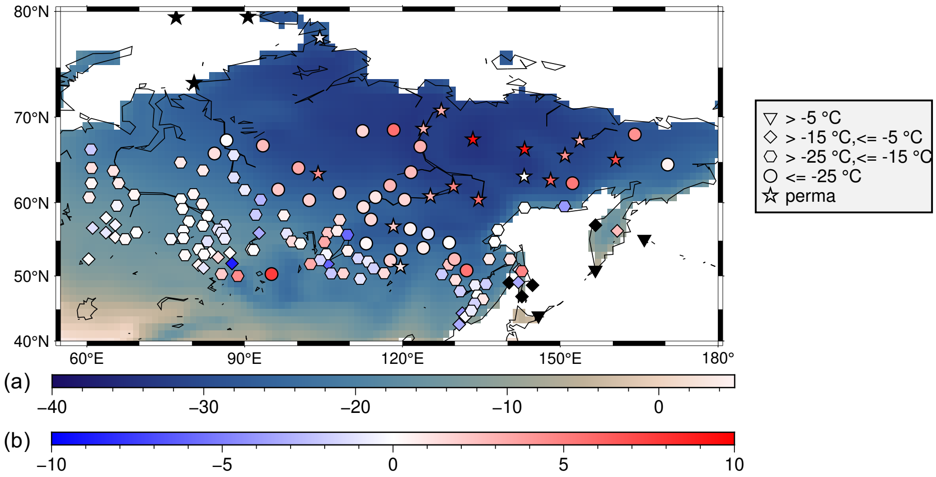

Figure 1. Mean near-surface air temperature (~~in °C~~a) as given by ~~ERA5-Land~~ the weighted multi-model ensemble from CMIP6, and the difference between the ensemble mean and the observational data (b) at sites for winter (DJF) in ~~1985-2014~~1985–2014 in °C. Symbols indicate the climate state at the observational sites (see legend), and symbol colors show the difference between the ensemble mean and the observational data.

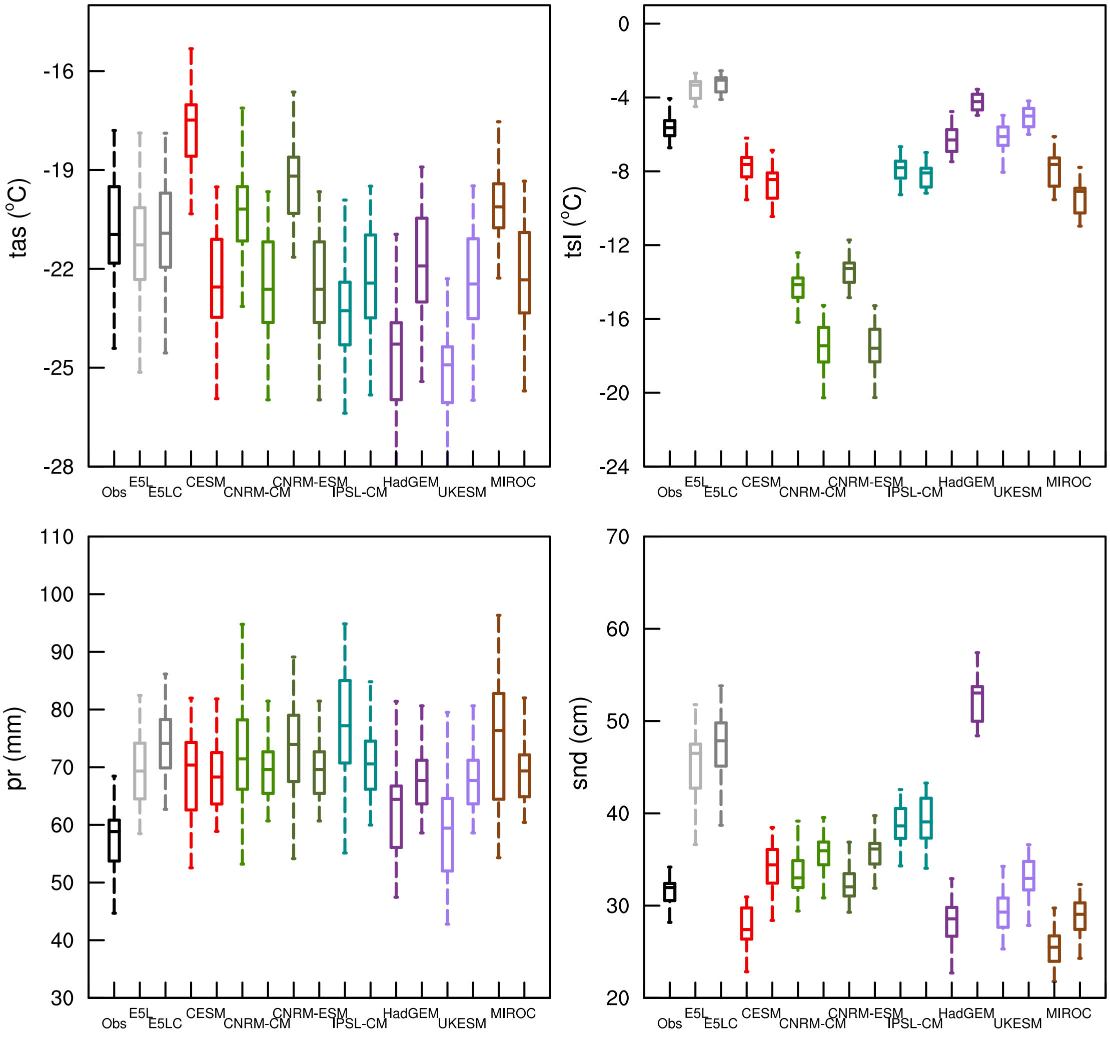

Figure 2. ~~Sites-averaged~~ Sites' averaged DJF~~-climates~~ -climates (~~1985–2014~~1985–2014) of hydrothermal variables as observed and simulated. Names ~~at~~ on the ~~abscissas~~ x-axis and the colors indicate the different data sources. ~~Observations~~ In addition to observations, ERA5-Land (~~Obs~~E5L)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

3.2 Model ~~climatologies~~Climatologies

~~Figure 2 shows the site-averaged winter , DJF climatologies from 1985 to 2014~~ The sites' averaged winter climatologies (DJF) of the four target variables for the different datasets ~~. The figure shows that~~ 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 ~~Obs climatologies than most of the models', where~~ observed climatologies than those of most models, though this is not the case for pr and snd.

~~The~~ Both ERA5-Land and E5LC were interpolated into the sites' locations using the nearest neighbor method. The coarser resolution resulted in slightly higher values of the four variables we analyzed, all less than one interquartile range of observations.

In LS3MIP~~'s~~, tas and pr ~~are interpolated~~ were derived from the same forcing ~~dataset. Still~~data. However, Fig. ~~2 shows slight differences between different land models because of interpolation uncertainties using different model grids with different setups. This illustrates how carefully a comparison of coarse-grid model output against point-like station data has to be interpreted. Additionally, the~~ 2 shows slight discrepancies of tas (less than 1 °C) and pr (less than 3 mm) among land-only models. The LS3MIPs' tas climatologies ~~are~~ were systematically colder by more than 1~~°C than Obs and ERA5~~ °C than observations and E5LC, and their pr climatologies ~~align better with ERA5-Land than with Obs (which has~~ aligned better with E5LC than with observations (which have on average about 15 % smaller values).

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

~~Variations of tsl are controlled by changes in overlying tas and by geothermal heat from below (Smith and Riseborough, 2002) .~~ Most models' tsl climatologies ~~are~~ were colder than observed~~(with the~~. 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~~climatologies~~. The strong cold bias of CNRM-CM6.1 and CNRM-ESM2.1 ~~more than −8 °C). Models belonging to the same family (the two CNRM models and the models HadGEM and UKESM, see Tab.1)exhibit similarities in their ability to simulate tsl. As to be~~ is also shown during spring and autumn (Fig. A2). As expected, the temporal variability, quantified by the box plots' IQR, ~~is~~ 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 ~~diversity in~~ spread of tas and tsl climatologies ~~, respectively , is~~ (Fig. A6) was respectively smaller than in ~~DJF (not shown). In June, July, and August (~~DJF. During JJA~~), the models' different~~, model differences in tas ~~variabilities are~~ variability were clearly reflected in tsl~~(with most sites lacking insulating snow~~, 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 ~~are higher than in observations by~~ were typically 10 mm ~~and thus closer to ERA5-Land. The diversity in the models' median values is within about ±10~~higher than in the observations. UKESM1.0-LL simulated relatively good pr in CMIP6, with less than 3 ~~mm. Some CMIP6 models show up to 1.5 times greater interannual~~

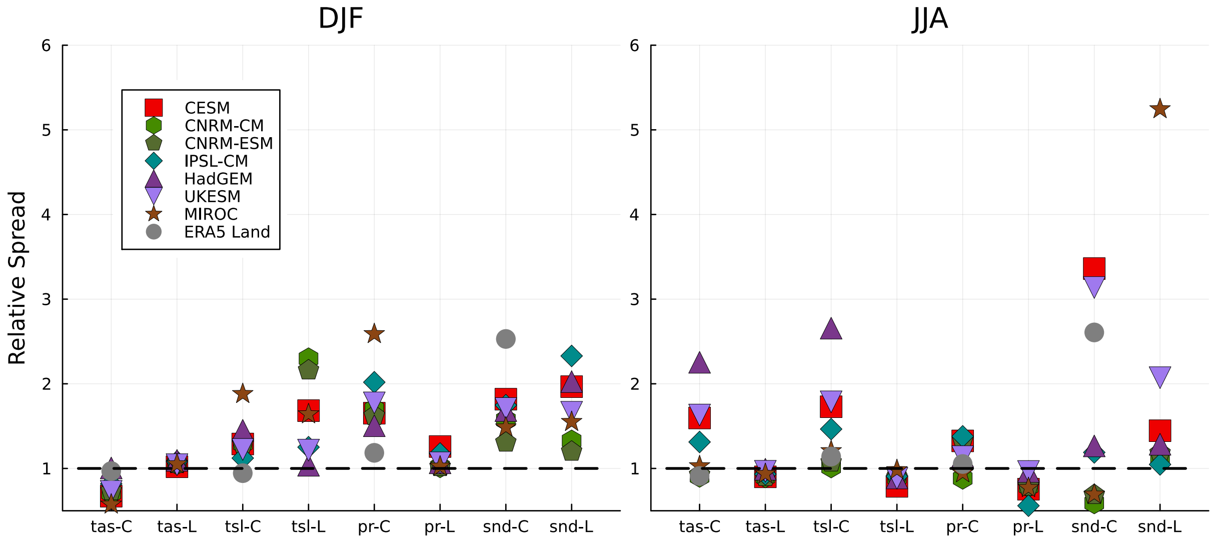

Figure 3. Relative spread (RS) of the sites-averaged climates (1985–2014) of the four variables in both ensembles and in ~~ERA5-Land~~ E5LC with reference ~~Obs~~ observations for all seasons. The colors indicate the models, and the ~~abscissas~~ x-axes show the variables and ensembles (C indicates CMIP6, and L indicates LS3MIP runs).

~~variability in DJF climatology compared to observation, suggesting an overestimation of the year-to-year fluctuations in winter climate~~mm overestimation in DJF average. The observed DJF snd ~~of about~~ was approximately 30 cm~~is~~. It was overestimated by 70~~% in the~~ % in HadGEM3-GC31-LL (LS3MIP ~~simulation~~ run) and by 50~~% in ERA5-Land. All the other model medians are within ±10 from Obs but with larger~~ % in E5LC. 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 ~~ERA5-Land~~ E5LC and the different models compared to ~~Obs~~ 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.

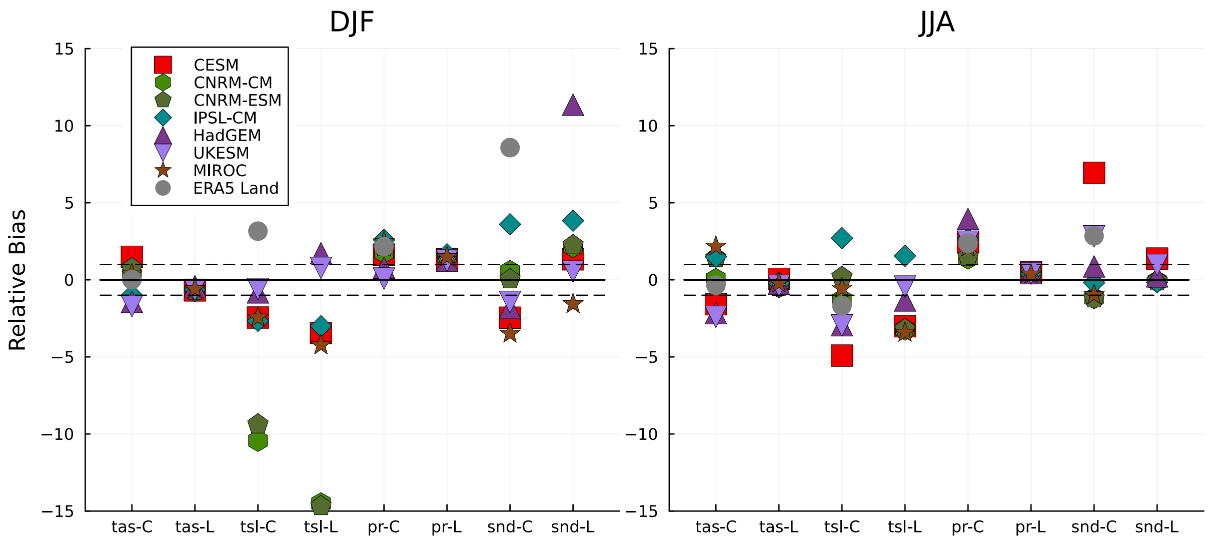

Figure 4. Same as in Fig. ~~3~~3, but for relative biases (RB). The ~~shaded areas~~ dashed lines (from -1 to 1) indicate the range of absolute median differences smaller than the observation's IQR.

~~Fig. 3 shows that most~~ Most CMIP6 models ~~underestimate~~ underestimated *tas* DJF climate variability ~~, besides in JJA with four models (CESM2, HadGEM3-GC31-LL, IPSL-CM6A-LR, UKESM1.0-LL)more than 1.5 times the variability of Obs. Most~~ and models overestimated *tsl* climate variability (Fig. 3). The CMIP6 models ~~overestimate *tsl* climate variability, but~~ showed a larger diversity in *tas*, and all values were below 1. However, the LS3MIP models ~~show~~ showed a large diversity ~~in winter (from 0.5 to almost 2, and it's a more extensive spread than the CMIP6 models despite their larger diversity in *tas*) and systematically underestimate variability in summer. For DJF, MAM, and SON~~for *tsl* and *pr* in winter, ranging from 1 to almost 2.7. The LS3MIP models performed well in simulating JJA variability of *tsl*, though there was a general underestimation. In DJF and SON, the *tas* ~~relative spreads~~ *RS* for all CMIP6 models ~~are~~ were within the range of 0.5 to ~~1.2.~~ 1. In MAM, they were within the range of 0.7 to 1.7. CESM2, HadGEM3-GC31-LL, ~~IPSL-CM6A-LR,~~ and UKESM1.0-LL ~~exceed~~ exceeded 1.5 ~~in JJA~~*RS* of *tas* in JJA. Larger variability differences in simulated *tsl* ~~are~~ were observed for LS3MIP simulations in DJF and SON, even though their *tas* share ~~an identical relative spread~~almost identical *RS*.

The ~~winter~~ DJF *pr* relative ~~spread~~ spreads of Group C ~~is~~ were all higher than ~~1.2. The *pr* in Group C exhibits more extensive group diversity than in Group L.~~ 1.3. In spring and autumn, most climate models simulated good ~~inter-annual~~ interannual variability of *pr*, with less than a 20 % difference to observation (~~not shown~~Fig. A1). Both groups and ~~ERA5-Land overestimate~~ E5LC overestimated the spread in ~~winter~~ DJF *snd*~~with *RS*-values larger than 1.2~~, having high discrepancies among models. The *RS* ~~-values scatter between 0.5 and about 2 in the other seasons~~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).

~~Fig. 4 shows the~~ The relative bias RB of all variables calculated in all seasons ~~. The shaded zone~~ 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, ~~the model's bias is considered relatively small~~its performance is considered adequate. For example, the winter RB of ~~MIROC is within the shaded area~~ all LS3MIP models was within the dashed line zone for tas, ~~tsl, and snd, but not for~~ and was at or slightly above 1 for the DJF pr ~~(note that Obs' pr is smaller than all models', including ERA5-Land's). Almost~~ of all LS3MIP models as they are derived from the same atmospheric forcing data. Nearly all CMIP6 and LS3MIP models ~~have a positive~~ exhibited a positive relative pr ~~-bias but a smaller relative and non-systematic~~ bias. The relative snd ~~-bias in winter~~bias was much more diverse; models only showed a consistent positive snd bias in MAM (Fig A2). The CMIP6 runs of HadGEM3-GC31-LL and UKESM1.0-LL ~~underestimate~~ underestimated the values of tas and tsl in all seasons. ~~For~~ The tsl ~~, the relative bias in DJF for LS3MIP CNRM-CM6.1 and CNRM-ESM2.1 exceeds the observed IQR-range~~ RB in DJF for CNRM simulations exceeded -15 to -9 times the observed IQR, while other models ~~are all within an acceptable range~~ were all within a range of ±5 times in both groups. ~~Negative~~ The RB ~~in~~ of CMIP6 and LS3MIP in tsl ~~is the largest~~ were all negative in the transition seasons, ~~which also shows a large spread~~ and also high discrepancies were shown in snd RB ~~-values. With better RS and RB of tas and pr, models still have diverse abilities simulating tsl and snd.~~ values.

3.2.2 Spatial Heterogeneity

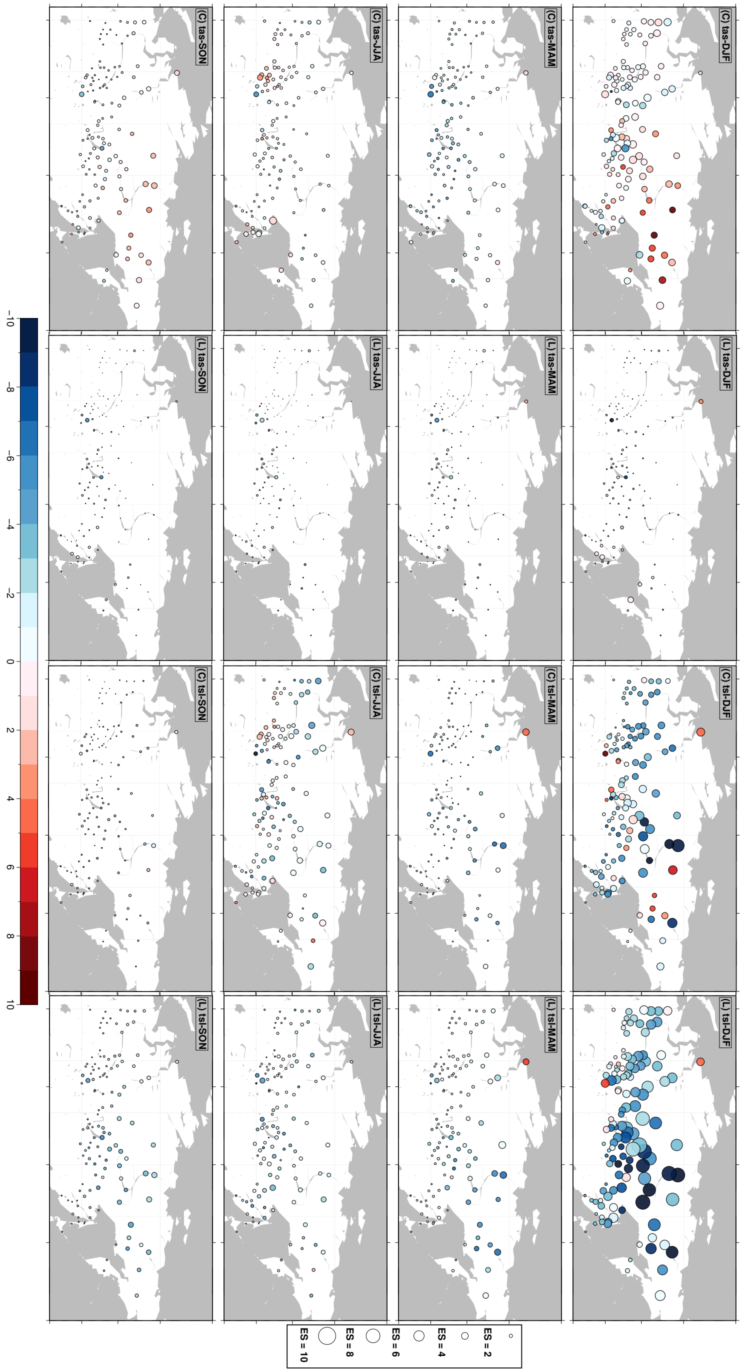

Figure 5. ~~Mean~~ Weighted multi-model mean biases (EB) and their standard deviations (~~°C~~ES) of ~~the models'~~ tas and tsl at ~~Obs~~ observational sites in °C. Panels A–H and I–P are for tas and tsl, respectively. The first and third columns are the results of CMIP6 models, and the second and 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 5 shows the model ensemble mean biases (~~In DJF, multi-model ensembles had the largest EB ~~) and the ensemble spread of biases (ES) for tas~~ and ~~tsl at the Obs sites. EB and bias diversity are largest in winter and larger in tsl than in tas. In northeastern Siberia's winter~~ES (Fig 5). In DJF, the CMIP6 models ~~overestimate~~ overestimated tas on average by more than ~~4 °C~~ 0.5 °C, but with a large ~~model bias standard deviation of 3.2 °C.~~ ES of 2.96 °C. The DJF average EB and ES for LS3MIP were -0.51 °C and 0.69 °C, respectively.

In the other seasons, EB and ES ~~are~~ were distinctly smaller, with most CMIP6 runs' tas slightly too cold in spring at most stations and slightly too warm in autumn ~~at~~ in northeastern Siberia. ~~Differences in grid cell scale among models can lead to biases in the tas state over the grid.~~ The tas EB of the forced models ~~are~~ was small at most sites (exceptions ~~are~~ were near water bodies, e.g., Lake Baikal and sea coasts, which ~~indicate interpolation artifacts because of~~ indicated interpolation artifacts due to the selected grids). ~~The magnitude of Group L's tas ES is almost unchanged among seasons, implying its bias mainly relates to the geographical characteristics of stations' sites.~~

The tsl EB and ES ~~are~~ were larger in magnitude than for tas, especially in winter~~with spatial~~, with spatial averaged EB of ~~-3.0 °C and -4.0 °C~~ -2.66 °C and -3.57 °C for CMIP6 and LS3MIP runs, respectively. Additionally, the LS3MIP runs ~~are notably more diverse~~ had a larger bias spread than the CMIP6 runs ~~throughout all seasons but in~~ in all seasons except summer. In winter, the ~~ensemble bias spreads are 3.6 °C~~ ES were 2.98 °C for the CMIP6 and as large as ~~5.7 °C~~ 4.55 °C for the LS3MIP runs. ~~Given the smaller tas EB and the larger tsl EB in the LS3MIP than in the CMIP6 simulations, the modeling diversity introduced by the land surface models is large and partly compensated for by atmospheric diversity in the CMIP6 simulations.~~

3.3 Permafrost Region

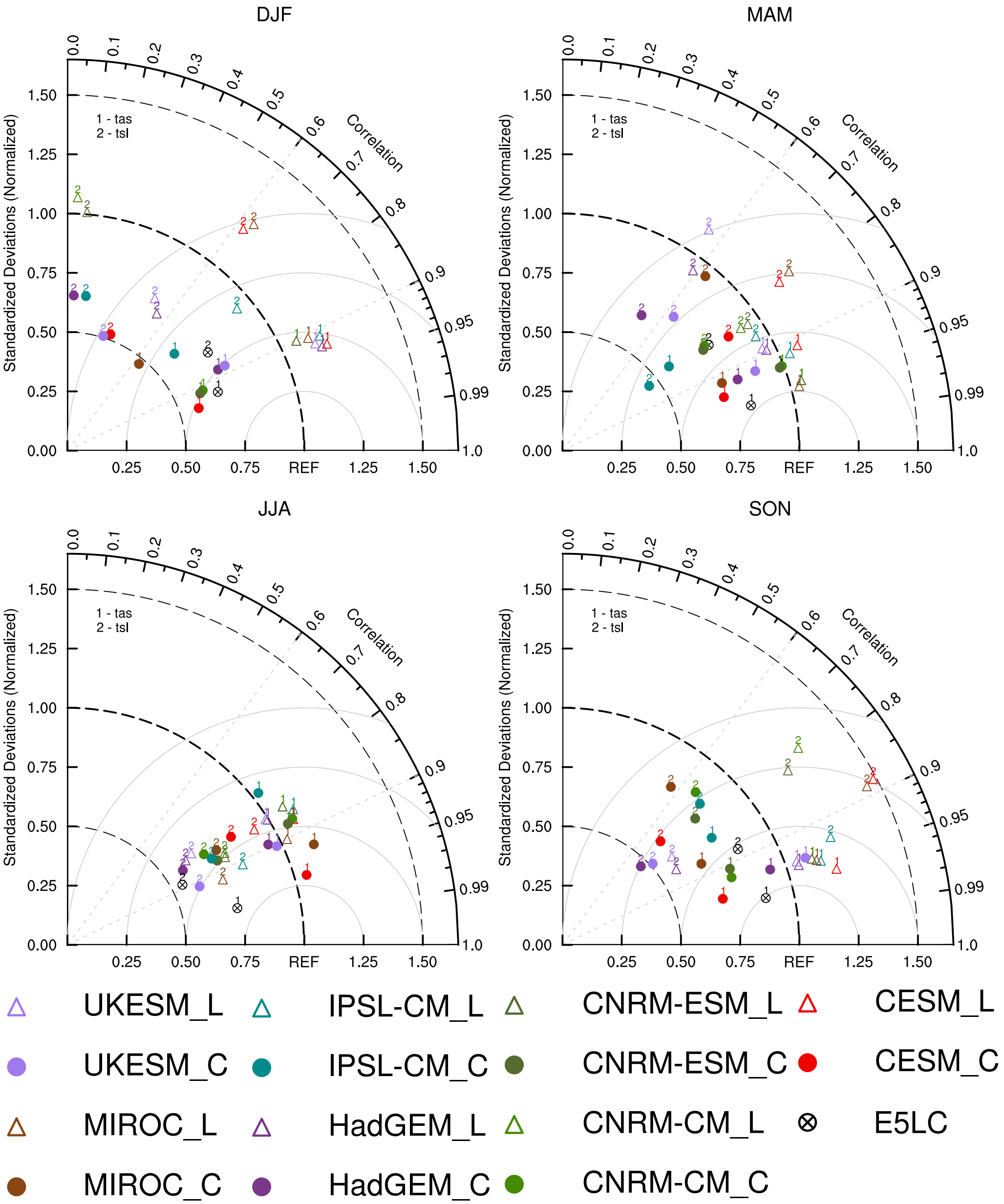

Figure 6. Taylor diagrams indicating the spatial correlation, standard deviation, and unbiased root-mean-square (RMS) difference (~~grey~~ gray circles) for the four seasons of the simulated variables tas, tsl ~~, pr, and snd~~ (labeled by numbers) against observations (REF) at permafrost sites (here: sites with mean winter ~~$tas < -25°C$~~$tas < -25\,^\circ$C). Normalization is applied using the standard deviations of the observations. Colors indicate different models and black crossed circles ~~ERA5-Land~~E5LC. Triangles show the results of the LS3MIP runs, and solid circles of the CMIP6 runs.~~Snow depth is not shown in JJA.~~

~~Here, we show the models' results at the 45 observational sites with average winter tas below −25 °C, at which we find the largest ES (see Fig. 5). This research focuses on the shallow soil response to atmospheric forcing, explicitly targeting a depth of 20 cm. Using the authentic~~ Using the classic definition of permafrost, which requires more than two consecutive years of temperatures below 0~~°C, only five of the sites qualify at this depth. However, under similar climatic conditions (where winter tas are below -25 °C), we observe 40 more sites distributed within permafrost areas as defined in Brown et al. (1997), indicating that permafrost may exist at deeper soil layers. Incorporating these sites for statistical analysis will enhance the robustness of the findings.~~

~~Figure 6 shows spatial~~ °C, 19 permafrost sites were selected. 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)~~, which compare the 30-year-average~~ . 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 standard deviation ranging from 0.95 to 1.1, while E5LC's standard deviation was lower than 0.75. In the other seasons, the tas simulations had correlations higher than 0.6; however, they all underestimated the standard deviation of tas. In MAM and SON, tas correlations were almost above 0.9, which was slightly higher than tas correlations in JJA.

Almost all CMIP6 and LS3MIP models struggled ~~with simulating soil temperature,~~ to simulate tsl ~~, with large RMS differences and low correlations~~ in permafrost sites, exhibiting low correlations and large RMSE, especially during DJF. ~~Correlations are generally better~~ The correlations were generally higher than 0.5 for all variables in the other seasons. ~~The~~ E5LC had similar DJF tas ~~correlations are better than ca. 0.8 for all models(except for MIROC and IPSL-CM of Group C in DJF), which indicates well-simulated spatial patterns by the~~ simulation performance to the CMIP6 ~~models~~models, but it simulated DJF tsl with the smallest RMSE compared to all models. When considering the same model across CMIP6 and LS3MIP, land-only simulations consistently outperformed others in DJF tsl. This is because tas and snd were superior in LS3MIP (Fig. A3).

~~In the spring and autumn(MAM, SON),~~ From spring to autumn, the CMIP6 models ~~tend~~ tended to overestimate precipitation variability (~~figure not shown~~see Fig. A3). Overall, the Group L models and ~~ERA5-Land tend to perform only slightly~~ E5LC performed better than the Group C models in ~~tsl and MAM~~ the MAM and SON snd ~~. In SON, there is a more considerable difference in correlation performance~~ simulations. In MAM, the standard deviation of snd ~~than in MAM, less than 0.6 for the~~ was considerably larger than in DJF and SON, exceeding 1.25 for Group C models and ~~higher than 0.7 for the~~ 1.5 for Group L models. ~~Meanwhile, the correlations of~~ Using the same forcing, the snd correlations improved in SON, with all values above 0.8 except for UKESM1.0-LL, which had a correlation of 0.7. The correlations of CMIP6 pr ~~are~~ were similarly low in ~~both seasons. This indicates that the accumulation of snd in climate models is better related to~~ all seasons. Despite LS3MIP showing a high correlation (higher than 0.8) and good standard deviation (0.9 to 1.1) of pr ~~condition than melting.~~

~~In JJA, tas are less correlated with observation in JJA than in MAM and SON, while tsl maintains the similar accuracy of tas simulations. The normalized RMS differences remain mainly within ± 0.25, while correlation coefficients range from 0.6 to 0.9~~in DJF and MAM, CMIP6 and LS3MIP were unable to simulate snd correlations higher than 0.8 in these two seasons.

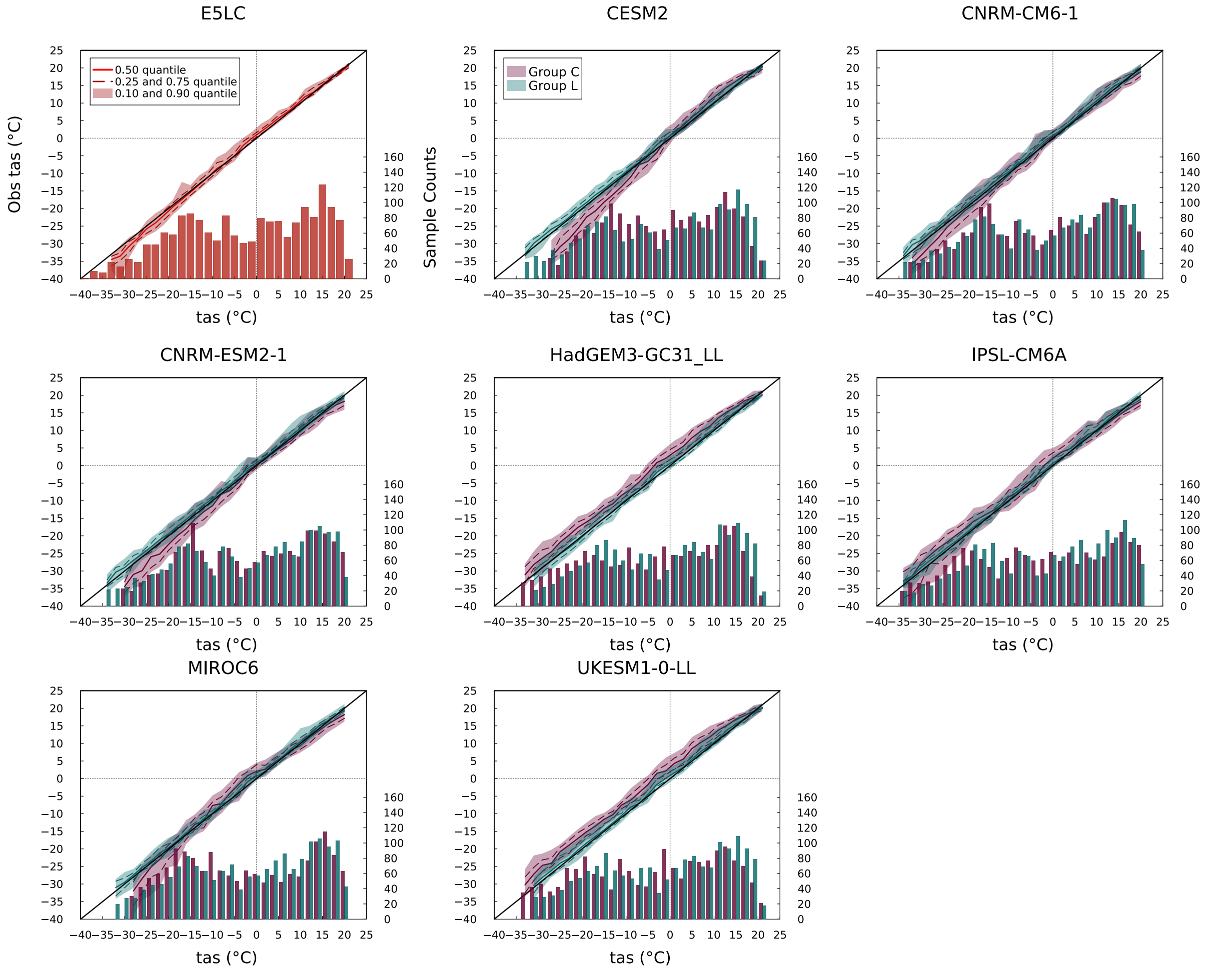

Figure 7. Quantile-Quantile (QQ) and histogram plots for tas (Wilks, 2019). The plots show 30-year mean monthly tas data of models at all sites and their observational values. The model simulation outputs are binned at 2~~°C~~ °C intervals. The colored solid and dash curves are the median and 1st/3rd quartiles of the corresponding observations for all data points in the temperature interval, and the shaded area is the inter-decide 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.

3.4 Climate Dependency of Modeled Temperatures

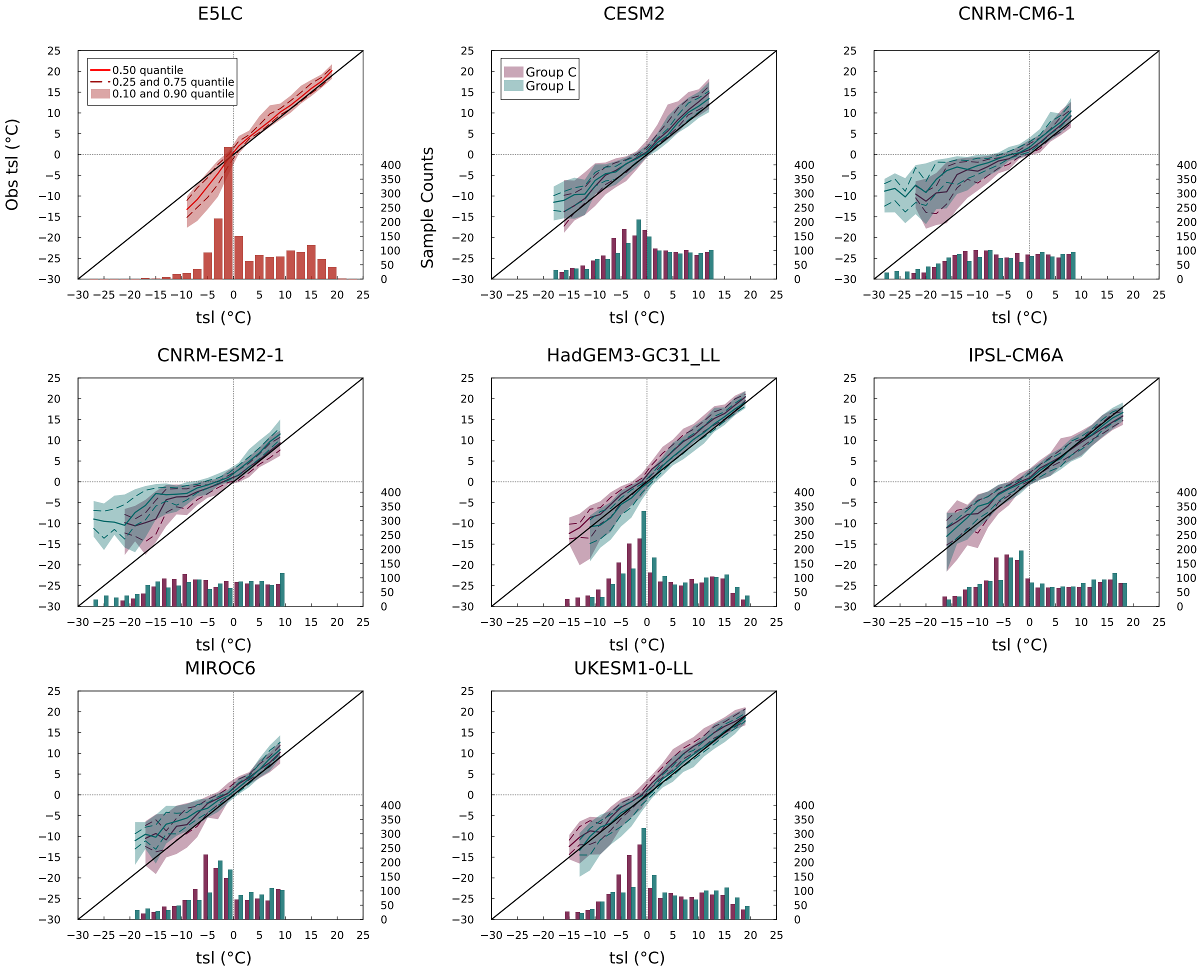

Figure 8. Same as Fig. ~~7~~7, but for tsl.

~~Since larger modeled~~ Larger model biases and discrepancies ~~are~~ were found in permafrost regions~~, we want to further evaluate how modeled~~ . To evaluate how model errors depend on the temperature ~~state. So, we categorize~~ conditions, we categorized all the model outputs of tas and tsl ~~by~~ based on their values and ~~then compare the relationship between model outputs and corresponding Obs~~compared these outputs to the corresponding observations. For each model, the monthly data ~~of every site is an average of~~ from every site was calculated as an average over 30 years. ~~Then, we look for the corresponding value of~~ We then examined the corresponding values for the same site and month with valid ~~observation. The mean value,~~ observations, calculating the median values, quartiles, and deciles ~~are calculated~~ for each interval.

As shown in Fig. ~~7~~7, Group L models ~~simulate similar histograms as observation and small~~ generally reproduced the observed sample distribution of tas and maintained narrow quantile spreads across ~~all temperature ranges. However, a slight cold bias below -30 °C exists in Group L. Conversely, ERA5-Land has almost no bias when tas is above -20 °C, but has a warm bias below -20 °C. Group C 's simulation results exhibit larger errors, with~~ the full temperature range. In contrast, Group C models exhibited larger deviations. For example, CESM2, CNRM-CM6.1, and CNRM-ESM2.1 ~~showing warm bias when~~ tended to simulate tas ~~is~~ higher than observed when the observed tas was below 0~~°C,and~~ ,°C, with median differences exceeding +3 °C. HadGEM3-GC31-LL and UKESM1.0-LL~~displaying a comprehensive cold bias of~~ , by contrast, consistently produced tas values approximately 5~~°C. In most temperature states,~~ °C lower than observed across a broad range of frozen conditions. IPSL-CM6A-LR ~~shows a slight cold bias of less than~~ showed persistent cold biases of over -5~~°C~~ °C in most subzero-temperature cases. MIROC6 ~~exhibits a cold~~ produced tas ~~bias from -10 to~~ values that were lower than the observed values between -10 and 2~~°C and a warm~~ °C, but the simulated tas ~~bias that gradually increases below -20 °C~~values were increasingly higher than the observed values when simulated tas dropped below -20 °C. A common ~~characteristic among the~~ feature of Group C models ~~in tas simulation is that a larger vertical quantile rangeappears~~ was a wider vertical spread of quantiles in the below-freezing temperature range, especially when tas ~~is below -15 °C, emphasizing the requirement to improve the accuracy under such surface air temperature conditions~~dropped below -15 °C.

~~Larger vertical quantile ranges appear in~~ Compared to tas, quantile spreads were generally larger in the tsl ~~results than in tas. The~~ simulations. For instance, the E5LC model exhibited a bimodal tas ~~distribution of ERA5-Land shows two main peaks~~ sample distribution (Fig. ~~7), one around~~ 7), with one peak near 15~~°C and the other around -15 °C. However, the~~ °C and another around -15 °C, whereas its tsl ~~of ERA5-Land has one major peak~~ sample distribution (Fig. ~~8) , which is within -5 to 0 °C. The ERA5-Land vertical quantile range is~~ 8) showed a single dominant peak between -5 and 2 °C. E5LC also displayed smaller and more ~~unbiased~~ centered tsl quantiles above 0~~°C. Still, it shows a warming bias and increased vertical quantile range below 0 °C. Cao et al. (2022) discussed an unreasonable warm~~ °C, but it showed an increasing warm bias and quantile spread as tsl ~~bias that is possibly due to the overestimation of permafrost snow depth in ERA5-Land. Similar warming of tsl due to excessive snd occurs~~ dropped below freezing. Similar near-surface soil warming was observed in the LS3MIP ~~simulation~~ run of HadGEM3-GC31-LL, which was associated with excessive snow depth. CNRM-CM6.1 and CNRM-ESM2.1 ~~are the only ones that do~~ were the only models that did not show a ~~peak in soil temperature samples~~ cluster of tsl values near 0~~°C. The~~ °C. Instead, their tsl ~~in CNRM-CM6.1 and CNRM-ESM2.1 are more likely distributed and show a strong cold bias from 2 to 15 °C~~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 ~~have an overall cold bias, -5 °C~~ also displayed systematic cold biases, with median tsl differences of -5 and ~~-6 °C at -10 °C of observed temperature, respectively. The~~ -6 °C, respectively, when the observed tsl was around -10 °C. Group L models ~~have varying warm bias~~ varied more in magnitude but generally showed positive tsl biases below 0~~°C tsl, indicating that the tsl simulated by these models have a larger bias~~ °C.

~~There are notable discrepancies between the two ensembles~~ The two groups also showed distinct differences in simulating cold temperature extremes, particularly in the minimum tas. For example, ~~the~~ CESM2's minimum tas in the ~~CMIP6 run for CESM2 is~~ coupled simulation was 6~~°C higher than that of the LS3MIP run, whereas the tas for CMIP6 IPSL-CM6A-LR is 4 °C lower than that of the~~ °C higher than in its land-only ~~run. Such~~ 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 ~~differences in the lower extreme directly impact that of tsl, creating a gap of~~ values that were 2~~to 6 °C in the lowest value between two ensembles. For models (CESM2, CNRM-CM6.1, CNRM-ESM2.1, and MIROC6) underestimate tsl while exhibiting warm biases of tas, one conceivable explanation is that they underestimate the snow insulation effect(in line with results by (Dutch et al., 2022)) and surface insulation effect (the thermal offset, especially if tas is below -15 °C), allowing a too large energy loss from the soil to the atmosphere.~~

~~There is an excessively low tsl shown in Fig.8, possibly due to insufficient geothermal (functions as upward energy flux from the bottom of soil columns). As the decrease~~ °C lower in the coupled runs than in their land-only counterparts. These discrepancies in tas ~~has a limited influence on the~~ at the cold end of the sample distribution were also reflected in the tsl ~~through high snd, the main source of error is likely from the other side of energy transportation (thermal conditions in the bottom of the soil column). The melting/accumulating snow in MAM and SON and the prevailing phase change processes in the soil are conceivably the major reason for higher~~ simulations. In particular, the minimum tsl ~~inter-annual variability in these seasons.~~ values differed by up to ±6 °C between models in the two groups.

~~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 °C (Set Frozen), -5 °C to 5 °C (Set Intermediate), and above 5 °C (Set Warm).~~

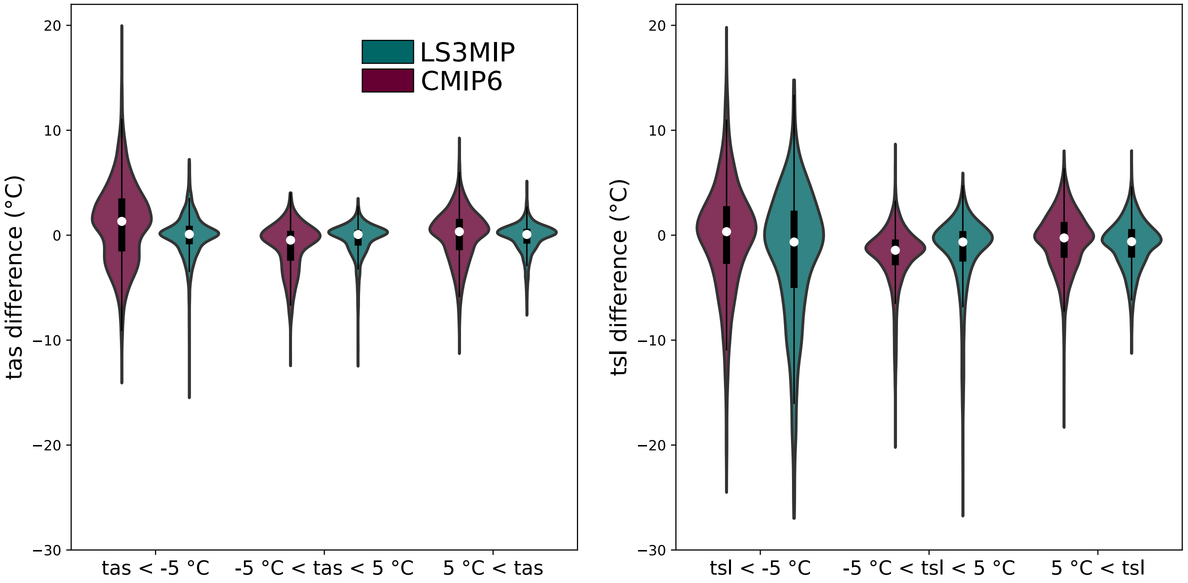

Figure 9. Difference between model tas (left), tsl (right) and corresponding observations. The x-axis represents different temperature intervals, and the y-axis is the 30-year average temperature difference ($T_{model} - T_{obs}$). Differences are categorized into three sets according to the 30-year average temperatures of every month at the observation sites: below -5 °C (Set Frozen), -5 °C to 5 °C (Set Intermediate), and above 5 °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.

Category	Sample Size	Ensemble	Q_1	Median	Q_3	Mean	Std. Dev.	Kurtosis
tas Set Warm	3475	CMIP6	-1.42	0.34	1.56	0.01	2.46	0.76
		LS3MIP	-0.81	0.07	0.63	-0.22	1.43	2.93
tas Set Intermediate	1480	CMIP6	-2.44	-0.47	0.41	-1.12	2.28	0.77
		LS3MIP	-0.99	0.06	0.52	-0.38	1.61	7.79
tas Set Frozen	3805	CMIP6	-1.55	1.32	3.51	1.16	3.85	0.59
		LS3MIP	-0.87	0.09	0.90	-0.04	2.06	6.64
tsl Set Warm	3295	CMIP6	-2.15	-0.26	1.26	-0.51	2.81	1.60
		LS3MIP	-2.13	-0.61	0.58	-0.86	2.35	1.21
tsl Set Intermediate	3655	CMIP6	-2.86	-1.44	-0.41	-1.95	2.87	4.33
		LS3MIP	-2.51	-0.65	0.40	-1.70	3.81	6.41
tsl Set Frozen	1445	CMIP6	-2.74	0.34	2.77	-0.12	5.26	1.68
		LS3MIP	-5.03	-0.65	2.34	-1.81	6.43	1.24

~~In addition, we categorized the model output by the freeze/thaw state of observation. This enables us to compare the performance of two groupsin different temperature states. In order to avoid assessment errors caused by different sample sizes, we discuss 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 occurs most frequently between the boundaries.~~

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

Statistical values of Fig. 9 were listed in Table 3. Set Frozen and Set Warm tas data sample sizes ~~are~~ were more than twice as large as in Set Intermediate. Set Frozen ~~has~~ had the largest standard deviations, ~~3.84 °C~~ 3.85 °C in CMIP6 runs, and ~~2.14 °C~~ 2.06 °C in LS3MIP runs. The mean values of the CMIP6 runs ~~are rather~~ were divided, with ~~-1.76 °C and 0.11 °C~~ -1.12 °C and 0.01 °C in Set Intermediate and Set Warm, respectively. ~~In contrast, the difference in LS3MIP runs is negligible, suggesting that the climate model will likely have a cold and small deviation under these temperature states.~~

The tsl samples ~~are~~ were mainly concentrated in Set Intermediate and Set Warm. In contrast, there ~~are~~ were also higher standard deviations in Set Frozen, ~~5.39 °C and 6.73 °C~~ 5.26 °C and 6.43 °C for CMIP6 and LS3MIP runs, respectively. The standard deviations of Set Frozen and Set Intermediate in LS3MIP runs ~~are~~ were higher than those in CMIP6 runs by ~~1.33 °C and 1.12 °C~~1.18 °C and 0.94 °C, respectively. ~~This again indicates that the land model alone caused higher variability.~~ The mean and minimum value of tsl bias ~~is~~ was much lower than that of tas bias, and this negative bias ~~is~~ was shown in all ~~Sets~~ corresponding sets of both groups ~~, suggesting that the land models tend to simulate tsl colder than it should be. Better accuracy of tas of Group L in Set Frozen and Set Intermediate does not lead to better tsl results for the models~~in Table 3. And tsl below -5~~°C shows~~ °C showed higher overall variability in Group L than in Group C, with the highest IQR of 7.37 °C, and highest standard deviation of 6.43 °C among all categories.

3.5 Snow Insulation

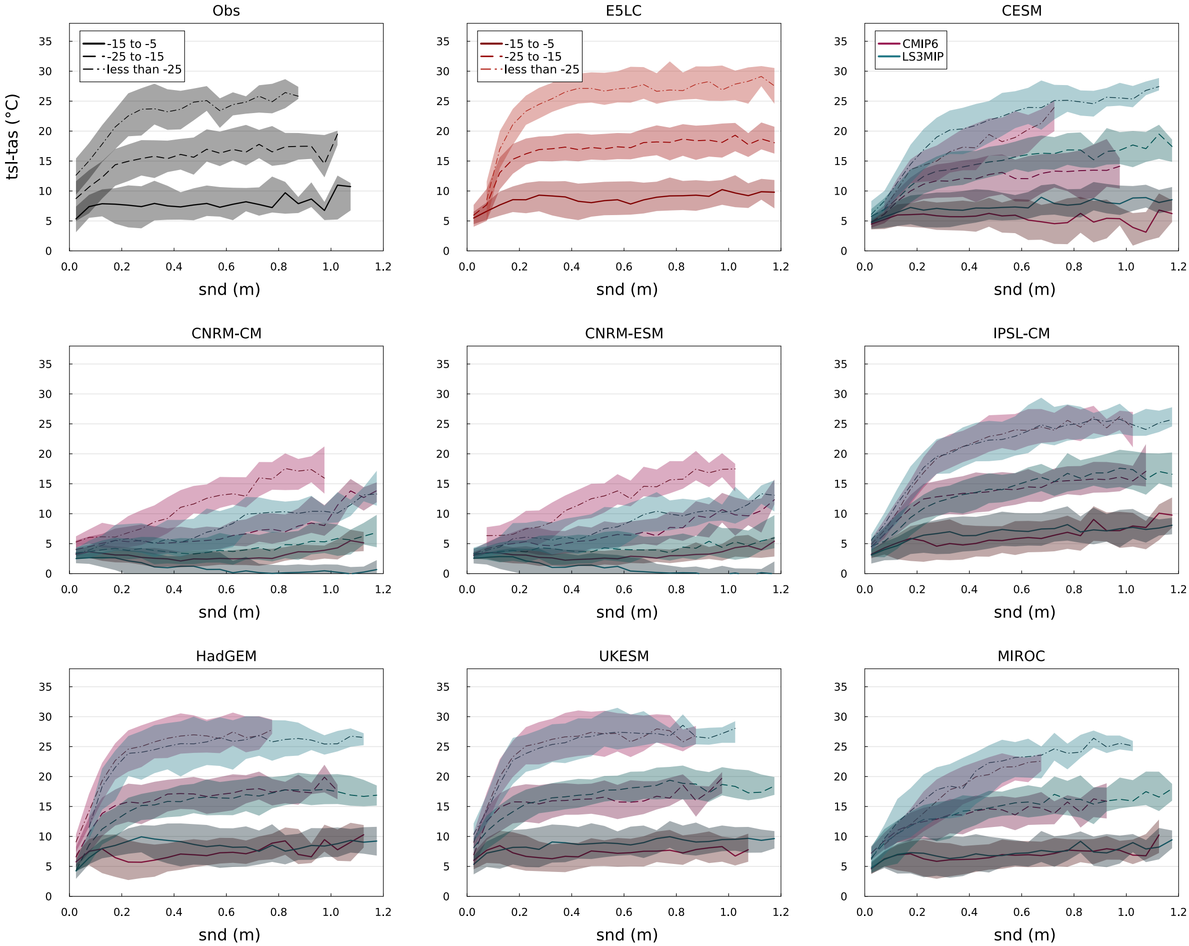

Figure 10. Snow depth and soil-surface temperature differences. The temperature ~~categorizing~~ categorization method ~~of~~ follows Wang et al. (2016) ~~and encompasses~~ but includes data from all seasons. The plots use different line styles and color schemes to ~~differentiate~~ distinguish between ~~the various~~ 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 ~~a~~ coverage of 0–1.2 m snd, and each interval contains a minimum of ten samples for accuracy. The ~~plot curves~~ lines represent the median values, ~~with~~ while the shaded areas ~~indicating~~ indicate the ~~25th–75th~~ interquartile range (25th–75th percentile).

To better understand the insulating effect of snow in the models, we ~~refer to the method applied~~ 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 ~~which contain valid~~ containing tas, tsl and snd ~~, characterize~~ and characterized them by three different tas intervals (-15 to -5~~°C~~°C, -25 to -15~~°C~~ °C, and less than -25~~°C~~°C). Analyzing the relationship between temperature difference and snow depth ~~allows~~ allowed us to understand each model's snow insulation effect under different tas and snd conditions.

According to the ~~observation~~ observations shown in Fig. ~~10, two key factors should be considered regarding the snow insulation effect. Firstly,~~ 10, even when snd is very thin, the insulating effect of the soil surface layer ~~itself is still influential~~was still large. Its intensity ~~can vary~~ could range from 3~~°C~~ °C to 15~~°C~~ °C according to tas. ~~Secondly, the~~ The impact of snow on ~~Δ~~ΔT ($tsl - tas$) ~~has the highest diversity~~ showed the greatest variability when the snow depth is shallow~~; under sufficiently thick snow, the~~ . However, once the snow depth exceeded about 0.2 m, tsl ~~gradually convergences~~ stabilized near 0~~°C and is primarily impacted~~ °C or lower, becoming nearly constant and minimally affected by tas~~in a limited manner~~. For temperatures ranging from -15 to -5~~°C~~°C, this inflection point ~~is~~ was reached when snd ~~is~~ was about 0.1~~m~~ m high. For colder temperatures, it ~~is~~ was reached at depths around 0.2~~m~~ m to 0.3~~m. However, snow shields most of the impacts from the overlying tas in thicker snow conditions, as shown by a strong relationship between ΔT~~ m. Thicker snow conditions limited the cooling of the underlying soils, as indicated by the strong relationship observed between ΔT and tas when snd ~~is higher than 0.3 m. The insulation effect of snow becomes stronger at lower tas. The~~ exceeded 0.3m. The median value of maximum ~~Δ~~ΔT for the -5~~°C~~ °C to -15~~°C category is~~ °C category was approximately 7.5~~°C~~°C, while for the -15~~°C~~ °C to -25~~°C~~ °C category, it ~~is~~ was 15~~°C~~°C, and it ~~reaches~~ reached 24~~°C~~ °C when tas ~~is~~ was below -25~~°C~~ °C. When snd ~~exceeds~~ exceeded 0.3 m, tsl drops by 1~~°C~~ °C for every 4~~°C~~ °C to 5~~°C decrease~~ °C decreased in tas, ~~taking into account~~ 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 snow and generally larger insulation in the coldest category.

Only HadGEM3-GC31-LL, IPSL-CM6A-LR, and UKESM1.0-LL ~~exhibit~~ exhibited ~~Δ~~ΔT curves ~~which are similar to observation~~that were similar to observations. Despite the generally good performance, HadGEM3-GC31-LL and UKESM1.0-LL ~~have~~ had deficiencies in simulating soil insulation below -15~~°C, with~~ °C, with ~~Δ~~ΔT 1 to 2~~°C~~ °C higher than observation when snd ~~is~~ was higher than 0.3 m. IPSL-CM6A-LR also ~~underestimates~~ underestimated the insulating effect when there ~~is~~ was almost no snow. ~~IPSL-CM6A-LR accurately reproduces~~ However, it accurately reproduced the inflection point of the snd-ΔT curve. After reaching ~~the inflection~~ this point, the insulation effect still ~~increases~~ continued to increase in IPSL-CM6A-LR. ~~While~~ In contrast, HadGEM3-GC31-LL and UKESM1.0-LL ~~show~~ showed a slower rate of increasing ~~Δ~~ΔT that ~~is~~ was more comparable with the observations. ~~All other models fail to reproduce the observation-like curve, underestimating the snow insulation effect under most conditions .~~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.

~~None of the models can accurately depict soil insulation . When snd is below 0.05 m, all models simulate ΔT that is lower than the observed value. This could be due to low resolutions of land surface models, which make it difficult to accurately determine the distribution of organic matter on the surface (de Vrese et al., 2023), resulting in a miscalculation~~

~~of the surface insulation effect. This shortage poses a potential issue where increasing model soil surface insulation could lead to overestimating the total ΔT. It is especially true when some models adequately reproduce the insulation magnitude at high snd.~~ Comparing the CMIP6 runs with the corresponding LS3MIP runs, UKESM1.0-LL~~simulates large insulation effects of 4 to 12 °C under low snd (lower than 0.05)~~, ~~which is closest to the benchmark and which might guide in further improvements.~~ 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. 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 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 LS3MIP runs.

~~Out of all the~~

4 Discussion

4.1 Winter 2 m Temperature in Target Area

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 resolution did not have a major contribution to bias, concerning differences between ERA5-Land and E5LC 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, may not represent the broader grid-scale climate. Therefore, higher consistency between the two 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~~tested,~~' ability to simulate snd and tsl with similar atmospheric forcing (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 due to differences between grid data and site-level data. The large inter-model spread in CMIP6 tas climatologies indicated substantial differences in model configuration or parameterization. The family resemblance in tsl biases (e.g., CNRM, 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.

The excessive spread in JJA tas and tsl for specific CMIP6 models (CESM2, HadGEM3-GC31-LL, UKESM1.0-LL~~consistently demonstrated similar snow insulation effects in both ensembles. However, the results~~) points to seasonal over-responsiveness to summer variability. The excessive DJF RS and generally positive RB of ~~the other models varied, suggesting that~~ 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 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. The positive winter relative pr bias and smaller, less systematic relative snd bias across most models points to possible compensation mechanisms or 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 snd. This indicates challenges in accurately capturing the variability and relative bias for these specific variables across different models and seasons. 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 tsl simulation.

Given the smaller tas EB and the larger tsl EB in the LS3MIP than in the 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 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, the multi-model EB of CMIP6 and LS3MIP tsl was negative, but it's positive for CMIP6 tas. And the ES of LS3MIP tsl was larger than ES of CMIP6 tas, although LS3MIP models were forced by the same atmospheric data. This indicates larger variability caused by the land surface model than the atmospheric model.

4.3 Permafrost Region

The large RMSE and low correlations of tsl in DJF, compared to better results in other seasons, highlight a seasonal dependency in simulation skill, possibly tied to limitations in frozen-process representations. The consistency of 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 difference in the correlation performance of SON snd between Group L and Group C highlights the importance of a good simulation pr during the snow accumulation period. Nevertheless, low correlations and high standard deviations of Group C pr from spring to autumn suggest that improvements are still needed in simulating precipitation seasonality.

4.4 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 models (CESM2, CNRM-CM6.1, CNRM-ESM2.1, and MIROC6) exhibited cold tsl biases despite warm tas, likely due to underestimated snow and soil insulation. This resulted in excessive energy loss from soil to atmosphere when $tas < -5\,^{\circ}$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 Fig. A5, at deeper soil layers (0.8 m and 1.6 m), the cold bias was also present when soil temperatures were above 0 °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 the land-only model parameterizations or low soil moisture content.

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. 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 $-5\,^{\circ}$C suggest that discrepancies in snow insulation representation, soil layering schemes, or freeze-thaw parameterizations contribute to greater simulation inconsistency within the model ensemble.

4.5 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 near 0 °C or lower once the snow depth exceeded about 0.2 m. This stabilization was minimally affected by tas, with inflection points reached at different depths depending on temperature ranges. The insulating effect of snow is due to its low thermal conductivity (about $0.3\,\mathrm{W\,m^{-1}\,K^{-1}}$), which reduces heat loss from the soil. This effect became more pronounced at lower tas as the temperature difference (ΔT) increases.

The results from Fig. 10 suggest that factors beyond tas and snd ~~also impact the snow layer's ability~~ 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 ~~example,~~ instance, the CESM2 ~~shows~~ LS3MIP run showed a snow insulation effect closer to observations~~in its LS3MIP run, but the~~ , whereas the CESM2 CMIP6 run ~~of CESM2 simulates~~ simulated a much lower ~~Δ~~ΔT curve. ~~Despite cold conditions, an increase in *snd* still affects the snow insulation effect of LS3MIP CESM2. Similar results are observed in MIROC6 as well. Both models show realistic snow insulation effect values when *snd* is above 0.4 .~~ The issue with their snow insulation dynamics ~~is the slow rise in~~ was the low initial insulation and slow increase in ~~Δ~~ΔT when snd ~~is~~ was below 0.2~~. On the other hand, CNRM-CM6.1 and CNRM-ESM2.1 have a low curve of ΔT-*snd*, even with a negative ΔT outcome for *snd* above 0.2 under the -5 °C to -15 °C category of LS3MIP runs.~~ m.

Four land models mentioned in this study are newer versions of the land models studied by Wang et al. (2016). ~~We evaluate the different versions by comparing their snow insulation effect. The study reveals that the newer modelshave made noteworthy progress~~Comparing Table A1 with findings of Wang et al. (2016) revealed notable advancements in newer models. CLM5~~(here indicated by CESM2) generates~~ .0 (indicated as CESM LS3MIP in Fig. 10) generated ~~Δ~~ΔT profiles that ~~are more similar to those observed in comparison to CLM4.5, particularly for temperatures ranging from~~ more closely resemble observed values within the range of -5~~°C to~~ °C to -15~~°C. The two~~ °C with an RMSE of 0.79 °C (Table A1) compared to CLM4.5 (RMSE of 1.46 °C in 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 UKESM1.0-LL~~) outperform their predecessors when~~ , provided better initial insulation at low snd ~~is less than 0.1~~ values and did not overestimate the insulation effect at higher snd values (lower RMSE in all 3 categories). Compared to its ~~older versions, the~~ previous version, ORCHIDEE (IPSL-CM6A-LR) exhibited a larger snow insulation effect ~~for ORCHIDEE is also improved. The most obvious improvement is observed~~ 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. Furthermore, the ~~IPSL-CM6A-LR version of ORCHIDEE used in LS3MIP, where a substantial reduction in the lack of snow insulation is seen.~~

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

~~All the selected modelsuse multi-layer snow schemes to calculate the accumulation and melting of snowpack. The multi-layer schemes show good performance in reproducing accurate *snd*. And better spatial correlation with observation in~~ study by Burke et al. (2020) also compared the insulation effect of various 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 the land-only simulations (CESM LS3MIP~~simulations indicates the snow layering schemes adequately respond to forcing~~) performed better.

~~Most models use spectral-related methods to calculate albedo. Although IPSL-CM6A-LR employs a simpler spectral-averaged albedo scheme than other land surface models, it does not have an observable impact on its tsl simulation. Albedo has a large uncertainty beyond methods employed by land models since the albedo calculation strongly relies on the simulation of snow . For example, if the coverage and thickness of snow are not well simulated in regions of low snow depth or forest area, albedo cannot be precisely calculated (Krinner et al., 2018; Menard et al., 2021). Considering snow conductivity, the Power Function could be why CNRM-CM6.1 and CNRM-ESM2.1 have a negative bias of larger than -6 °C in the SON (figure not shown).The uncertainty of CNRM-CM6.1 and CNRM-ESM2.1 soil temperatures in spring and autumn is the largest. In autumn, all models underestimate soil temperature, which proves again models are poor in simulating snowthermal insulation effect when snow depths (also snow density) are low. IPSL-CM6A-LR simulates the most stable snd, but with the Quadratic Equation, its snow thermal conductivity is strong when snd is under 0.2 m, reflected by its lower snow insulation effect than Obs. Using similar schemes of snow does not guarantee similar performances.~~

4.6 Impact of Land Model Features on Performance

The strong cold bias of CNRM models may be linked to how the models handle snow cover fraction. These models allow 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 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, 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.

Better snow and soil temperature ~~simulation are also related to other processes in land models, including soil bottom schemes~~(under snow) simulations are related not only to snow parameterization but also to other land model processes, such as soil boundary conditions. HadGEM3-GC31-LL and UKESM1.0-LL ~~have~~ had good performance in simulating tsl with their shallow soil column. Moreover, the zero-flux assumption is possibly more influential on the soil surface when used in a shallower soil column~~bottom~~; it constrains the RS of soil temperature in winter. ~~Generally speaking,~~ In general, defining a bottom flux showed better results than providing fixed values when considering the impact of bottom boundary conditions on soil temperature~~, defining a bottom flux shows better results than just giving fixed values.~~.

~~Taking account of RS and RB of summer tsl,~~ None of the models could accurately depict soil insulation. When snd was below 0.05 m, all models simulated 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. The shortage could be due to insufficient representation of the thermal offset, which is controlled by factors such as soil texture, soil moisture, and surface organic matter.

Organic layer can accumulate up to 15 cm on top of the surface of frozen soil with porosity greater than 0.95 (Boike et al., 2013) . It is an important factor in the thermal dynamics of the soil surface (Zhu et al., 2019). 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. 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. 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 ~~exhibit the best organic layer thermal insulation~~ exhibited the best soil surface insulation when considering the RS and RB of summer tsl. However, the significance of ~~organic layers continues to be underestimated. Current considerations of the hydro-thermodynamic in CLM5 and Surfex 8.0c do not improve their insulation performances.~~ 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 ~~becomes difficult to ascertain~~ is difficult to determine whether simulation errors ~~are the result of~~ 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 ~~forcings~~ forcing or coupled models were quantified and discussed.

Except in summer months, inaccurate ~~inter-annual~~ interannual variability in the simulation of soil temperature by CMIP6 models is mainly caused by deficiencies in the land surface models and is less inherited from atmospheric components. Biases in the land surface ~~model~~ models even partially compensate for the influence of air temperature biases. Similarly, ~~better~~ improved precipitation simulation does not ~~ensure~~ necessarily lead to better snow depth results ~~improve, especially~~ in winter and spring. However, in autumn, better snd was simulated with better pr in LS3MIP. Land surface models performed better when coupled to an atmospheric model (CMIP6). Balancing tuning in coupled climate models with achieving physically accurate 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.

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 ~~sites in permafrost regions~~permafrost sites. The largest ~~model biases~~ bias standard deviation disagreements of tas and tsl are ~~witnessed under~~ observed in Set Frozen (temperature lower than -5~~°C~~ °C) for both ensembles (see Table 3). These indicate a ~~weakness for~~ 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.

The deficiency of land surface models is reflected in the ~~near-surface energy transport process, particularly during winter. The primary sources are the simulated snow amount and the snow insulation effect . Land models tend to simulate lower tsl when overlying snowexists. Further improvement of parameterizationof the land model surface insulationprocess is necessary.~~ ability to simulate snow depth and/or to represent the effect of thermal insulation. Snow insulation plays a critical role in modulating soil temperatures. Updating the parameterization of snow thermal conductivity, as demonstrated in recent models such as HadGEM3-GC31-LL and UKESM1.0-LL, can enhance the insulation effect of snow. 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.

Note that the main scope of this study is 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 hydrothermal processes within the deeper soil, including thermal offset, permafrost active layers, seasonally frozen soil freezing depth, and heat and water transport. ~~Additionally, the impact of the applied bottom boundary conditions is not investigated.~~ These factors play a critical role in capturing frozen soil dynamics and ~~have to be further investigated~~ must be investigated further.

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

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

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

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. ~~Thanks to Danny Risto and~~ The authors thank Professor Duoying Ji from Beijing Normal University for providing valuable insights and support for this research. Thanks to Mittal Parmar for ~~their~~ her contribution to this research. 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.

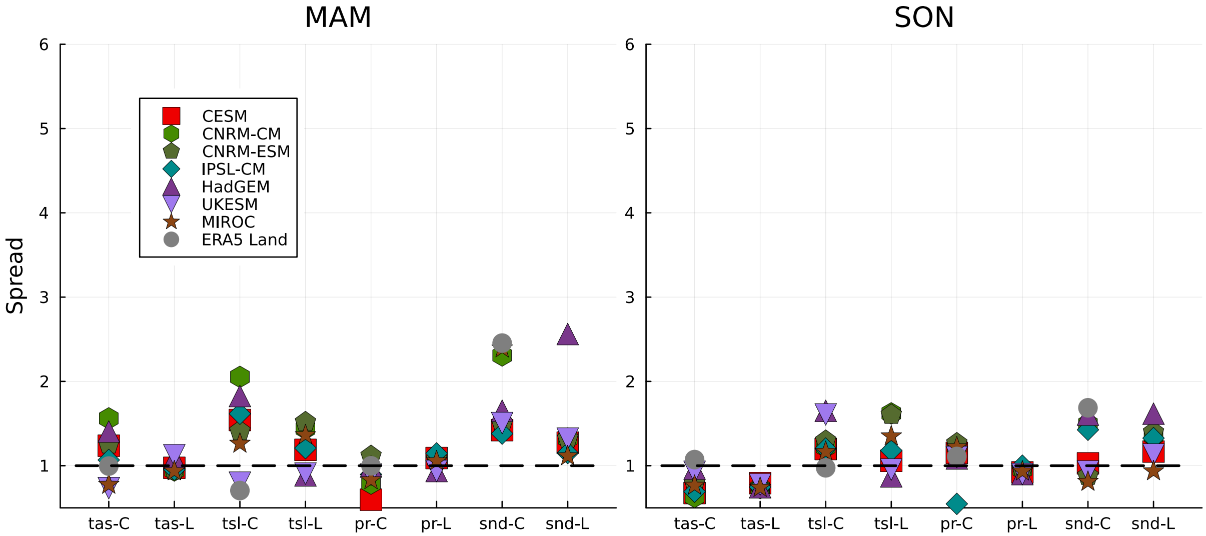

Figure A1. Relative spread (RS) of the sites-averaged climates (1985–2014) of the four variables in both ensembles and in E5LC with reference observations for all seasons. The colors indicate the models, and the x-axes show the variables and ensembles (C indicates CMIP6, and L indicates LS3MIP runs).

Appendix A: Additional Figures

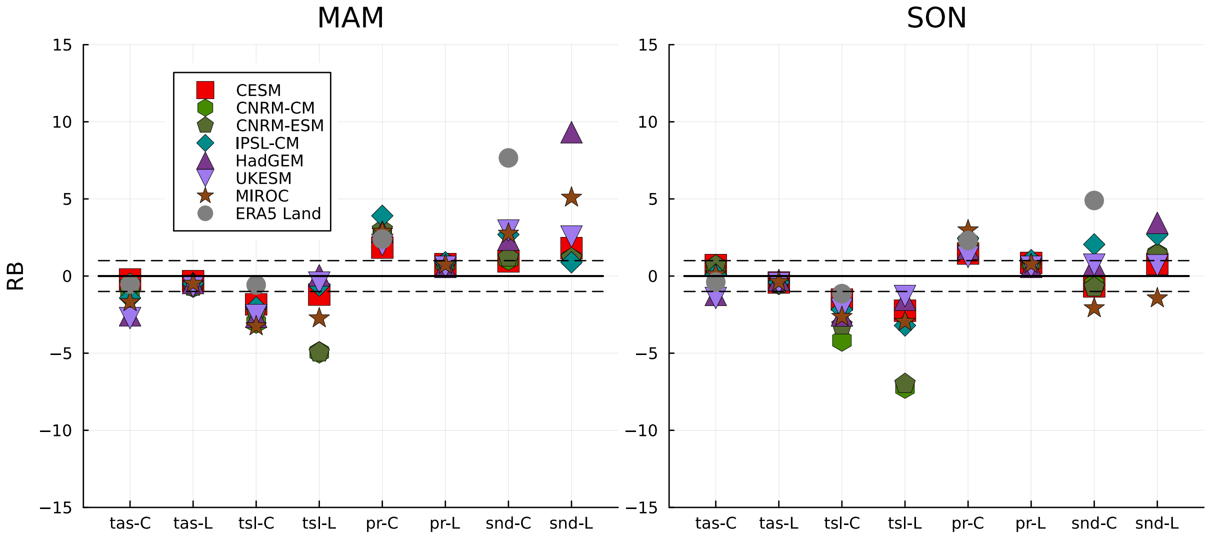

Figure A2. Same as in Fig. A1, but for relative biases (RB). The dashed lines (from -1 to 1) indicate the range of absolute median differences smaller than the observation's IQR.

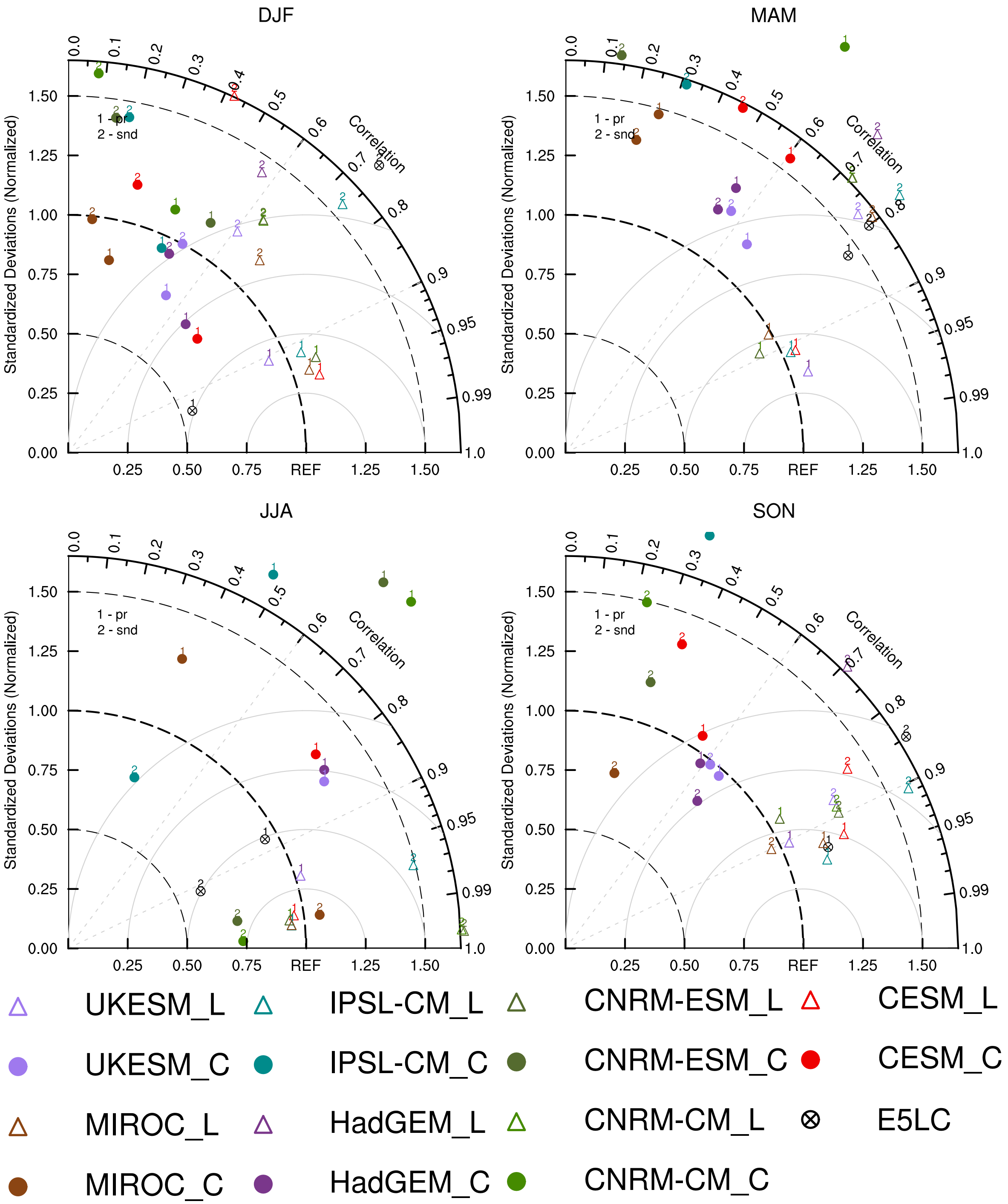

Figure A3. Taylor diagrams indicating the spatial correlation, standard deviation, and unbiased root-mean-square (RMS) difference (gray circles) for the four seasons of the simulated variables *pr*, *snd* (labeled by numbers) against observations at permafrost sites (here: sites with mean winter $tas < -25\,^{\circ}\mathrm{C}$). Normalization is applied using the standard deviations of the observations. Colors indicate different models and black crossed circles E5LC. Triangles show the results of the LS3MIP runs, and solid circles of the CMIP6 runs.

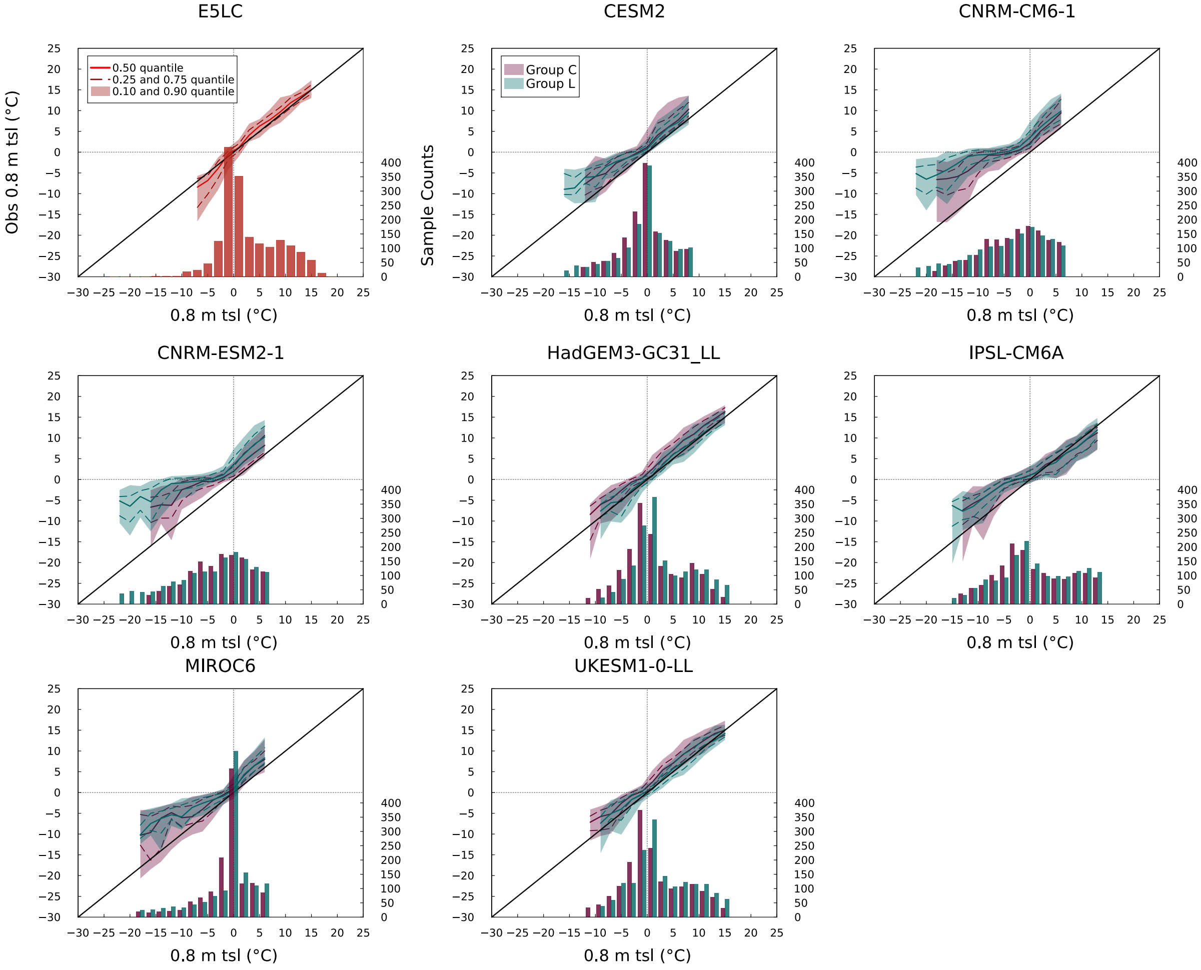

Figure A4. Quantile-Quantile (QQ) and histogram plots for soil temperature in depth 0.8 m. The plots show 30-year mean monthly 0.8 m *tsl* data of models at all sites and their observational values. The model simulation outputs are binned at 2 °C intervals. The colored solid and dash curves are the median and 1st/3rd quartiles of the corresponding observations for all data points in the temperature interval, and the shaded area is the inter-decide 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.

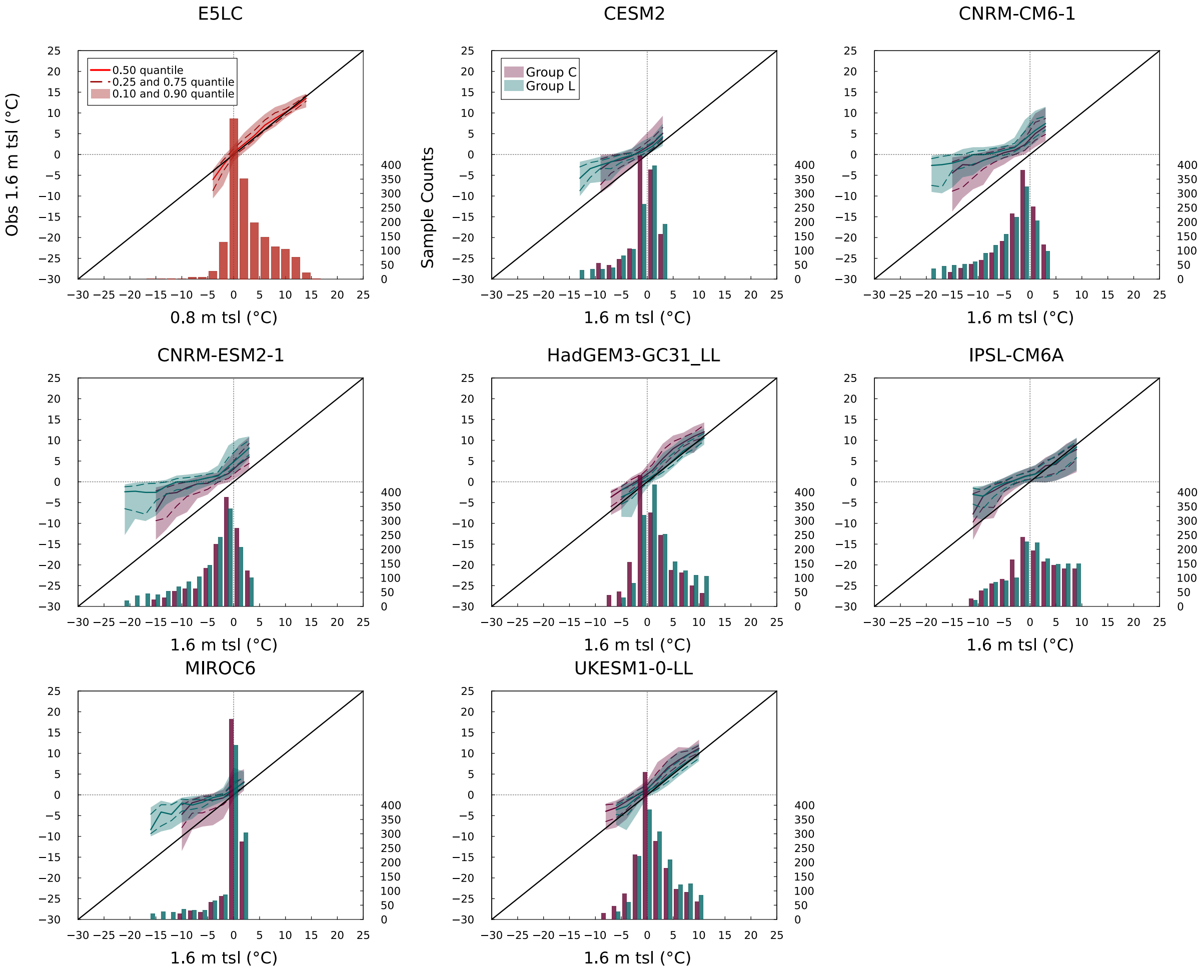

Figure A5. Same method as Fig. A4, but for soil temperature in depth 1.6 m.

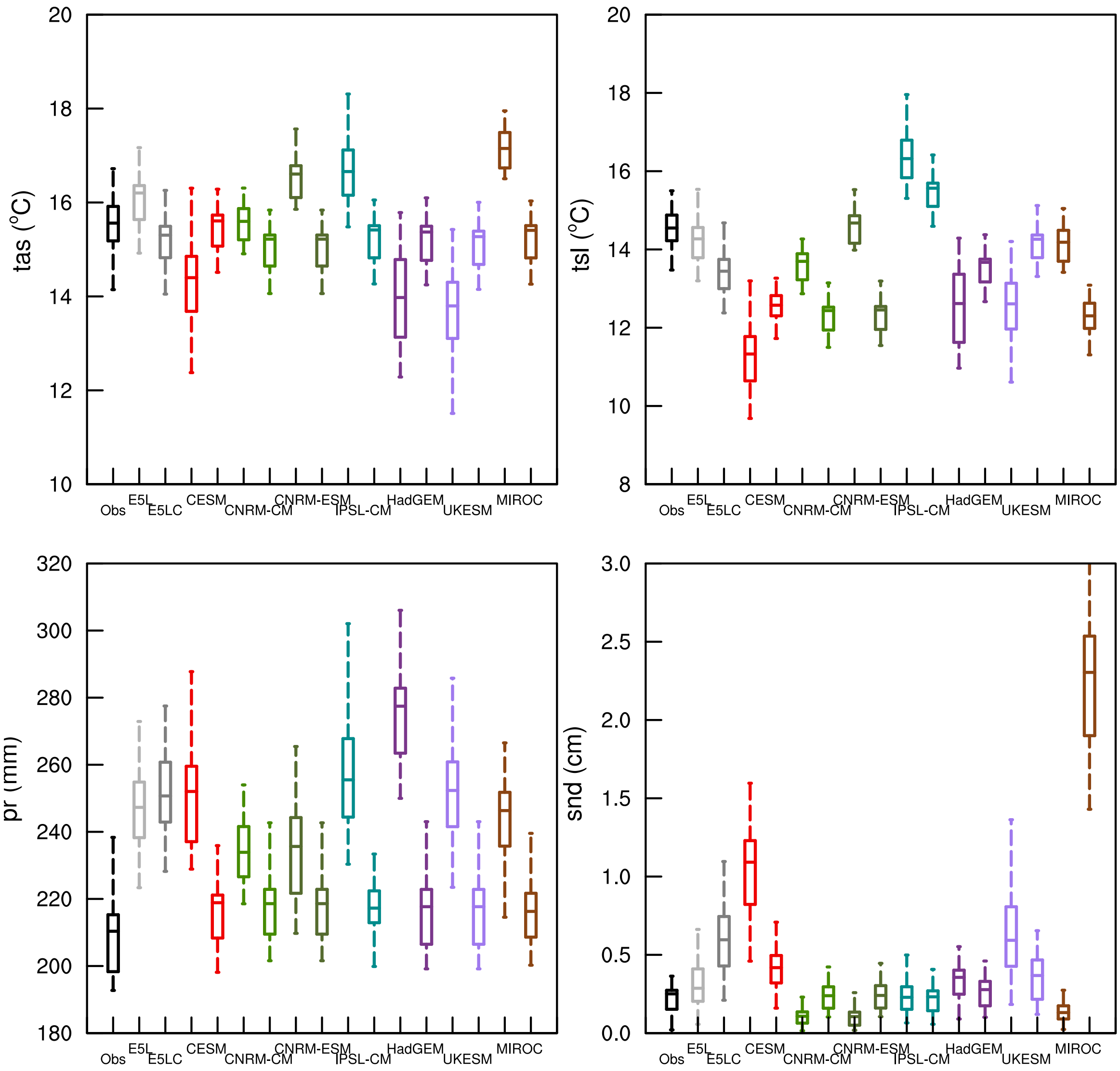

Figure A6. Sites' locations averaged JJA-climates (1985–2014) of hydrothermal variables as observed and simulated. Names on the x-axis and the colors indicate the different data sources. Model names indicate CMIP6 model output (left), and LS3MIP (right; see Table 1). Each CMIP6-LS3MIP pair shares the same color. The boxes represent the medians, first and third quartiles; the ± 1.5×IQR or the maximum and minimum values, if within the former range, are taken as the whiskers' length.

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	1.52	2.73
CESM (CMIP6)	2.09	3.97	6.22
CESM (LS3MIP)	**0.79**	2.18	4.09
CNRM-CM (CMIP6)	4.83	9.58	12.39
CNRM-CM (LS3MIP)	6.30	11.47	16.01
CNRM-ESM (CMIP6)	4.75	9.67	12.15
CNRM-ESM (LS3MIP)	6.20	11.41	15.75
IPSL-CM (CMIP6)	2.26	2.79	4.34
IPSL-CM (LS3MIP)	1.56	3.25	3.94
HadGEM (CMIP6)	1.09	**1.05**	**2.33**
HadGEM (LS3MIP)	1.26	1.43	**1.57**
UKESM (CMIP6)	**0.71**	**1.12**	2.56
UKESM (LS3MIP)	1.20	**1.16**	**2.15**
MIROC (CMIP6)	1.11	2.14	5.41
MIROC (LS3MIP)	**0.79**	1.72	4.32

References

Abbott, B. W. and Jones, J. B.: Permafrost collapse alters soil carbon stocks, respiration, <span style="font-variant:small-caps;">CH</span>$_4$, 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.

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.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V., Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li, W., Marke, T., Marks, D., Ménégoz, M., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T. G., Stacke, T., Strasser, U., Svenson, S., Turkov, D., Wang, T., Wever, N., Yuan, H., Zhou, W., and Zhu, D.: ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, Geoscientific Model Development, 11, 5027–5049, https://doi.org/10.5194/gmd-11-5027-2018, 2018.

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., Slater, A. G., Tomas, R. A., Holland, M. M., and Deser, C.: Accelerated Arctic land warming and permafrost degradation during rapid sea ice loss, Geophysical Research Letters, 35, 2008GL033 985, https://doi.org/10.1029/2008GL033985, 2008.

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.

Miner, K. R., Turetsky, M. R., Malina, E., Bartsch, A., Tamminen, J., McGuire, A. D., Fix, A., Sweeney, C., Elder, C. D., and Miller, C. E.: Permafrost carbon emissions in a changing Arctic, Nature Reviews Earth & Environment, 3, 55–67, https://doi.org/10.1038/s43017-021-00230-3, 2022.

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, M. W. and Riseborough, D. W.: Climate and the limits of permafrost: a zonal analysis, Permafrost and Periglacial Processes, 13, 1–15, https://doi.org/10.1002/ppp.410, 2002.

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.

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.