

Seasonal soil freeze/thaw variability across North America via ensemble land surface modeling

Mahsa Moradi, Eunsang Cho, Jennifer M. Jacobs, Carrie M. Vuyovich



PII: S0165-232X(23)00036-8

DOI: <https://doi.org/10.1016/j.coldregions.2023.103806>

Reference: COLTEC 103806

To appear in: *Cold Regions Science and Technology*

Received date: 24 June 2022

Revised date: 30 December 2022

Accepted date: 20 February 2023

Please cite this article as: M. Moradi, E. Cho, J.M. Jacobs, et al., Seasonal soil freeze/thaw variability across North America via ensemble land surface modeling, *Cold Regions Science and Technology* (2023), <https://doi.org/10.1016/j.coldregions.2023.103806>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Seasonal Soil Freeze/Thaw Variability Across North America via Ensemble Land Surface Modeling**

**Mahsa Moradi<sup>1,2,\*</sup>, Eunsang Cho<sup>3,4</sup>, Jennifer M. Jacobs<sup>1,2</sup>, Carrie M. Vuyovich<sup>3</sup>**

<sup>1</sup>Department of Civil and Environmental Engineering, University of New Hampshire, Durham, NH, USA

<sup>2</sup>Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA

<sup>3</sup>Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

<sup>4</sup>Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

Journal Pre-proof

---

\* \*\*\*\*\*Corresponding author: Mahsa Moradi, Email: [mm1631@wildcats.unh.edu](mailto:mm1631@wildcats.unh.edu), Phone: 6037676584

**Abstract**

Land surface modeling provides the opportunity to investigate winter soil processes such as soil freezing and thawing over large domains. However, the variability in simulated winter soil characteristics among land surface models and forcing datasets is not well understood. In this study, a nine-member ensemble was employed to characterize the spatial and inter-annual variability of three winter soil characteristics, annual number of frozen days, annual minimum temperature and annual number of freeze-thaw cycles over North America during the water years 2010-2016. The ensemble, including three land surface models (JULES, Noah2.7.1 and Noah-MP) and three forcing datasets (ECMWF, GDAS and MERRA2), was developed through the Snow Ensemble Uncertainty Project (SEUP). In many regions, there was remarkably good agreement across the ensemble for the winter soil temperatures. However, the variability among the ensemble's annual number of frozen days as quantified with standard deviation exceeded 150 days at the northern Pacific coastline. While the ensemble members' annual minimum temperature differences were typically less than 3 °C, the differences exceeded 6 °C north of 50 °N. The ensemble members generally agreed within one to three freeze-thaw cycles in mid latitude regions, Alaska, and the south and west coasts of the USA. High variability among the ensemble, more than six freeze-thaw cycles, occurred in the Great Plains, northern Pacific coastline, and along the Appalachian Mountains. Differences in winter soil temperature characteristics were more apparent among the LSMs rather than the meteorological forcing datasets. Except for maritime regions, the Noah2.7.1 members had the fewest annual number of frozen days and the least number of freeze-thaw cycles. Noah2.7.1 also had the coldest annual minimum temperatures except for ephemeral regions. Comparisons between in-situ observations and the SEUP estimates of winter soil characteristics revealed that the modeled frozen period was much longer than observed, that the modeled annual minimum temperatures were much colder than observed, and the modeled freeze-thaw cycles occurred more frequently than observed. Excluding the high latitude sites,

the observed frozen period is less than two months with the minimum temperature above  $-5^{\circ}\text{C}$  at most of the studied sites, while the ensemble members simulated, on average, a four month frozen period with  $-10^{\circ}\text{C}$  minimum temperature. Errors in capturing snow during the accumulation period appear to impact differences in modeled versus observed soil temperature throughout the entire winter.

**Keywords:** Soil freeze/thaw, Winter soil characteristics uncertainty, Land surface modeling, Snow ensemble uncertainty project

## 1 Introduction

The frequency and severity of soil freeze and thaw processes have thermodynamic, hydrological, geochemical and ecological significance for the Earth's land-atmosphere system. Winter soil temperatures modulate regional climate and land-atmospheric boundary processes by controlling surface and subsurface energy fluxes (Zhang and Sun, 2011; Guo et al., 2011; Yang et al., 2019), evapotranspiration (Zhang et al., 2011; Liu et al., 2019) and albedo (Guo et al., 2011). From a hydrologic perspective, the soil freeze/thaw state strongly alters partitioning of infiltration and overland flow due to the reduced hydraulic conductivity of frozen soil (Kane 1980; Seyfried and Murdock, 1997; Cherkauer and Lettenmaier, 1999; McCauley et al., 2002; Roy et al., 2021), and regulates flood timing and intensity over the course of winter and early spring (Bayard et al., 2005; Niu and Yang, 2006). Also, during cold seasons, the soil temperature gradient induces soil water migration to colder layers resulting in the redistribution of soil moisture (Perfect and Williams 1980). Minimum soil temperature and FT state also contribute to soil biotic activities (Kreyling et al., 2012; Kreyling et al., 2019), root mortality (Repo et al., 2014), ecosystem diversity and productivity (Euskirchen et al., 2006, Kreyling et

al., 2008), plant species compositions (Joseph and Henry, 2008), nutrient cycling (Wang et al., 2017), and carbon dynamics and mineralization (Vestgarden et al., 2009; Fuss et al., 2016).

Despite the importance of winter soil temperature, very few reliable long-term in-situ soil temperature observations are available and even fewer soil temperature networks exist (Xia et al., 2013). In addition to the limited spatial distribution of soil temperature observations, temporal inconsistency of these point measurements restricts their application to investigations of soil thermal processes at large spatial scales. Even though remote sensing techniques offer an alternative to in-situ point measurements that addresses the spatial scale issues, remote sensing of soil state is challenging and complex due to the presence of vegetation and snow layers on the ground (Kim et al., 2019, Johnston et al., 2021). Therefore, model-derived estimations of soil temperature are often employed as alternatives when soil temperature data with large spatiotemporal coverage is required. Previous studies used land surface model (LSM) estimates of soil temperature to study the severity of freeze-thaw (FT) cycles (Xia et al., 2011), sub-pixel variability of FT remote sensing products (Johnston et al., 2021), improvement of the synthetic brightness temperatures (Han et al., 2014), carbon cycles uncertainty (Wieder et al., 2018) and carbon and nitrogen dynamics (Bonan et al, 2013; Shi et al., 2016).

Land surface models have shown promise for data assimilation frameworks to achieve more skillful estimations of soil states. For example, Johnston et al. (2021) used an ensemble of LSMs to investigate the relationship among satellite-based FT products and ensemble mean soil temperatures. They concluded that a new FT classification framework can be achieved through fusion of high-resolution LSM derived temperature data and remote sensing products. However, the uncertainties associated with

LSMs' estimations could impact the estimates' skills of data assimilation procedures or multi-model ensembles (Liu and Gupta, 2007; Bohn et al., 2010; Dumedah et al., 2014). Thus, understanding and quantifying such uncertainties is indispensable before any application of LSMs.

The commonly used multi-model approach provides an opportunity to explore the variabilities in model simulations across a range of models with different structures and forcing input (Bohn et al., 2010; Mudryk et al., 2015; Kumar et al., 2017; Vuyovich et al., 2019; Kim et al., 2021). Despite the value of this method in characterizing the uncertainty in hydrologic simulations and climate modeling, to the best of our knowledge, there has been no attempt to characterize LSMs derived winter soil characteristics and their uncertainties among different LSMs over North America. Previous studies have typically evaluated an individual land surface model's ability to simulate a soil's winter thermodynamic regime at various spatial scales (Zhu and Liang, 2005; Godfrey and Stensrud, 2008; Xia et al., 2011; Xia et al., 2013; Li et al., 2021) but have not provided insights as to the variabilities among different models. Model performance is linked to forcing data and other state variables including snow characteristics which substantially modulate winter soil processes (Stieglitz et al., 2003; Zhang, 2005; Lawrence and Slater, 2010; Wang et al., 2016; Li et al., 2021). Such state variables and the complex interaction of soil and snowpack vary among different models, depending upon their physics and snow schemes. Previous snow models intercomparisons, such as SnowMIP (Etchevers et al., 2002) and ESM-SnowMIP (Krinner et al., 2018), and SEUP (Kim et al., 2021) have shown that snow water equivalent (SWE), snow depth, snow albedo, and snow cover duration can vary significantly among the models, depending on the locations, yet the ensemble mean may closely follow

the observations. Thus, an understanding of similarities and differences among models is as necessary as comparisons between models and observations.

Motivated by the current gap in understanding LSMs' performance and consistency in cold season modeled soil characteristics, this study seeks to quantify the variabilities in such characteristics over North America through a multi-model ensemble evaluation across models and in comparison to observations. The cold season soil characteristics used here refer to the annual number of frozen days, the annual number of FT cycles and the annual minimum temperature. This study employs an ensemble of three land surface models (Noah 2.7.1, Noah with Multi-Parameterization [Noah-MP], Joint UK Land Environment Simulator [JULES]) and three meteorological forcings (European Centre for Medium-Range Weather Forecasts [ECMWF], Global Data Assimilation System [GDAS], and Modern-Era Retrospective Analysis for Research and Applications, version 2 [MERRA2]), developed through Snow Ensemble Uncertainty Project (SEUP) (Kim et al., 2021), to provide an estimate of variability and uncertainty in modeled winter soil temperature characteristics. In addition to ensemble evaluation, we aim to assess the performance of the ensemble members with respect to in-situ observations. This study also explores how models' snow simulations impact soil temperature estimates through the course of winter. Therefore, the primary goal is to address the following questions:

- Where, when, and how do the SEUP ensemble members agree and disagree on the winter characteristics of soil and experience high inter-annual variability?
- How do the model derived soil metrics compare with observations?

- How does the temporal evolution of modeled soil temperature relate to snow water equivalent (SWE) accumulation and ablation in the SEUP?

## 2 Data and Methods

### 2.1 Snow Ensemble Uncertainty Project and Ensemble Configuration

Snow Ensemble Uncertainty Project (SEUP) was designed to quantify SWE uncertainty across North America using an ensemble-based land surface modeling approach (Kim et al, 2021). The original SEUP ensemble consists of 12 members, generated by the combination of four land surface models (JULES, Noah 2.7.1, Noah-MP3.6 and Catchment version LSM [CLSM-F2.5]) and three meteorological forcing datasets (ECMWF, GDAS and MERRA2). In this study, CLSM-F2.5 model is excluded from the original SEUP ensemble because its 3 soil layers (0-2, 2-100 and 100-200 cm) configuration is not consistent with the other models (described in section 2.2). The selected land surface models and forcing datasets have been employed at key operational centers and systems, and they are able to provide a valuable basis to study uncertainty associated with winter temperature characteristics of soil.

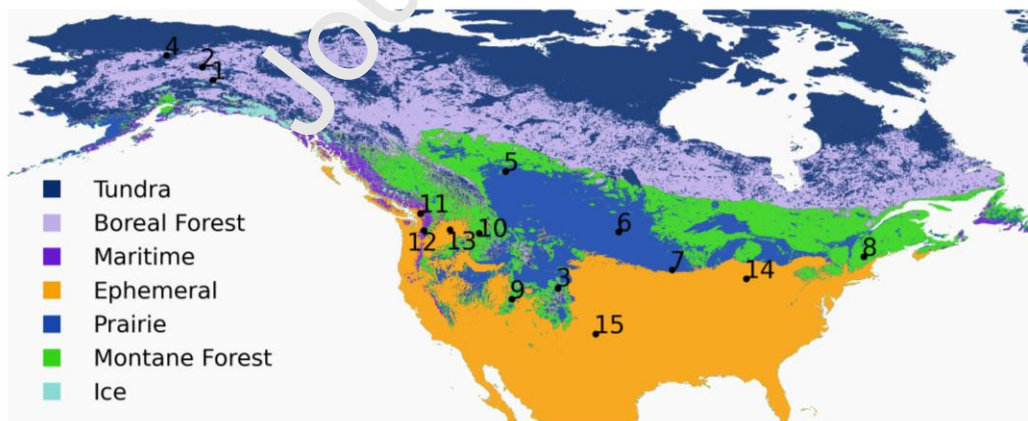


Figure 1. Snow Classes over Snow Ensemble Uncertainty Project (SEUP) domain. Black circles and their associated numbers illustrate location of the in-situ sites in this study. These sites are : 1-Granite Creek, (AK) 2-Upper Nome Creek (AK), 3-Arapaho Ridge (CO), 4-Gobblers Knob(AK), 5-Mundare AGDM (Canada, Alberta),



6-Streeter (ND), 7-South Fork (IA), 8-Hubbard Brook (NH), 9-Beaver Dams (UT), 10-Crater Meadows (ID), 11-MF Nooksack (WA), 12-Cayuse Pass (WA), 13-Lind (WA), 14-St Joseph's (IN), 15-Bushland (TX)

The SEUP study domain is North America, covering from 4.875°N to 71.875°N and 168.625°W to 51.875°W on a 0.05° equidistant cylindrical grid (Figure 1). The models were run with 3 hour time steps from 2000 to 2017. The first nine years (2000-2009) were served as spin-up time for LSM initialization and the rest were used for analysis in this study.

### 2.1.1 Land Surface Models

Originally developed from the Met Office Surface Exchange Scheme (MOSES) (Cox et al., 1999), the JULES (Best et al., 2011; Clark et al., 2011) is a key model of the Met Office's modeling infrastructure and NERC's Earth System Modeling Strategy. This community land model is widely used in the climate and weather forecast models both as a standalone model and as the land surface component. JULES typically has 4 soil layers. To calculate subsurface temperature in JULES model, one dimensional heat diffusion equations are used including the heat fluxes of solid-liquid phase transition of water. The Dharssi's equation (Dharssi et al., 2009) was used to describe soil thermal conductivity. For the SEUP ensemble, the model was run with depths of 10, 25, 65, and 100 cm. JULES's multi-layer snow scheme and diagnostic snow albedo were used in SEUP to simulate snow. This snow scheme estimates snow density, ice content and liquid water content for each layer. Using snow density, JULES provides snow thermal conductivity estimates (Best et al., 2011). JULES assumes freezing temperature equal to 273.15°K.

The community Noah land surface model (Ek et al. 2003), developed based on the Oregon State University (OSU) LSM (Mahrt and Pan 1984), was first executed in National Centers for Environmental Prediction (NCEP) Eta Data Assimilation System (EDAS), followed by implementation in the NCEP Global Forecast System (GFS) (Mitchell et al. 2005) and later in Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010). Since then, the Noah model has been widely employed as a

numerical weather prediction (NWP) and climate model for operational and research applications. For example, it has been adopted at the NCEP as NWP model, in the North America Land Data Assimilation System (NLDAS) (Mitchell et al., 2004) and the Land Information System (Peters-Lidard et al., 2007). The SEUP ensemble employs version 2.7.1 of Noah. Structurally, the model consists of four soil layers (here with depths of 10, 30, 60, and 100 cm from top to bottom, resulting in a 2 meters soil column), a single canopy layer, and a single layer snowpack. Noah model employs Jordan's equation (Jordan 1991) for snow thermal conductivity. Frozen ground processes are described by Koren et al. (1999). Soil moisture and soil temperature are calculated from the one dimensional diffusive form of Richard's equation and the heat diffusion equation respectively. The soil heat flux formulation of Noah model takes account of the heat flux through thin patchy snow cover since the model allows exposed ground for the snow depth below a certain threshold. More detailed descriptions of the governing equations and the parameterizations of Noah model can be found in Ek et al. (2003) and Koren et al. (1999). Freezing temperature in Noah model is equal to 273.15°K.

Originating from the Noah LSM, Noah-MP benefits from multiple parameterization options for the key land-atmosphere interaction processes (Niu et al., 2011). In addition to its global applicability, it currently is in the core of National Water Model (NWM) and RAL/HAP WRF convection-permitting climate modeling efforts. While the depths of Noah soil layers are retained, its bulk snow layer is improved to a multiple layer snowpack in Noah-MP model. Depending on the total depth of snow, up to three snow layers can be present on the top of Noah-MP soil column. In the SEUP ensemble, the Noah-MP3.6 was run with the default options. For the frozen soil processes, the default scheme is Niu and Yang (2006) which employs a more general form of freezing-point depression equation in comparison to the Koren et al. (1999) method. In this scheme, freezing temperature is assumed to be 273.16°K. Yen's scheme (Yen, 1965) and Peters-Lidard's scheme (Peters-Lidard et al., 1998) were used

to describe snow and soil thermal conductivity, respectively. Details on Noah-MP physics and schemes can be found in Niu et al. (2011) and Niu and Yang (2006).

Although all three models use the same principle, surface energy balance equation and one dimensional heat diffusion equation, to simulate soil temperature, they compute the components of the energy balance equation differently (Best et al., 2011; Ek et al., 2003; Niu et al., 2011). The models also employ different approaches to describe soil freezing processes and supercooled liquid water in soil (Cox et al., 1999; Koren et al., 1999; Niu and Yang, 2006). Moreover, representation of frozen ground characteristics including thermal conductivity and permeability differ. The other key difference among the SEUP ensemble models is the representation of snow processes including partitioning precipitation into snow and rain, snowpack layering, thermal conductivity and albedo.

### 2.1.2 Forcing Datasets

The European Center for Medium-Range Weather Forecasts (ECMWF; Molteni et al., 1996) global atmospheric reanalysis is produced by assimilating vast amounts of surface observations and high resolution satellite datasets. Gridded meteorological data from the ECMWF-Integrated Forecast System [IFS], used in the SEUP, are generated by assimilating available atmospheric observations every 12 hours into a forecast model with surface meteorological fields.

Developed at the Environmental Modeling Center [EMC] of the National Oceanic and Atmospheric Administration [NOAA] / NCEP, the Global Data Assimilation System [GDAS] (Derber et al., 1991) assimilates various types of observations such as ground observations, balloon data, wind profiler data, airborne data and satellite-derived data into an atmospheric model that can be employed for initializing weather forecast systems.

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA2; Gelaro et al., 2017; Molod et al., 2015) produced by NASA's Global Modeling and Assimilation

Office [GMAO] is a state-of-the-art global atmospheric reanalysis providing data. To produce MERRA2, Goddard Earth Observing System [GEOS-5] data assimilation system is used to assimilate in-situ and satellite observations including modern hyperspectral radiance, microwave observations and GPS-Radio Occultation datasets. Further details and descriptions on the data assimilation system and data production procedures can be found at Gelaro et al. (2017).

All the three original coarse resolution meteorological datasets were downscaled to a 5 km grid, first, prior to forcing the SEUP land surface models. Details on the downscaling methods for the forcing variables can be found in Kim et al. (2021).

## 2.2 Ancillary Datasets

In this study, the Liston and Sturm (2021) snow classification with  $2.5 \text{ arcmin} \times 2.5 \text{ arcmin}$  (~5 km) spatial resolution is used to evaluate variability in cold seasons soil characteristics of SEUP ensemble at a continental scale across different snow classes. The Liston and Sturm (2021) seasonal snow cover classification system employs a binary decision tree to link air temperature, precipitation, and wind speed to stratigraphic and textural properties of snow covers (e.g. sequence of snow layers, thickness, density, etc.) in order to categorize terrestrial snow into seven classes: tundra, boreal forest, maritime, ephemeral, prairie, montane forest, and ice.

To evaluate the relationship between SEUP estimation of soil temperature and observed soil temperature as well as impacts of snow cover (if available), 15 stations are selected from North Dakota Agricultural Weather Network (NDAWN), Alberta Climate Information Service (ACIS), United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) network and Soil Climate Analysis Network (SCAN) (Table 1). These 15 stations were chosen to provide a reasonable spatial coverage over different snow classes (Figure 1). Hourly observations of soil temperature at 5 cm below the surface, air temperature, SWE or snow depth

(if available) at these stations are obtained. To control consistency of the in-situ measurements at all stations, years with more than 20 days of missing data were excluded from this study (Table 1).

Table 1, Details of the 15 sites, chosen for comparison against SEUP simulations

Station Name	Latitude (°N)	Longitude (°W)	State	Network	Elevation (m)	Data record used in this study (WY)	Excluded years	Snow class	Snow data availability
1.Granite Creek	63.95	145.4	AK	SNOTEL	377.95	2010-2017	-	Tundra	Yes
2.Upper Nome Creek	42.65	146.6	AK	SNOTEL	768.1	2010-2017	-	Tundra	Yes
3.Arapaho Ridge	40.35	106.38	CO	SNOTEL	3342.7	2010-2017	2011	Forest	Yes
4.Gobblers Knob	66.75	150.67	AK	SNOTEL	618.744	2010-2017	2012	Forest	Yes
5.Mundare AGDM	53.56	112.29	Canada, Alberta	ACIS	690	2010-2017	-	Prairie	Yes
6.Streeter	46.72	99.46	ND	NDAWN	180.59	2015-2017	-	Prairie	No
7.South Fork	42.426	93.417	IA			2016-2017	-	Prairie	No
8.Hubbard Brook	43.93	71.72	NH	SCAN	451.1	2010-2017	2013, 2014	Montane Forest	Yes
9.Beaver Dams	39.14	111.56	UT	SNOTEL	2435.35	2010-2017	-	Montane Forest	Yes
10.Crater Meadows	46.56	115.29	ID	SNOTEL	1816.6	2012-2017	-	Forest	Yes
11.MF Nooksack	48.82	121.93	WA	SNOTEL	1514.86	2013-2017	-	Maritime	Yes
12.Cayuse Pass	46.87	125.53	WA	SNOTEL	1597.15	2012-2017	2016	Maritime	Yes
13.Lind	47	118.57	WA	SCAN	499.87	2010-2017	-	Ephemeral	No
14.St Joseph's	41.449	85.011	IN			2016-2017	-	Ephemeral	No
15.Bushland	35.17	102.1	TX	SCAN	1164.34	2010-2017	-	Ephemeral	No

## 2.3 Methods

The annual number of frozen days, annual minimum soil temperature, and annual number of FT cycles are the three soil metrics used in this study to represent soil processes in the cold season. These variables are selected due to their importance for ecological and hydrological processes (Kreyling et al., 2019; Repo et al., 2014). All metrics of interest are calculated on a cell-by-cell basis using the 6 AM 3-hourly temperature in the top 10 cm of soil. The 6 AM period was selected for analysis because the top soil layer undergoes nighttime cooling prior to daytime warming and thus is cold enough to represent freezing processes in soil. It is also coincident with the Soil Moisture Active Passive (SMAP) descending overpass which allows the comparison between the two datasets in future work. Hereafter,

the term soil temperature refers to the 6AM soil temperature for the top soil layer. The use of the 6 AM period does not identify diurnal FT cycles. Thus, the study's FT cycles counts are lower than if diurnal cycles were counted. Water years (WYs), October 1 to September 30, are used in this study to obtain the annual soil metrics.

Freezing temperature ( $T_{frz}$ ), used as the threshold between frozen and thawed soil states, was set to 0.01°C, which is the temperature utilized by Noah-MP model in its freezing scheme. Detecting freeze state based on any  $T_{frz}$  less than 0.01°C noticeably changes the number of frozen days and FT cycles of Noah-MP soil layer. In contrast, Noah2.7.1 and JULES set  $T_{frz}$  to 0°C. Our preliminary tests indicated that Noah2.7.1 and JULES freeze state estimates do not change when  $T_{frz} = 0.01^\circ\text{C}$  (Figures S-1 and S-2).

The annual number of frozen days (FD) is total number of days in each WY when the soil temperature was below the freezing temperature.

$$FD = \sum_{i=1}^n \begin{cases} 1 & \text{if } T_i < T_{frz} \\ 0 & \text{if } T_i \geq T_{frz} \end{cases}$$

where  $T_i$  is soil temperature on day  $i$  and  $n$  is the number of days for each WY.

The annual minimum soil temperature was determined for each WY and cell. The annual number of FT cycles is the number of times when the soil temperature cools from a temperature above the freezing temperature to below the freezing temperature. Therefore, for each cell:

$$FT = \sum_{i=2}^n \begin{cases} 1 & \text{if } T_i < T_{frz} \wedge T_{i-1} \geq T_{frz} \\ 0 & \text{else} \end{cases}$$

where  $T_i$  and  $T_{i-1}$  are soil temperature at day  $i$  and  $i-1$ , respectively.

The annual ensemble mean for each year is the average of the variable of interest for the nine ensemble members on a cell-by-cell basis. The overall mean was calculated by averaging of the annual

ensemble means over the WYs 2010-2016. Because SEUP data is only available until 30 May 2017, this partial year was excluded from mean and standard deviation analysis to eliminate its potential impacts on the statistics of high latitude regions. The deviation from the mean for each year is the annual ensemble mean for that year minus overall mean. Similarly, the annual standard deviation was calculated over the nine ensemble members for each WY. These three statistics are used to assess overall variability of soil characteristics over the ensemble.

The snow class-averaged soil variable is the average of the soil variable of interest over the entire area with a specific snow regime, calculated separately for each ensemble member and year. This variable was employed to investigate the uncertainty by different snow classes.

### 3 Results

#### 3.1 Ensemble Mean and Standard Deviation

##### 3.1.1 Annual Number of Frozen Days

Figure 2-A shows the ensemble mean of the annual number of frozen days generally increases with latitude. Not surprisingly, latitude is a key underlying factor in SEUP estimations of frozen days. However, latitude is not the only driver. The pattern of the ensemble mean at the regions with latitude greater than 50°N suggests that there are other additional factors impacting this pattern. The Rocky Mountain region can experience up to 260 frozen days annually while other regions at similar latitudes usually have less than 160 frozen days. The northern Pacific coastline's ensemble mean, there are fewer than 120 frozen days, and the USA west coast's ensemble mean, frozen days less than 40 days, are noticeably fewer than other regions at the same latitude.

Figures 2-B to 2-H display annual differences from the overall ensemble mean of frozen days. Deviations from the average have considerable spatial coherence particularly in the central USA. The



Northern Pacific coastline has considerable year-to-year variability. High interannual variations also occur around the Great Lakes and the northeast of the USA. The southern Canadian prairies (near the USA and Canadian border) experienced many more frozen days in 2013 and 2014. For other years, that region is fairly consistent.

Figure 2-I presents the average variability across the model ensemble quantified by the standard deviation of the annual number of frozen days. Over much of the study domain the variations among the ensemble members are typically on the order of one month. In contrast, there is considerably more

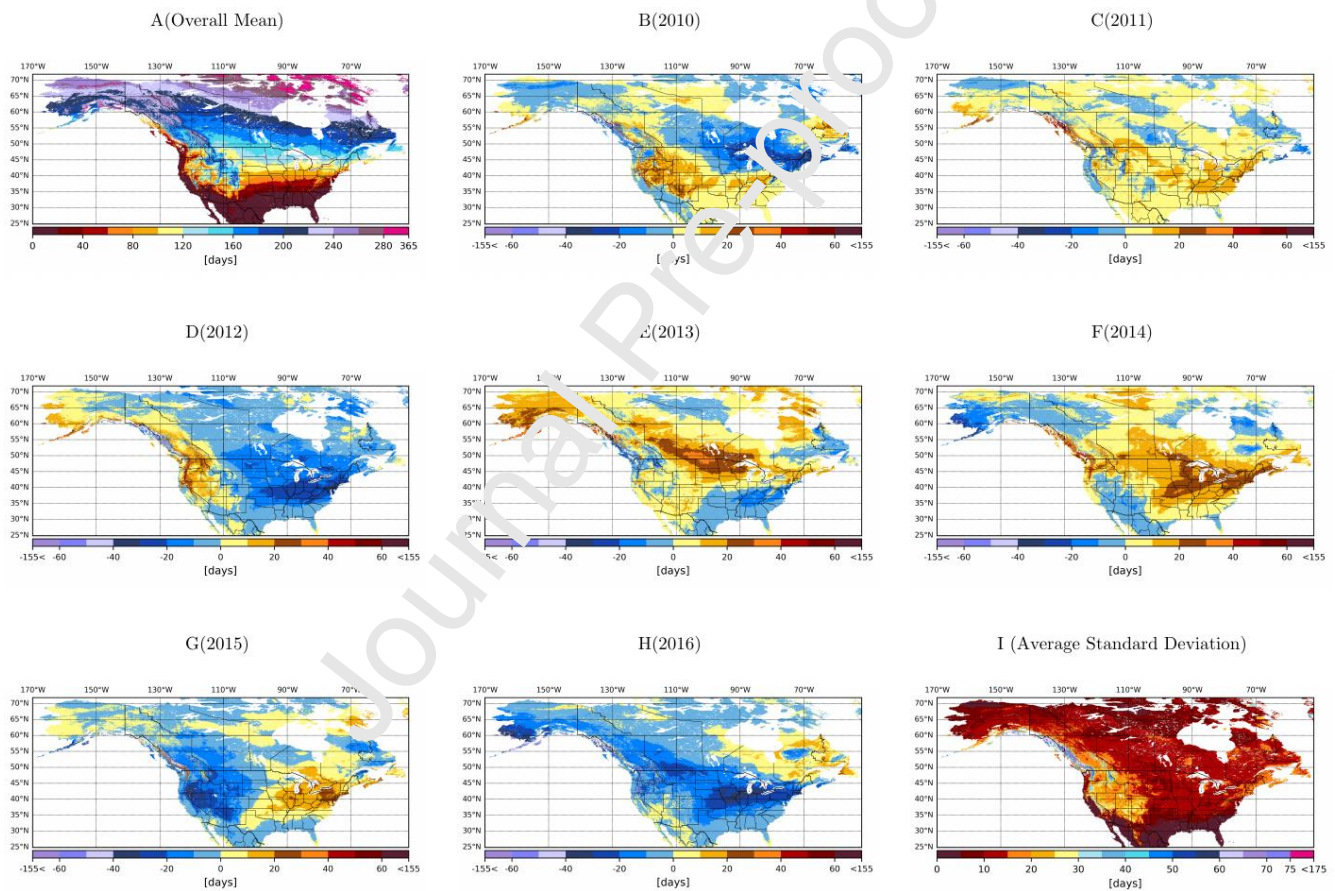


Figure 2. Spatial distribution of overall ensemble mean (A), differences between annual ensemble mean and overall ensemble mean (B-H), and average standard deviation (I) of number of frozen days computed using 3-hourly soil temperature at 6AM over 9 ensemble members for each year from WT 2010-2016

disagreement in the northern Pacific coastline, the northern Rocky Mountains, and the prairie regions of Alaska. In the Northern Pacific coastline, the standard deviation can be as high as 150 days.



Coefficients of variation (standard deviation divided by mean) exceed one in this region indicating a lack of consensus among the ensemble members for this area. The northern Pacific coastline had the highest and lowest agreement among ensemble members in 2011 and 2010, respectively (Figure S-3). A localized area in the Rocky Mountains also has a high standard deviation ( $\approx 60$  days) across all the years. However, considering the lengthy frozen period for this region ( $\approx 200$  days), this high standard deviation is less notable than the Pacific regions. In general, the variability among the ensemble members is higher for the western regions of the USA as compared with the rest of the study domain.

### 3.1.2 Annual Minimum Temperature

Figure 3-A illustrates the overall ensemble mean of annual minimum temperature, averaged over the seven WYs (2010-2016). In a typical year, the temperature of the top soil layer never freezes in the southern coastal regions of North America (latitude  $< 35^{\circ}\text{N}$ ), as well as the west coast of the USA (latitude  $< 40^{\circ}\text{N}$ ). The coldest minimum temperatures,  $T_{\min}$  colder than  $-15^{\circ}\text{C}$ , are found primarily in the northern Great Plains, the Arctic regions above  $66.5^{\circ}\text{N}$  and along the Brooks Range in Alaska.

Figures 3-B to 3-H show the deviation of annual ensemble mean minimum temperature from the overall ensemble mean of the minimum temperature. Comparing these figures reveals that most of the study domain experiences at least one year that is at least  $2^{\circ}\text{C}$  colder or warmer than usual. There are a few regions including California that have low interannual variability of annual minimum temperature. The figures also suggest that the spatial extent of the areas with a very low minimum soil temperature is not consistent across years. There are years, such as 2014, in which an extended area across the northern Great Plains and Arctic regions where  $T_{\min}$  is colder than  $-15^{\circ}\text{C}$ . For other years, such as 2011, only a small portion of these regions experienced very low minimum temperatures. In fact, minimum temperatures in most of the north were between  $-5^{\circ}\text{C}$  and  $-10^{\circ}\text{C}$ .

Figure 3-I shows the average standard deviation of annual minimum temperature across nine ensemble members. For much of the CONUS, the variability is less than 3°C. Higher variability in minimum soil temperature over the ensemble members occurs north of 50°N. The maximum variability, larger than 6°C, is found in these high latitude regions but the locations and the spatial extent of these extremes vary among the years (Figure S-4). For example, there was large disagreement among the ensemble members' annual minimum temperature for much of Canadian Prairies in 2010 (Figure S-4). For other years, very limited areas in this region experienced such a high variability. For regions south of 50°N, the highest standard deviation occurs in the southern Rocky Mountains, regions around the Great Lakes and the northeastern USA. Disagreement among ensemble members also occurred in an extended region in 2011 over the central parts of the USA.

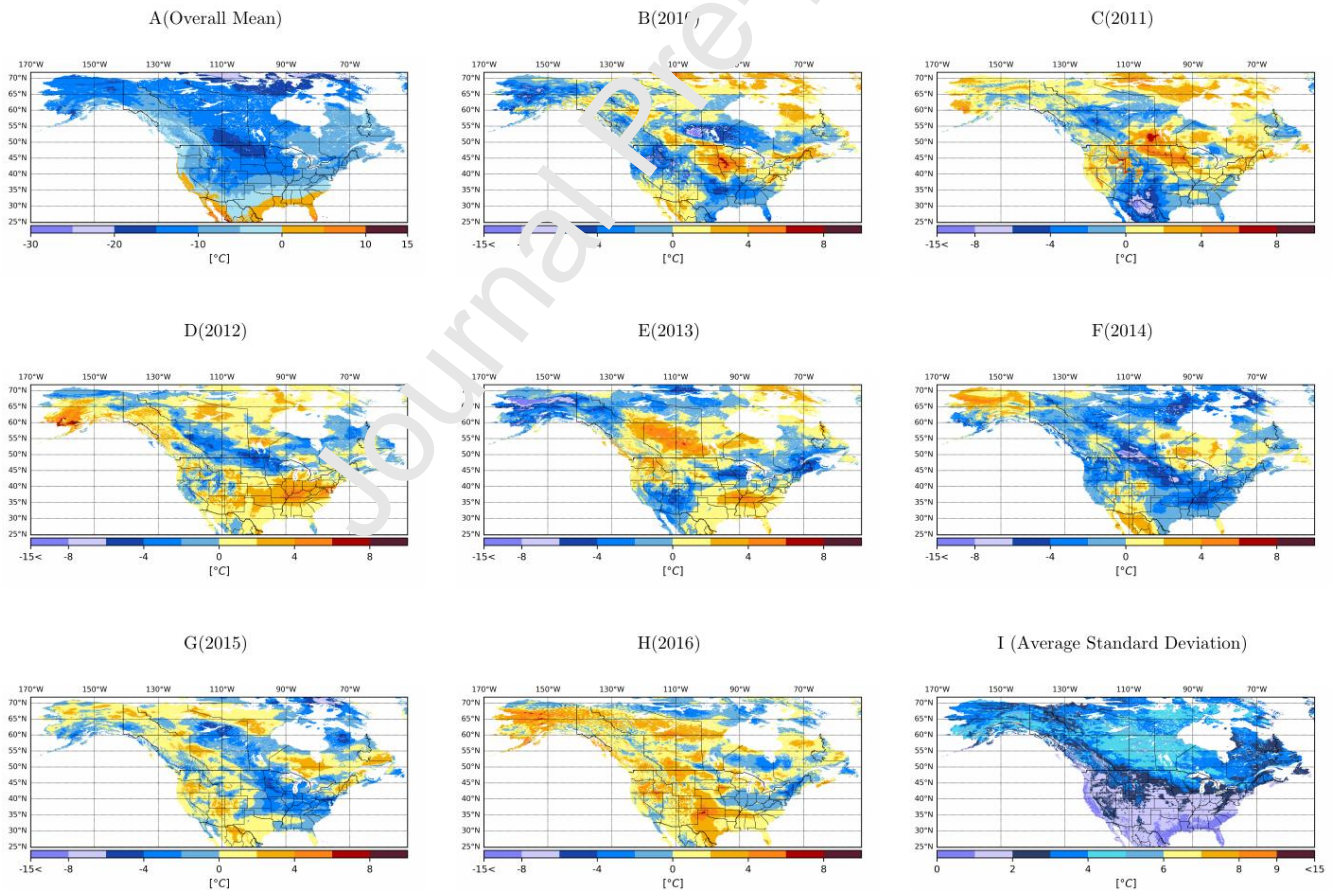


Figure 3. Spatial distribution of overall ensemble mean (A), differences between annual ensemble mean and overall ensemble mean (B-H), and average standard deviation (I) of minimum soil temperature computed using 3-hourly soil temperature at 6AM over 9 ensemble members for each year from WT 2010-2016

### 3.1.3 Annual Number of Freeze-Thaw Cycles

Figure 4-A shows the overall ensemble mean of the annual number of FT cycles. Regions with latitude between 32.5°N and 50°N typically experience high number of FT cycles (6 to 16 cycles per year). The ensemble mean of FT cycles can reach to 18 cycles per year in western Nebraska and eastern Colorado. Alaska and other Arctic regions have less than six cycles annually with the number of cycles decreasing with increasing latitudes. The south and west coast of the USA usually have less than two cycles per year.

Figures 4-B to 4-H examine the difference between the annual ensemble mean of FT cycles and the overall ensemble mean. High latitude regions seem to have a consistent pattern across years with a few exceptions. The ensemble mean is noticeably different from the overall mean in 2010 and 2014 for the Hudson Bay's southeastern region in central Canada. Also, in general, there is high interannual variability in the ensemble mean of FT cycles for Alaska's prairie regions for all the years. In mid-latitude regions, deviations between 2 and 6 cycles can be found around the Great Lakes and the northern Central Lowlands. In most years, the areas where FT cycles exceed 12 are limited to the central Great Plains. There are few years, including 2012 and 2013, in which that area greatly expands. In southern regions, there are large year-to-year variation in the northern part of ephemeral regions east of 100°W. Other areas of North America are usually relatively consistent over years with a few notable exceptions such as the northeastern USA in 2010 and 2016.

The average variations of FT cycles across the nine ensemble members are shown in Figure 4-I. The poorest agreement among ensemble members is found in the Great Plains, northern Pacific coastline, and along the Appalachian Mountains, the Sierra Nevada, and the southern Cascade Range. A small area in south of 30°N also has poor agreement among the ensemble members on the number of FT cycles. In contrast, models agree within one to two cycles in Alaska, as well as the south and west coasts of the USA. Two to three cycle variations are typical in the mid latitudes. It is evident from

Figure 4 that the highest variability among the models, exceeding six cycles, is not necessarily collocated with locations of high FT cycles. There are also regions where the magnitude of ensemble variability changes greatly year-to-year (Figure S-5). For example, ensemble members had good agreement around the Lake Superior in 2013 and 2015 but strong disagreement in 2010 and 2012.

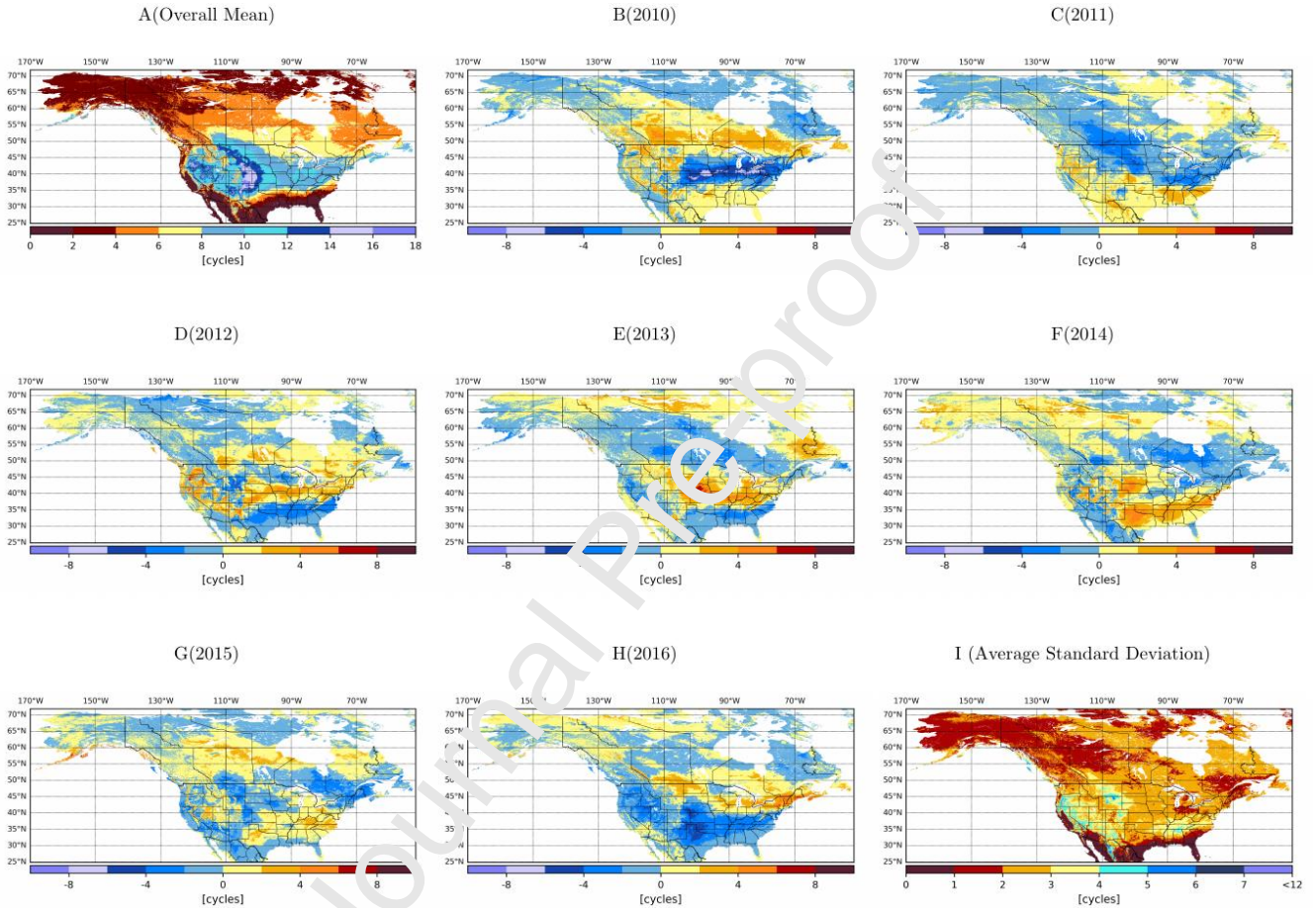


Figure 4, Spatial distribution of overall ensemble mean (A), differences between annual ensemble mean and overall ensemble mean (B-H), and average standard deviation (I) of number of FT cycles computed using 3-hourly soil temperature at 6AM over 9 ensemble members for each year from WT 2010-2016

## 3.2 Ensemble variability across snow classes

### 3.2.1 Annual Number of Frozen Days

Figure 5 shows the distributions of the class-averaged annual number of frozen days by snow classes and ensemble members. Each box consists of seven points, corresponding to the seven WYs (2010-2016), each of which represents the annual number of frozen days averaged over the given snow

class. In each box, the black line indicates median. Top and bottom of each box are bounded to the 75<sup>th</sup> (Q3) and 25<sup>th</sup> (Q1) percentiles. The upper whisker extends to the last point that is less than  $Q3 + 1.5*(Q3-Q1)$ . Similarly, the lower whisker extends to the first point that is greater than  $Q1 - 1.5*(Q3-Q1)$ . Beyond the whiskers, data were considered outliers and are shown with black color as individual points.

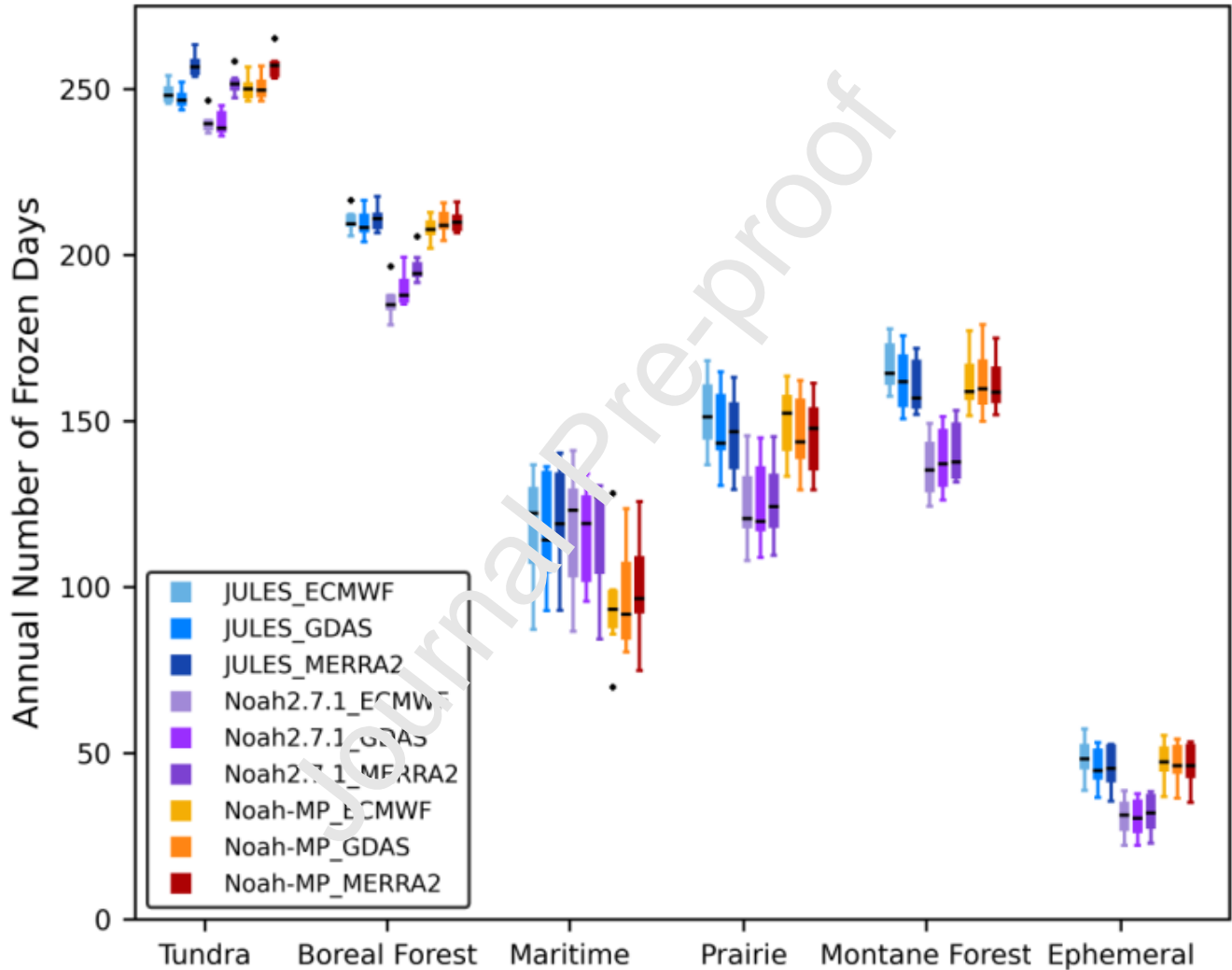


Figure 5, Distribution of annual number of frozen days, averaged over different snow classes. Each box-plot consists of seven points, each of them representing the annual number frozen days at a WY (2010-2016), averaged over a snow class. The black line indicates FD median; top and bottom of box are the 75<sup>th</sup> and 25<sup>th</sup> percentiles, and top and bottom of whiskers represent the maximum and minimum FD without outliers. Outliers are shown as black dots.

There are distinct differences in the annual number of frozen days across snow classes. Tundra and ephemeral classes have the highest (248.6 days) and the lowest (41.4 days) median of annual number of frozen days, respectively. There are also notable differences among models. Noah2.7.1 members'



simulations yield the fewest median of class-averaged annual number of frozen days for all the snow classes except maritime. While the Noah-MP and JULES estimations of the annual number of frozen days are very similar at almost all snow classes, Noah-MP estimates shorter frozen period at maritime. The forcing datasets that drive a common LSM do not have conspicuous differences. It appears that differences among LSM model physics cause a larger discrepancy in soil frozen days than differences in the meteorological forcing datasets.

Including the outliers, tundra regions experience the least year-to-year variation in annual number of frozen days across all the ensemble members. The difference between the largest and smallest annual number of frozen days (range) is typically 9.8 days for this class. Maritime regions have the most year-to-year variation, typically 47.6 days, in the annual number of frozen days. The interannual variability of boreal forest (12.3 days) and ephemeral regions (17.1 days) are almost similar. Within a snow class, there is not a notable difference in year-to-year variation across the ensemble members.

### 3.2.2 Annual Minimum Temperature

The distributions of the class-averaged, annual minimum temperature vary notably among ensemble members and snow classes (Figure 6). The warmest simulated minimum temperatures occur in ephemeral and maritime regions. While tundra regions have the longest frozen period, their minimum temperatures are not cooler than several other classes. The annual minimum temperatures estimated by the Noah2.7.1 are cooler than the corresponding values of the other members across all snow classes, except ephemeral. Ephemeral regions are noted for the consistency in minimum temperature estimates across the ensemble. The Noah-MP's annual minimum temperatures are warmer than JULES's in some classes such as boreal forest, maritime and montane forest while they are colder in tundra.

Differences in annual minimum temperature across the forcing datasets were evident but differed by snow class. The warmest and coldest values are estimated by MERRA2 and ECMWF, respectively, in tundra and boreal forest. In contrast, MERRA2 has the coldest minimum temperature in the maritime and ephemeral classes and there is not a notable difference between ECMWF's and GDAS's minimum temperatures. The montane forest and prairie have the least and greatest average year-to-year variation of annual minimum temperature, respectively. The average year-to-year variations are similar in other regions (range from  $\approx 3^{\circ}\text{C}$  to  $3.8^{\circ}\text{C}$ ) but can differ by ensemble member. For example, all ensemble members within the ephemeral class have a similar variation of annual minimum temperature ( $\approx 3^{\circ}\text{C}$ ). Whereas, year-to-year variations in maritime regions range from  $2.9^{\circ}\text{C}$  for Noah-MP to  $5.1^{\circ}\text{C}$  for Noah2.7.1.

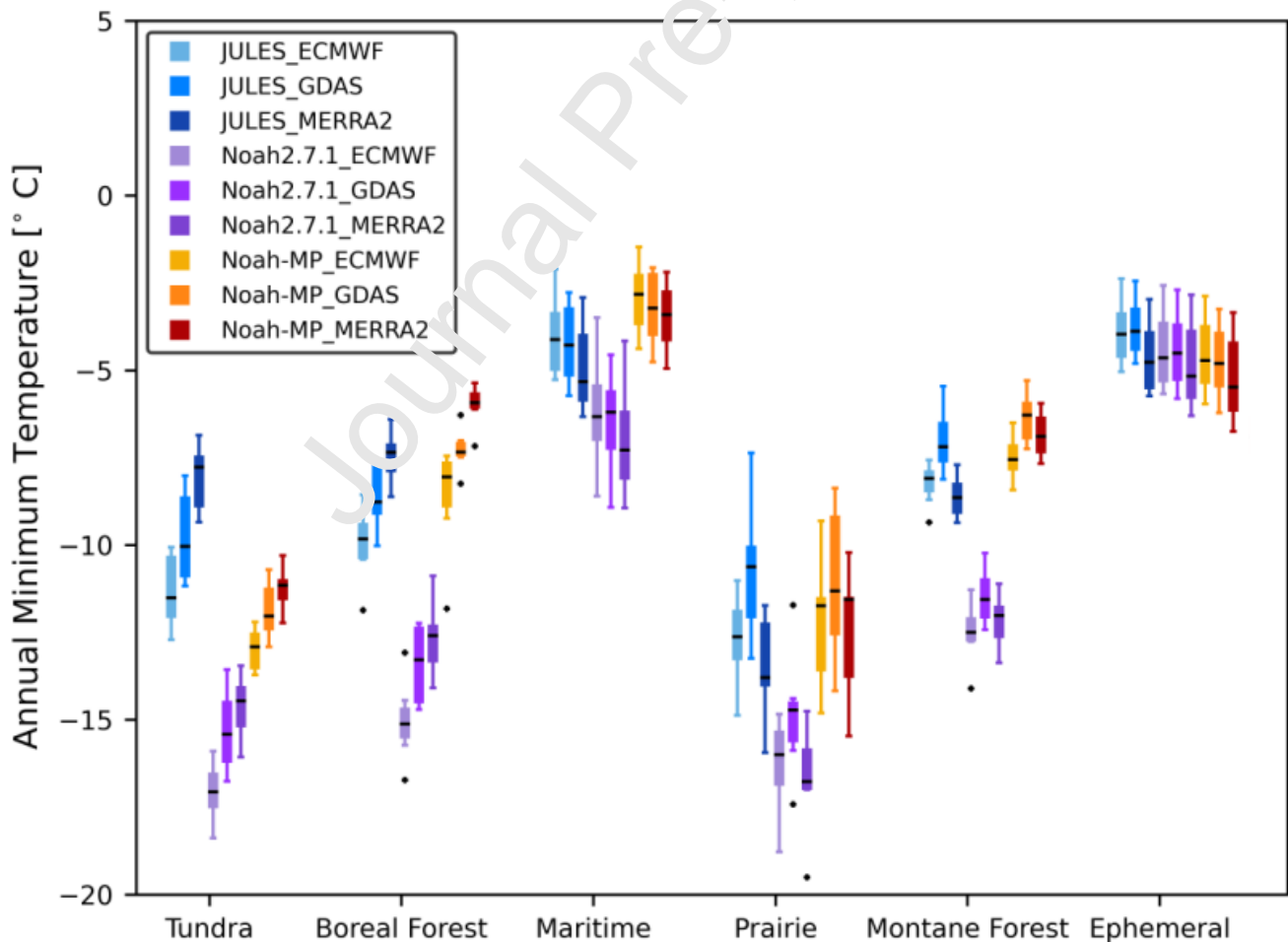


Figure 6, Distribution of annual minimum temperature, averaged over different snow classes. Each box-plot consists of seven points each of them representing the annual minimum temperature at a WY (2010-2016),

averaged over a snow class. The black line indicates median; top and bottom of box are the 75<sup>th</sup> and 25<sup>th</sup> percentiles, and top and bottom of whiskers represent the maximum and minimum  $T_{\min}$  without outliers. Outliers are shown as black dots.

The differences in the annual minimum temperature due to forcing datasets when driven with a common LSM are smaller than differences among the LSMs when driven with common forcing data. Therefore, it can be concluded that the model physics in LSMs, rather than forcing data, resulted in the greater variability in both modeled minimum temperature and frozen days.

### 3.2.3 Annual Number of Freeze-Thaw Cycles

The class-averaged, annual number of FT cycles for the ensemble members across snow classes show different patterns than the annual number of frozen days and minimum temperature (Figure 7). Prairie regions experience the most FT cycles annually (typically 9 cycles). The fewest FT cycles are found in the cold tundra and boreal forest regions (1.2 and 1.1 cycles, respectively). A striking difference in FT cycles is evident among the models. Noah2.7.1 and Noah-MP members simulate the lowest and highest median number of freeze thaw cycles, respectively, in all classes except maritime. In contrast, there is not a consistent pattern in FT cycle distributions due to their forcing dataset. In the prairie and ephemeral classes, MERRA2 members estimate the most FT cycles, but they have slightly fewer cycles than ECMWF and GDSAS members in tundra and boreal forest classes. While LSMs are the main source of FT cycles variability for all snow classes, in some classes such as prairie, cycles are also sensitive to the forcing data.

Including the outliers, all of the ensemble members across all snow classes show high year-to-year variation in the FT cycles counts. If the outliers are excluded, the highest year-to-year variations in the class-averaged annual number of FT cycles are found in the prairie class. In the maritime region, Noah-MP LSM yields a large spread over the seven WYs with Noah-MP-MERRA2 having large interannual variability of about 3.7 FT cycles.



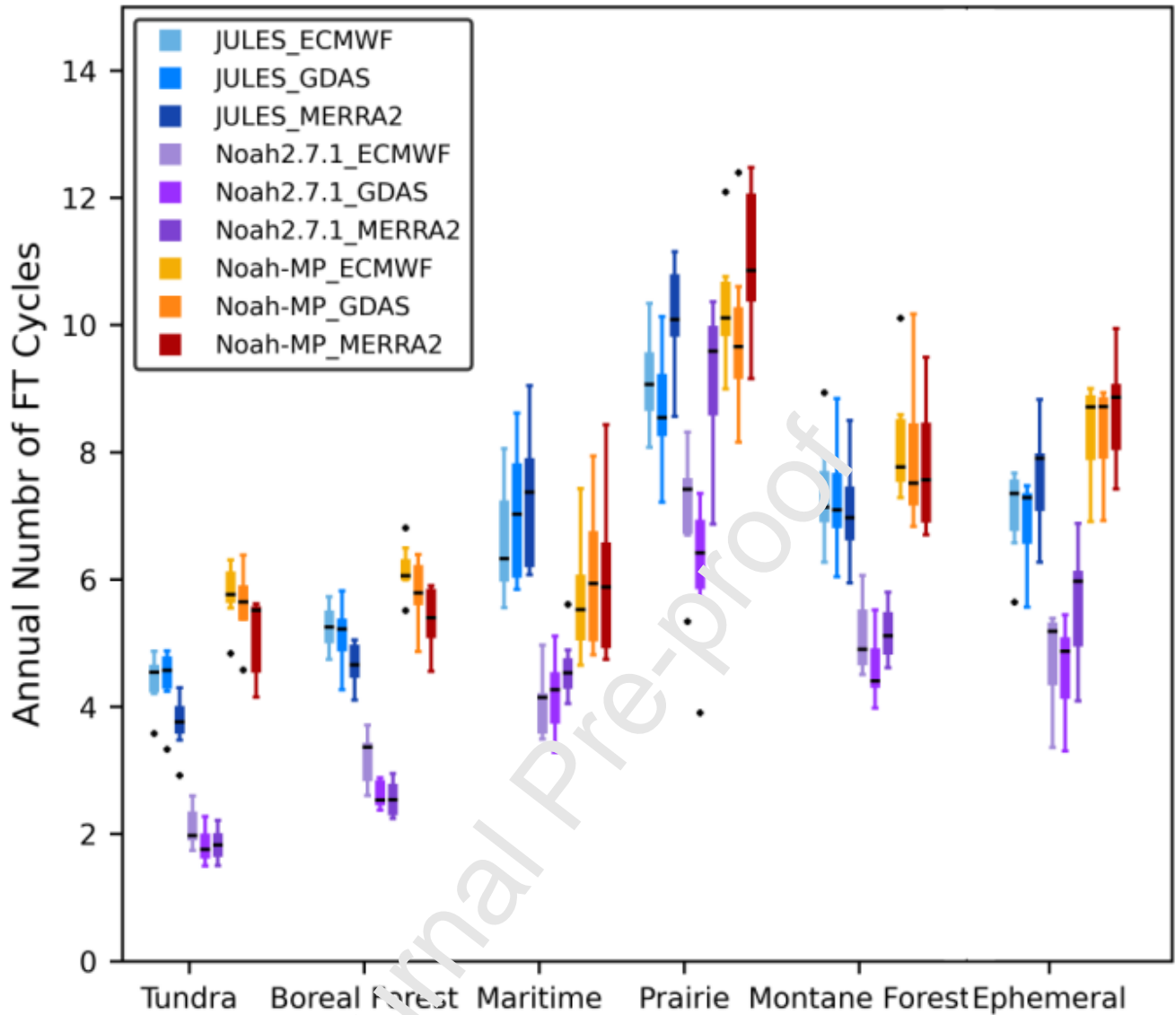


Figure 7, Distribution of annual number of FT cycles, averaged over different snow classes. Each box-plot consists of seven points, each of them representing the annual number FT cycles at a WY (2010-2016), averaged over a snow class. The black line indicates FT median; top and bottom of box are the 75<sup>th</sup> and 25<sup>th</sup> percentiles, and top and bottom of whiskers represent the maximum and minimum FT without outliers. Outliers are shown as black dots.

### 3.3 SEUP ensemble and in-situ observations

To investigate the relationship between the SEUP ensemble simulations and the in-situ observations, the three variables of interest were calculated using the SEUP output and the in-situ data at 15 sites. While the spatial representativeness discrepancy between 5 km simulations and point observations impacts the uncertainty of these comparison, here the results are reported assuming that the study sites are spatially representative of their surroundings. Table 2, Summary of annual number of

frozen days, annual minimum temperature and annual number of FT cycles based on in-situ observations, averaged over WYs 2010-2017 at the 15 studied sites

Table 2, Summary of annual number of frozen days, annual minimum temperature and annual number of FT cycles based on in-situ observations, averaged over WYs 2010-2017 at the 15 studied sites

Station Name	Annual number of frozen soil days (days)	Annual minimum temperature (°C)	Annual number of FT cycles (cycles)
1.Granite Creek	201	-5.7	2
2.Upper Nome Creek	224	-5.8	2
1.Arapaho Ridge	130	-1.1	2
4.Gobblers Knob	221	-12.7	2
5.Mundare AGDM	110	-7.9	4
6.Streeter	119	-6	3
7.South Fork	60	-5.2	2
8.Hubbard Brook	<=1	0.4	<=1
9.Beaver Dams	<=1	0.1	<=1
10.Crater Meadows	35	-0.2	2
11.MF Nooksack	<=1	0.2	<=1
12.Cayuse Pass	0	0.24	0
13.Lind	28	-7.2	4
14.St Joseph's	24	-1.4	3
15.Bushland	6	-0.6	3

The 15 study sites were chosen from locations in all snow regimes across the North America (Figure 1). The observed number of frozen days is typically less than 40 days per year in the montane forest, maritime and ephemeral sites. The sites located in tundra, boreal forest and prairie routinely have more than 110 frozen days although South Fork, IA (#7) experiences shorter frozen period (less than 80 days) generally. The minimum top-layer soil temperature observed across all the studied sites is generally warmer than -9°C with a few exceptions. In Upper Nome Creek, AK (#2), the minimum temperature was -11.1°C in 2013. The coldest minimum top-layer soil in Mundare AGDM, Canada, Alberta (#5) was -13.5°C in 2010. The soils cooled to -12.1°C and -13.7°C in Lind, WA (#13), in 2010 and 2014, respectively. Among all the sites, Gobblers Knob, AK (#4) experiences the coldest minimum top-layer soil, frequently less than -12°C. All sites except Cayuse Pass, WA (#12) experienced at least one year when annual number of FT cycles exceeded zero. Most sites had less than five FT cycles annually. However, there are sites that underwent more than five freeze-thaw cycles for some years. Southern sites Bushland, TX (#15), Mundare AGDM, Canada, Alberta (#5), experienced 11 cycles in

2014 and more than eight cycles in two years, respectively. Table 2 summarizes the observed three soil states, averaged for the study period by site. WY 2017 was included in this section. Although it is a partial year, the observations at all of the 15 sites indicate that soil was well thawed by the end of May 2017 and no freezing events occurred afterward.

### 3.3.1 Annual Number of Frozen Days

Figure 8 compares the observed and estimated annual number of frozen days at the in-situ locations by year. The SEUP ensemble members showed conspicuous biases in annual number of frozen days. While SEUP members generally overestimated the annual number of frozen days, there are relatively good agreements at the Alaskan sites. Moreover, at Streeter, ND (#6), the number of frozen days simulated by the ensemble members closely followed the observation driven annual number of frozen days consistently over all period of the two winters and South Fork, IA (#7), had good agreement in WY 2017.

While in most sites, low variability in the number of frozen days can be seen over years, Arapaho Ridge, CO (#3) had considerable year-to-year variability. At this site, the SEUP models agreed better with observations for years with the longest frozen period, but had a high bias for the years with fewer frozen days. Also, the observations had higher year-to-year differences in the number of frozen days than the models. The SEUP members were not able to discern a year with fewer frozen days from one with a longer frozen period. For example, at Mundare AGDM, Canada, Alberta (#5) where the observed number of frozen days ranged from 84 days to 134 days over the seven WYs, the SEUP simulated frozen days range was nearly fixed (approximately 130 to 180 days) across the years.

There are also sites that did not generally freeze, yet the ensemble members estimated high numbers of frozen days. For example, Hubbard Brook, NH (#8) had no observed frozen days in five years, but the SEUP members simulated between 88 to 171 frozen days for those years.

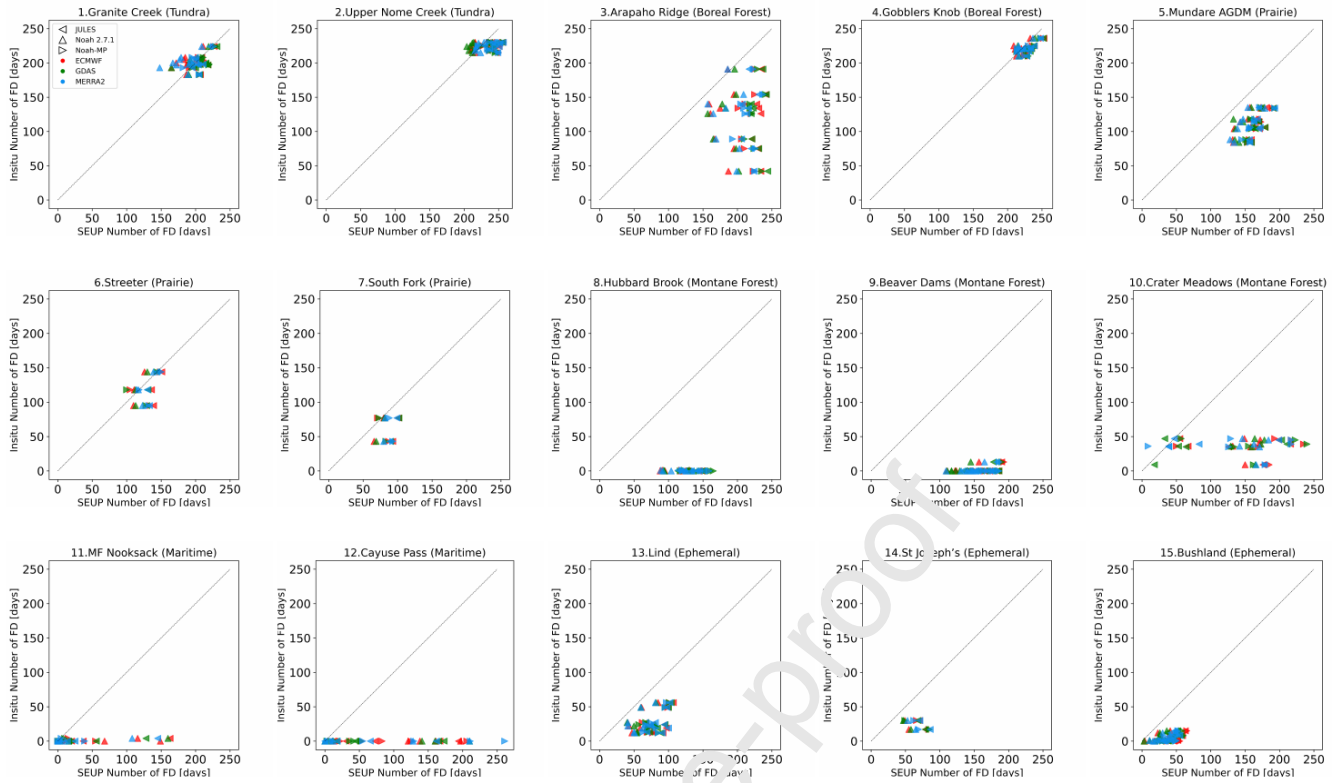


Figure 8, The observation based annual number of frozen days (y axis) versus the SEUP ensemble annual number of frozen days (x axis).

### 3.3.2 Annual Minimum Temperature

Figure 9 shows the observed and estimated annual minimum temperature for the 15 studied sites by year. The SEUP simulated annual minimum temperatures are much colder than was observed at all studied sites except Lind, WA (#13) and Gobblers Knob, AK (#4). The coldest minimum temperature,  $-31.8^{\circ}\text{C}$ , was simulated at Granite Creek, AK (#1) when observed minimum temperature was  $-7.2^{\circ}\text{C}$ . The ensemble's underestimations of annual minimum temperature are also noticeable for sites where the minimum soil temperature fluctuates around  $0^{\circ}\text{C}$ . At these sites, the soil never cooled below  $-0.7^{\circ}\text{C}$ , but the modeled minimum temperature was as low as  $-15^{\circ}\text{C}$ .

At most sites, there was less than a  $5^{\circ}\text{C}$  year-to-year difference in the observed annual minimum temperature. Among all of them, montane forest and maritime sites had the least variation in the annual minimum temperature over years (less than  $2^{\circ}\text{C}$ ). In contrast, the observed minimum temperature at Upper Nome creek, AK (#2), Gobblers Knob, AK (#4), Mundare AGDM, Canada, Alberta (#5) and

Lind, WA (#13) had the highest year-to-year variation ( $\approx 10^{\circ}\text{C}$ ). At these sites, the SEUP ensemble did capture the variations although they still underestimated the minimum temperature. Gobblers Knob, AK (#4) agreed best with the warmest observed temperature. Gobblers Knob, AK (#4) agreed best with the warmest observed temperature. Lind, WA (#13), had the best agreement for the coldest years.

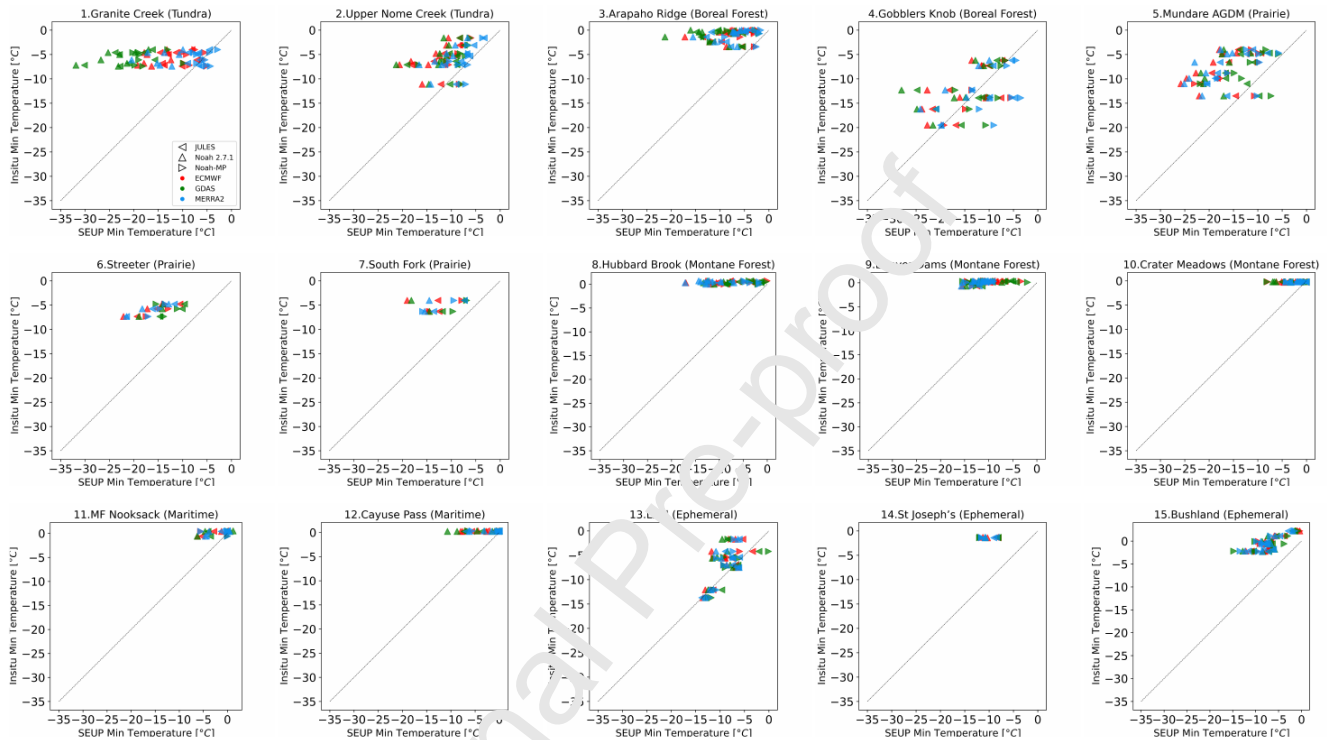


Figure 9, The observation based annual minimum temperature (y axis) versus the SEUP ensemble annual minimum temperature (x axis).

### 3.3.3 Annual Number of Freeze-Thaw Cycles

Figure 10 depicts the in-situ and SEUP driven annual number of freeze-thaw cycles for the study sites by year. The SEUP members generally overestimate the number of FT cycles. At most of the studied sites, none of the SEUP estimations of FT cycles match the observed FT cycles. Sites such as Hubbard Brook, NH (#8), and Beaver Dams, UT (#9) that usually have no FT cycles, have modeled estimates as high as 15 cycles. Although year-to-year variation in the observed number of FT cycles is evident at some sites, the SEUP estimated annual number of FT cycles does not capture these year-to-

year variations at any site. For example, the ensemble members' range is nearly fixed for all the years at Granite Creek, AK (#1), and Arapaho Ridge, CO (#3).

In summary, the SEUP ensemble members' ability to capture the three variables of interest, annual number of frozen days, annual minimum temperature and annual number of FT cycles, is generally poor over the studied sites. Even for sites such as Streeter, ND (#6), where the SEUP members provide a reasonable estimation of annual number of frozen days, there are considerable biases in the annual FT cycles and minimum temperature estimates.

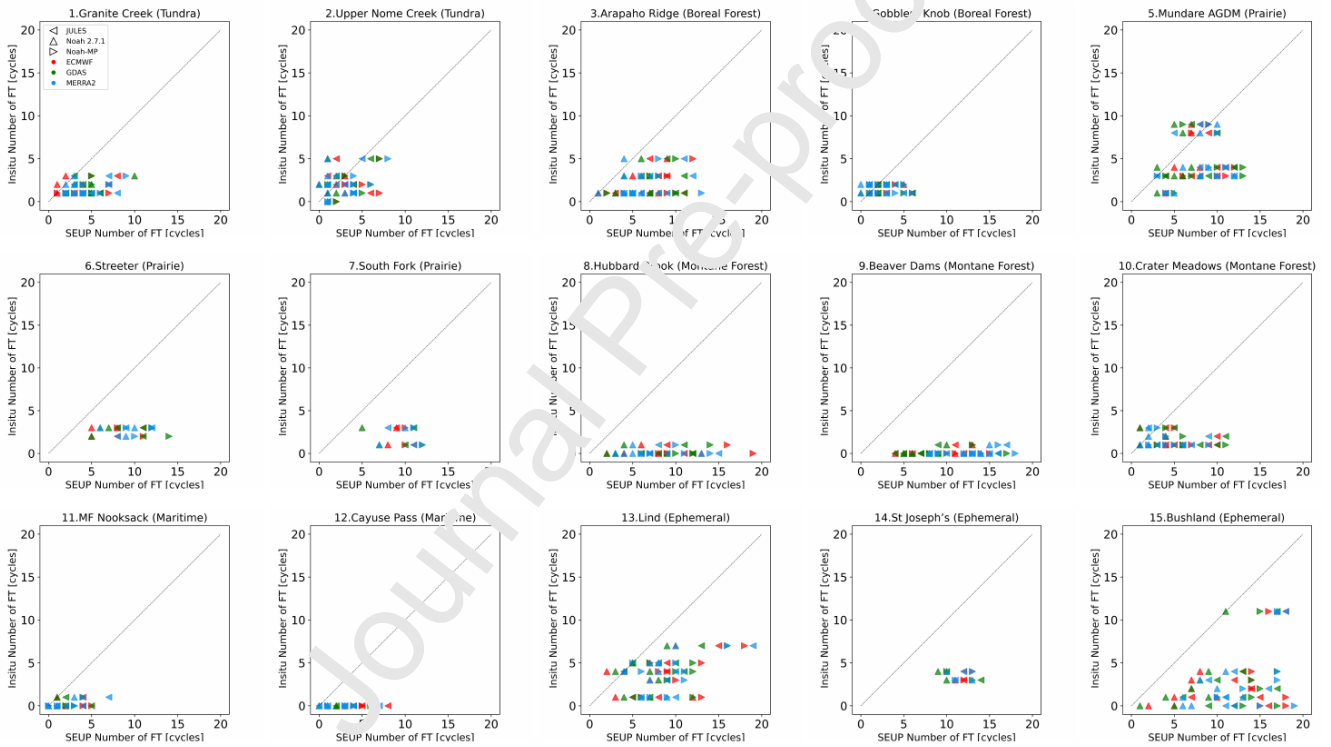


Figure 10, The observation based annual number of FT cycles (y axis) versus the SEUP annual number of FT cycles (x axis).

### 3.3.4 Temporal Evolution of Soil Temperature

This section investigates how the snowpack contributes to the soil temperature temporal evolution throughout the course of winter in the SEUP simulations. The observed and simulated SWE or snow depth, air and soil temperatures were examined for the 10 studied sites where all soil, air and snow

observations are available (Table 1). The WYs 2016 and 2017 results for four sites, Arapaho Ridge, CO (#3), Beaver Dams, UT (#9), Granite Creek, AK (#1), Mundare AGDM, Canada, Alberta (#5), and MF Nooksack, WA (#11) are provided here for the illustration purposes.

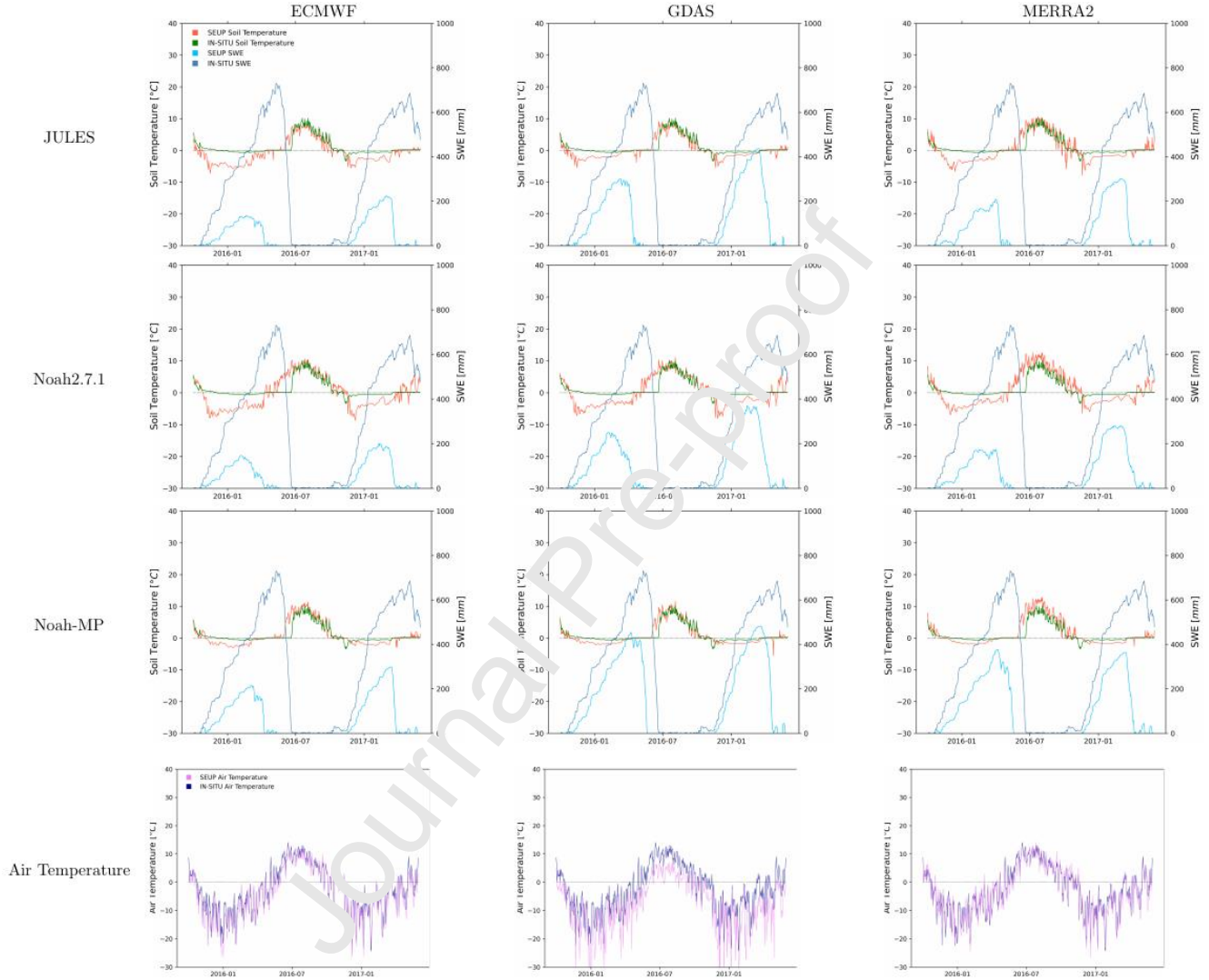


Figure 11, Observed and SEUP simulated SWE, soil temperature and air temperature at 3. Arapaho Ridge (boreal forest, CO) during WYs 2016 and 2017.

Figures 11-15 show that the observed air temperature generally agrees better with the SEUP ensemble than the soil temperature. The modeled and observed soil temperatures also generally agree prior to the onset of snow. The observed soil temperatures reach a steady state value near zero when the snow arrives. The modeled snow accumulation generally starts later than observed. As a result, the SEUP soil



layer is not insulated by the snow till later in the winter. During this period, the modeled soil temperature continues to cool, and reaches its minimum when the modeled snow begins to accumulate.

The magnitude of the day-to-day fluctuations in modeled winter soil temperature is considerably larger than observations. The observed data have relatively smooth fluctuations as compared to the larger, abrupt changes in the SEUP soil temperatures. The differences in the ensemble members' soil temperatures, in early winter, before a snow layer is established, are more evident among the models than forcing datasets. For example, Noah2.7.1 and JULES' soil temperatures at Arapaho Ridge, CO (#3) dropped sharply below freezing in November while Noah-MP's soil temperatures had a smoother transition and less dramatic decrease (Figure 11). After the modeled snowpack accumulates, the ensemble members with higher SWE values have lower day-to-day variations in soil temperature. For instance, the Noah-MP members have higher SWE at Beaver Dams, UT (#9) for all forcing datasets (Figure 12). The Noah-MP members' soil temperature magnitude and fluctuations were in much better agreement with observations than the other models.

Although snow is a key contributor to soil thermal processes, it should be noted that the variabilities in SWE alone cannot explain the variabilities among the models and the errors in comparison with the observations. For example, soil temperature and SWE plots at Granite Creek, AK (#1) (Figure 13) reveal that even when the models agree on SWE estimation their simulated soil temperatures could evolve differently. For example, all the GDAS SWE estimates were closely



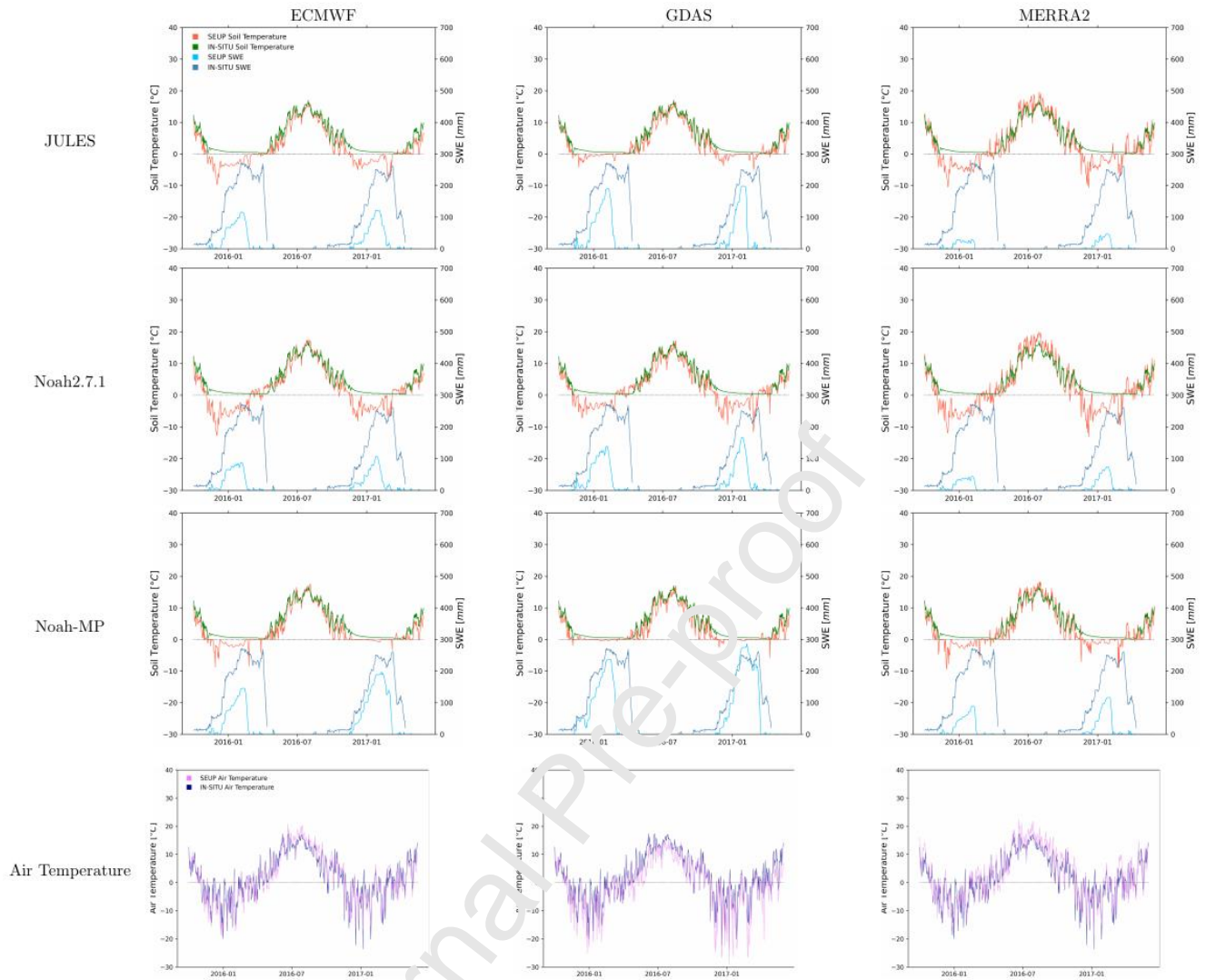


Figure 12, Observed and SEUP simulated SWE, soil temperature and air temperature at 9. Beaver Dams (montane forest, UT) during WYs 2016 and 2017.

followed observed SWE at 2017, yet the models behaved differently in estimating soil temperature. Comparing to Noah-MP, sharper drops and higher variations can be found in soil temperature of JULES and Noah2.7.1 even in the presence of snow which resulted in colder soil temperature and thus larger errors. Mundare AGDM, Canada, Alberta (#5) is another example confirming the possible contribution of other factors in addition to SWE and snow depth to the soil thermal processes (Figure 14). At this site, Noah2.7.1 simulated higher snow depth than JULES, Noah-MP and observations but even more snow was not able to prevent soil layer from cooling down and its soil layer still had colder soil temperature through the course of winter.

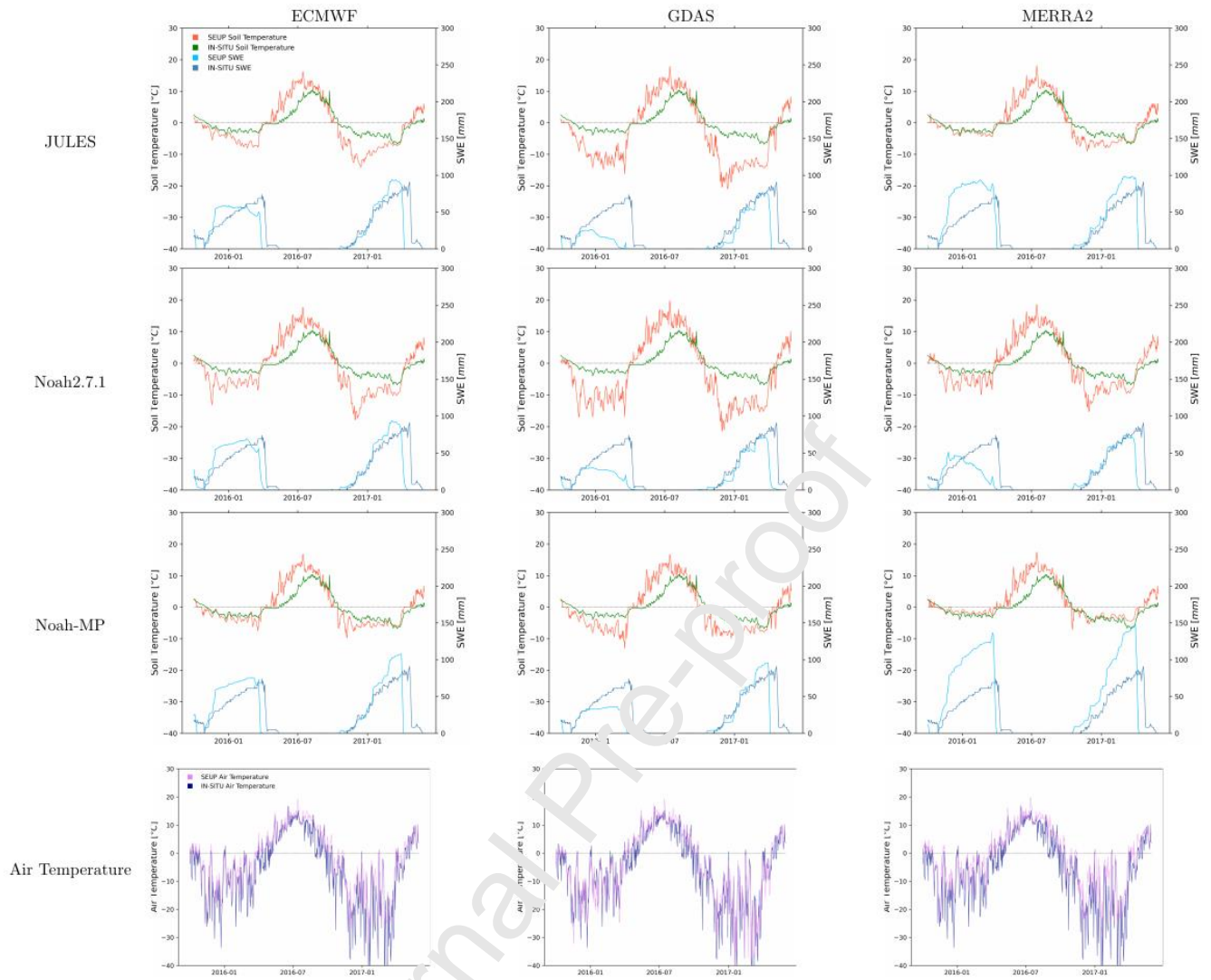


Figure 13, Observed and SEUP simulated SWE, soil temperature and air temperature at 1. Granite Creek (tundra, AK) during WYs 2016 and 2017.

Soil temperatures and snow plots also show that the timing of snow ablation and the rate of snow melt differs among the ensemble members and appears to impact spring soil warming. SEUP's modeled snowpacks generally melt and disappear sooner than the observed snow layer, both decreasing the frozen period and warming the modeled soil temperature at the end of the cold season. For example, the Noah2.7.1 snowpack melted sooner at Arapaho Ridge, CO (#3), which resulted in an abbreviated frozen periods for Noah2.7.1 members (Figure 11).

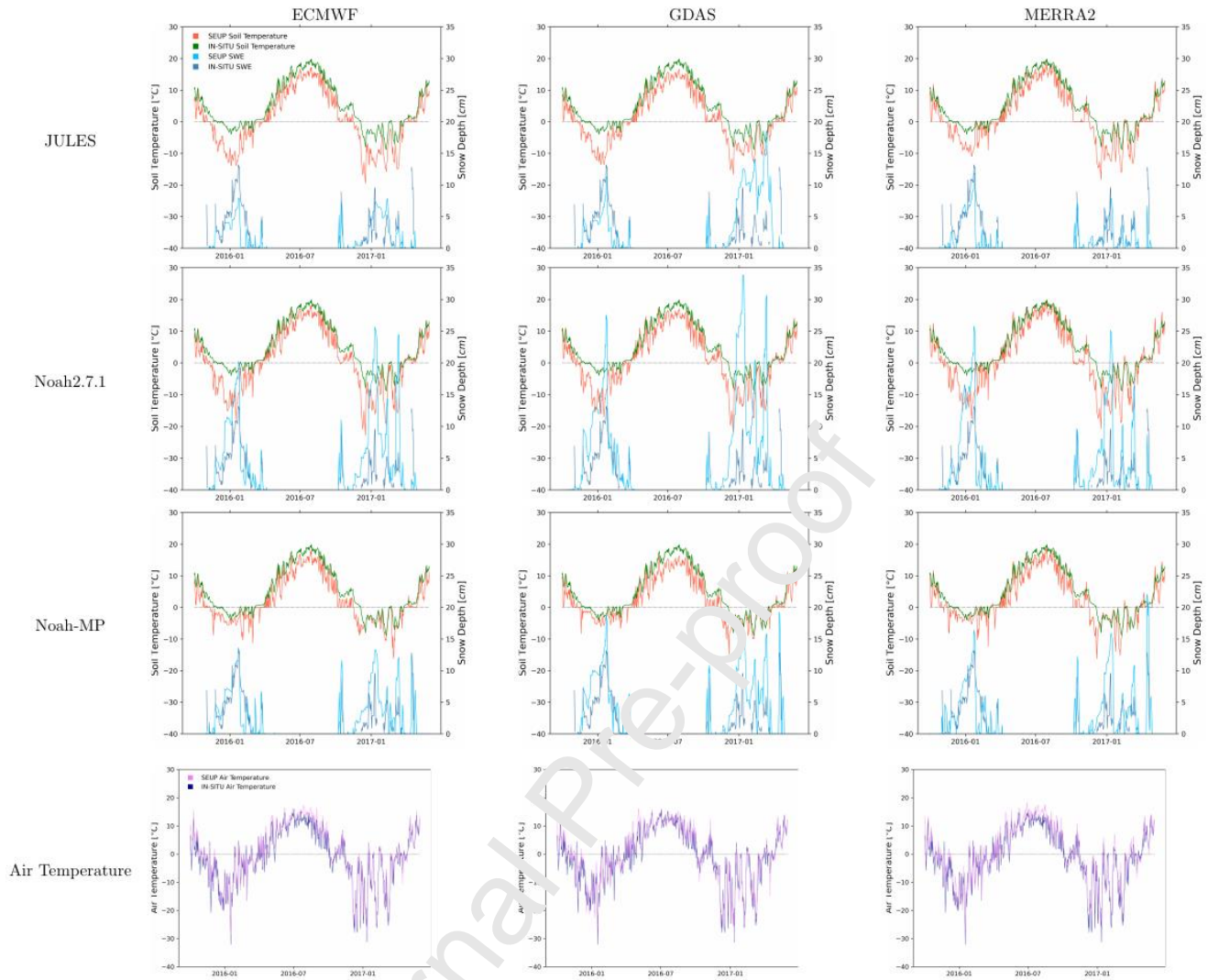


Figure 14, Observed and SEUP simulated snow depth, soil temperature and air temperature at 5. Mundare AGDM (prairie, Canada, Alberta) during WYs 2016 and 2017.

## 4 Discussion

### 4.1 Ensemble Variability

Our study agreed with earlier findings that the ensemble members' simulated soil metrics can vary widely, but that the magnitude of this variability is not consistent across North America. The standard deviation of annual number of frozen days, annual minimum temperature and annual number of FT cycles vary in range of 0-175 days, 0-15 °C, and 0-12 cycles over the North America, respectively. Consistent with our findings, a wide range of winter soil temperatures simulated by six LSMs (Jena

Scheme for Biosphere–Atmosphere Coupling in Hamburg (JSBACH), Organising Carbon and Hydrology In Dynamic Ecosystems (ORCHIDEE), JULES, COUP, HYBRID8, Lund-Potsdam-Jena General Ecosystem Simulator (LPJ- GUESS)) was also reported by Ekici et al. (2015) across their four alpine, high Arctic, wet polygonal tundra and non-permafrost Arctic study sites. The wide ranges of soil metrics among the SEUP ensemble over our large study domain with extremely different climate and surface characteristics have two implications. First, the processes impacting winter soil metrics such as snow accumulation and ablation, and heat transfer through soil layer and snowpack are location-specific. Second, models simulate such dominant processes differently due to different representations and equations employed by models.

We found that the regions having the greatest disagreement among the ensemble members depend on the specific soil metrics. Disagreement over one soil metric doesn't necessarily lead to variance in other soil metrics. The northern Pacific coastline is the only region where the models are not relatively consistent in both the length of frozen period and the number of FT cycles. Kim et al. (2021) found that this region also experiences the largest spread in ensemble mean SWE. They concluded that the differences in rain–snow partitioning schemes of the land surface models could partially contribute to the large spread in mean SWE at the northern Pacific coastline. The variability in the amount of snow over the ground during the accumulation period can have serious impacts on insulating the ground and thus on FT processes. Comparing our results and Kim et al. (2021) revealed that the large spread in the simulated soil metrics and the mean SWE are not necessarily found at the same areas. This means that the ensemble variabilities in annual mean SWE do not entirely explain the ensemble variabilities in winter soil metrics. Our results suggest that the uncertainty in snow characteristics at the beginning of winter may be a more important indicator of the soil simulation performance over the entire winter season rather than the seasonal mean or maximum SWE. While a recent study showed that atmospheric forcings (e.g., precipitation) led the largest uncertainties in SWE estimations (Cho et al., 2022), we found the variabilities in winter soil metrics appear to be more rooted in the differences among the

LSMs' model physics rather than forcing datasets. This indicates that efforts to improve models' estimation of winter soil characteristics should first focus on the LSMs' physics.

Regional differences in the ensemble mean of annual number of frozen days across snow classes can partially be explained by latitude and air temperature. The soil layer at northern tundra and boreal forest regions stays frozen for a longer time because air temperature in these regions is frequently below freezing (Figure S-6). In contrast, the lower latitude, ephemeral regions experience few days with air temperature below freezing (Figure S-6), which results in an abbreviated frozen period for the soil layer. In the mid latitude regions, including maritime, prairie and montane forest classes, other factors than air temperature such as the regional snowpack magnitude, timing, and duration could be more important controls for number of frozen days. Maritime regions have the deepest snowpack over the North America (Kim et al., 2021) as well as a persistent snowpack (Figure S-9). A deep and persistent snowpack would prevent the soil layer from freezing and reduce the number of frozen days in maritime regions as compared to prairie and montane forest areas. Across all the snow classes except maritime, the Noah2.7.1's soil layer has the shorter frozen period. Based on the soil temperature time series at the 15 study sites, a reasonable hypothesis is that the abrupt changes in soil temperature of Noah2.7.1 model at the beginning and end of winter season do not allow the soil layer to stabilize at an isothermal state and fluctuate around freezing point. This hypothesis could also explain why Noah2.7.1 simulated fewer FT cycles for all snow classes.

The higher ensemble mean annual minimum temperature in maritime and ephemeral classes are linked to higher minimum air temperature at these classes. While tundra regions experience colder minimum temperature comparing to prairie (Figure S-7), there is not much difference in the ensemble mean minimum soil temperature of these two classes. In prairie regions, the soil layer is not insulated by snow due to a shallower snow (Kim et al., 2021) and more snow-free days (Figure S-9) as compared to tundra and boreal forest areas. Annually, there were more than 30 snow free days with air temperature below freezing in prairie regions (Figure S-9). The same conditions can explain the



regional differences between boreal forest and montane forest. The reason that the coldest minimum temperature for all snow classes were simulated by Noah2.7.1 appears to be due to its relatively high response in soil temperature to changes in air temperature. Although it needs further assessment, Noah2.7.1 may have a higher thermal conductivity for its snowpack and soil layer.

Despite notable difference among maritime, ephemeral and montane forest classes' frozen period, SWE and persistence of snowpack, no significant differences are evident among their ensemble means of number of FT cycles. This interesting result however does not necessarily mean that FT processes are similar in these regions. Less snow in ephemeral region means that soil has an increased the chance of freezing and thawing even in the middle of winter. On the other hand, in maritime and montane forest areas, FT cycles appear to occur most frequently during the shoulder periods of winter season, before snowpack establishment or during the melt season.

The quantification of the spatial and temporal variability in simulated winter soil characteristics enabled us to identify the regions with extreme values, considerable year-to-year variability and variability among models. The latter directs the modeling community to specific regions where there is a need to better understand the source of variability. The application of modeled winter soil temperature in regions having high variability in time or across models should recognize the uncertainties in that output and consider using a longer time period or an ensemble of models to capture the region's variability.

## **4.2 Factors leading to uncertainties in LSM's cold soil metrics**

Cold biases in minimum soil temperature and overestimates of the number of FT cycles and frozen days were found at almost all of the 15 studied sites. Also, the ensemble simulations couldn't reflect the observed year-to-year variation in the soil metrics. While, the simulated frozen period is on average 20 days longer than observations in northern latitude tundra, boreal forest and prairie sites, it is generally more than 110 days longer in the mid-latitude montane forest sites. In the maritime and ephemeral

sites, 50 more frozen days were simulated by models on average. The ensemble mean minimum soil temperature is generally about 6 °C colder than observation in all the sites except the maritime and prairie sites. Overall, the lowest ( $\approx 2^{\circ}\text{C}$ ) and highest differences ( $\approx 8^{\circ}\text{C}$ ) between the observed and ensemble mean minimum soil temperature can be found in maritime and prairie sites respectively. The overestimation of FT cycles is more significant in prairie and montane forest sites ( $\approx 6$  more simulated cycles on average) compared to that in maritime and northern tundra and boreal forest sites ( $\approx 2$  more simulated cycles). These results are consistent with previous studies in which cold biases in soil temperature modeled by Noah model (Godfrey and Stensrud, 2008; Xia et al., 2013), Noah-MP (Li et al., 2021) and JULES (Ekici et al., 2015) were noted. Warm biases were also found in the early spring period (Xia et al., 2011) and warm part of day (Godfrey and Stensrud, 2008).

Similar to our study, these previous studies also relied on comparing point scale observations with gridded simulations to assess the performance of the models. It is important to acknowledge the potential spatial representative issues that may be raised in such comparisons. Despite the relatively high resolution of the SEUP ensemble, full representation of local heterogeneity for microphysical features and soil winter processes is not possible over a 5-km grid. Thus, differences in the scale between model grid cells and an in-situ station likely account for some of the biases. This is an inevitable limitation when comparing model simulations against point observations. However, the collective use of comparisons to identify consistent differences between the modeled output and observations is still valuable. The other limitation here that may contribute to some of the high biases is the binary method with a fixed threshold ( $0.01^{\circ}\text{C}$ ) used in this study to differentiate between frozen and thawed states of soil.

Various factors may contribute to the errors in simulated cold season soil metrics by LSMs. Previous studies found that errors in the snow simulations are the primary control on the ground temperature estimates (Gubler et al., 2013; Ekici et al., 2015; Wang et al., 2016; Li et al., 2021), which are supported by our findings. One of the critical properties of snow that modulates winter soil

temperature is snow's thermo-insulation. Our results suggest that there are errors associated with the rate of accumulation, timing, duration, snow depth and SWE simulated by the SEUP ensemble members which impact the snow insulating property and, consequently, the soil temperature. The comparison of the in-situ and the SEUP estimates of snow depth revealed that snow depth was generally underestimated by the SEUP ensemble members at the beginning of winter which could result in limited insulation. Using snow depth measurements, Sharratt et al. (1992) found that 15 to 42.5 cm of snow depth was the threshold for which soil temperature fluctuations became independent of air temperature fluctuations. While the dampening of in-situ soil temperature fluctuations observed at our 15 sites was consistent with Sharratt et al.'s finding, the modeled time series of the SEUP snow depth, soil and air temperatures showed otherwise. In SEUP simulations with deeper snowpacks, the soil layer is still responsive to air temperature. This points to a potential overestimate of snow thermal conductivity. Accurately estimating snow thermal properties such as thermal conductivity and heat capacity is challenging but important to estimate snow insulation (Cook et al., 2008; Gouttevin et al., 2012; Wang et al. 2016). The three LSMs used in this study, JULES, Noah and Noah-MP, simulate snow thermal conductivity as a function of snow density. Therefore, any error in snow density simulations impacts heat transfer through the snowpack and has consequences for soil temperature. Also, the lack of depth hoar in the land surface models could contribute to soil temperature errors. In the non-permafrost regions, Zhang (1996) found that an increase in the depth hoar fraction from 0.0 to 0.6 can lead to an 8.4°C increase in the daily ground surface temperature.

While not specifically examined in this study, albedo is another snow property that has a pronounced impact on soil thermal budget (Zhang, 2005). If albedo is overestimated, less solar energy is available for the snowpack-soil system, which could lead to cold biases in soil temperature estimates. Previous studies showed that there is a general tendency in land surface models, including Noah and Noah-MP, to overestimate snow albedo (Mölders and Romanovsky, 2006; Roesch, 2006; Chen et al., 2014; He et al., 2019, Rai et al., 2019).



Other potential contributing factors to the errors in winter soil temperature simulations are related to the soil layer itself. Underestimation of soil moisture and consequently underestimation of soil layer volumetric heat capacity could contribute to soil temperature underestimation (Godfrey and Stensrud, 2008; Chen and Dudhia 2001). Moreover, soil moisture directly alters surface albedo and regulates the surface energy balance during snow-free periods of winter (Idso et al. 1975). Because the surface albedo decreases by up to 50% as soil moisture increases (Idso et al. 1975; Wang et al., 2005; Liu et al., 2008; Guan et al., 2009), the underestimation of soil moisture could lead to a higher surface albedo, reduced net solar radiation and a colder soil temperature. In general a soil moisture error of  $0.1 \text{ m}^3/\text{m}^3$  may lead to an error of more than  $1.6^\circ\text{K}$  for the maximum or minimum daily soil temperature (Godfrey and Stensrud, 2008).

Misrepresentation of soil thermal properties such as thermal conductivity are also potential sources of uncertainty in soil temperature estimations (Peters-Lidard et al., 1998; Mölders and Walsh, 2004; Lawrence and Slater, 2008; Dai et al., 2019; Zhang et al., 2021). LSMs rely on thermal conductivity of mineral components as well as thermal conductivity of water, ice, and air to estimate soil thermal conductivity. The mineral thermal conductivity could be computed based upon clay, silt and sand fractions of soil (JULES) or Quartz content (Noah, Noah-MP) (Dharssi et al., 2009; Peters-Lidard et al., 1998). The Dharssi scheme used in JULES model assumed thermal conductivity of silt and sand is 1.36 times higher than clay (Dharssi et al., 2009) so any uncertainty in the soil texture classification used to force the model could propagate to soil heat transfer and temperature simulations. The uncertainties in Quartz content of Noah and Noah-MP soil layer could impact soil thermal conductivity because Quartz thermal conductivity is almost four times higher than thermal conductivity of other soil components (Farouki, 1981; Peters-Lidard et al., 1998). The other critical component impacting soil thermal conductivity is the ice content of soil. Ice has higher thermal conductivity that is almost four time higher than liquid water. Thus, overestimating the ice content could result in higher thermal conductivity.

Finally, errors in forcing data such as an underestimate of radiative fluxes (Godfrey and Stensrud, 2008; Ekici et al., 2015) and errors in surface air temperature (Zhu and Liang, 2005) can result in cold biases in soil temperature. According to Cho et al. (2022), ECMWF, GDAS and MEERA2 winter air temperatures were colder than air temperatures ( $-4^{\circ}\text{C} < \text{mean differences}$ ) at 68, 86, and 53% of SNOTEL stations over the western U.S. for the eight water years from 2010 to 2017. Such cold biases could partially contribute to the underestimation of soil temperature by the SEUP ensemble members.

While all models have considerable biases in the annual number of frozen days, the annual minimum soil temperature and the annual number of FT cycles, comparisons between the temporal evolution of soil temperature at the 15 study sites and across models suggests that Noah-MP's soil temperature simulations were the most comparable to the observations. Both Noah2.7.1 and JULES had abrupt temperature changes and relatively high day-to-day soil temperature fluctuations that were rarely seen in the observations.

## 5 Summary and conclusion

Accurate and spatially consistent estimates of winter soil temperature and FT cycles are important for hydrology, geochemistry and ecology. In the absence of in-situ networks with spatially and temporally consistent observations of soil temperature and the challenges in remote sensing of winter soil temperatures, land surface modeling seems to be the most promising option for characterizing winter soil temperature over large domains. In contrast to previous studies which primarily focused on soil temperatures simulated by one LSM over a small region with specific topography, land cover or snow characteristics, our study employed a nine members ensemble, three land surface models (JULES, Noah2.7.1, and Noah-MP) and three forcing datasets (ECMWF, GDAS, and MERRA2), over North America for eight winters. The simulations were compared to observations from 15 sites covering a wide range of snow regimes. The comprehensive assessment done in this study to quantify spatial and temporal variability in simulated winter characteristics provides valuable insights on the

regions with extreme values, considerable year-to-year variability and notable disagreement among models. It also sheds light on the potential sources of errors in winter soil temperature estimates and enables the modeling community to move toward a better representation of winter processes in land surface models.

Overall, the uncertainties in winter soil characteristics are more rooted in the differences among the LSMs' than the meteorological forcing data across the North America. For almost all the snow classes except for maritime class, the shorter frozen period and the least number of FT cycles were simulated by Noah2.7.1. The Noah2.7.1 also had the lowest annual minimum temperature at all classes except ephemeral.

Compared to in-situ observations, large biases were observed in all three studied soil characteristics, the annual number of frozen days (overestimated by SEUP members), annual minimum temperature (underestimated) and the annual number of FT cycles (overestimated). The observed and simulated soil temperature evolution during winter highlights the importance of snow during accumulation period. In the absence of adequate insulation from snowpack, soil responds to air temperature fluctuations which could lead to freezing condition. Later, when the snowpack establishes on the ground, soil is already frozen and stays frozen for the rest of season. Therefore, errors in snowpack simulations, accumulation rate, timing and duration of snow appear to play a key role in errors of soil winter characteristics.

The findings of this study point to the need to prioritize modeling efforts to improve LSM performance during the onset of winter as compared to forcing datasets. Processes that impact snow accumulation over ground such partitioning precipitation into rainfall and snowfall as well as soil and snow thermal properties that impact heat transfer appear to be the important initial targets to diagnose the sources of errors in winter soil temperature simulations and improve model performance.

## **Data Availability**

While due to size of SEUP simulation outputs it is not possible to publicly archive them with the available resources, data can be made available from the corresponding author upon request.

Information on data availability to replicate the model simulations is found in Kim et al., 2021.

### Acknowledgement

This research has been supported by the National Aeronautics and Space Administration (grant nos. NNX16AN34G and NNH16ZDA001N). The authors are grateful to Lee Friess for providing a technical review of the draft manuscript.

### References

- Bayard, D., Stähli, M., Parriaux, A. and Flüeler, H., 2005. The influence of seasonally frozen soil on the snowmelt runoff at two Alpine sites in southern Switzerland. *Journal of Hydrology*, 309(1-4), pp.66-84.
- Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. J. H., Ménard, C. B., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*. <https://doi.org/10.5194/gmd-4-677-2011>
- Bohn, T.J., Sonessa, M.Y. and Lettenmaier, D.P., 2010. Seasonal hydrologic forecasting: Do multimodel ensemble averages always yield improvements in forecast skill?. *Journal of Hydrometeorology*, 11(6), pp.1358-1372.
- Bonan, G.B., Hartman, M.D., Parton, W.J. and Wieder, W.R., 2013. Evaluating litter decomposition in earth system models with long-term litterbag experiments: an example using the Community Land Model version 4 (CLM 4). *Global change biology*, 19(3), pp.957-974.
- Chen, F., Barlage, M., Tewari, M., Koschnussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G.Y. and Wen, L., 2014. Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study. *Journal of Geophysical Research: Atmospheres*, 119(4), pp.13-795.
- Chen, F. and Dudhia, J., 2001. Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly weather review*, 129(4), pp.569-585.
- Cherkauer, K.A. and Lettenmaier, D.P., 1999. Hydrologic effects of frozen soils in the upper Mississippi River basin. *Journal of Geophysical Research: Atmospheres*, 104(D16), pp.19599-19610.
- Cho, E., Vuyovich, C. M., Kumar, S. V., Wrzesien, M. L., Kim, R. S., & Jacobs, J. M., 2022. Precipitation Biases and Snow Physics Limitations Drive the Uncertainties in Macroscale Modeled Snow Water Equivalent. *Hydrology and Earth System Sciences*. <https://doi.org/10.5194/hess-2022-136>.

- Clark, M.P., Hendrikx, J., Slater, A.G., Kavetski, D., Anderson, B., Cullen, N.J., Kerr, T., Örn Hreinsson, E. and Woods, R.A., 2011. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, 47(7).
- Cook, B.I., Bonan, G.B., Levis, S. and Epstein, H.E., 2008. The thermoinsulation effect of snow cover within a climate model. *Climate dynamics*, 31(1), pp.107-124.
- Dai, Y., Wei, N., Yuan, H., Zhang, S., Shangguan, W., Liu, S., Lu, X. and Xin, Y., 2019. Evaluation of soil thermal conductivity schemes for use in land surface modeling. *Journal of Advances in Modeling Earth Systems*, 11(11), pp.3454-3473.
- Derber, J.C., Parrish, D.F. and Lord, S.J., 1991. The new global operational analysis system at the National Meteorological Center. *Weather and Forecasting*, 6(4), pp.538-547.
- Dharssi, I., Vidale, P.L., Verhoef, A., Macpherson, B., Jones, C. and Best, M., 2009. New soil physical properties implemented in the Unified Model at PS18.
- Dumedah, G. and Walker, J.P., 2014. Assessment of land surface model uncertainty: A crucial step towards the identification of model weaknesses. *Journal of Hydrology*, 519, pp.1474-1484.
- Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. and Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, 108(D22).
- Ekici, A., Chadburn, S., Chaudhary, N., Hajdu, L.H., Marmy, A., Peng, S., Boike, J., Burke, E., Friend, A.D., Hauck, C. and Krinner, G., 2015. Site level model intercomparison of high latitude and high altitude soil thermal dynamics in tundra and barren landscapes. *The Cryosphere*, 9(4), pp.1343-1361.
- Etchevers, P., Martin, E., Brown, R., Ferz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.J., Essery, R., Fernandez, A. and Gusev, Y., 2004. Validation of the energy budget of an alpine snowpack simulated by several snow models (Snow MIT project). *Annals of Glaciology*, 38, pp.150-158.
- Euskirchen, E.S., McGuire, A.D., Kicklighter, D.W., Zhuang, Q., Klein, J.S., Dargaville, R.J., Dye, D.G., Kimball, J.S., McDonald, K.C., Melillo, J.M. and Romanovsky, V.E., 2006. Importance of recent shifts in soil thermal dynamics on growing season length, productivity, and carbon sequestration in terrestrial high-latitude ecosystems. *Global Change Biology*, 12(4), pp.731-750.
- Farouki, O.T., 1981. *Thermal properties of soils*. Cold Regions Research and Engineering Lab Hanover NH.
- Fuss, C.B., Driscoll, C.T., Groffman, P.M., Campbell, J.L., Christenson, L.M., Fahey, T.J., Fisk, M.C., Mitchell, M.J., Templer, P.H., Durán, J. and Morse, J.L., 2016. Nitrate and dissolved organic carbon mobilization in response to soil freezing variability. *Biogeochemistry*, 131(1), pp.35-47.
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R. and Wargan, K., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate*, 30(14), pp.5419-5454.

- Godfrey, C.M. and Stensrud, D.J., 2008. Soil temperature and moisture errors in operational Eta Model analyses. *Journal of Hydrometeorology*, 9(3), pp.367-387.
- Gouttevin, I., Krinner, G., Ciais, P., Polcher, J. and Legout, C., 2012. Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology. *The Cryosphere*, 6(2), pp.407-430.
- Guan, X., Huang, J., Guo, N., Bi, J. and Wang, G., 2009. Variability of soil moisture and its relationship with surface albedo and soil thermal parameters over the Loess Plateau. *Advances in Atmospheric Sciences*, 26(4), pp.692-700.
- Gubler, S., Endrizzi, S., Gruber, S. and Purves, R.S., 2013. Sensitivities and uncertainties of modeled ground temperatures in mountain environments. *Geoscientific Model Development*, 6(4), pp.1319-1336.
- Guo, D., Yang, M. and Wang, H., 2011. Characteristics of land surface heat and water exchange under different soil freeze/thaw conditions over the central Tibetan Plateau. *Hydrological Processes*, 25(16), pp.2531-2541.
- Han, X., Franssen, H.J.H., Montzka, C. and Vereecken, H., 2014. Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations. *Water resources research*, 50(7), pp.6081-6105.
- He, C., Chen, F., Barlage, M., Liu, C., Newman, A., Tang, W., Ikeda, K. and Rasmussen, R., 2019. Can convection-permitting modeling provide decent precipitation for offline high-resolution snowpack simulations over mountains?. *Journal of Geophysical Research: Atmospheres*, 124(23), pp.12631-12654.
- Idso, S.B., Jackson, R.D., Reginato, R.J., Kimball, B.A. and Nakayama, F.S., 1975. The dependence of bare soil albedo on soil water content. *Journal of Applied Meteorology and Climatology*, 14(1), pp.109-113.
- Johnston, J.M., Houser, P.R., Maggioni, V., Kim, R.S. and Vuyovich, C., 2021. Informing Improvements in Freeze/Thaw State Classification Using Subpixel Temperature. *IEEE Transactions on Geoscience and Remote Sensing*.
- Jordan, R.E., 1991. A one-dimensional temperature model for a snow cover: Technical documentation for SNTherm. 89.
- Joseph, G. and Henry, H.A., 2008. Soil nitrogen leaching losses in response to freeze–thaw cycles and pulsed warming in a temperate old field. *Soil Biology and Biochemistry*, 40(7), pp.1947-1953.
- Kane, D.L., 1980. Snowmelt infiltration into seasonally frozen soils. *Cold Regions Science and Technology*, 3(2-3), pp.153-161.
- Kim, R.S., Kumar, S., Vuyovich, C., Houser, P., Lundquist, J., Mudryk, L., Durand, M., Barros, A., Kim, E.J., Forman, B.A. and Gutmann, E.D., 2021. Snow Ensemble Uncertainty Project (SEUP): quantification of snow water equivalent uncertainty across North America via ensemble land surface modeling. *The Cryosphere*, 15(2), pp.771-791.

- Koren, V., Schaake, J., Mitchell, K., Duan, Q. Y., Chen, F., & Baker, J. M., 1999. A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research Atmospheres*.
- Kreyling, J., 2019. The ecological importance of winter in temperate, boreal, and arctic ecosystems in times of climate change. In *Progress in Botany Vol. 81* (pp. 377-399). Springer, Cham.
- Kreyling, J., Beierkuhnlein, C., Pritsch, K., Schlöter, M. and Jentsch, A., 2008. Recurrent soil freeze–thaw cycles enhance grassland productivity. *New Phytologist*, 177(4), pp.938-945.
- Kreyling, J., Peršoh, D., Werner, S., Benzenberg, M. and Wöllecke, J., 2012. Short-term impacts of soil freeze-thaw cycles on roots and root-associated fungi of *Holcus lanatus* and *Calluna vulgaris*. *Plant and Soil*, 353(1), pp.19-31.
- Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H. and Ménard, C.B., 2018. ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks. *Geoscientific Model Development*, 11(12), pp.5027-5049.
- Kumar, S.V., Wang, S., Mocko, D.M., Peters-Lidard, C.D. and Xia, Y., 2017. Similarity assessment of land surface model outputs in the North American Land Data Assimilation System. *Water Resources Research*, 53(11), pp.8941-8965.
- Lawrence, D.M. and Slater, A.G., 2008. Incorporating organic soil into a global climate model. *Climate Dynamics*, 30(2-3), pp.145-160.
- Lawrence, D.M. and Slater, A.G., 2010. The contribution of snow condition trends to future ground climate. *Climate dynamics*, 34(7-8), pp.961-981.
- Li, X., Wu, T., Wu, X., Chen, J., Zhu, X., Hu, G., Li, R., Qiao, Y., Yang, C., Hao, J. and Ni, J., 2021. 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(3), pp.1751-1761.
- Liston, G. E. and M. Sturm. 2021. *Global Seasonal-Snow Classification, Version 1*. [Indicate subset used]. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: <https://doi.org/10.5067/99FTCYYLAQ0>
- Liu, H., Wang, B. and Fu, C., 2008. Relationships between surface albedo, soil thermal parameters and soil moisture in the semi-arid area of Tongyu, northeastern China. *Advances in Atmospheric Sciences*, 25(5), pp.757-764.
- Liu, X., Yang, W., Zhao, H., Wang, Y. and Wang, G., 2019. Effects of the freeze-thaw cycle on potential evapotranspiration in the permafrost regions of the Qinghai-Tibet Plateau, China. *Science of the total environment*, 687, pp.257-266.
- Liu, Y. and Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water resources research*, 43(7).



- Mahrt, L. and Pan, H., 1984. A two-layer model of soil hydrology. *Boundary-Layer Meteorology*, 29(1), pp.1-20.
- McCauley, C.A., White, D.M., Lilly, M.R. and Nyman, D.M., 2002. A comparison of hydraulic conductivities, permeabilities and infiltration rates in frozen and unfrozen soils. *Cold Regions Science and Technology*, 34(2), pp.117-125.
- Mitchell, K.E., Lohmann, D., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L. and Higgins, R.W., 2004. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, 109(D7).
- Mitchell, K., Wei, H., Lu, S., Gayno, G. and Meng, J., 2005. NCEP implements major upgrade to its medium-range global forecast system, including land-surface component. *GEWEX News*, 15(4), pp.8-9.
- Mölders, N. and Walsh, J.E., 2004. Atmospheric response to soil-frost and snow in Alaska in March. *Theoretical and applied climatology*, 77(1), pp.77-105.
- Molod, A., Takacs, L., Suarez, M. and Bacmeister, J., 2015. Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MEK2. *Geoscientific Model Development*, 8(5), pp.1339-1356.
- Molteni, F., Buizza, R., Palmer, T. N., & Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1256/smsqj.52904>
- Mudryk, L.R., Derksen, C., Kushner, P.J. and Brown, R., 2015. Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. *Journal of Climate*, 28(20), pp.8037-8051
- Niu, G.Y. and Yang, Z.L., 2006. Effect of frozen soil on snowmelt runoff and soil water storage at a continental scale. *Journal of Hydro-meteorology*, 7(5), pp.937-952.
- Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*. <https://doi.org/10.1029/2010JD015139>
- Perfect, E. and Williams, P.J., 1980. Thermally induced water migration in frozen soils. *Cold Regions Science and Technology*, 3(2-3), pp.101-109.
- Peters-Lidard, C.D., Blackburn, E., Liang, X. and Wood, E.F., 1998. The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *Journal of the Atmospheric Sciences*, 55(7), pp.1209-1224.
- Peters-Lidard, C.D., Houser, P.R., Tian, Y., Kumar, S.V., Geiger, J., Olden, S., Lighty, L., Doty, B., Dirmeyer, P., Adams, J. and Mitchell, K., 2007. High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering*, 3(3), pp.157-165.

- Rai, A., Saha, S.K. and Sujith, K., 2019. Implementation of snow albedo schemes of varying complexity and their performances in offline Noah and Noah coupled with NCEP CFSv2. *Climate Dynamics*, 53(3), pp.1261-1276
- Repo, T., Sirkiä, S., Heinonen, J., Lavigné, A., Roitto, M., Koljonen, E., Sutinen, S. and Finér, L., 2014. Effects of frozen soil on growth and longevity of fine roots of Norway spruce. *Forest ecology and Management*, 313, pp.112-122.
- Robock, A., Luo, L., Wood, E.F., Wen, F., Mitchell, K.E., Houser, P.R., Schaake, J.C., Lohmann, D., Cosgrove, B., Sheffield, J. and Duan, Q., 2003. Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season. *Journal of Geophysical Research: Atmospheres*, 108(D22).
- Roesch, A., 2006. Evaluation of surface albedo and snow cover in AR4 coupled climate models. *Journal of Geophysical Research: Atmospheres*, 111(D15).
- Roy, D., Jia, X., Chu, X. and Jacobs, J.M., 2021. Hydraulic Conductivity Measurement for Three Frozen and Unfrozen Soils in the Red River of the North Basin. *Transactions of the ASABE*, 64(3), pp.761-770.
- Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H.M., Pan, H.L., Moorthi, S., Behringer, D. and Stokes, D., 2006. The NCEP climate forecast system. *Journal of Climate*, 19(15), pp.3483-3517.
- Seyfried, M.S. and Murdock, M.D., 1997. Use of air permeability to estimate infiltrability of frozen soil. *Journal of Hydrology*, 202(1-4), pp.95-107.
- Sharratt, B.S., Baker, D.G., Wall, D.B., Skaggs, R.H. and Ruschy, D.L., 1992. Snow depth required for near steady-state soil temperatures. *Agricultural and Forest Meteorology*, 57(4), pp.243-251.
- Shi, M., Fisher, J.B., Brzostek, E.R. and Phillips, R.P., 2016. Carbon cost of plant nitrogen acquisition: global carbon cycle impact from an improved plant nitrogen cycle in the Community Land Model. *Global change biology*, 22(3), pp.1299-1314.
- Stieglitz, M., Déry, S.J., Romanovsky, V.E. and Osterkamp, T.E., 2003. The role of snow cover in the warming of arctic permafrost. *Geophysical Research Letters*, 30(13).
- Vestgarden, L.S. and Austnes, K., 2009. Effects of freeze–thaw on C and N release from soils below different vegetation in a montane system: a laboratory experiment. *Global Change Biology*, 15(4), pp.876-887.
- Vuyovich, C., Kumar, S., Mudryk, L., Kim, R., Lundquist, J., Durand, M., Derksen, C., Barros, A., Houser, P. and Kim, E., 2019, July. Evaluation of Seasonal Water Budget Components Over the Major Drainage Basins of North America Using an Ensemble-Based Land Surface Model Approach. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5617-5620). IEEE.
- Wang, Q., Liu, J. and Wang, L., 2017. An experimental study on the effects of freeze–thaw cycles on phosphorus adsorption–desorption processes in brown soil. *RSC advances*, 7(59), pp.37441-37446.

- Wang, W., Rinke, A., Moore, J.C., Ji, D., Cui, X., Peng, S., Lawrence, D.M., McGuire, A.D., Burke, E.J., Chen, X. and Decharme, B., 2016. Evaluation of air–soil temperature relationships simulated by land surface models during winter across the permafrost region. *The Cryosphere*, 10(4), pp.1721-1737.
- Wang, K., Wang, P., Liu, J., Sparrow, M., Haginoya, S. and Zhou, X., 2005. Variation of surface albedo and soil thermal parameters with soil moisture content at a semi-desert site on the western Tibetan Plateau. *Boundary-Layer Meteorology*, 116(1), pp.117-129.
- Wieder, W.R., Hartman, M.D., Sulman, B.N., Wang, Y.P., Koven, C.D. and Bonan, G.B., 2018. Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global change biology*, 24(4), pp.1563-1579.
- Xia, K., Luo, Y. and Li, W., 2011. Simulation of freezing and melting of soil on the northeast Tibetan Plateau. *Chinese Science Bulletin*, 56(20), pp.2145-2155.
- Xia, Y., Ek, M., Sheffield, J., Livneh, B., Huang, M., Wei, H., Feng S., Luo, L., Meng, J. and Wood, E., 2013. Validation of Noah-simulated soil temperature in the North American land data assimilation system phase 2. *Journal of applied meteorology and climatology*, 52(2), pp.455-471.
- Yang, K. and Wang, C., 2019. Water storage effect of soil freeze-thaw process and its impacts on soil hydro-thermal regime variations. *Agricultural and Forest Meteorology*, 265, pp.280-294.
- Yen, Y.C., 1965. Heat transfer characteristics of naturally compacted snow.
- Zhang, K., Kimball, J.S., Kim, Y. and McDonald, K.C., 2011. Changing freeze-thaw seasons in northern high latitudes and associated influences on evapotranspiration. *Hydrological Processes*, 25(26), pp.4142-4151.
- Zhang, G., Chen, Y. and Li, J., 2021. Effects of organic soil in the Noah-MP land-surface model on simulated skin and soil temperature profiles and surface energy exchanges for China. *Atmospheric Research*, 249, p.105284.
- Zhang, T., 2005. Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics*, 43(4).
- Zhang, T., Osterkamp, T.E. and Stamnes, K., 1996. Influence of the depth hoar layer of the seasonal snow cover on the ground thermal regime. *Water Resources Research*, 32(7), pp.2075-2086.
- Zhang, X. and Sun, S., 2011. The impact of soil freezing/thawing processes on water and energy balances. *Advances in Atmospheric Sciences*, 28(1), pp.169-177.
- Zhu, J. and Liang, X.Z., 2005. Regional climate model simulation of US soil temperature and moisture during 1982–2002. *Journal of Geophysical Research: Atmospheres*, 110(D24).

## 5.1 Author Statement:

**Mahsa Moradi:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing

**Eunsang Cho** Conceptualization, Methodology, Writing - Review & Editing

**Jennifer M. Jacobs:** Conceptualization, Methodology, Writing - Review & Editing, Funding acquisition, Supervision

**Carrie M. Vuyovich:** Writing - Review & Editing

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

**Highlights:**

- The uncertainties in winter soil characteristics are more rooted in the differences among the land surface models than the meteorological forcing data across North America.
  - Notable biases were found in the annual number of frozen days (overestimated), annual minimum temperature (underestimated) and the annual number of FT cycles (overestimated) in comparison with in-situ observations.
  - Snow processes especially in the accumulation period play a critical role in temporal evolution of soil thermal processes.