Permafrost Variability over the Northern Hemisphere Based on the MERRA-2 Reanalysis

Jing Tao[1,2,a], Randal D. Koster[2], Rolf H. Reichle[2], Barton A. Forman[3], Yuan Xue[3,b], Richard H. Chen[4], Mahta Moghaddam[4]

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA.

[2]Global Modelling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

[3]Department of Civil and Environmental Engineering, University of Maryland, College Park, Maryland, USA.

[4]Department of Electrical Engineering, University of Southern California, Los Angeles, California, USA.

[a]Now at Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; and Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington, USA.

[b]Now at George Mason University, Fairfax, Virginia, USA.

Correspondence to: Jing Tao (jingtao@lbl.gov)

Abstract. This study introduces and evaluates a comprehensive, model-generated dataset of Northern Hemisphere permafrost conditions at 81-km^2 resolution. Surface meteorological forcing fields from the Modern-Era Retrospective Analysis for Research and Applications-2 (MERRA-2) reanalysis were used to drive an improved version of the land component of MERRA-2 in middle-to-high northern latitudes from 1980 to 2017. The resulting simulated permafrost distribution across the Northern Hemisphere mostly captures the observed extent of continuous and discontinuous permafrost but misses the ecosystem-protected permafrost zones in western Siberia. Noticeable discrepancies also appear along the southern edge of the permafrost regions where sporadic and isolated permafrost types dominate. The evaluation of the simulated active layer thickness (ALT) against remote sensing retrievals and in-situ measurements demonstrates reasonable skill except in Mongolia. The RMSE (bias) of climatological ALT is 1.22 m (-0.48 m) across all sites and 0.33 m (-0.04 m) without the Mongolia sites.

In northern Alaska, both ALT retrievals from airborne remote sensing for 2015 and the corresponding simulated ALT exhibit limited skill versus in-situ measurements at the model scale. Although the simulated ALT agrees well with the remotely sensed ALT when considering measurement uncertainty, it demonstrates larger spatial variability than the ALT retrievals. Controls on the spatial variability of ALT are examined with idealized numerical experiments focusing on northern Alaska; meteorological forcing and soil types are found to have dominant impacts on the spatial variability of ALT, with vegetation also playing a role through its modulation of snow accumulation. A correlation analysis further reveals that accumulated above-freezing air temperature and maximum snow water equivalent explain most of the year-to-year variability of ALT nearly everywhere over the model-simulated permafrost regions.

1 Introduction

Permafrost is an important component of the climate system, and its variations can have significant impacts on climate and society. Of deep concern is a potential positive feedback loop by which carbon stored within permafrost regions is released through global warming, thereby adding greenhouse gases to the atmosphere that accelerate the warming further (Dorrepaal et al., 2009; Schuur et al., 2009; MacDougall et al., 2012; Schuur et al., 2015). Communities and infrastructure in ice-rich permafrost regions are particularly vulnerable to land subsidence and infrastructure damage caused by permafrost thaw (Nelson et al., 2001; Liu et al., 2010; Guo and Sun, 2015).

Permafrost variations, including pronounced permafrost degradation due to a warming climate, have been reported for many regions, including Alaska (Nicholas and Hinkel, 1996; Osterkamp and Romanovsky, 1996; Jorgenson et al., 2001; Hinkel and Nelson, 2003; Jafarov et al., 2012; Liu et al., 2012; Jones et al., 2016; Batir et al., 2017), Canada (Chen et al., 2003; James et al., 2013), Norway (Gisnas et al., 2013), Sweden (Pannetier and Frampton, 2016), Russia (Romanovsky et al., 2007; Romanovsky et al., 2010), Mongolia (Sharkhuu and Sharkhuu, 2012), and the Qinghai–Tibet Plateau (Zhou et al., 2013; Wang et al., 2016a; Lu et al., 2017; Ran et al., 2018). For the entire Northern Hemisphere, rapidly accelerated permafrost degradation in recent years has been reported by Luo et al. (2016) based on in-situ measurements at a point-scale or a spatially-aggregated

scale (up to 1000m×1000m) from the Circumpolar Active Layer Monitoring (CALM) network. However, the current state and evolution of global permafrost (including permafrost temperature, ice content, and degradation rates) are still largely unknown across much of the Northern latitudes.

5   The impact of a changing climate on permafrost dynamics must depend on local site characteristics. Subsurface heat transfer processes and active layer thickness (ALT; the maximum thaw depth at the end of the thawing season) are influenced by more than surface meteorological forcing – they are also influenced by vegetation type, surface organic layer characteristics, soil properties and soil moisture (Stieglitz et al., 2003; Shur and Jorgenson, 2007; Yi et al., 2007; Luetschg et al., 2008; Dankers et al., 2011; Johnson et al., 2013; Jean and Payette, 2014; Yi et al., 2015; Fisher et al., 2016; Matyshak et al., 2017; Tao et al.,

10   2017). Understanding the contributions from the different controls on ALT (and permafrost conditions in general) is crucial for assessing permafrost behaviour and its resilience to a warming climate.

Physically-based numerical model simulations are potentially useful for quantifying and understanding these dynamics at large spatial scales; they can also provide insights into associated impacts on the global carbon cycle. Permafrost dynamics can be

15   modelled, for example, by driving a land surface model (LSM) offline (i.e., uncoupled from an atmospheric model) with meteorological forcing data (including air temperature, radiation, precipitation, etc.) from some credible source. LSMs that have been used to quantify large-scale permafrost patterns (i.e., distributions and thermal states) and their interactions with a warming climate include, for example, the Joint UK Land Environment Simulator (JULES, Dankers et al., 2011), the ORganizing Carbon and Hydrology in Dynamic EcosystEms (ORCHIDEE) - aMeliorated Interactions between Carbon and

20   Temperature (ORCHIDEE-MICT, Guimberteau et al., 2018), the Catchment Land Surface Model (CLSM, Tao et al., 2017), and the Community Land Model (Alexeev et al., 2007; Nicolsky et al., 2007a; Yi et al., 2007; Lawrence and Slater, 2008; Lawrence et al., 2008; Lawrence et al., 2012; Koven et al., 2013; Chadburn et al., 2017; Guo and Wang, 2017). Most of these land models were run at coarse spatial resolutions, e.g., ranging from $0.5° \times 0.5°$ to $1.8° \times 3.6°$ for LSMs participating in the Permafrost Carbon Network (PCN) (Wang et al., 2016a) and from $0.188° \times 0.188°$ to $4.10° \times 5°$ for the models participating

in the Coupled Model Intercomparison Project phase 5 (CMIP5) (Koven et al., 2013; https://portal.enes.org/data/enes-model-data/cmip5/resolution).

Differences in the permafrost behaviour simulated with these models reflect model-specific process representations as well as biases associated with different meteorological forcing datasets (Barman and Jain, 2016; Wang et al., 2016a; Wang et al., 2016b; Guo et al., 2017; Guimberteau et al., 2018). Such forcing biases are difficult to avoid given the sparsity of direct observations of meteorological variables in most parts of the high latitudes. Even reanalyses, which assimilate a variety of global observations, inevitably have biases in high latitudes due to observation sparsity in cold regions combined with the many challenges of physical process modelling. Nevertheless, despite these issues, permafrost behaviour simulated with LSMs driven offline by reanalysis forcing fields can still be useful for understanding the impacts of climate variability on permafrost. The present paper utilizes this approach. Specifically, we generate here a dataset of Northern Hemisphere permafrost conditions by driving an updated version of NASA's Catchment Land Surface Model (CLSM) with Modern-Era Retrospective Analysis for Research and Applications-2 (MERRA-2; Gelaro et al., 2017) surface meteorological forcing fields for the middle-to-high latitudes across the Northern Hemisphere over the period 1980-2017. We perform the simulations at 81 km^2 resolution encompassing permafrost areas in the middle-to-high latitudes of the Northern Hemisphere. This resolution is high relative to most existing modelling studies at the global scale; published simulations at higher resolution are limited to plot scales (e.g., CALM-site scale in Shiklomanov et al. (2010)), landscape scales (e.g., polygonal tundra landscape scale in Kumar et al. (2016)), or regional scales (e.g., 4 km^2 in Jafarov et al. (2012) covering Alaska; 1 km^2 in Gisnas et al. (2013) covering Norway).

Due to the sparsity of in-situ measurements at the regional to global scale, evaluating the spatial pattern of ALT produced by any such simulation remains challenging. Indeed, it is difficult to compare the simulated values at model resolutions with in-situ observations taken at the point scale unless the measurement point is uniformly representative of the area covered by the model grid cell or the representation errors associated with the point-to-grid comparison are well defined. Remotely sensed permafrost products, which provide a unique source of spatially distributed ALT at the landscape-scale, may provide help in

this regard. Existing remote sensing ALT products have been retrieved from ground-based Ground Penetrating Radar (GPR) (Chen et al., 2016a; Jafarov et al., 2017), airborne polarimetric Synthetic Aperture Radar (SAR), and spaceborne interferometric SAR (Liu et al., 2012; Li et al., 2015; Schaefer et al., 2015). These ALT products are available at the landscape-scale and can complement our modelling analysis. In this study, we use remote sensing information from the NASA Airborne

5 Microwave Observatory of Subcanopy and Subsurface (AirMOSS) mission. In 2015, AirMOSS acquired P-band (420-440 MHz) SAR observations over portions of northern Alaska from which Chen et al. (2019) retrieved regional estimates of ALT and soil layer dielectric properties that are related to soil moisture and freeze/thaw states. In their study, Chen et al. (2019) mainly focus on the development and improvement of the ALT retrieval algorithm, whereas the present study uses the ALT retrievals in combination with in-situ measurements to aid in assessing the (fully independent) ALT simulations.

In the present paper, we evaluate our simulated permafrost extent and ALTs against an observations-based permafrost distribution map and against multi-year in-situ observations. We also compare the skill of our model estimates to that of the AirMOSS ALT retrievals. In these comparisons, we account for uncertainty to the extent possible. Overall, we pursue three scientific objectives: 1) evaluate the relative importance of the factors that determine the spatial variability of ALT, 2) evaluate

15 CLSM-simulated ALT and permafrost extent against observations, and 3) quantify and assess the large-scale characteristics of ALT (in terms of means and interannual variability) in Northern Hemisphere permafrost regions from 1980 through 2017. As a side benefit, the side-by-side comparison of modelled and remotely sensed ALT estimates is an important first step toward combining this information effectively in future model-data fusion efforts. Section 2 below describes the model and datasets used in this study, Section 3 describes methods, and Section 4 provides results. Our findings are summarized and discussed

20 in Section 5.

2 Model and data sets

2.1 NASA Catchment Land Surface Model (CLSM)

CLSM is the land model component of NASA's Goddard Earth Observing System (GEOS) Earth system model and was part of the model configuration underlying the MERRA-2 reanalysis product (Reichle et al., 2017a; Gelaro et al., 2017). CLSM explicitly accounts for sub-grid heterogeneity in soil moisture characteristics with a statistical approach (Koster et al., 2000; Ducharne et al., 2000). The land fraction within each computational unit (or grid cell) is partitioned into three soil moisture regimes, namely the wilting (i.e., non-transpiring), unsaturated, and saturated area fractions. Over each of the three moisture regimes, a distinct parameterization is applied to estimate the relevant physical processes (e.g., runoff and evapotranspiration). This version of CLSM includes a three-layer snow model that estimates the evolution of snow water equivalent (SWE), snow depth, and snow heat content (Stieglitz et al., 2001) in response to the forcing data. The snow model accounts for key physical mechanisms that contribute to the growth and ablation of the snowpack, including snow accumulation, aging, melting, and refreezing. The model also includes the insulation of the ground from the atmosphere by the snowpack. The CLSM subsurface heat transfer module uses an explicit finite difference scheme to solve the heat diffusion equation for six soil layers (0-0.1m, 0.1-0.3m, 0.3-0.7m, 0.7-1.4m, 1.4-3m, and 3-13m). The soil layer thicknesses increase with depth following a geometric series for consistency with the linear heat diffusion calculation (Koster et al., 2000). A no-heat-flux condition is employed at 13m depth.

The updated version of CLSM used here includes modifications aimed at improving permafrost simulation. It accounts, for example, for the impact of soil carbon on the soil thermal properties with soil porosity, thermal conductivity, and specific heat capacity calculated separately for mineral soil and soil carbon, after which the two are averaged using a carbon-weighting scheme. Higher (lower) soil carbon content, therefore, results in lower (higher) soil thermal conductivity. The updated version produces more realistic subsurface thermodynamics in cold regions than does the original scheme (Tao et al., 2017). This version of CLSM, however, does not include dynamic soil carbon pools.

Particularly relevant to the present analysis is our calculation of ALT from CLSM simulation output. We compute ALT from the simulated soil temperature profile and the ice content within the soil layer that contains the thawed-to-frozen transition. Precisely, the thawed-to-frozen depth is calculated as:

$$z_{bottom}(l) - f_{ice}(l, t) \times \Delta z(l), \tag{1}$$

where layer l is the deepest layer that is fully or partially thawed, $z_{bottom}(l)$ represents the depth at the bottom of layer l, $f_{ice}(l, t)$ is the fraction of ice in layer l at time t (i.e., $f_{ice}(l, t) \in [0 \ 1]$), and $\Delta z(l)$ is the thickness of layer l. To identify layer l, we use a 0°C degree temperature threshold. Specifically, T > 0°C degree indicates that a layer is fully thawed, T = 0°C degree indicates that a layer is partially thawed, and T < 0°C degree indicates that a layer is fully frozen. That is, layer l is the deepest layer that satisfies $T(l) \geq 0$°C. Equation (1) then expresses that the thawed-to-frozen depth is equal to the bottom depth of the layer l but adjusted upward according to the ice fraction within the partially thawed layer l. This upward adjustment, by the way, allows the thawed-to-frozen depth to be a continuous variable; it is not quantized to the imposed layer depths. We search for the deepest l if multiple thawed-to-frozen transitions are present (e.g., if a seasonal frost at the surface is separated from the permafrost below by a thawed soil layer). The annual ALT for a given year, then, is defined as the deepest depth at which a thawed-to-frozen transition occurs within that year. Note that the calculation of equation (1) is made at the scale of a model grid cell, and thus features such as talik are not represented if they occur at sub-grid cell scale.

We drive the improved CLSM version of Tao et al. (2017) in a land-only (offline) configuration across permafrost areas in the Northern Hemisphere. The simulation domain, shown in Figure 1a, covers the major permafrost regions of the Northern Hemisphere middle-to-high latitudes for which soil carbon data are available from the Northern Circumpolar Soil Carbon Database version 2 (NCSCDv2, https://bolin.su.se/data/ncscd/) (Hugelius et al., 2013a; Hugelius et al., 2013b). The NCSCDv2 data are used to calculate the CLSM soil thermal properties used in the simulations (Tao et al., 2017). The model simulation covered the period from 1980 to 2017 and was performed at an 81-km^2 spatial resolution on the 9-km Equal-Area Scalable Earth grid, version 2 (Brodzik et al., 2012).

Surface meteorological forcings were extracted from the MERRA-2 reanalysis data, which are provided at a resolution of 0.5° latitude × 0.625° longitude (Global Modeling and Assimilation Office (GMAO), 2015a, b). At latitudes south of 62.5°N within our simulation domain, the MERRA-2 precipitation forcing used here is informed by gauge measurements from the daily 0.5° global Climate Prediction Center Unified gauge product (Chen et al., 2008) as described in (Reichle et al., 2017b). We further rescaled the precipitation to the long-term, seasonally varying climatology of the Global Precipitation Climatology Project version 2.2 product (Huffman et al., 2009). Further details regarding model parameters and forcing inputs are found in Tao et al. (2017).

The model was spun-up for 180 years by looping five successive times through the 36-year period of MERRA-2 forcing from 1 January 1980 to 1 January 2016 in order to achieve a quasi-equilibrium state. The spatial terrestrial state variables at the end of the fifth loop were used to initialize the model for the final simulation experiment from 1980 to 2017.

2.2 Remotely Sensed ALT from AirMOSS

Radar backscatter measurements are sensitive to changes in the soil dielectric constant (or relative permittivity) which in turn are associated with changes in soil moisture and the soil freeze-thaw state. Based on this relationship, Chen et al. (2019) used the AirMOSS airborne P-band (420-440 MHz) synthetic aperture radar (SAR) observations collected during two campaigns in 2015 to estimate ALT in northern Alaska. As shown in Figure 2a, the AirMOSS flights originated from Fairbanks International Airport and headed west toward the Seward Peninsula (HUS, KYK, COC), then turned back east (KGR) prior to heading north towards the Arctic coast overpassing Ambler (AMB), Ivotuk (IVO), and Atqasuk (ATQ). From there, the flights turned south again, flying over Barrow (BRW), Deadhorse (DHO), and Coldfoot (CFT) en route to Fairbanks. In the present paper, the remotely-sensed ALT retrievals are compared with in-situ observations and CLSM-simulated ALT.

Chen et al. (2019) used AirMOSS P-band SAR observations at two different times to retrieve active layer properties: (1) acquisitions on 29 August 2015 when the downward thawing process approximately reached its deepest depth (i.e., the bottom

of the active layer), and (2) acquisitions on 1 October 2015 when the active layer started to refreeze from the surface while the bottom of the active layer remained thawed. ALT was assumed constant from late August to early October because over this period changes in thawing depth are found typically negligible (Carey and Woo, 2005; Chen et al., 2016b; Zona et al., 2016). Strictly speaking, the radar retrievals represent the approximate thaw depth of the thawed-to-frozen boundary on 29 August

5 2015 and 1 October 2015. The unknown, true ALT for 2015 might occur later if the thawing continued and the maximum thaw depth occurred after the October flight time. Based on an analysis of in-situ observations (not shown), however, it is rare that this occurs, and the subsequent impact on the estimated ALT value would be relatively small in any case. We, therefore, equate the retrieved thaw depth with ALT.

10 In the retrieval algorithm, used a three-layer dielectric structure to represent the active layer and underlying permafrost. In their algorithm, the two uppermost layers together constitute the active layer that accounts for a top, unsaturated zone and an underlying, saturated zone. The bottommost (third) layer of the retrieval model structure represents the permafrost. Because the soil moisture at saturation only depends on the porosity of the soil medium, the dielectric constant of the saturated zone in the active layer is assumed constant over the time window. An iterative forward-model inversion scheme was used to

15 simultaneously retrieve the dielectric constants and layer thicknesses of the three-layer dielectric structure from the SAR observations collected on 29 August 2015 and 1 October 2015. Note that the retrieved ALT cannot exceed the radar sensing depth of about 60 cm. This is the depth below which the AirMOSS radar is expected to lose sensitivity to subsurface features, and it is calculated based on the radar system noise floor and calibration accuracy. Therefore, any retrieved ALT larger than 60 cm is expected to have large uncertainties, and the error is further expected to grow linearly as the retrieved values of ALT

20 essentially "saturate." This limitation may also lead to underestimates of the actual thaw depth.

In this study, we focus on the retrievals of four flight lines across the Alaska North Slope, including IVO (Ivotuk), ATQ (Atqasuk), BRW (Barrow), and DHO (Deadhorse) as shown in Figure 2a. These four transects cover areas with light to moderate vegetation. Since the radar scattering model is only applicable to bare surfaces or lightly vegetated tundra areas

25 (Chen et al., 2019), the ALT estimates derived for IVO, ATQ, BRW, and DHO are considered more accurate than ALT

retrievals for the remaining transects, which include more vegetated areas. Moreover, some of the southern transects cover discontinuous permafrost where the ALT often exceeds the P-band radar sensing depth of about 60 cm and thus the retrievals have large uncertainty (Chen et al., 2019).

5   **2.3 Circum-Arctic Permafrost Conditions and In-situ Observations of ALT**

The permafrost distribution simulated by CLSM is evaluated against the observations-based Circum-Arctic Map of Permafrost and Ground-Ice Conditions (Brown et al., 2002) shown in Figure 1b. The map is based on the distribution and character of permafrost and ground ice using a physiographic approach. Permafrost conditions are categorized into four classes: continuous (90-100%), discontinuous (50-90%), sporadic (10-50%), and isolated (0-10%), where the numbers in parentheses indicate the
10   area fraction of permafrost extent.

In-situ observations of ALT obtained by the CALM network (https://www2.gwu.edu/~calm/; Brown et al., 2000) were used to evaluate both the AirMOSS ALT retrievals and CLSM-simulated ALT results. The CALM network provides observations from 1990 to 2017, but few sites have records in the early 1990s. We did not use measurements that were flagged as having
15   been taken too early in the season or under unusual conditions (e.g., after the site was burned or covered with lava, which occurred at sites R30A and R30B in Kamchatka). In total, there are 220 sites located within the CLSM simulation domain (Figure 1b), and we use 213 sites to evaluate results. Thaw depth measurements are usually made at the end of the thawing season. Most of the CALM sites (129 out of the 213 sites used here) employ a spatially-distributed mechanical probing method to measure thaw depths along a transect or across a rectangular grid ranging in size from 10m×10m to 1000m×1000m. At 20
20   sites, thaw tubes or boreholes are used to measure the thaw depth. At 63 sites, ground temperature measurements from boreholes are used to infer thaw depth. For the remaining site, no information about the measurement method is available. Only point-scale measurements are available from the thaw tube/borehole and ground temperature sites (including, e.g., the sites in Mongolia).

In addition, daily in-situ observations of soil temperature profiles at ten Alaskan sites from the Permafrost Laboratory at the University of Alaska Fairbanks (UAF) (http://permafrost.gi.alaska.edu/sites_map; Romanovsky et al., 2009) were used to infer thawed-to-frozen depth using the 0ºC degree threshold and to complement the CALM ALT observations in Alaska. Table 1 provides the coordinates and measuring methods of the UAF in-situ sites. The UAF measurements were used along with the CALM data to evaluate the ALT estimates derived from the CLSM simulation and the AirMOSS radar observations for the North Slope of Alaska in section 4.1.

3 Methods

3.1 Comparing ALT from In-situ Observations, AirMOSS Retrievals, and CLSM Results in Alaska

First, we compare AirMOSS radar retrievals and CLSM simulation results of ALT for 2015 against each other and against in-situ observations: (i) we compare the spatial patterns of the AirMOSS retrievals with those of the model-simulated ALT over northern Alaska; and (ii) we evaluate the simulated ALT against both the AirMOSS retrievals and in-situ observations from the CALM and UAF networks. We rely on several metrics to evaluate the model and radar-retrieval performance, including bias, root mean square error (RMSE), and correlation coefficient (R). The results are discussed in section 4.1.

We conducted the intercomparison at the model scale. The radar retrievals were provided at 2-arcsec $\times$ 2-arcsec (roughly 20 m x 60 m in the Arctic) resolution whereas the CLSM-simulated ALTs are at 81 km^2. We thus aggregated the AirMOSS retrievals to the CLSM model grid by averaging all the retrieval data points within each 81 km^2 model grid cell. Only model grid cells that were at least 30% covered by radar retrievals were used in the comparison. The AirMOSS transects cover several different regions with different climatologic regimes, topography, vegetation, and soil type (Figure 2). Note that although the vegetation class used in the model (Figure 2b) suggests the presence of dwarf trees over the Alaska North Slope, the actual satellite-based LAI, vegetation height, greenness fraction and albedo will still instruct the model that the tree cover there is extremely sparse. The data sources for these vegetation-related boundary conditions can be found in Table 1 of Tao et al.

(2017). Overall, the variability of ALT along these transects encompasses the influence of a variety of factors at the regional scale.

The daily UAF in-situ soil temperature profile observations on the AirMOSS flight date (29 August 2015) were used to
5   calculate the thawed-to-frozen depth (i.e., approximated ALT). The ALT measurements at all of the 13 CALM sites covered by the AirMOSS transects were obtained in August of 2015 (Table 1). Among them, eight CALM sites obtained ALT measurements slightly earlier than the overflight date (within at most 18 days from 29 August 2015). Nevertheless, we assume that these earlier measurements still represent the thaw depth at the end of August reasonably well. Prior to comparison with the model results and the aggregated radar retrievals, the distributed measurements for a given CALM site (see sampling
10 methods in Table 1) were averaged into a single value. If multiple CALM or UAF sites lay within a single CLSM grid cell, a single "spatially-averaged" observed value was computed for the cell.

We employed the strategy of Schaefer et al. (2015) to handle the uncertainty propagation, i.e., adding in quadrature the uncertainty components from each scale/level involved (see the supplementary file for a detailed description). For AirMOSS
15 retrievals, the sampling uncertainty of mean ALT at the 81 km^2 model grid-cell scale is negligible given the large sampling size and the fact that the retrieval uncertainty dominates the overall uncertainty (see supplementary file). Here, we use a nominal estimate of 0.15 m to represent the AirMOSS uncertainty (i.e., the average of the lower and upper bound of the actual retrieval uncertainty for individual radar pixels as discussed by Chen et al. (2019)).

20 When comparing in-situ measurements with model results at the 81 km^2 scale (i.e., a point-to-grid comparison), the ultimate measurement uncertainty propagated from the point-scale measurements to the 81 km^2 scale is, for all intents and purposes, unknown due to a lack of sufficient measurements over the 81 km^2 scale to compute upscaling errors (see supplementary file). We thus show instead the standard deviation of CALM measurements to illustrate, in a highly approximate way, the spatial representativeness error of the in-situ measurements – a small (large) standard deviation represents a homogeneous
25 (heterogeneous) area in terms of ALT, meaning that the in-situ mean likely can (cannot) represent an average over a larger

scale, assuming the site-scale heterogeneity is somewhat transferable to the larger scale. Such transferability might only apply to the largest in-situ site scales (e.g., 1000 m × 1000 m) to the model grid-scale (81 km^2) and is thus, in general, questionable. We thus make no claim here that the standard deviations shown represent true uncertainty levels.

5 **3.2 Idealized Experiments**

After comparing the spatial patterns of the AirMOSS retrievals with the CLSM-simulated ALT results, we then investigate the factors that affect the spatial variability of ALT through a series of idealized experiments. Specifically, we repeated the simulation along the AirMOSS transects multiple times, each time removing the spatial variation in some aspect of the model forcing or parameters and then quantifying the resulting impact on ALT variability.

For these supplemental simulations, we first identified a grid cell within the IVO transect (shown in Figure 2a) that represents roughly average (typical) conditions across the ten different transects. In the first idealized experiment, we then modified the baseline configuration by applying the surface meteorological forcing data from the selected representative grid cell within the IVO transect to all grid cells along all AirMOSS transects. Thus, in this modified simulation (HomF, for homogenized forcing),

15 spatial variability in meteorological forcing is artificially removed. All model parameters related to soil type and vegetation, however, remain spatially variable, matching those in the baseline simulation. In the next idealized experiment (HomF&Veg), we further replaced the vegetation-related parameters (including vegetation class, vegetation height, and time-variable Leaf Area Index (LAI) and greenness) along the AirMOSS transects using the corresponding parameters from the representative grid cell, which is characterized by dwarf tree vegetation cover. Thus, in this simulation, spatial variability in both forcing

20 and vegetation is artificially removed.

In a third idealized experiment (HomF&Veg&Soil), spatial variability in soil type and topography-related model parameters is removed along with that of the forcing and vegetation. The homogenized parameters include soil organic carbon content, porosity, saturated hydraulic conductivity, Clapp-Hornberger parameters, wilting point, soil class, sand and clay fraction,

vertical decay factor for transmissivity, baseflow parameters, area partitioning parameters, and timescale parameters for moisture transfer (Ducharne et al., 2000; Koster et al., 2000). Here we use an intermediate soil carbon content value (i.e., 40 kg/m^2) for the homogenization; recall that the carbon content impacts the soil thermal properties (see section 2.1). Our investigation reveals that the model sensitivity to soil carbon content is much larger for lower soil organic carbon content

5   (SOC) than for higher SOC, and easily gets saturated for high SOC (i.e., larger than ~100 kg/m^2) (not shown). Thus, we trust that 40 kg/m^2 is an appropriate value representing an intermediate SOC condition. All other soil parameters are homogenized to those at the representative grid cell.

Finally, we investigate potential nonlinearities by conducting two additional experiments: one in which we homogenized both

10   the vegetation and soil parameters (HomVeg&Soil) and another in which we homogenized both forcing and soil parameters (HomF&Soil). Put differently, in experiment HomVeg&Soil only the forcing varies along the transects, whereas in experiment HomF&Soil, only the vegetation parameters vary along the transects. Combined with the experiment HomF&Veg (in which only soil properties vary along the transects), these three experiments show in a different way how each individual factor (forcing, vegetation, or soil) can contribute to ALT variability. Table 2 provides a summary of these idealized experiments.

15   Taken together, the six experiments (including the baseline) allow us to identify the individual contribution of each factor to the ALT variability along the AirMOSS transects. The results are discussed in section 4.2.

3.3 Quantifying ALT Spatiotemporal Characteristics

In section 4.3 we quantify the large-scale characteristics of ALT over the Northern Hemisphere for the current climate (1980

20   - 2017) as determined by the response of the land model to 38 years of MERRA-2 forcing (section 2.1). The output from this multi-decadal, offline simulation allows the characterization of permafrost dynamics at each grid cell. In particular, we can compute a number of relevant ALT statistics, including mean, standard deviation, and skewness, from the diagnosed yearly values at each cell, and we can examine how these statistics relate to those of MERRA-2 forcing data (particularly the mean annual air temperature, MAAT) over the last 38 years.

Besides MAAT statistics, we also consider the evolution of the air temperature during the warm season in terms of the energy it could provide to the land surface and thus to the determination of ALT. A simple surrogate for the total warm-season energy in year N can be computed from daily-averaged air temperature, $T_{air}(t)$, and the freezing temperature, T_f (0°C degree), as follows:

$$T_{cum}(N) = \sum_{t=1}^{t=M} T_{pos}(t) , \tag{2}$$

where

$$T_{pos}(t) = \begin{cases} T_{air}(t) - T_f & if\ T_{air}(t) > T_f \\ 0 & if\ T_{air}(t) \leq T_f \end{cases}, \tag{3}$$

The index t in equation (2) for year N starts with a value of 1 on 1 September of the year (N-1) and ends with a value of M on 31 August of year N. The number of days M is 365 or 366 depending on the presence of a leap year. Note the air temperature throughout this study means the near-surface air temperature (i.e., 2 m above the displacement height) derived from MERRA-2.

We first computed the correlation coefficient (R) between the annual time series of ALT and $\sqrt{T_{cum}}$ and between the annual time series of ALT and maximum SWE (SWE_{max}) to quantify the degree to which variations of ALT can be explained solely by air temperature or by snow mass. Then, to quantify the joint contributions of $\sqrt{T_{cum}}$ and SWE_{max}, we performed a multiple linear regression analysis by fitting the equation

$$ALT = a_0 + a_1\sqrt{T_{cum}} + a_2 SWE_{max} , \tag{4}$$

to the available data. The correlation coefficient relating ALT to $\sqrt{T_{cum}}$ and SWE_{max} is the square root of the coefficient of multiple determination (R^2) obtained through fitting Equation (4). This equation is similar in form to the common degree-day model for predicting ALT from accumulated degree days of thaw based on the Stefan solution (e.g., Shiklomanov and Nelson, 2002; Zhang et al., 2005; Riseborough et al., 2008; Shiklomanov et al., 2010). Here, however, we constructed equation (4) for a different purpose: to explore how much of the temporal variability of ALT can be jointly explained by snow mass and

above-freezing air temperature. Before calculating these correlation coefficients, we removed the linear trend within ALT, T_{cum}, and SWE$_{max}$ to avoid potentially exaggerating the correlation due to an underlying trend. The results are discussed in section 4.3.

5   **3.4 Evaluating Simulated Northern Hemisphere Permafrost Extent and ALT**

We first evaluated the simulated permafrost extent against the observation-based permafrost map (Brown et al., 2002, as shown in Figure 1b). Note the model's description of permafrost is binary – either permafrost exists across a grid cell or it is completely absent. We cannot then expect an exact comparison to a specification of isolated permafrost (0-10% of area by definition) or even, to a lesser extent, sporadic permafrost (10-50% of area by definition). Therefore, we compared our

10   simulated permafrost area with that of the total area of continuous, discontinuous, and sporadic permafrost area together from Brown et al. (2002) and computed the percentage error relative to the observation-based area (i.e., the total area of continuous, discontinuous and sporadic permafrost regions). We also compared our simulated permafrost area against the total area of only continuous and discontinuous permafrost regions.

15   Further, the CALM network of in-situ ALT measurements (section 2.3) allows a quantitative evaluation of the simulated ALTs for the grid cells containing measurement sites. Our comparisons here focus on both multi-year annual ALTs and the average (climatological) ALT at the 81 km^2 scale of CLSM data. To ensure a consistent comparison, we average the simulated ALTs only over the years for which observations are available. As noted in section 3.1 and the supplementary file, the uncertainty of the CALM ALT measurements in the context of evaluating grid cell-scale model results theoretically involves uncertainty

20   derived from probing point measurement uncertainty, site-scale mean uncertainty, and upscaling errors in going from the site-scale to the model-scale. This latter uncertainty, in particular, is unknown. In our figures (in section 4.4) we show the standard deviation of the observed ALT as a very crude surrogate for the spatial representativeness error associated with the point-to-grid comparison. As before, we make no claim here that the standard deviations shown represent the relevant statistical uncertainty. The results are discussed in section 4.4.

4 Results

4.1 Simulated ALT versus In-situ Measurements and AirMOSS Retrievals in Alaska

In this section, we compare the simulated ALT and the AirMOSS ALT retrievals at the 81-km^2 model resolution. Note that Chen et al. (2019) provide maps of the AirMOSS retrievals and an evaluation versus in-situ measurements at the native (20 m by 60 m) scale of the retrievals.

Figure 3 compares the spatial pattern of AirMOSS ALT retrievals and CLSM-simulated results. Generally, the patterns of the AirMOSS retrievals and CLSM results are quite different. For example, the AirMOSS-retrieved ALT is greater in the northern portion of the DHO transect than in the southern portion (Figure 3a), whereas this pattern is largely reversed in the simulated ALT for DHO (Figure 3b). Across all transects, there are portions where the AirMOSS ALT is less than the CLSM-simulated ALT and portions where the AirMOSS ALT is greater (Figure 3c), though it should be noted that the differences in Figure 3c are generally less than the assumed uncertainty of 0.l5 m (see section 3.1). Generally, the CLSM-simulated ALT shows relatively larger spatial variability (0.35 - 0.85 m) than the AirMOSS retrievals (0.4 – 0.6 m). The AirMOSS ALT exhibits some spatial variability at the native resolution (see Chen et al., 2019), but much of this variability averages out during the aggregation to the coarse model grid (Figure 3a). Variations of the simulated ALT within a single transect (Figure 3a) are predominantly induced by changes in soil type (indicated in Figure 2c and 2d). In essence, the higher the organic carbon content within the soil, the smaller the simulated ALT due to slower heat transfer associated with lower thermal conductivity, higher porosity, heat capacity, etc. (Tao et al., 2017). See also section 4.2 for a discussion of the influence of soil texture on the spatial pattern of ALT.

Next, we compare the simulated ALT in 2015 with in-situ observations from the CALM and UAF sites that are collocated with the AirMOSS transects (section 3.1). Figures 4a and 4b show that the CLSM-simulated ALTs agree with the in-situ observations with an overall mean bias of -0.05 m and an RMSE of 0.17 m. The most significant discrepancies between the CLSM-simulated ALT and in-situ measurements are at U6, U31, FB1&FBD&FBW (Figure 4a), where the simulated ALT

underestimates the in-situ measurements by 0.25-0.28 m, and at U28 where the simulated ALT overestimates the in-situ ALT by 0.27 m. Nevertheless, the scatter in Figure 4b is large, and the corresponding correlation coefficient is quite weak (0.27).

The AirMOSS ALT radar retrievals, for their part, again averaged to the 81-km^2 model resolution (section 2.2), show less

5   spatial variability than the observations (Figure 4a). The largest error for the AirMOSS retrievals at the model scale is also at FB1&FBD&FBW, where the retrievals significantly underestimate the observed in-situ ALT by 0.38m. Note that radar retrievals at the 81 km^2 scale are not available at some sites because of our imposed 30% filling restriction. Although the AirMOSS ALT retrievals generally underestimate the in-situ ALT measurements (as shown in Figure 4a), the retrievals tend to be more consistent with the observations when the in-situ measurements are within the ~60 cm sensing depth of the P-band

10 radar data, as indicated in Table 3. Specifically, excluding the sites with in-situ ALT measurements that exceed the AirMOSS sensing depth of ~60 cm, the overall mean bias for the AirMOSS retrievals at the 81 km^2 scale drops to -0.01 m, and the correlation coefficient increases to 0.64. In contrast, the CLSM simulation results show a bias of 0.01 m and a zero correlation coefficient at these sites.

15 Nevertheless, as noted in section 3.1, given that the upscaling errors in going from the CALM site-scale to the model-scale is unknown and the fact that the standard deviation of these measurements (as shown by error bars in Figure 4a and 4b) indicates large representativeness errors of the in-situ measurements, the point-to-grid comparison result is hard to quantify. In this regard, the AirMOSS retrievals aggregated to the same scale as model results provide a comparable counterpart for evaluation. Figures 4c further shows that the CLSM-simulated ALT agrees well with the AirMOSS ALT retrievals to within the

20 measurement uncertainty of 0.15 m at all the site-located model grid cells. Indeed as Figure 3c illustrated, the differences between simulated ALT and the AirMOSS retrievals over all the transects examined here are generally below the measurement uncertainty of 0.15 m.

4.2 Sources of ALT Spatial Variability: Results from Idealized Experiments

Here we investigate the specific factors that drive ALT spatial variability along all ten of the AirMOSS transects (Figure 2a). For this analysis, the simulated ALT estimates were aggregated across the width of the radar swath (compare Figure 3). Figure 5a illustrates that the simulated ALT captures the spatial variability exhibited by the in-situ measurements. This conclusion is, however, very tentative given the limited number of in-situ ALT observations.

The simulated ALT is shallowest in the northern transects (ATQ, BRW, and DHO) and deepest in the southeastern transects (KYK, COC, KGR, and AMB). This pattern correlates somewhat (R = 0.46) with that of the mean screen-level (2-meter) air temperature (Tair) for the preceding 12-month period (i.e., from 1 September 2014 to 31 August 2015) from MERRA-2 (green line in Figure 5a). The soil carbon content, by contrast, appears anti-correlated (R = -0.59) with the simulated ALT, as exemplified by the transect portions within the red box (Figure 5a and 5b). Such a correlation presumably reflects the fact that soil with high organic carbon content has low thermal conductivity, which hinders heat transfer from the surface to the deeper soil in the summertime, thus resulting in a relatively smaller ALT. In addition, heat transfer is slowed by a higher effective heat capacity associated with higher organic carbon content – not from the carbon itself, but from the extra water that can be held in the soil due to the increased porosity. The maximum snow depth (Figure 5c) displays a positive correlation with ALT (R=0.47), reflecting, at least in part, the fact that subsurface soil temperatures remain relatively insulated under thick and persistent snow cover, which reduces heat transfer out of the soil column during the wintertime and hence facilitates a deeper thawing during the summer and thus a deeper ALT.

The correlations in Figure 5 suggest (without proving causality) that for the model, surface meteorological forcing (including air temperature and precipitation), as well as soil type, are important drivers of ALT variability along the AirMOSS transects. However, the relatively low values of the correlations indicate that a simple linear relationship cannot explain the mutual control that these variables exert on ALT spatial variability. In the remainder of this section, we use a series of idealized model simulations (as described in section 3.2) to better quantify the relative impacts of these driving factors along the AirMOSS transects.

The results of the idealized experiments are shown in Figure 6. The above-mentioned, large-scale spatial variation of ALT in the baseline simulation, with larger values in the southeastern transects (KYK, COC, and KGR) and lower values in the northern transects (ATQ, BRW, and DHO), is absent after homogenizing the meteorological forcing (HomF; Figure 6a). Experiment HomF correspondingly has much less spatial variation in the temperature of the top soil layer than does the baseline simulation (Figure 6b). In addition, homogenizing the forcing (which includes snowfall) significantly reduces the variability in maximum snow depth along the AirMOSS transects (Figure 6c). These results indicate that in the model, meteorological forcing exerts the dominant control over the spatial patterns of ALT, the temperature in the top soil layer, and snow depth at the regional scale, as expected.

Homogenizing the vegetation attributes in addition to the forcing (HomF&Veg) results in ALT differences (relative to HomF) primarily along the northern transects (ATQ, BRW, and DHO). Along these transects, homogenizing the vegetation parameters (including LAI and tree height) to those of the representative grid cell within the IVO transect results in generally shallower ALT. This is because the generally lower albedo of the taller and leafier trees (representative of the IVO transect) during the snow season resulted in increased snowmelt and thus reduced snowpack during the snow season (compare the green and red curves in Figure 6c), thereby reducing the thermal insulation of the wintertime ground. With reduced insulation, cold season ground temperatures dropped, making it more difficult for temperatures to recover during summer(Tao et al., 2017).

As might be expected, the simulation in which soil properties are homogenized in conjunction with forcing and vegetation (i.e., HomF&Veg&Soil) essentially eliminates all remaining spatial variability in ALT, snow depth, and soil temperature. Owing to the strong control of soil type-related parameters (see section 3.2 and Table 2) on soil moisture, spatial variability in soil moisture remains high in HomF and HomF&Veg and is only eliminated once these soil type-related parameters are homogenized (Figure 6d), which explains the abrupt changes shown in Figure 3c as mentioned in section 3.1. (Note that to maintain consistency with the hardwired scaling factors for snow-free albedo within the model (Mahanama et al., 2015), we

still used the original, vegetation-related parameters to calculate surface albedo during snow-free conditions along the transects. This is likely the cause of the few tiny bumps seen in the Figure 6a for HomF&Veg&Soil.)

An alternative view of these results is provided in Figure 7a, which shows the (spatial) standard deviation of ALT along the AirMOSS transects for each of the above experiments. Homogenizing the meteorological forcing data results in a significant reduction of the ALT standard deviation (from 0.16 to 0.10). Additionally homogenizing the vegetation only reduces the ALT standard deviation slightly (from 0.10 to 0.09). The remaining ALT variability is eliminated through the additional homogenization of the soil type-related parameters (HomF&Veg&Soil), which emerge as another important driver of ALT variability along the AirMOSS transects. Note that the ALT variability associated with soil type is generally realized at smaller spatial scales than that associated with the meteorological forcing discussed earlier regarding Figure 6a. The impact of potential nonlinearities is examined in Figure 7b, which shows the individual impact of vegetation, soil, and forcing heterogeneity on the ALT standard deviation along the transects, with the other inputs having been homogenized. The graphic confirms that the meteorological forcing is the dominant driver of ALT spatial variability in our modelling system, followed by the soil type-related parameters and the vegetation parameters.

Note that in Figure 6a, the soil impact on ALT (the difference between HomF&Veg&Soil in black and HomF&Veg in red) appears smaller than that of the vegetation (the difference between HomF in green and HomF&Veg in red) over the northern transects (ATQ, BRW, and DHO). Even so, Figure 7b shows that, in terms of the integrated impact along all the transects, the soil influence clearly outweighs the influence of vegetation – at several other transects, including HUS, KYK, COC, AMB, IVO and the first half of ATQ (where vegetation conditions might be similar to those used for homogenizing), the changes in vegetation parameters do not have much impact.

4.3 Spatiotemporal Characteristics of ALT across the Northern Hemisphere

Figure 8a shows the distribution of mean ALT over the modelling domain, and Figure 8b shows the ALT standard deviation in time over the 38-year period. As might be expected, ALT tends to increase with distance from the pole, with the largest values found in Mongolia and near the southern portion of Hudson Bay, though there are areas (e.g., just north of 60°N at ~120°E) with local minima that break this pattern. The largest ALT standard deviations (red color in Figure 8b) are found mainly in discontinuous and sporadic permafrost regions (see Figure 1b) where ALTs are deeper on average than that in continuous permafrost region. Figure 8c provides the skewness of the temporal distribution. Though there are some exceptions, by and large, the skewness is positive in most permafrost regions, suggesting that the largest positive ALT anomalies tend to be of greater magnitude than the largest negative anomalies.

Figure 8d displays the average of annual mean 2-meter air temperature as derived from MERRA-2. The observed continuous and discontinuous permafrost areas shown in Figure 1b are well confined within the cold side of the 0°C (273.15K) isotherm in the mean air temperature map (Figure 8d). For the most part, the observed sporadic and isolated permafrost regions of Figure 1b also lie on the cold side of the 0°C isotherm. The consistency with this isotherm, however, is not as clearly present in the simulated permafrost extent (i.e., the extent of the non-grey and non-white areas in Figure 8a).

The relationship between the spatiotemporal characteristics of simulated ALT and air temperature forcing has been investigated before in many studies at the site to landscape scale (e.g., Klene et al., 2001; Shiklomanov and Nelson, 2002; Zhang et al., 2005; Juliussen and Humlum, 2007) and at the regional scale (e.g., Anisimov et al., 2007). Here we simply analyze the correlation coefficient between ALT and two variables: the proxy of total energy input into the ground (i.e., $\sqrt{T_{cum}}$, see section 3.3) and the maximum SWE. Our goal is to explore how much of the spatiotemporal variability of ALT across the globe can be jointly explained by these two variables.

Figure 9a shows a map of the correlation coefficient between the 37-year time series (i.e., from September 1980 through August 2017) of $\sqrt{T_{cum}}$ and the corresponding time series of simulated ALT. The areas with p values larger than 0.05, which

indicate correlations that are not statistically different from zero at the 95% confidence level, are shown as green. Figure 9a demonstrates that most permafrost regions indeed have significant positive correlations (red colours) between ALT and $\sqrt{T_{cum}}$. Clearly, in these regions, air temperature exerts a dominant control on year-to-year ALT variability.

5   However, not all regions exhibit a significant correlation; other variable(s) must also be exerting control on interannual ALT variability. One reasonable candidate variable is snowpack. As noted above, snow acts as a thermal insulator -- regions with thicker snowpack are better able to insulate the ground from becoming too cold during winter, thereby supporting higher subsurface temperatures during non-winter months. Variable, but often thick, snowpack is in fact common in the areas of Figure 9a that show a low (green) or negative (blue) correlation between ALT and $\sqrt{T_{cum}}$ – areas such as Central Siberia, the

10   Southern part of eastern Siberia, and a vast region in Canada surrounding the Hudson Bay, as well as other small areas that appear in high mountains or on the windward side of the mountains (e.g., locations B, C and D in Figure 1a).

    In Figure 9b we show the correlation coefficient between the time series of ALT and the maximum SWE (SWE$_{max}$) during the preceding winter. A positive correlation is seen in many areas, most notably in areas with a poor or negative correlation

15   between ALT and $\sqrt{T_{cum}}$ (Figure 9a) – for example, just west of Hudson Bay and along a zonal band at 60°N in Russia. Apparently, in these areas, the impacts of snow physics on ALT outweigh the impacts of lumped energy input ($\sqrt{T_{cum}}$). In some other areas ALT correlates positively with both $\sqrt{T_{cum}}$ and SWE$_{max}$. Figure 9c shows how the resulting coefficient of multiple correlation varies in space. High correlations largely blanket the modelled area. That is, over most of the area examined, a substantial portion of the year-to-year variability of ALT can be explained by joint variations in $\sqrt{T_{cum}}$ and

20   SWE$_{max}$. Even so, a few limited areas still exhibit low correlations (p>0.05, green colour in Figure 9c). Some of these areas are in mountainous regions, for instance the Eastern Siberian (Ostsibirisches) Bergland, where more complex environmental controls might be playing a dominant role. In addition, MERRA-2 snow forcing might be severely erroneous in these regions.

4.4 Evaluation of Simulated Permafrost Extent and ALT across the Northern Hemisphere

Qualitatively, the simulated permafrost extent (Figure 8a) generally shows reasonable agreement with the observation-based permafrost map in Figure 1b, especially for the continuous permafrost regions. This is shown explicitly in Figure 10a. The main deficiency in the simulation results is the failure to capture a large area of permafrost in western Siberia (labelled as A in Figure 1a). The reasons for this particular deficiency are unclear. One possible reason is that the permafrost in western Siberia is characterized as an ecosystem-protected permafrost zone (Shur and Jorgenson, 2007) where a thick moss-organic layer (i.e., moss-dominated mires (Anisimov and Reneva, 2006; Anisimov, 2007; Peregon et al., 2009)) protects the permafrost below from thawing under a warm air temperature. This is mainly attributed to the low thermal conductivity of the organic layer in summer, which strongly insulates the permafrost from the warm atmosphere, and the high thermal conductivity of the frozen organic layer in winter, which allows cold temperature penetration from above, provided the snowpack is not too thick (Nicolsky et al., 2007b; Jafarov and Schaefer, 2016). This mechanism is lacking in the current version of CLSM (Tao et al., 2017). Thus, improving the model through a better representation of thermal processes in an organic layer above the soil column in combination with initializing the simulation with a sufficiently cold soil temperature should improve the simulation results. This work is reserved for a future study.

Another possible reason for the poor skill in western Siberia is that the model initial conditions there were too warm, although MERRA-2 appears to underestimate summer air temperatures in this region (Draper et al., 2018; their Figure 7e). Note that some other global models, such as CLM3 and the Community Climate System Model version 3 (CCSM3) as reported in Lawrence et al. (2012), also missed this area of permafrost and that updated versions of these models (i.e., CLM4 and CCSM4) showed improved performance in this regard (Lawrence et al., 2012). Guo et al. (2017) reported underestimated permafrost extent simulated in western Siberia using CLM4.5 driven by three different reanalysis forcings (i.e., CFSR, ERA-I and MERRA), and they showed an improved simulation of permafrost extent in this area when using another reanalysis forcing, the CRUNCEP (Climatic Research Unit - NCEP) (Guo and Wang, 2017). Guimberteau et al. (2018) found similar improvements stemming from the use of CRUNCEP forcing. We leave for further study whether the MERRA-2 forcing data is responsible for the western Siberia deficiency seen in our results.

The disagreements between the simulated and observed permafrost extents (covering about a few degrees latitude) toward the south in Figure 10a (green and blue areas at the southern edge of permafrost regions) are less of a concern, since the comparison in such areas is muddied by the interpretation of "isolated" permafrost in the observational map (Figure 1b). The specific areas of each type shown in Figure 10a are listed in Table 4. The simulated permafrost extent covers 81.3% of the observation-based area (i.e., the total area of continuous, discontinuous and sporadic permafrost regions), and misses 18.7% of the observed permafrost area. When comparing simulated permafrost extent with only continuous and discontinuous types, these metrics change to 87.7% and 12.3%, respectively. Meanwhile, the permafrost extent is overestimated by $3.2\times10^6\,\mathrm{km}^2$.

To produce Figure 10b, multi-year averages of CLSM-simulated ALT values were spatially averaged over each of the four permafrost types outlined in Figure 1b. (As is appropriate, permafrost is only occasionally simulated over the fourth, "isolated", permafrost type. The ALT average shown for this type is thus based on a particularly limited number of grid cells.) The average ALT is smallest in the continuous permafrost zone, higher in the discontinuous zone, and higher still in the sporadic permafrost zone; it is highest in areas of isolated permafrost. The progression, of course, is in qualitative agreement with expectations – larger breaks in permafrost coverage imply a greater amount of available energy, which should also act to increase ALT.

The observed and CLSM-simulated annual ALT and multi-year ALT averages are compared in Figure 11. Generally, the simulated annual ALT and the averages agree reasonably well with observations for shallow permafrost regions, that is, for smaller ALT. A large bias, however, is found for most of the Mongolia sites; in Mongolia, the observed annual ALT and the climatological ALTs tend to be much larger than the simulated ALTs (light purple dots in Figure 11). Overall, the RMSE, bias, and R are all significantly improved when the Mongolian sites are excluded from consideration. Specifically for the climatological ALTs, the RMSE (and bias) of simulated ALT climatological means is 1.22 m (and -0.48 m), and it drops to 0.33 m (and -0.04 m) if the Mongolia sites are excluded (Figure 11d). Given simplifications in the model, uncertainties in boundary conditions (e.g., vegetation types, soil properties, etc.), and upscaling issues stemming from the coarse-scale nature

of the forcing data relative to the point-scale and plot-scale nature of the observations (i.e., the representative errors as indicated by the large standard deviation shown in Figure 11a), these results seem encouraging. The correlation coefficient metric (R), however, is somewhat less encouraging, amounting to only 0.5 when considering all sites. The correlation coefficient is in fact lower (0.3) when the Mongolian sites are excluded; the correlation coefficient is 0.39 for the Mongolian sites considered

5   in isolation. Note that the existing literature on simulated ALT fields (e.g., Dankers et al. (2011), Lawrence et al. (2012) and Guo et al. (2017)) reveals a general tendency for models to overestimate ALT climatology at the global scale. In light of this, our results suggest that the CLSM-simulated ALT fields are perhaps among the better simulation products, especially for shallow permafrost.

10   Comparing the observed and simulated spatial distributions of the ALT averages provides a further test of the accuracy of the simulation results (as shown in Figure 12). The model successfully simulates the large-scale spatial patterns in ALT, capturing, for example, the variations in Siberia, Svalbard, northern Canada, and northern Alaska (see Figure 12a, b). Figure 12c, d show the differences between the observed and estimated values in middle latitudes (45ºN to 60ºN) and high latitudes (60ºN to 90ºN), respectively; in agreement with Figure 11a, the model clearly performs better in high-latitude regions, i.e., outside of

15   Mongolia. Many of the sites north of 60ºN (Figure 12d) are coloured grey, indicating a small error in the simulation of ALT at these sites – the errors at these sites range from only -0.10m to 0.10m.

The significant underestimation of ALT in Mongolia might result from errors in the meteorological forcing provided by MERRA-2. However, a comparison (not shown) of MERRA-2 air temperatures with measurements at six weather stations

20   collocated with CALM sites in Mongolia calls this explanation into question. While MERRA-2 summer temperatures are indeed too low at four of the weather stations examined, they are too high at the other two weather stations. An additional reason for the underestimation of ALT in Mongolia might be a mismatch between the land surface parameter values used in the model and the actual conditions at each site. For instance, detailed soil information (https://www2.gwu.edu/~calm/data/webforms/mg_f.html) indicate that some Mongolian sites have special "rocky" soil types

25   including limestones (e.g., M04), slatestones (e.g., M05), gravelly sand (e.g., M06 and M08), etc. that are not well represented

in the model. As another example, sites on south-facing slopes presumably have much deeper ALT than those on slopes with less exposure to the sun, which is not captured by CLSM. The large representative errors of Mongolian sites are clearly illustrated by the standard deviation (although computed only with 3 to 5 measurements) as shown by the error bars in Figure 11a.

5 Conclusion and Discussion

We produced a dataset (effectively a derivative of MERRA-2) of permafrost variations in space and time across middle-to-high latitudes. This dataset can be considered unique in terms of its daily temporal resolution combined with a relatively high spatial resolution at the global scale (i.e., 81 km^2). The dataset, which is derived from a state-of-the-art reanalysis (MERRA-

10 2), shows reasonable skill in capturing permafrost extent (87.7% of the total area of continuous and discontinuous types, according to one validation dataset) and in adequately estimating ALT climatology (with a RMSE of 0.33m and a mean bias of -0.04m), excluding Mongolian sites. We note that our MERRA-2-driven permafrost simulation results, while potentially better than those we might have obtained with MERRA forcing, are still lacking (e.g., in western Siberia). Still, with its resolution and available variables (ALT, subsurface temperature and ice content at different depths), the dataset could prove

15 valuable to many future permafrost analyses.

This work also provides a first comparison between two highly complementary approaches to estimating permafrost: model simulation and remote sensing. In northern Alaska, excluding sites that have ALT measurements exceeding the radar sensing depth (~ 60cm), the evaluation metrics for ALT retrievals against in-situ measurements are better than those for simulated

20 ALT at the 81 km^2 scale. However, the remotely sensed ALT estimates generally show lower levels of spatial variability relative to the simulated ALT estimates (and relative to the in-situ observations), and the spatial patterns of the simulated and retrieved values differ considerably. The remote sensing approach is still relatively new, with many aspects still requiring development. It is important, though, to begin considering the modeling and remote sensing approaches side by side, as both should play important roles in permafrost quantification in the years to come. Indeed, once the science fully develops, joint

use of modeling and remote sensing (e.g., through the application of downscaling methods) should allow the generation of more accurate permafrost products at higher resolutions.

It is important to note that the retrieved ALT was determined by the dielectric transition from thawed to frozen conditions whereas the modelled ALT and the ALT for some of the in-situ measurements was based on a freezing temperature of 0ºC (see sections 2.1 and 2.3). Depending on local conditions, soil does not typically freeze at 0ºC but rather at slightly lower temperatures (e.g., around -1ºC) due to the presence of dissolved compounds that depress the freezing point (Watanabe and Wake, 2009). The sharp drop in conductivity and dielectric constant is much more accurately tied to a frozen state than to a temperature threshold. These and other differences in the various ALT measurement methods (section 2.3) introduce considerable uncertainty into our comparisons. The use of the 0ºC degree threshold in CLSM for determining the thawed or frozen layer may explain in part the model's underestimation of ALT, as may the lack of an explicit treatment of local aspect, errors in assigned model parameters, and so on.

Analysis of the CLSM-simulated data, along with data produced in idealized experiments with specific homogenized controls, show how the statistics of permafrost variability in space are controlled by forcing variability and by variability in the imposed surface boundary conditions. In the idealized experiments, we employ successive homogenization of controls to quantify how meteorological forcing, soil type, and vegetation cover affect the underground thermodynamic processes associated with the variability of ALT along the AirMOSS flight paths in Alaska. Meteorological forcing and soil type are found to be the two dominant factors controlling ALT variability along these transects. Vegetation plays a smaller role by modulating the accumulation of snow. A multiple regression analysis relating yearly ALT jointly to accumulated air temperature and maximum SWE shows that time variations in these two latter quantities explain most of the time variability of ALT in the CLSM-identified permafrost regions.

Many aspects of the modelling framework may contribute to the noted errors in the simulated ALTs. For example, the observed climatological ALTs at the Mongolia sites are all larger than 3m. This depth falls well within the 6th soil layer of the model,

which has a thickness of 10m; the subsurface vertical resolution in the CLSM may be too coarse to capture these deeper ALTs. Test simulations (not shown) with alternative model configurations indicate that increasing the number of soil layers may act to decrease somewhat the simulated ALT, suggesting that our values may be a little overestimated; however, based on results from a new study by Sapriza-Azuri et al.(2018), our use of a no-heat-flux condition at the bottom boundary rather than a

5   dynamic geothermal flux may lead to underestimates of ALT. Such uncertainties should naturally be kept in mind when interpreting our results. Our supplemental simulations (not shown) also suggest that increasing the total modelled soil depth has only a small impact on simulated ALT. Uncertainty in our description of soil organic carbon, i.e., both soil carbon content and vertical carbon distribution, leads to corresponding uncertainty in our ALT simulations. We indeed find a significant improvement in simulated ALT at several Mongolian sites when we arbitrarily impose less total soil carbon content and

10   concentrate less soil carbon in top layers (not shown). Besides the vertical distribution of soil carbon, the vertical variation in other soil hydrological properties (e.g., soil texture and porosity) should also play a significant role since they all affect soil thermal conductivity and heat capacity. In addition, the lack of a necessary organic layer on top of soil column and the related thermal processes is also a major deficiency for the model, especially in ecosystem-protected performant regions.

15   Another issue affecting our ALT comparisons is the climatological representation of vegetation parameters such as LAI used in CLSM. An additional investigation (not shown) revealed large differences between the LAI climatology used in CLSM and more realistic, time-varying, satellite-based LAI products at several Mongolian sites. In addition, while we did exclude from our analyses any measurements that were affected by notable disturbance (e.g., wildfire), the impacts of other potential land changes on ALT, including overgrazing in Mongolia (Sharkhuu and Sharkhuu, 2012; Liu et al., 2013), were not explicitly

20   treated in the model. The model also lacks the vertical advective transport of heat in the subsurface due to downward flowing liquid water, which can significantly affect permafrost thawing (Kane et al., 2001; Rowland et al., 2011; Kurylyk et al., 2014). Also relevant are potential errors in the MERRA-2 forcing. The MERRA-2 reanalysis is known to have problems capturing trends in high latitudes (Simmons et al., 2017).

Such modelling deficiencies must always be kept in mind when evaluating a product like the one examined here. That said, as long as appropriate caution is employed, the product could have significant value for further analyses of permafrost. The product features daily subsurface temperatures and depth-to-freezing estimates over middle-to-high latitudes in the Northern Hemisphere at an 81 km^2 resolution, covering the period 1980-2017. It is, in a sense, a value-added derivative product of the MERRA-2 reanalysis and will be available via the National Snow and Ice Data Center (NSIDC). The comparisons against observations discussed above, along with the intuitively sensible connections shown between permafrost variability, forcing variability, and boundary condition variability, gives confidence that this dataset contains useful information. These data can potentially contribute, for example, to ecological studies focused on the dynamics of microbial activity and soil respiration in cold regions, on vegetation migration/adaptation in response to climate change, and so on.

Acknowledgments

Funding for this work was provided by the NASA Interdisciplinary Science program (NNX14AO23G). We thank Qing Liu at GMAO/GSFC/NASA for providing us corrected MERRA-2 precipitation. We thank the anonymous reviewers for their helpful comments. We acknowledge the University of Maryland supercomputing resources (http://www.it.umd.edu/hpcc) made available for the research reported in this paper.

Data availability

The CALM ALT observations used in this study are available at https://www2.gwu.edu/~calm/. The UAF observations are available at http://permafrost.gi.alaska.edu. The AirMOSS ALT retrievals are available at https://daac.ornl.gov/ABOVE/guides/ABoVE_PBand_SAR.html. The MERRA-2 land forcing fields are available at https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/. The produced permafrost product, including ALT, subsurface soil temperature and ice content, will be available at the National Snow and Ice Data Center (NSIDC), and are also available from the corresponding author upon request.

Author contributions

The first author obtained all the data, performed model simulations, analyzed results, wrote and revised the paper. RDK and RHR contributed to experiment design, editing the original and revised manuscript, providing valuable discussion and guidance. RHC and MM produced and provided the AirMOSS ALT retrievals, and contributed to revisions regarding
5   AirMOSS retrievals. BAF and YX provided comments and discussion. BAF also provided computational resources at the University of Maryland.

Competing interests

The authors declare that they have no conflict of interest.

References

10   Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E., and Lawrence, D. M.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, Geophys Res Lett, 34, 10.1029/2007gl029536, 2007.

Anisimov, O., and Reneva, S.: Permafrost and changing climate: The Russian perspective, Ambio, 35, 169-175, 10.1579/0044-7447(2006)35[169:Pacctr]2.0.Co;2, 2006.

Anisimov, O. A.: Potential feedback of thawing permafrost to the global climate system through methane emission,
15   Environmental Research Letters, 2, 10.1088/1748-9326/2/4/045016, 2007.

Anisimov, O. A., Lobanov, V. A., Reneva, S. A., Shiklomanov, N. I., Zhang, T., and Nelson, F. E.: Uncertainties in gridded air temperature fields and effects on predictive active layer modeling, Journal of Geophysical Research: Earth Surface, 112, 2007.

Barman, R., and Jain, A. K.: Comparison of effects of cold-region soil/snow processes and the uncertainties from model forcing
20   data on permafrost physical characteristics, J Adv Model Earth Sy, 8, 453-466, 2016.

Batir, J. F., Hornbach, M. J., and Blackwell, D. D.: Ten years of measurements and modeling of soil temperature changes and their effects on permafrost in Northwestern Alaska, Global and Planetary Change, 148, 55-71, 2017.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., and Savoie, M. H.: EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets, ISPRS International Journal of Geo-Information, 1, 32-45, 2012.

Brown, J., Hinkel, K. M., and Nelson, F.: The circumpolar active layer monitoring (CALM) program: research designs and initial results, Polar geography, 24, 166-258, 2000.

5 Brown, J., Ferrians, O., Heginbottom, J. A., and Melnikov, E.: Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2. [Permafrost Extent], NSIDC: National Snow and Ice Data Center, Boulder, Colorado USA, https://nsidc.org/data/ggd318, 2002.

Carey, S. K., and Woo, M.-k.: Freezing of subarctic hillslopes, Wolf Creek Basin, Yukon, Canada, Arctic, Antarctic, and Alpine Research, 37, 1-10, 2005.

10 Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., and Westermann, S.: An observation-based constraint on permafrost loss as a function of global warming, Nat Clim Change, 7, 340-+, 2017.

Chen, A., Parsekian, A. D., Schaefer, K., Jafarov, E., Panda, S., Liu, L., Zhang, T. J., and Zebker, H.: Ground-penetrating radar-derived measurements of active-layer thickness on the landscape scale with sparse calibration at Toolik and Happy Valley, Alaska, Geophysics, 81, H9-H19, 10.1190/Geo2015-0124.1, 2016a.

15 Chen, M., Shi, W., Xie, P., Silva, V., Kousky, V. E., Wayne Higgins, R., and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, Journal of Geophysical Research: Atmospheres, 113, 2008.

Chen, R. H., Tabatabaeenejad, A., and Moghaddam, M.: A time-series active layer thickness retrieval algorithm using P-and L-band SAR observations, Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, 2016b, 3672-3675,

20 Chen, R. H., Tabatabaeenejad, A., and Moghaddam, M.: Retrieval of permafrost active layer properties using time-series P-band radar observations, IEEE Transactions on Geoscience and Remote Sensing, 10.1109/TGRS.2019.2903935, 2019.

Chen, W. J., Zhang, Y., Cihlar, J., Smith, S. L., and Riseborough, D. W.: Changes in soil temperature and active layer thickness during the twentieth century in a region in western Canada, Journal of Geophysical Research-Atmospheres, 108, 2003.

Dankers, R., Burke, E. J., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, 25 Cryosphere, 5, 773-790, 2011.

De Lannoy, G. J. M., Koster, R. D., Reichle, R. H., Mahanama, S. P. P., and Liu, Q.: An updated treatment of soil texture and associated hydraulic properties in a global land modeling system, J Adv Model Earth Sy, 6, 957-979, 10.1002/2014ms000330, 2014.

Dorrepaal, E., Toet, S., van Logtestijn, R. S. P., Swart, E., van de Weg, M. J., Callaghan, T. V., and Aerts, R.: Carbon respiration from subsurface peat accelerated by climate warming in the subarctic, Nature, 460, 616-U679, 10.1038/nature08216, 2009.

Draper, C. S., Reichle, R. H., and Koster, R. D.: Assessment of MERRA-2 Land Surface Energy Flux Estimates, Journal of Climate, 31, 671-691, 2018.

Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model 2. Parameter estimation and model demonstration, Journal of Geophysical Research-Atmospheres, 105, 24823-24838, Doi 10.1029/2000jd900328, 2000.

Fisher, J. P., Estop-Aragones, C., Thierry, A., Charman, D. J., Wolfe, S. A., Hartley, I. P., Murton, J. B., Williams, M., and Phoenix, G. K.: The influence of vegetation and soil characteristics on active-layer thickness of permafrost soils in boreal forest, Global Change Biol, 22, 3127-3140, 2016.

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419-5454, 2017.

Gisnas, K., Etzelmuller, B., Farbrot, H., Schuler, T. V., and Westermann, S.: CryoGRID 1.0: Permafrost Distribution in Norway estimated by a Spatial Numerical Model, Permafrost Periglac, 24, 2-19, 2013.

Global Modeling and Assimilation Office (GMAO): MERRA-2 inst1_2d_lfo_Nx: 2d,1-Hourly,Instantaneous,Single-Level,Assimilation,Land Surface Forcings V5.12.4, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed [03/13/2017] Greenbelt, MD, USA, 2015a.

Global Modeling and Assimilation Office (GMAO): MERRA-2 tavg1_2d_lfo_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Forcings V5.12.4, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed [03/13/2017] Greenbelt, MD, USA, 2015b.

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

Guo, D. L., and Sun, J. Q.: Permafrost Thaw and Associated Settlement Hazard Onset Timing over the Qinghai-Tibet Engineering Corridor, Int J Disast Risk Sc, 6, 347-358, 2015.

Guo, D. L., and Wang, H. J.: Simulated Historical (1901-2010) Changes in the Permafrost Extent and Active Layer Thickness in the Northern Hemisphere, Journal of Geophysical Research-Atmospheres, 122, 12285-12295, 10.1002/2017jd027691, 2017.

Guo, D. L., Wang, H. J., and Wang, A. H.: Sensitivity of Historical Simulation of the Permafrost to Different Atmospheric Forcing Data Sets from 1979 to 2009, Journal of Geophysical Research-Atmospheres, 122, 12269-12284, 2017.

Hinkel, K., and Nelson, F.: Spatial and temporal patterns of active layer thickness at Circumpolar Active Layer Monitoring (CALM) sites in northern Alaska, 1995–2000, Journal of Geophysical Research: Atmospheres, 108, 2003.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Gu, G.: Improving the global precipitation record: GPCP version 2.1, Geophys Res Lett, 36, 2009.

Hugelius, G., Bockheim, J. G., Camill, P., Elberling, B., Grosse, G., Harden, J. W., Johnson, K., Jorgenson, T., Koven, C. D., Kuhry, P., Michaelson, G., Mishra, U., Palmtag, J., Ping, C. L., O'Donnell, J., Schirrmeister, L., Schuur, E. A. G., Sheng, Y., Smith, L. C., Strauss, J., and Yu, Z.: A new data set for estimating organic carbon storage to 3m depth in soils of the northern circumpolar permafrost region, Earth Syst Sci Data, 5, 393-402, 10.5194/essd-5-393-2013, 2013a.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst Sci Data, 5, 3-13, 10.5194/essd-5-3-2013, 2013b.

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

Jafarov, E. E., Marchenko, S. S., and Romanovsky, V. E.: Numerical modeling of permafrost dynamics in Alaska using a high spatial resolution dataset, Cryosphere, 6, 613-624, 10.5194/tc-6-613-2012, 2012.

Jafarov, E. E., Parsekian, A. D., Schaefer, K., Liu, L., Chen, A. C., Panda, S. K., and Zhang, T.: Estimating active layer thickness and volumetric water content from ground penetrating radar measurements in Barrow, Alaska, Geosci Data J, 4, 72-79, 10.1002/gdj3.49, 2017.

James, M., Lewkowicz, A. G., Smith, S. L., and Miceli, C. M.: Multi-decadal degradation and persistence of permafrost in the Alaska Highway corridor, northwest Canada, Environmental Research Letters, 8, 2013.

Jean, M., and Payette, S.: Effect of Vegetation Cover on the Ground Thermal Regime of Wooded and Non-Wooded Palsas, Permafrost Periglac, 25, 281-294, 2014.

Johnson, K. D., Harden, J. W., McGuire, A. D., Clark, M., Yuan, F. M., and Finley, A. O.: Permafrost and organic layer interactions over a climate gradient in a discontinuous permafrost zone, Environmental Research Letters, 8, 10.1088/1748-9326/8/3/035028, 2013.

Jones, B. M., Baughman, C. A., Romanovsky, V. E., Parsekian, A. D., Babcock, E. L., Stephani, E., Jones, M. C., Grosse, G., and Berg, E. E.: Presence of rapidly degrading permafrost plateaus in south-central Alaska, Cryosphere, 10, 2673-2692, 2016.

Jorgenson, M. T., Racine, C. H., Walters, J. C., and Osterkamp, T. E.: Permafrost degradation and ecological changes associated with a warmingclimate in central Alaska, Climatic change, 48, 551-579, 2001.

Juliussen, H., and Humlum, O.: Towards a TTOP ground temperature model for mountainous terrain in central-eastern Norway, Permafrost Periglac, 18, 161-184, 2007.

Kane, D. L., Hinkel, K. M., Goering, D. J., Hinzman, L. D., and Outcalt, S. I.: Non-conductive heat transfer associated with frozen soils, Global and Planetary Change, 29, 275-292, 2001.

Klene, A. E., Nelson, F. E., Shiklomanov, N. I., and Hinkel, K. M.: The n-factor in natural landscapes: variability of air and soil-surface temperatures, Kuparuk River Basin, Alaska, USA, Arctic, Antarctic, and Alpine Research, 33, 140-148, 2001.

5 Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure, Journal of Geophysical Research-Atmospheres, 105, 24809-24822, Doi 10.1029/2000jd900327, 2000.

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

10 Kumar, J., Collier, N., Bisht, G., Mills, R. T., Thornton, P. E., Iversen, C. M., and Romanovsky, V.: Modeling the spatiotemporal variability in subsurface thermal regimes across a low-relief polygonal tundra landscape, Cryosphere, 10, 2241-2274, 2016.

Kurylyk, B. L., McKenzie, J. M., MacQuarrie, K. T., and Voss, C. I.: Analytical solutions for benchmarking cold regions subsurface water flow and energy transport models: One-dimensional soil thaw with conduction and advection, Adv.

15 Water Resour., 70, 172-184, 2014.

Lawrence, D. M., and Slater, A. G.: Incorporating organic soil into a global climate model, Clim Dynam, 30, 145-160, 10.1007/s00382-007-0278-1, 2008.

Lawrence, D. M., Slater, A. G., Romanovsky, V. E., and Nicolsky, D. J.: Sensitivity of a model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter, Journal of Geophysical Research-

20 Earth Surface, 113, 10.1029/2007jf000883, 2008.

Lawrence, D. M., Slater, A. G., and Swenson, S. C.: Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4, Journal of Climate, 25, 2207-2225, 10.1175/Jcli-D-11-00334.1, 2012.

Li, Z. W., Zhao, R., Hu, J., Wen, L. X., Feng, G. C., Zhang, Z. Y., and Wang, Q. J.: InSAR analysis of surface deformation over permafrost to estimate active layer thickness based on one-dimensional heat transfer model of soils, Sci Rep-Uk, 5,

25 2015.

Liu, L., Zhang, T., and Wahr, J.: InSAR measurements of surface deformation over permafrost on the North Slope of Alaska, Journal of Geophysical Research: Earth Surface, 115, 2010.

Liu, L., Schaefer, K., Zhang, T., and Wahr, J.: Estimating 1992–2000 average active layer thickness on the Alaskan North Slope from remotely sensed surface subsidence, Journal of Geophysical Research: Earth Surface, 117, 2012.

5 Liu, Y. Y., Evans, J. P., McCabe, M. F., De Jeu, R. A., van Dijk, A. I., Dolman, A. J., and Saizen, I.: Changing climate and overgrazing are decimating Mongolian steppes, Plos One, 8, e57599, 2013.

Lu, Q., Zhao, D. S., and Wu, S. H.: Simulated responses of permafrost distribution to climate change on the Qinghai-Tibet Plateau, Sci Rep-Uk, 7, 2017.

Luetschg, M., Lehning, M., and Haeberli, W.: A sensitivity study of factors influencing warm/thin permafrost in the Swiss

10 Alps, J Glaciol, 54, 696-704, 2008.

Luo, D. L., Wu, Q. B., Jin, H. J., Marchenko, S. S., Lu, L. Z., and Gao, S. R.: Recent changes in the active layer thickness across the northern hemisphere, Environmental Earth Sciences, 75, 2016.

MacDougall, A. H., Avis, C. A., and Weaver, A. J.: Significant contribution to climate warming from the permafrost carbon feedback, Nat Geosci, 5, 719-721, 2012.

15 Mahanama, S. P., Koster, R. D., Walker, G. K., Takacs, L. L., Reichle, R. H., De Lannoy, G., Liu, Q., Zhao, B., and Suarez, M. J.: Land Boundary Conditions for the Goddard Earth Observing System Model Version 5 (GEOS-5) Climate Modeling System: Recent Updates and Data File Descriptions, 2015.

Matyshak, G. V., Goncharova, O. Y., Moskalenko, N. G., Walker, D. A., Epstein, H. E., and Shur, Y.: Contrasting Soil Thermal Regimes in the Forest-Tundra Transition Near Nadym, West Siberia, Russia, Permafrost Periglac, 28, 108-118, 2017.

20 Nelson, F. E., Anisimov, O. A., and Shiklomanov, N. I.: Subsidence risk from thawing permafrost, Nature, 410, 889, 2001.

Nicholas, J. R. J., and Hinkel, K. M.: Concurrent permafrost aggradation and degradation induced by forest clearing, Central Alaska, USA, Arctic Alpine Res, 28, 294-299, 1996.

Nicolsky, D. J., Romanovsky, V. E., Alexeev, V. A., and Lawrence, D. M.: Improved modeling of permafrost dynamics in a GCM land-surface scheme, Geophys Res Lett, 34, 2007a.

Nicolsky, D. J., Romanovsky, V. E., Alexeev, V. A., and Lawrence, D. M.: Improved modeling of permafrost dynamics in a GCM land-surface scheme, Geophys Res Lett, 34, 2007b.

Osterkamp, T. E., and Romanovsky, V. E.: Characteristics of changing permafrost temperatures in the Alaskan Arctic, USA, Arctic Alpine Res, 28, 267-273, 1996.

5 Pannetier, R., and Frampton, A.: Air warming trends linked to permafrost warming in the sub-Arctic catchment of Tarfala, Sweden, Polar Res, 35, 2016.

Peregon, A., Maksyutov, S., and Yamagata, Y.: An image-based inventory of the spatial structure of West Siberian wetlands, Environmental Research Letters, 4, 10.1088/1748-9326/4/4/045014, 2009.

Ran, Y., Li, X., and Cheng, G.: Climate warming over the past half century has led to thermal degradation of permafrost on

10 the Qinghai–Tibet Plateau, The Cryosphere, 12, 595-608, 2018.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., and De Lannoy, G. J. M.: Assessment of MERRA-2 land surface hydrology estimates, Journal of Climate, 0, null, 10.1175/jcli-d-16-0720.1, 2017a.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., and Partyka, G. S.: Land Surface Precipitation in MERRA-2, Journal of Climate, 30, 1643-1664, 10.1175/jcli-d-16-0570.1, 2017b.

15 Riseborough, D., Shiklomanov, N., Etzelmuller, B., Gruber, S., and Marchenko, S.: Recent advances in permafrost modelling, Permafrost Periglac, 19, 137-156, 10.1002/ppp.615, 2008.

Romanovsky, V. E., Sazonova, T. S., Balobaev, V. T., Shender, N. I., and Sergueev, D. O.: Past and recent changes in air and permafrost temperatures in eastern Siberia, Global and Planetary Change, 56, 399-413, 2007.

Romanovsky, V. E., Kholodov, A. L., Cable, W. L., Cohen, L., Panda, S., Marchenko, S., Muskett, R. R., and Nicolsky, D.:

20 Network of Permafrost Observatories in North America and Russia, NSF Arctic Data Center, 2009.

Romanovsky, V. E., Drozdov, D. S., Oberman, N. G., Malkova, G. V., Kholodov, A. L., Marchenko, S. S., Moskalenko, N. G., Sergeev, D. O., Ukraintseva, N. G., Abramov, A. A., Gilichinsky, D. A., and Vasiliev, A. A.: Thermal State of Permafrost in Russia, Permafrost Periglac, 21, 136-155, 2010.

Rowland, J. C., Travis, B. J., and Wilson, C. J.: The role of advective heat transport in talik development beneath lakes and

25 ponds in discontinuous permafrost, Geophys Res Lett, 38, 2011.

Schaefer, K., Liu, L., Parsekian, A., Jafarov, E., Chen, A., Zhang, T. J., Gusmeroli, A., Panda, S., Zebker, H. A., and Schaefer, T.: Remotely Sensed Active Layer Thickness (ReSALT) at Barrow, Alaska Using Interferometric Synthetic Aperture Radar, Remote Sens-Basel, 7, 3735-3759, 2015.

Schuur, E. A. G., Vogel, J. G., Crummer, K. G., Lee, H., Sickman, J. O., and Osterkamp, T. E.: The effect of permafrost thaw on old carbon release and net carbon exchange from tundra, Nature, 459, 556-559, 10.1038/nature08031, 2009.

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

Sharkhuu, N., and Sharkhuu, A.: Effects of climate warming and vegetation cover on permafrost of Mongolia, in: Eurasian Steppes. Ecological Problems and Livelihoods in a Changing World, Springer, 445-472, 2012.

Shiklomanov, N. I., and Nelson, F. E.: Active-layer mapping at regional scales: A 13-year spatial time series for the Kuparuk region, north-central Alaska, Permafrost Periglac, 13, 219-230, 2002.

Shiklomanov, N. I., Streletskiy, D. A., Nelson, F. E., Hollister, R. D., Romanovsky, V. E., Tweedie, C. E., Bockheim, J. G., and Brown, J.: Decadal variations of active-layer thickness in moisture-controlled landscapes, Barrow, Alaska, Journal of Geophysical Research: Biogeosciences, 115, 2010.

Shur, Y. L., and Jorgenson, M. T.: Patterns of permafrost formation and degradation in relation to climate and ecosystems, Permafrost Periglac, 18, 7-19, 2007.

Simmons, A., Berrisford, P., Dee, D., Hersbach, H., Hirahara, S., and Thépaut, J. N.: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets, Quarterly Journal of the Royal Meteorological Society, 143, 101-119, 2017.

Stieglitz, M., Ducharne, A., Koster, R., and Suarez, M.: The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales, J. Hydrometeorol., 2, 228-242, 2001.

Stieglitz, M., Dery, S. J., Romanovsky, V. E., and Osterkamp, T. E.: The role of snow cover in the warming of arctic permafrost, Geophys Res Lett, 30, 2003.

Tao, J., Reichle, R. H., Koster, R. D., Forman, B. A., and Xue, Y.: Evaluation and Enhancement of Permafrost Modeling With the NASA Catchment Land Surface Model, J Adv Model Earth Sy, 9, 2771-2795, 10.1002/2017MS001019, 2017.

Wang, W., Rinke, A., Moore, J. C., Cui, X., Ji, D., Li, Q., Zhang, N., Wang, C., Zhang, S., Lawrence, D. M., McGuire, A. D., Zhang, W., Delire, C., Koven, C., Saito, K., MacDougall, A., Burke, E., and Decharme, B.: Diagnostic and model dependent uncertainty of simulated Tibetan permafrost area, Cryosphere, 10, 287-306, 10.5194/tc-10-287-2016, 2016a.

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

Watanabe, K., and Wake, T.: Measurement of unfrozen water content and relative permittivity of frozen unsaturated soil using NMR and TDR, Cold Reg Sci Technol, 59, 34-41, 2009.

Yi, S. H., Woo, M. K., and Arain, M. A.: Impacts of peat and vegetation on permafrost degradation under climate warming, Geophys Res Lett, 34, 2007.

Yi, Y., Kimball, J. S., Rawlins, M. A., Moghaddam, M., and Euskirchen, E. S.: The role of snow cover affecting boreal-arctic soil freeze-thaw and carbon dynamics, Biogeosciences, 12, 5811-5829, 2015.

Zhang, T., Frauenfeld, O. W., Serreze, M. C., Etringer, A., Oelke, C., McCreight, J., Barry, R. G., Gilichinsky, D., Yang, D., and Ye, H.: Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin, Journal of Geophysical Research: Atmospheres, 110, 2005.

Zhou, J., Kinzelbach, W., Cheng, G. D., Zhang, W., He, X. B., and Ye, B. S.: Monitoring and modeling the influence of snow pack and organic soil on a permafrost active layer, Qinghai-Tibetan Plateau of China, Cold Reg Sci Technol, 90-91, 38-52, 10.1016/j.coldregions.2013.03.003, 2013.

Zona, D., Gioli, B., Commane, R., Lindaas, J., Wofsy, S. C., Miller, C. E., Dinardo, S. J., Dengel, S., Sweeney, C., and Karion, A.: Cold season emissions dominate the Arctic tundra methane budget, Proceedings of the National Academy of Sciences, 113, 40-45, 2016.

Table 1 – In-situ permafrost measurement sites covered by the AirMOSS transects in 2015.

AirMOSS flight (Official full name)	Permafrost Site (CALM or UAF)*	Latitude (degree)	Longitude (degree)	Sampling Method@	Measurement Date
COC (Council)	U27 (CALM)	64.8333	-163.7000	4	8/30/2015
	U28 (CALM)	65.4500	-164.6167	4	8/29/2015
IVO (Ivotuk)	IV4 (UAF)	68.4803	-155.7437	1[#]	8/29/2015
ATQ (Atqasuk)	U3 (CALM)	70.4500	-157.4000	4	8/25/2015
BRW (Barrow)	U1 (CALM)	71.3167	-156.6000	4	8/21/2015
	U2 (CALM)	71.3167	-156.5833	2	8/24/2015
	BR2 (UAF)	71.3090	-156.6615	1	8/29/2015
DHO (Deadhorse)	U4 (CALM)	70.3667	-148.5500	3	8/25/2015
	U5 (CALM)	70.3667	-148.5667	4	8/11/2015
	U6 (CALM)	70.1667	-148.4667	3	8/26/2015
	U31 (CALM)	69.6969	-148.6821	3	8/15/2015
	U8 (CALM)	69.6833	-148.7167	3	8/27/2015
	U32A (CALM)	69.4410	-148.6703	3	8/16/2015
	U32B (CALM)	69.4010	-148.8056	3	8/16/2015
	U9A (CALM)	69.1667	-148.8333	3	8/25/2015
	WD1 & WDN (UAF)	70.3745	-148.5522	1	8/29/2015
	DH2 (UAF)	70.1613	-148.4653	1	8/29/2015
	FB1 (UAF)	69.6739	-148.7219	1	8/29/2015
	FBD (UAF)	69.6741	-148.7208	1[%]	8/29/2015

	FBW (UAF)	69.6746	-148.7196	1	8/29/2015
	SG1 (UAF)	69.4330	-148.6738	1	8/29/2015
	SG2 (UAF)	69.4283	-148.7001	1	8/29/2015
	HV1 (UAF)	69.1466	-148.8483	1[%]	8/29/2015

* CALM: sites from the Circumpolar Active Layer Monitoring (CALM) network; UAF: sites from the Permafrost Laboratory at the University of Alaska Fairbanks (UAF).

@Sampling method: 1. Single point; 2. 320 random sampling points within 10m × 10m area; 3. 100m × 100m grid with a 10m sampling interval; 4. 1000m × 1000m grid with a 100m sampling interval.

5   # Two sensors are installed at IV4.

%Observations were taken from two conditions, including a frost-boil and an inter-boil area.

Table 2 – List of idealized simulation experiments along the AirMOSS transects.

Experiment Name	Meteorological forcing	Vegetation	Soil parameters*
Baseline	Original	Original	Original
HomF	Homogenized	Original	Original
HomF&Veg	Homogenized	Homogenized	Original
HomF&Veg&Soil	Homogenized	Homogenized	Homogenized
HomVeg&Soil	Original	Homogenized	Homogenized
HomF&Soil	Homogenized	Original	Homogenized

*CLSM soil parameters include soil organic carbon content, porosity, saturated hydraulic conductivity, Clapp-Hornberger parameters, wilting point, soil class, sand and clay fraction, vertical decay factor for transmissivity, baseflow parameters, area partitioning parameters, and time scale parameters for moisture transfer (Koster et al., 2000; Ducharne et al., 2000; Tao et al., 2017).

Table 3 – Evaluation metrics for model-simulated ALT and AirMOSS retrievals for 2015.

Metric	All sites		Sites with ALT measurements within AirMOSS sensing depth (~60 cm)	
	CLSM-simulated ALT	AirMOSS ALT retrievals	CLSM-Simulated ALT	AirMOSS ALT retrievals
RMSE (m)	0.17	0.17	0.12	0.06
Bias (m)	-0.05	-0.12	0.01	-0.01
R	0.27	0.61	-0.00	0.64

Table 4 – Evaluation results for simulated permafrost extent against the permafrost map by Brown et al. (2002). The calculation was based on the comparison between simulated permafrost area and the total area of continuous, discontinuous and sporadic permafrost regions from Brown's map. The number in the brackets was calculated against the total area of continuous and discontinuous permafrost regions.

Case	CLSM	Obs.	Simulated Area ($\times 10^6 \, \text{km}^2$)	Percentage Relative to Observation
4	No	No	48.8	-
3	Yes	No	1.9	-
2	No	Yes	3.2 (1.7)	18.7% (12.3%)
1	Yes	Yes	13.8 (12.3)	81.3 % (87.7%)
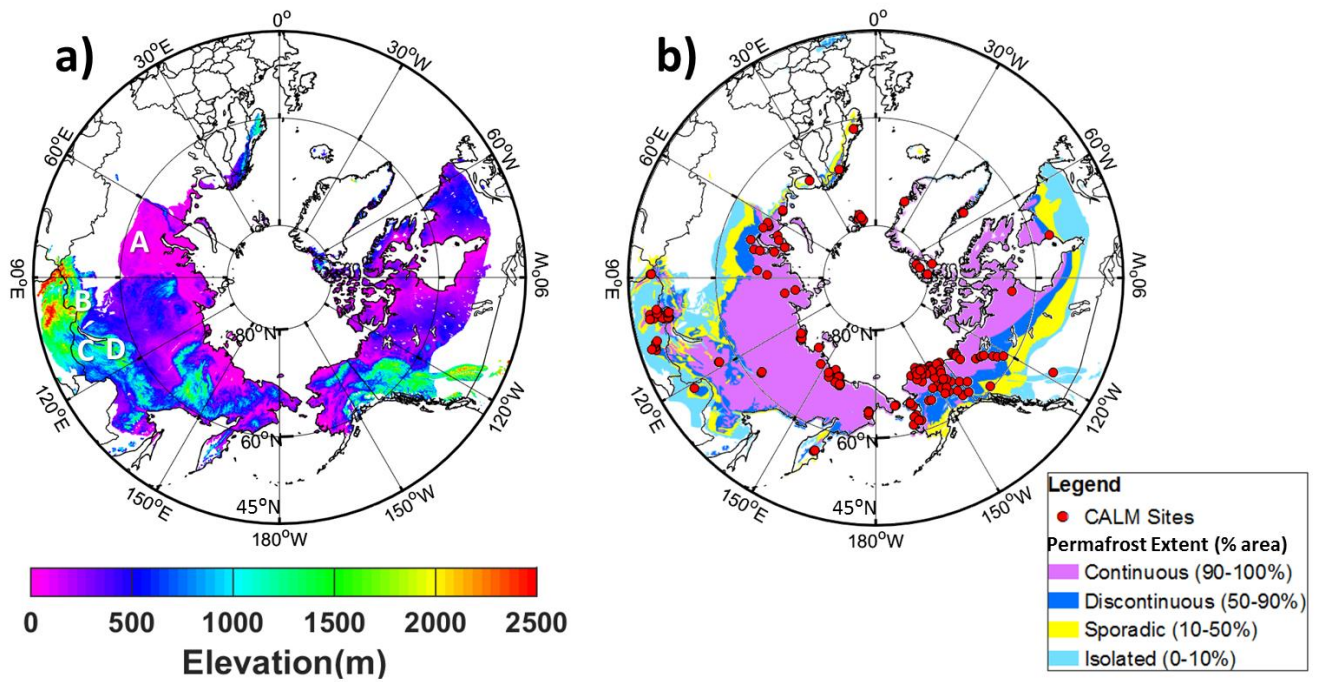
Figure 1: a) Elevation above mean sea level in the simulation domain, which is defined by the area for which NCSCDv2 data are available. Regions A, B, C, and D are discussed in the text. **b)** Permafrost and ground ice conditions adapted from (Brown et al., 2002). Red dots represent CALM sites.
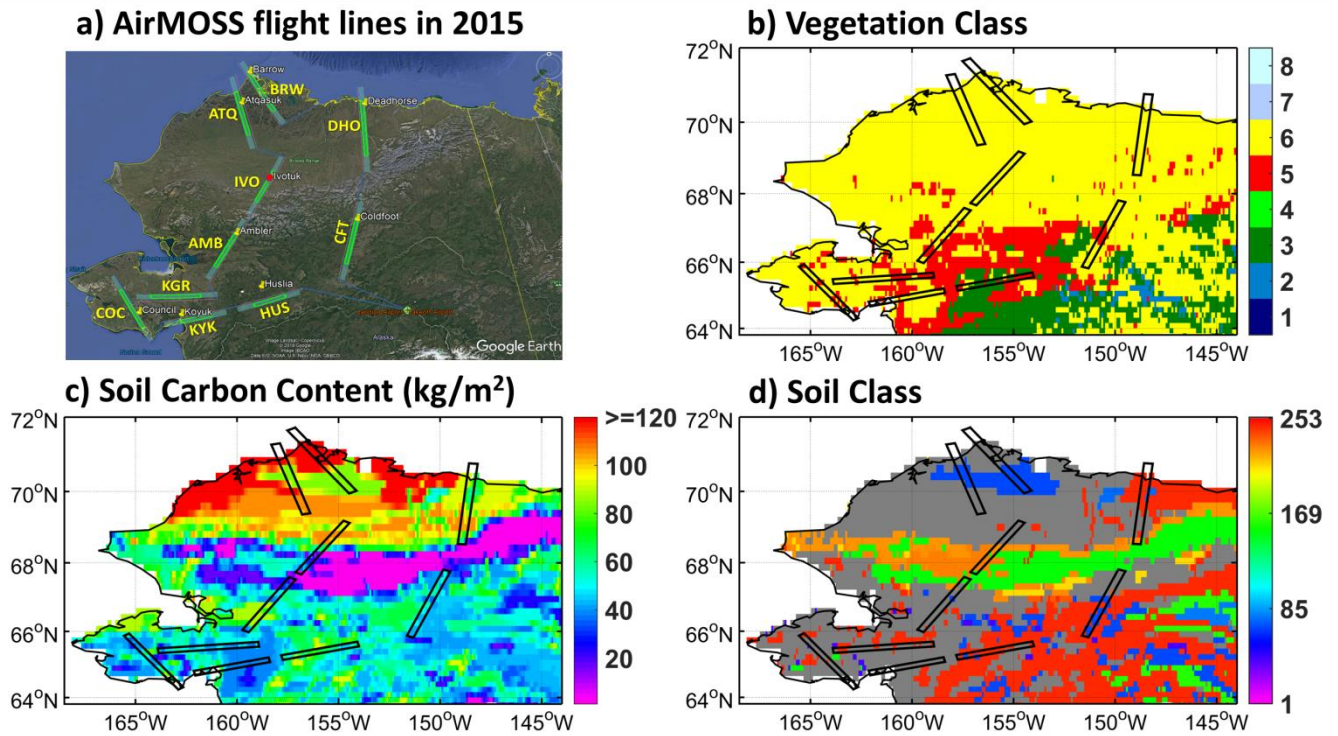
Figure 2: a) Ten transects of AirMOSS flights conducted in Alaska on 29 August 2015 and 1 October 2015, including HUS (Huslia), KYK (Koyuk), COC (Council), KGR (Kougarok), AMB (Ambler), IVO (Ivotuk), ATQ (Atqasuk), BRW (Barrow), DHO (Deadhorse), and CFT (Coldfoot). Each flight swath width is approximately 15 km. The red dot on IVO illustrates the location of the representative grid cell used and discussed in section 3.2. b) Vegetation class, c) soil organic carbon content, and d) soil class used in CLSM. The eight vegetation classes are 1-broadleaf evergreen trees, 2-broadleaf deciduous trees, 3-needleleaf trees, 4-grassland, 5-broadleaf shrubs, 6-dwarf trees, 7-bare soil, and 8-desert soil, respectively. The 253 soil classes include one "peat" class (#253), which is shown in dark grey, and 252 mineral soil classes (De Lannoy et al., 2014).

Figure 3: a) Radar retrievals of ALT derived from P-band radar observations on 29 August 2015 and 01 October 2015 for IVO, ATQ, BRW, and DHO, aggregated to 81 km² model grid cells. b) CLSM-simulated ALT. c) Difference between the aggregated ALT retrievals and the CLSM-simulated results. Magenta squares represent CALM sites covered by the flight swath whereas black circles represent UAF sites.
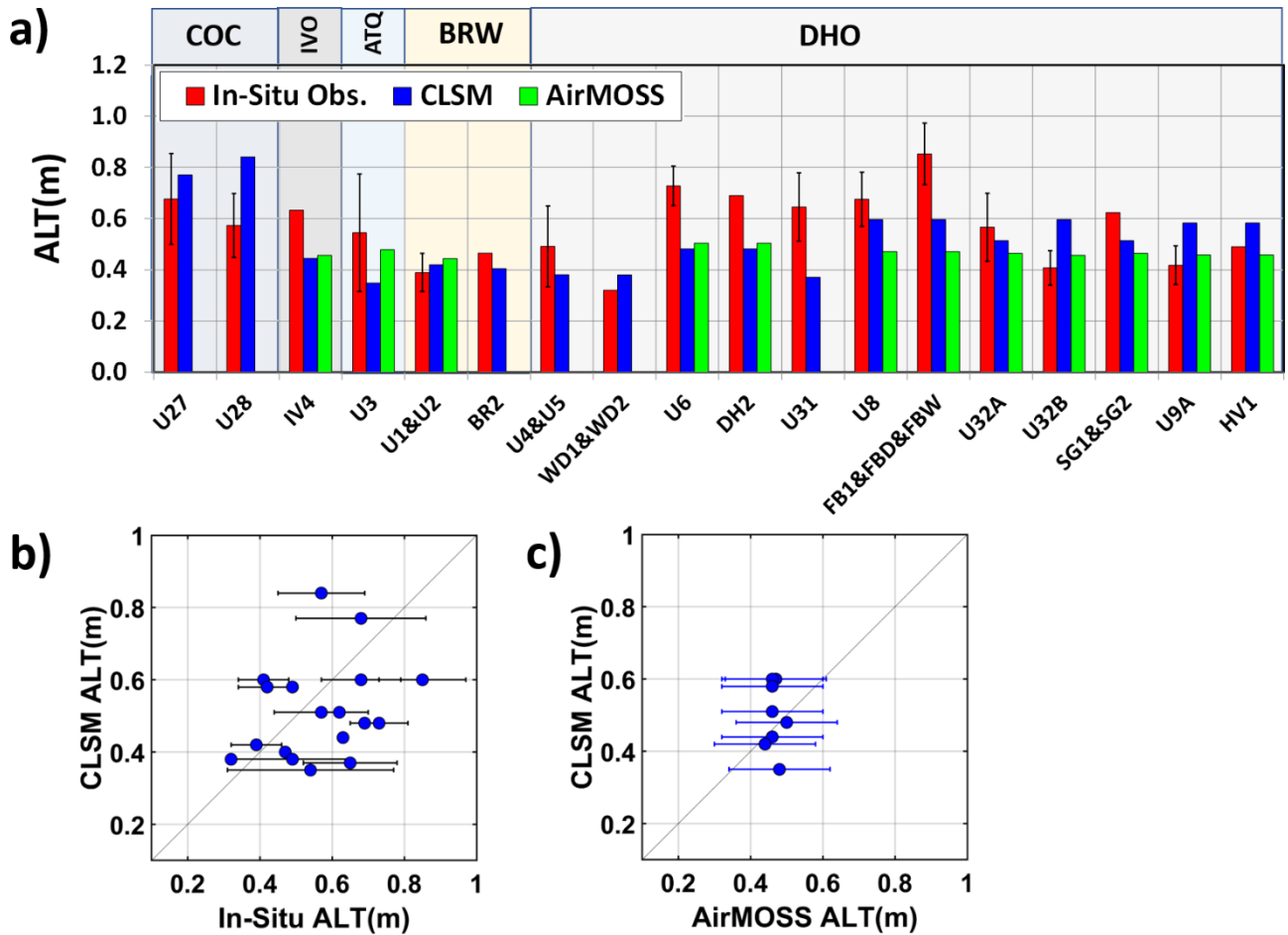
Figure 4: a) ALT observations (red) for 2015 from CALM and UAF sites covered by AirMOSS swaths and from radar retrievals aggregated to 81 km² grid cells (green), and CLSM-simulated ALT at 81 km² (blue). The short name of the corresponding covering swath is shown on the top (see also Figure 2a). Error bars represent the standard deviation for multiple observations at in-situ sites. No standard deviations are provided for UAF sites since single-point measurements were deployed. Averaged values were provided if multiple sites appear within a same model grid cell (e.g., U1&U2, U4&U5, WD1&WD2, FB1&FBD&FBW, and SG1&SG2). The sites are arranged aligning with the flight direction. b) CLSM estimates of ALT for 2015 versus in-situ measurements with error bars indicating the standard deviation as in a). c) Same as b) but versus aggregated AirMOSS ALT at model scale. The error bars here represent the uncertainty for radar retrievals mean within each 81 km² grid cell as explained in section 3.1. Corresponding estimates of CLSM uncertainty, which are presumably large, are not shown in the figure.
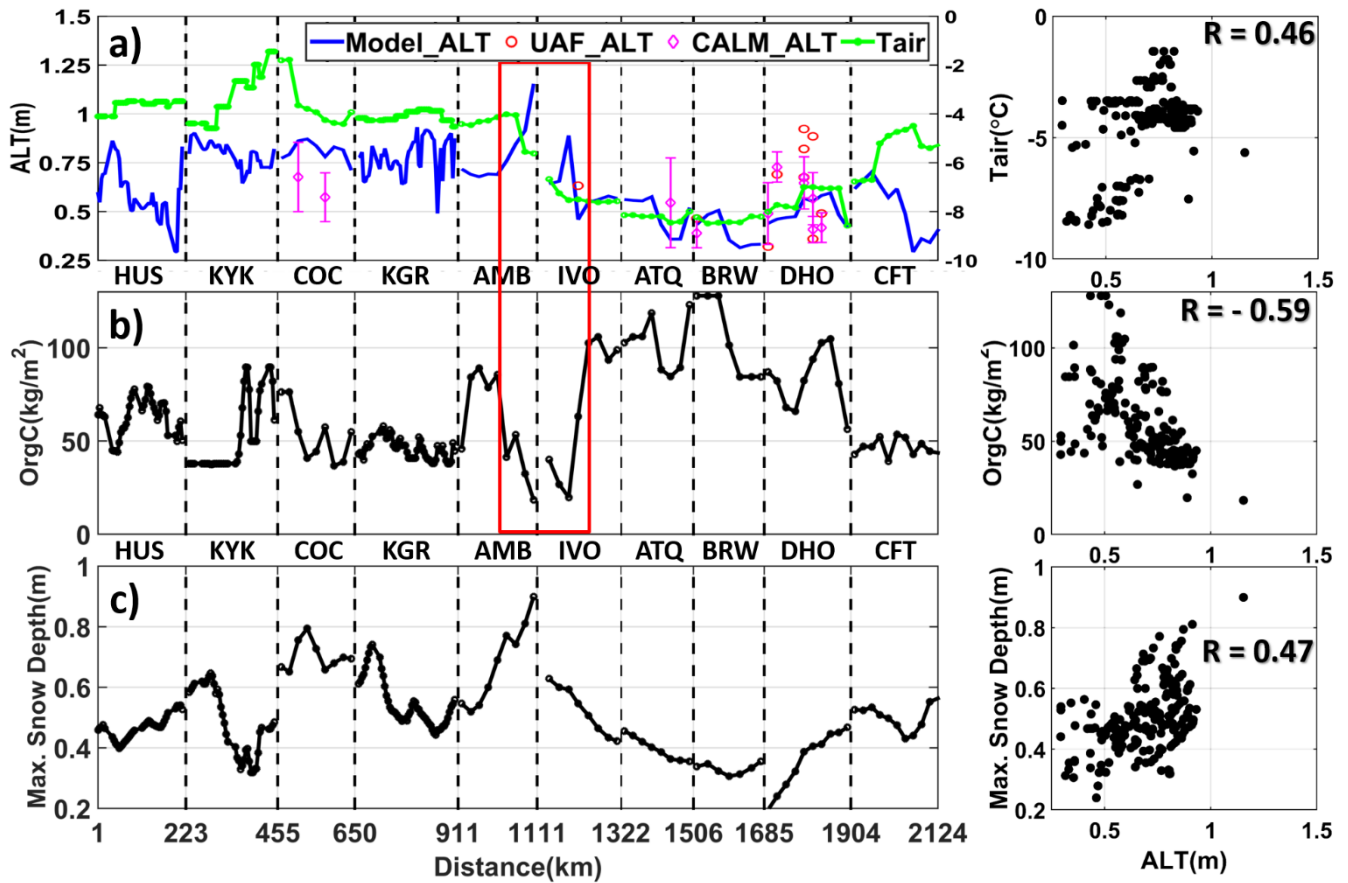
Figure 5: a) CLSM-simulated ALT (thawed-to-frozen depth) on 29 August 2015 along the AirMOSS flight transects. In-situ ALT observations from UAF and CALM are shown as red circles and magenta diamonds, respectively. Averaged air temperature at 2 m (Tair) from the preceding annual period (i.e., 01 September 2014 to 31 August 2015) is shown in green with the scale on the right ordinate. b) Organic carbon content and c) maximum snow depth during the preceding annual period (again from 01 September 2014 to 31 August 2015). The red rectangle across a) and b) highlights a portion of the domain that shows anti-correlation between organic carbon content and modelled ALT (see Section 4.2). The abscissa in c) provides cumulative distances in units of km along the transects.
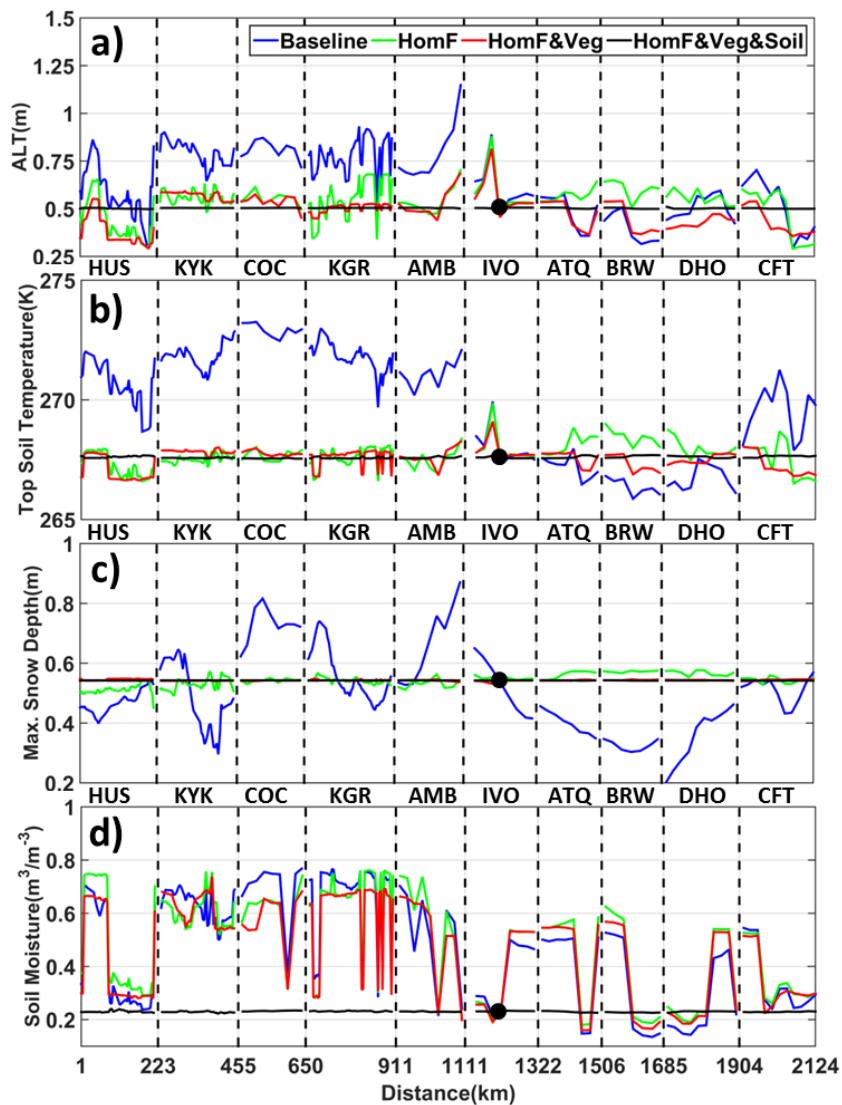
Figure 6: a) CLSM-simulated ALT (thawed-to-frozen depth) on the flight date (i.e., 29 August 2015) from the top four experiments listed in Table 2; b) simulated top layer soil temperature on the flight date, c) maximum snow depth the during the preceding annual period (i.e., from 01 September 2014 to 31 August 2015), and d) soil moisture within the soil profile on the flight date along the connected transects for the four experiments. The black dot indicates the representative location within the IVO transect from which the forcing, vegetation and/or soil data are used to homogenize the inputs in the idealized experiments. By construction, all simulations provide identical results at this representative location.
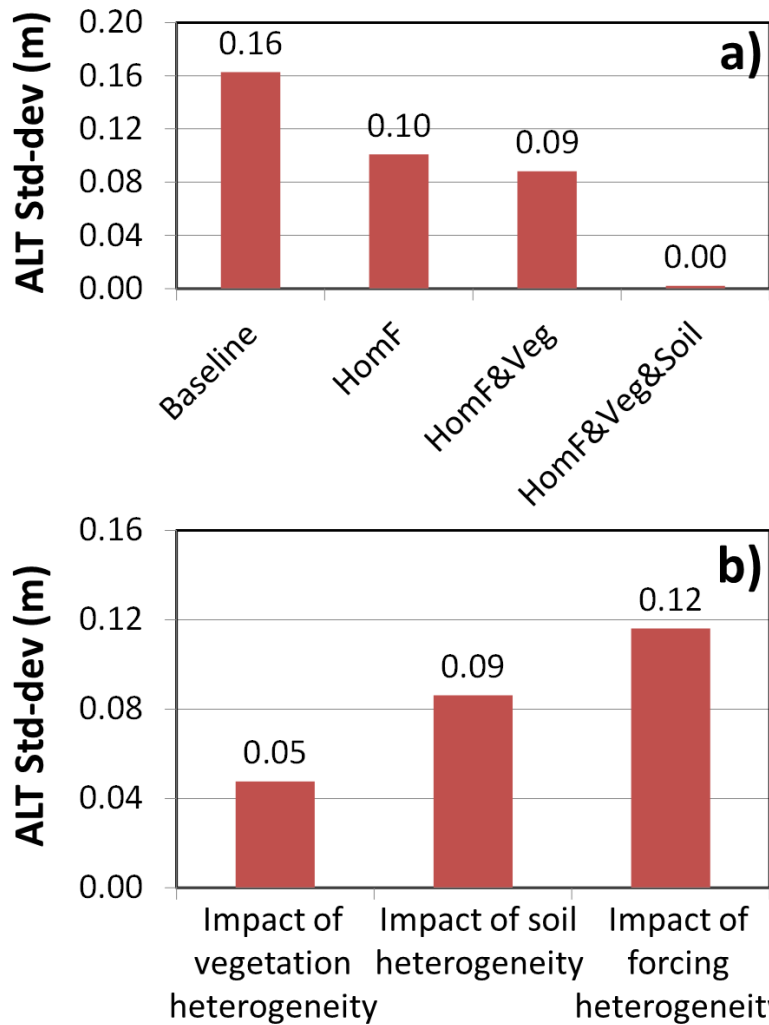
Figure 7: a) Standard deviation of ALT along the AirMOSS transects from the top four experiments listed in Table 2. b) The individual impact (or contribution) from heterogeneous vegetation, soil type and meteorological forcing, respectively. For instance, the impact of vegetation (or soil, or forcing) heterogeneity is the ALT standard deviation along the transects from HomF&Soil (or HomF&Veg, or HomVeg&Soil).
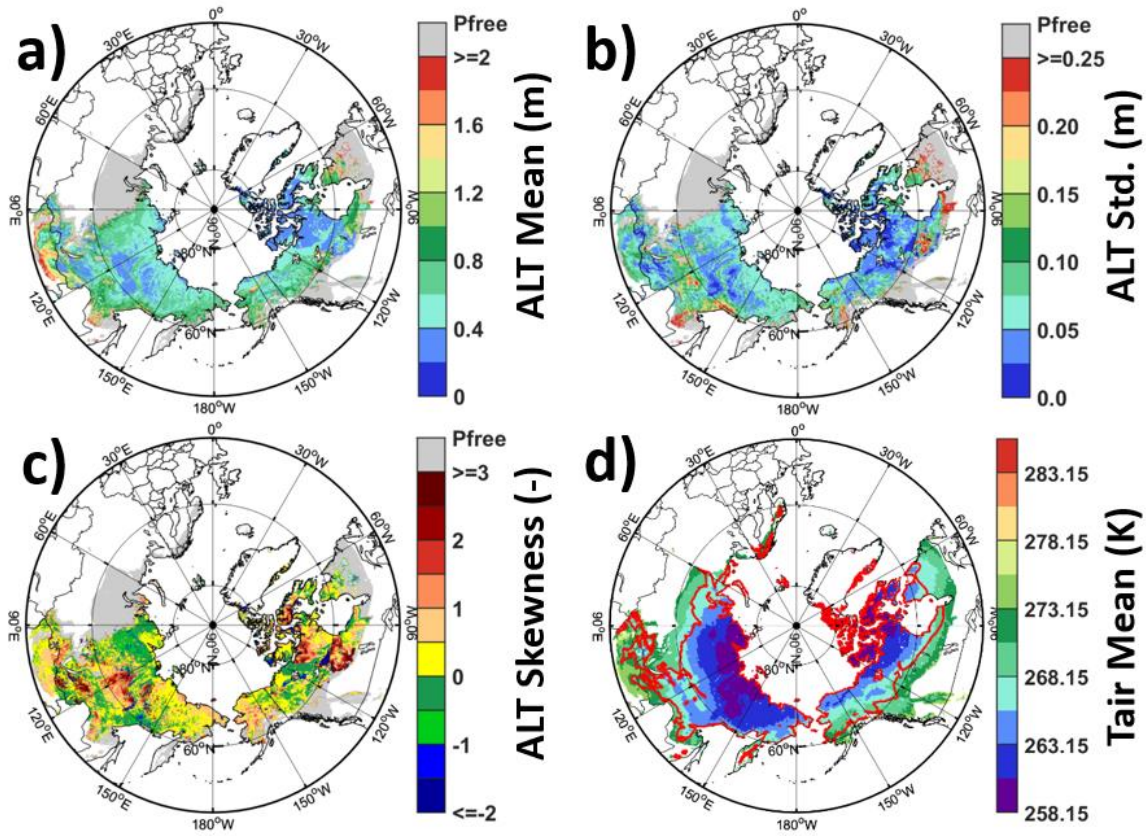
Figure 8: a) Mean, b) standard deviation, and c) skewness of CLSM-simulated ALT over the 38 years (1980 - 2017). Grey indicates permafrost-free (Pfree) areas in the simulation. d) 38-year averaged MERRA-2 annual atmospheric temperature at 2 m above displacement height (Tair). The red boundary outlines the continuous and discontinuous permafrost regions according to Brown et al. (2002).
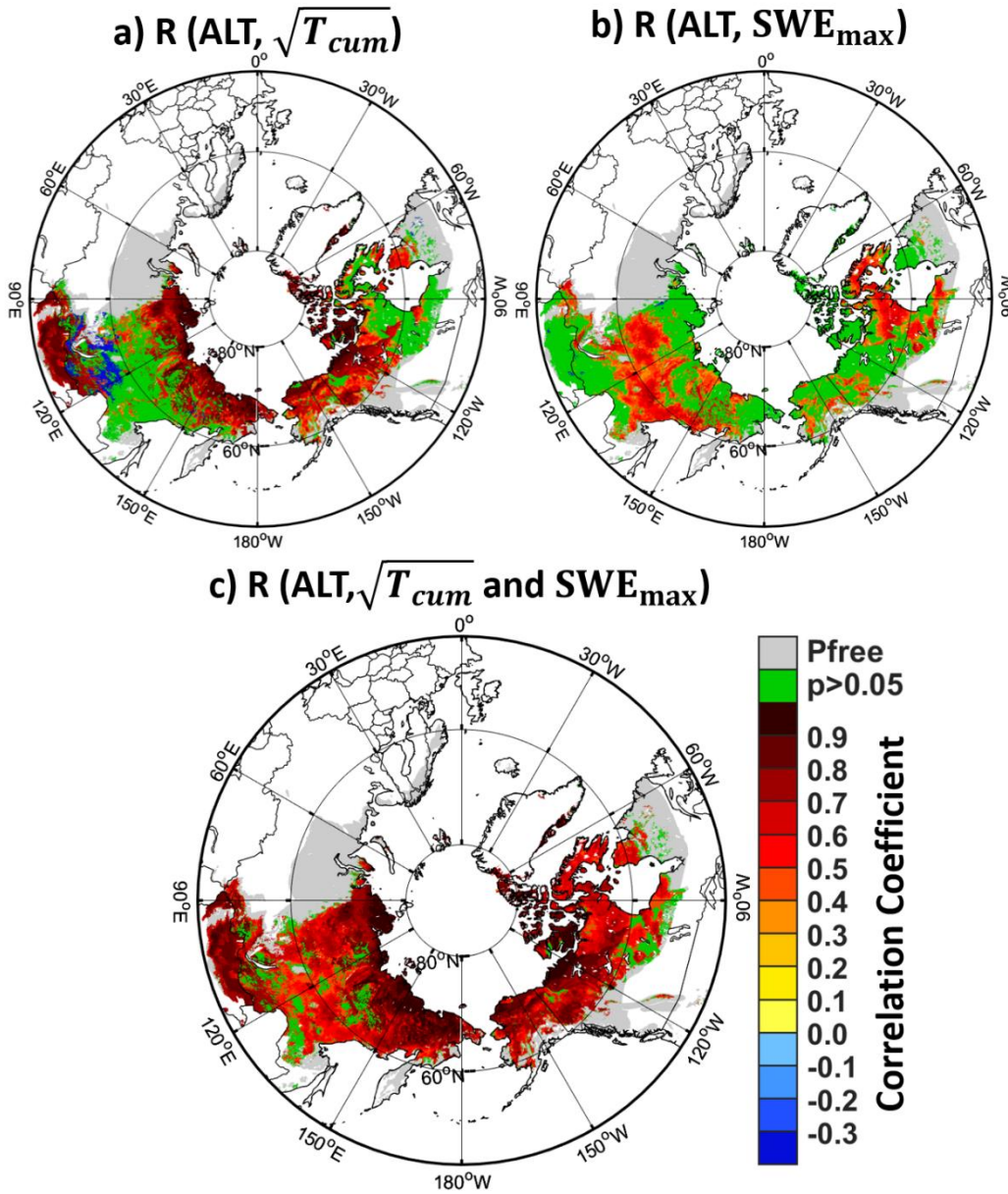
Figure 9: Correlation coefficient between a) ALT and square root of the effective accumulated air temperature ($\sqrt{T_{cum}}$) and b) ALT and maximum SWE (SWE_{max}) from the preceding September to the present August over the period 1980-2017. c) Multi-variable coefficient of correlation for a fitted multiple linear regression model between ALT and $\sqrt{T_{cum}}$ and SWE_{max}. Areas that have a p-value larger than 0.05 (i.e., statistically insignificant correlation) are masked in green. Grey indicates permafrost-free (Pfree) areas in the simulation.
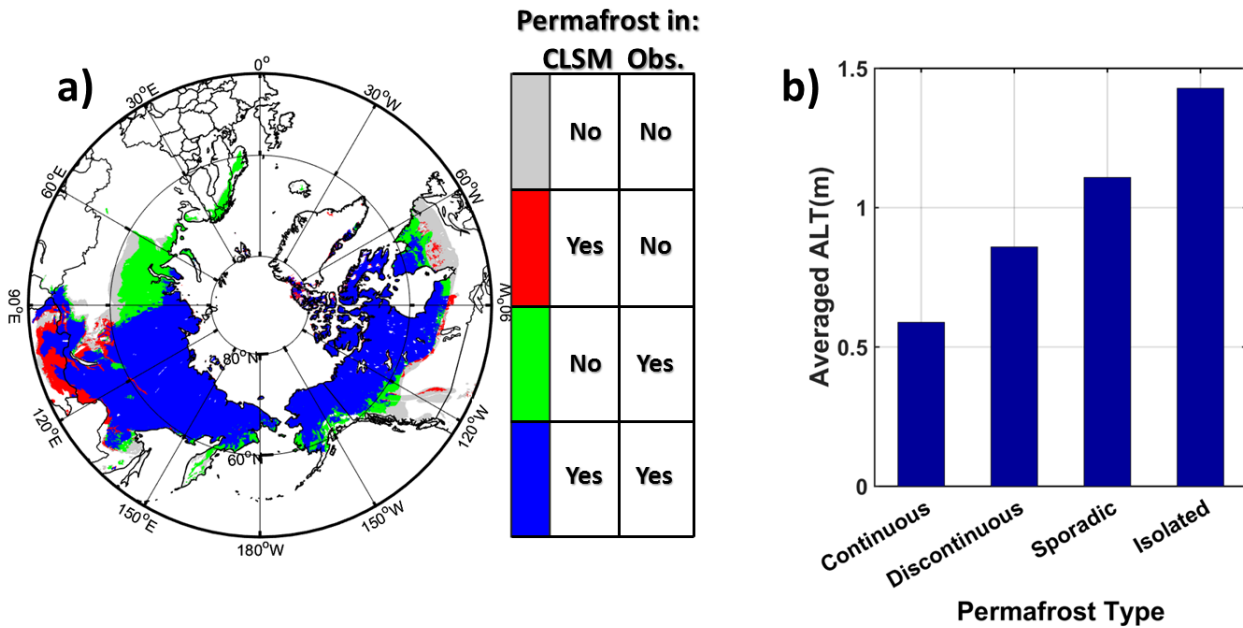
Figure 10: a) Four comparison categories include: 1) blue - CLSM collocates permafrost with the observation-based permafrost map of Brown et al. (2002) as either continuous, discontinuous, or sporadic permafrost; 2) green - CLSM has no permafrost, but the observation-based permafrost map does as either continuous, discontinuous, or sporadic types; 3) red - CLSM does have permafrost, but the observation-based permafrost map does not or contains isolated permafrost; and 4) grey - CLSM has no permafrost and neither does the observation-based permafrost map (except for isolated permafrost). **b)** area-weighted average of ALT as simulated by CLSM for the four different permafrost types.
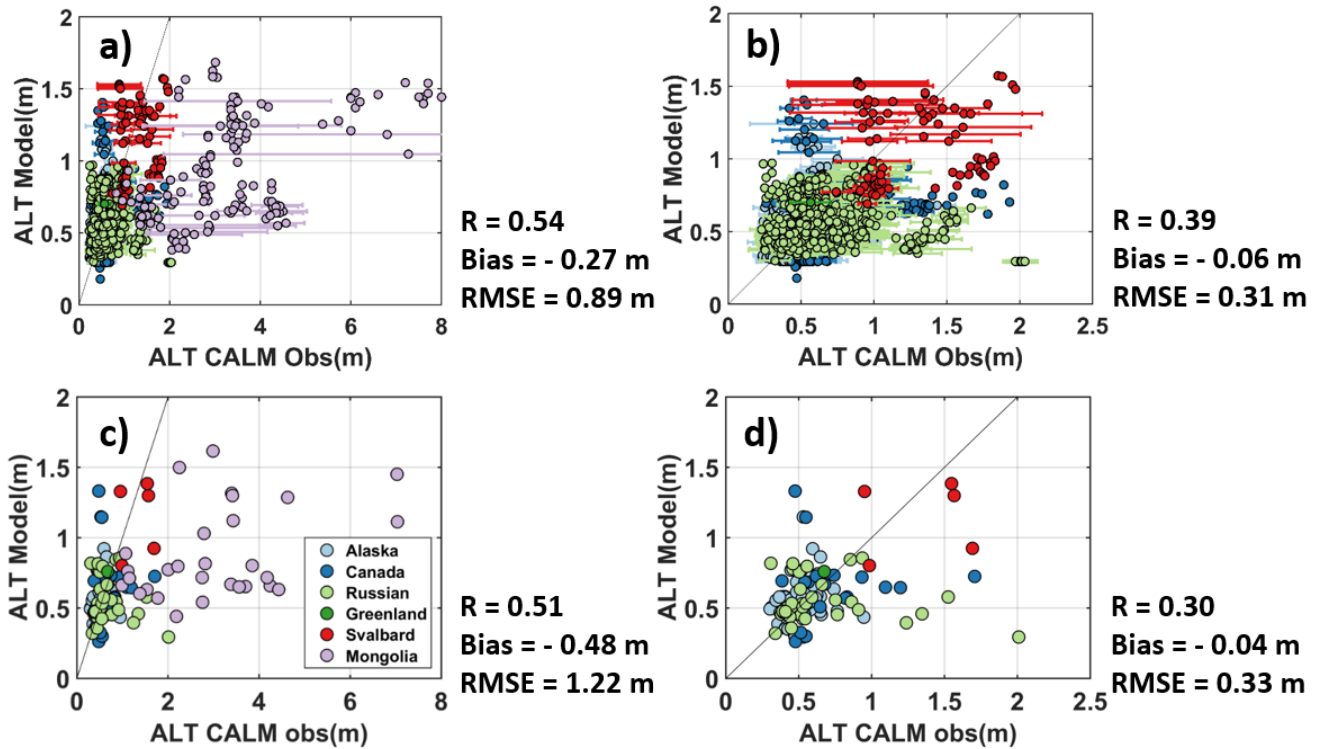
Figure 11: a) Annual ALT from CLSM simulation vs. CALM observations with horizontal error bars indicating standard deviations of measurements within the model grid cell. Error bar is absent if the number of measurements within a 81 km² grid cell is less than three. b) As in a) but excluding the Mongolia sites. c) 38-yr average ALT for the period 1980-2017 from CLSM simulation vs. CALM observations. d) As in c) but without the Mongolia sites. The correlation coefficient (R), bias, and root mean squared error (RMSE) are provided next to each subplot.
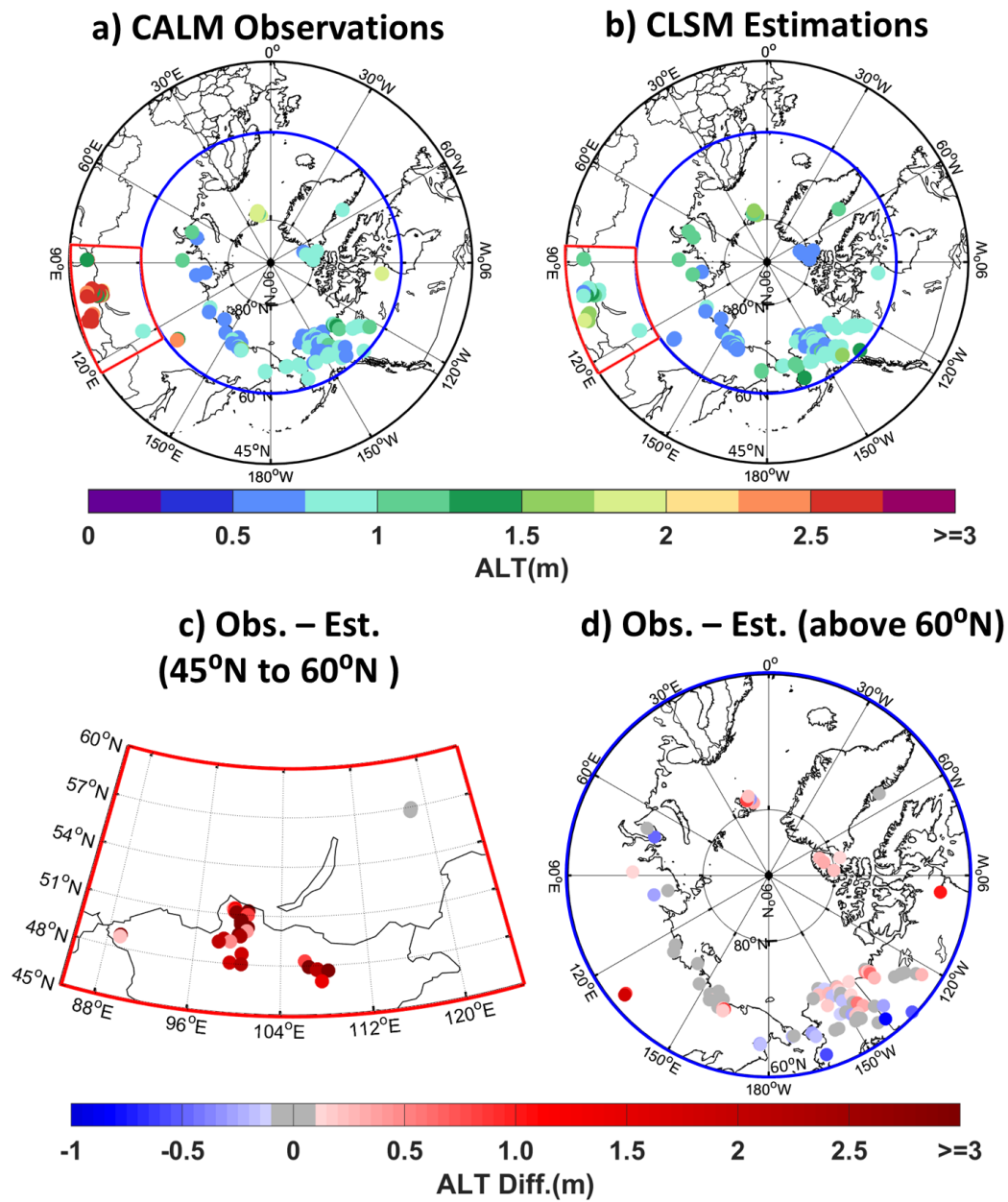
Figure 12: Multi-year average ALT at CALM site locations for a) CALM observations and b) CLSM results. c) ALT difference between observations and model results for locations within 45ºN- 60ºN latitude and 85ºE-125ºE longitude. d) Same as c) but for locations poleward of 60ºN latitude. In c) and d) grey indicates absolute ALT differences less than 0.10 m.