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

Jing Tao[1], Randal D. Koster[2], Rolf H. Reichle[2], Barton A. Forman[3], Yuan Xue[3], 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. Now at George Mason University, Fairfax, Virginia, USA.
[4]Department of Electrical Engineering, University of Southern California, Los Angeles, California, USA

Correspondence to: Jing Tao (jingtao@umd.edu)

Abstract. This study introduces and evaluates a comprehensive, model-generated dataset of Northern Hemisphere permafrost conditions. Surface meteorological forcing fields from the Modern-Era Retrospective Analysis for Research and Applications-2 (MERRA-2) reanalysis data 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 captures well the observed extent of continuous permafrost except in Western Siberia, which is permafrost-free in the simulation. Noticeable discrepancies also appear along the southern edge of the permafrost region where sporadic and isolated permafrost types dominate. The evaluation of the simulated active layer thickness (ALT) climatology against in-situ measurements demonstrates reasonable skill except in Mongolia. In northern Alaska, both ALT retrievals from airborne remote sensing for 2015 and the corresponding simulated ALT exhibit reasonable accuracy vs. in-situ measurements. However, the remotely sensed ALT retrievals generally demonstrate lower levels of spatial variability than both the observed and simulated ALT. Controls on the spatial variability of ALT are examined with idealized numerical experiments focusing on northern Alaska; meteorological forcing and soil type 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 air temperature and maximum snow water equivalent (SWE) explain most of the year-to-year variability of ALT nearly everywhere over the model-simulated permafrost regions. Simulated ALT trends from 1980 to 2017 indicate that some permafrost areas are experiencing significant degradation, with ALT increasing up to 0.5 cm/year.

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 (Schuur et

al., 2015; MacDougall et al., 2012; Dorrepaal et al., 2009; Schuur et al., 2009). 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).

5   Permafrost variations, including pronounced permafrost degradation due to a warming climate, have been reported for many regions, including Alaska (Jorgenson et al., 2001; Liu et al., 2012; Nicholas and Hinkel, 1996; Batir et al., 2017; Osterkamp and Romanovsky, 1996; Hinkel and Nelson, 2003; Jafarov et al., 2012; Jones et al., 2016), Canada (Chen et al., 2003; James et al., 2013), Mongolia (Sharkhuu and Sharkhuu, 2012), Norway (Gisnas et al., 2013), the Qinghai–Tibet Plateau (Zhou et al., 2013; Lu et al., 2017), Russia (Romanovsky et al., 2010; Romanovsky et al., 2007) and Sweden (Pannetier and Frampton, 10   2016). Some of these findings are based on in-situ measurements at a point-scale or at a spatially-aggregated scale (up to 1000m×1000m), such as through the Circumpolar Active Layer Monitoring (CALM) network. In particular, rapidly accelerated permafrost degradation in recent years has already been reported at CALM in-situ sites over the Northern Hemisphere (Luo et al., 2016).

15   For larger spatial scales, numerical model simulations are potentially useful. Simulations and/or predictions with the Land Surface Model (Dankers et al., 2011; Guimberteau et al., 2018; Tao et al., 2017), the Community Land Model (Lawrence and Slater, 2008; Lawrence et al., 2008; Nicolsky et al., 2007; Alexeev et al., 2007; Lawrence et al., 2012; Koven et al., 2013; Lawrence and Slater, 2005; Chadburn et al., 2017; Yi et al., 2007) and other numerical models (Jafarov et al., 2012; Gisnas et al., 2013), for example, have been used to quantify large-scale permafrost patterns (i.e., distributions and thermal states) 20   and their interactions with a warming climate. Given that the average air temperature in the Arctic in 2017 (ending in September) was the second warmest on record since 1900 (Arctic Report Card; http://www.arctic.noaa.gov/Report-Card/Report-Card-2017) and that 2017 was the warmest year on record for global ocean temperatures (Cheng and Zhu, 2018), changes in permafrost conditions can be expected. However, some aspects of the current global permafrost thermal states are arguably still unknown, a deficiency with relevance to ecosystem management and policy decisions.

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 (Yi et al., 2007; Fisher et al., 2016; Shur and Jorgenson, 2007; Tao et al., 2017; Johnson et 30   al., 2013; Jean and Payette, 2014; Yi et al., 2015; Stieglitz et al., 2003; Luetschg et al., 2008; Matyshak et al., 2017; Dankers et al., 2011). 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.

Again, such understanding can potentially be derived from models. Permafrost dynamics can be modelled, for example, by driving a land surface model offline (i.e., uncoupled from an atmospheric model) with meteorological forcing data (including air temperature, radiation, and precipitation) from some credible source. During the course of the simulation, the model produces estimates of ALT and permafrost thermal characteristics. A wide range of simulated permafrost behaviour has been

5 reported in the literature, with differences reflecting model-specific process representations and biases associated with different meteorological forcing datasets (Barman and Jain, 2016; Slater and Lawrence, 2013; Guimberteau et al., 2018; Guo et al., 2017). The latter source of bias is particularly difficult to reconcile given that direct observations of meteorological variables in most parts of the high latitudes are sparse. In addition, reanalysis datasets that assimilate a variety of observations provide global coverage but still inevitably have biases in high latitudes due to this observation sparsity in cold

10 regions combined with extensive challenges in physical process modelling.

Despite these issues, permafrost behaviour simulated with a land surface model driven by reanalysis forcing fields can still be useful for understanding the impacts of climate variability on permafrost, and, in turn, to evaluate the performance of the reanalysis data. The present paper utilizes this approach in the context of the Modern-Era Retrospective Analysis for

15 Research and Applications-2 (MERRA-2), an atmospheric reanalysis system that assimilates a wide range of conventional and satellite observations (Gelaro et al., 2017). MERRA-2 has been found skilful in its simulation of near-surface atmospheric conditions (Reichle et al., 2017a; Reichle et al., 2017b; Bosilovich et al., 2015; Bosilovich et al., 2017). For the present paper, we generated a dataset of Northern Hemisphere permafrost conditions by driving an updated version of NASA's Catchment Land Surface Model (CLSM) with MERRA-2 surface meteorological forcing fields for the middle-to-

20 high latitudes across the Northern Hemisphere over the period 1980-2017.

Detailed observations are another obvious source of understanding, and here, to complement our modelling analysis, we also make use of remote sensing information from the NASA Airborne Microwave Observatory of Subcanopy and Subsurface (AirMOSS) mission. In 2015, AirMOSS acquired P-band (420-440 MHz) radar observations over portions of northern

25 Alaska from which Chen et al. (2018) retrieved regional estimates of ALT and soil layer dielectric properties which are related to soil moisture and freeze/thaw states.

In the present paper we evaluate our simulated ALT and permafrost thermal conditions against an observations-based permafrost distribution map, against in-situ observations from CALM, and against the ALT retrievals derived from

30 AirMOSS. Overall we pursue three scientific objectives: 1) evaluate the relative importance of the factors that determine the spatial variability of ALT, 2) evaluate CLSM-simulated ALT characteristics against observations, and 3) quantify and assess the large-scale characteristics of ALT (in terms of both means and interannual variability) in Northern Hemisphere permafrost regions from 1980 through 2017. Section 2 below describes the datasets used and our analysis approach, and Section 3 provides results. Our findings are summarized in Section 4.

2 Approach 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

5   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). CLSM also includes a three-layer snow model that estimates the evolution of snow water equivalent

10   (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 updated version of CLSM used here (Tao et al., 2017) 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.

20   The updated version produces more realistic subsurface thermodynamics in cold regions than does the original scheme (Tao et al., 2017).

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

25   transition. Precisely, the thawed-to-frozen depth is calculated as $z_{bottom}(l) - f_{ice}(l,t) \times \Delta z(l)$, 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 where $f_{ice}(l,t) \in [0 \quad 1]$, and $\Delta z(l)$ is the thickness of layer l. Here we use the freezing point (0^0C degree) as the threshold for a thawed or frozen state; i.e., $T(l) > 0^0$C degree indicates that layer l is fully thawed. The annual ALT for a given year, then, is defined as the maximum thawed-to-frozen depth within that year.

30   2.2 Model Simulations

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). 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 a 81-km^2 spatial resolution on the 9-km Equal Area Scalable Earth grid, version 2 (Brodzik et al., 2012).

Surface meteorological forcing were extracted from the MERRA-2 reanalysis data, which are provided at a resolution of $0.5\,°$ latitude $\times 0.625\,°$ longitude (Global Modeling and Assimilation Office (GMAO), 2015b, a). 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).

10 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).

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

15 the end of the fifth loop were used to initialize the model for the final simulation experiment from 1980 to 2017. Further details regarding model parameters and forcing inputs are found in Tao et al. (2017).

2.3 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.(2018) used

20 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

25 present paper, the remotely-sensed ALT retrievals are compared with in-situ observations and CLSM-simulated ALT.

Chen et al. (2018) 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

30 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 (Chen et al., 2016; Carey and Woo, 2005; Zona et al., 2016). Strictly speaking, the radar retrievals represent the approximate thaw depth of the thawed-to-frozen boundary on 29 August 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.

5   In the retrieval algorithm, (Chen et al., 2018) 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 account 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
10   inversion scheme was used to 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 (i.e., saturation occurs at an approximate depth of 60 cm) and may thus underestimate the actual thaw depth.

15   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 (Chen et al., 2018), 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
20   discontinuous permafrost where the ALT often exceeds the P-band radar sensing depth of about 60 cm and thus cannot be retrieved from AirMOSS observations.

2.4 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
25   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%) permafrost, where the numbers in parentheses indicate the 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
30   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 been taken too early in the season or under unusual conditions (e.g., after the site was burned or covered with lava). 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 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 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.

The daily UAF in-situ soil temperature profile observations on the AirMOSS flight date (29 August 2015) were used to 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 radar retrievals, the distributed measurements for a given CALM site (see sampling methods in Table 1) were averaged into a single value. Similarly, we also aggregated the radar retrievals (20 m x 60 m) up to the respective site-scale (from 10 m × 10 m to 1000 m × 1000 m) if the CALM sampling grid for a given site included more than one AirMOSS pixel. Otherwise, the radar retrievals closest to the UAF (single-point) or CALM site are used in the site-scale comparison.

3 Results

3.1 Comparisons with observations across retrieval transects

In this section, we compare AirMOSS radar retrievals and CLSM simulation results against each other and against in-situ observations of ALT for 2015. The AirMOSS transects cover several different regions with different climatologic regimes, topography, vegetation and soil type (Figure 2). Thus, the variability of ALT along these transects encompasses the influence of a variety of factors at the regional scale.

First, we compare the spatial pattern of ALT estimates for 2015 between AirMOSS and CLSM (Figure 3). The radar retrievals were provided at 2-arcsec × 2-arcsec (roughly 20 m x 60 m in the Arctic) resolution (Fig. 3a) whereas the CLSM-

simulated ALTs are at 81 km^2 (Fig. 3c). We thus also aggregated the AirMOSS retrievals to the CLSM model grid (Fig. 3b) by averaging all the retrieval data points within each 81 km^2 model grid cell (Fig. 3b). Only grid cells that were at least 30% covered by radar retrievals were used in the comparison.

5 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 (Fig. 3a,b), whereas this pattern is largely reversed in the simulated ALT (Fig. 3c). The CLSM-simulated ALT (Figure 3c) is greater than that from AirMOSS (Figure 3b) in portions of each transect. Across all four transects, 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 10 spatial variability at the native resolution (Figure 3a), but some of this variability averages out during the aggregation to the relatively coarse model grid (Figure 3b). The abrupt changes in CLSM-simulated ALT shown in Figure 3c are predominantly controlled by soil type (see discussion below).

Next, we compare the retrieved and simulated ALT in 2015 with in-situ observations from the CALM and UAF sites that are 15 collocated with the AirMOSS transects. Figure 4b, c demonstrates that the CLSM-simulated results agree, to first order, with the in-situ observations – the model results in Fig. 4c, for example, are roughly clustered about the 1:1 line, with errors generally less than 0.3 m. The overall mean bias of simulated ALT relative to the in-situ measurements is -0.05 m. Nevertheless, the scatter in Fig. 4c is large. The most significant discrepancies between the CLSM-simulated ALT and in-situ measurements are at U6, U31, FB1&FBD&FBW (Fig. 4b), where the simulated ALT underestimates the in-situ 20 measurements by 0.25-0.28 m, and at U28 where the simulated ALT overestimates the in-situ ALT by 0.27 m.

The AirMOSS ALT radar retrievals, for their part, show less spatial variability than the observations (Fig. 4c). The largest error for these 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 model scale (green) are not available at some sites because 25 of our imposed 30% filling restriction (Figure 4b), whereas the retrievals at the site scale (black) are available at all sites within the IVO, ATQ, BRW and DHO transects (Figure 4a). The largest radar retrieval errors at the site scale are at FBD, FBW, and SG1, where the ALT retrievals underestimate the in-situ measurements by 0.32 - 0.53 m. For the AirMOSS retrievals, the overall ALT bias is -0.11 m at the site scale and -0.12 m at the model scale.

30 Although the AirMOSS ALT retrievals generally underestimate the in-situ ALT measurements, the retrievals are broadly consistent with the observations when the in-situ measurements are within the ~60 cm sensing depth of the P-band radar data. This is the case at sites in the northernmost area, e.g., U3, U1, U2, BR2, and U5. Excluding the sites with in-situ ALT measurements that exceed 60 cm, the overall mean bias for the AirMOSS retrievals at the site scale (model scale) drops to 0.02 m (-0.01 m), compared to a bias of 0.01 m for the CLSM simulation at the same sites.

It is important to note that the retrieved ALT was determined by the dielectric transition from thawed to frozen conditions while the modelled ALT and the ALT for some of the in-situ measurements are based on a freezing temperature of 0^0C (see sections 2.1 and 2.4). However, depending on local conditions, soil does not typically freeze at 0^0C but rather at slightly

5   lower temperatures (e.g., around -1^0C) 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 2.4) introduce considerable uncertainty into our comparisons. The use of the 0^0C degree threshold in CLSM for determining the thawed or frozen state of the soil may explain in part the model's underestimation of ALT. Further investigation of these

10   issues is left for future study.

3.2 Sources of ALT Variability

Here we investigate the specific factors that drive ALT spatial variability along all ten of the AirMOSS transects (Fig. 2a). For this analysis, the simulated ALT estimates shown in Fig. 5a were aggregated across the width of the radar swath (compare Fig. 3). Figure 5a illustrates that, in general, the simulated ALT captures the spatial variability exhibited by the in-

15   situ measurements. This conclusion is, however, very tentative given the limited number of in-situ ALT observations.

Generally, 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

20   MERRA-2 (green line in Figure 5a). (In our simulations, CLSM is forced using the MERRA-2 air temperature at the lowest model level, which is typically at 40-60 m above the surface.) 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

25   addition, heat transfer is also 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=0.45), 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 Fig. 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 to better quantify the relative impacts of these driving factors along the AirMOSS transects. Specifically, we repeated the simulation along the AirMOSS transects multiple times, each time removing the spatial variation in some aspects of the model forcing and parameters and then quantifying the resulting impacts on ALT variability.

For these supplemental simulations, we first identified a grid cell within the IVO transect (shown in Fig. 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), 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 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 and vegetation character is artificially removed.

In the final idealized experiment (HomF&Veg&Soil), spatial variability in soil type and topography-related model parameters is removed along with that of forcing and vegetation character. 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 (Koster et al., 2000; Ducharne 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). All other soil parameters are homogenized to those at the representative grid cell. Table 2 provides a summary of these idealized experiments. Taken together, the four experiments (including the baseline) allow us to identify the individual contribution of each factor to the ALT variability 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; Fig. 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.

5   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.   We speculate that this is because the generally lower albedo of the somewhat taller and leafier trees (representative of the IVO transect) during the snow season resulted in increased snowmelt and thus reduced snowpack

10   during the snow season (compare the green and red curves in Fig. 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.

As might be expected, the simulation in which soil properties are homogenized in conjunction with forcing and vegetation

15   (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 on soil moisture, spatial variability in soil moisture remains high in HomF and HomF&Veg and is only eliminated once the soil type-related parameters are homogenized (Fig. 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,

20   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 Fig. 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

25   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 Fig. 6a.

We investigated potential nonlinearities by conducting two additional experiments: one in which we homogenized both the vegetation and soil parameters (HomVeg&Soil) and another in which we homogenized both forcing and soil parameters (HomF&Soil) (Table 2). Put differently, in experiment HomVeg&Soil only the forcing varied along the transects, whereas in experiment HomF&Soil, only the vegetation parameters varied along the transects. Combined with the experiment

HomF&Veg (in which only soil properties varied along the transects), these three experiments show in a different way how each individual factor (forcing, vegetation, or soil) can contribute to ALT variability. The results, shown in Figure 7b, confirm 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.

5   3.3 ALT Characteristics over the Northern Hemisphere

In this section we quantify the large-scale characteristics of ALT over the Northern Hemisphere for the current climate (1980 - 2017) as determined by the response of the land model to 38 years of MERRA-2 forcing. Again, this forcing was applied to 81 km^2 grid cells in the middle-to-high latitude area defined by the existence of NCSCDv2 soil carbon data (see area outlined in Fig. 1a). Output diagnostics saved from this multi-decadal, offline simulation allow the characterization of

10 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 over the last 38 years.

Figure 8a shows the distribution of mean ALT over the modelling domain, and Fig. 8b shows the ALT standard deviation

15 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^{0}N at ~120^{0}E) with local minima that break this pattern. The largest ALT standard deviations (red colour in Figure 8b) are found mainly in discontinuous and sporadic permafrost regions (see Fig. 1b) where ALTs are deeper on average. Figure 8c provides the skewness of the temporal distribution. Though there are some exceptions, by and large the skewness is positive in most

20 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 near-surface air temperature (i.e., 2 m above the displacement height) as derived from MERRA-2. The observed continuous and discontinuous permafrost areas shown in Fig. 1b are well confined

25 within the cold side of the 0℃ (273.15K) isotherm in the mean air temperature map (Figure 8d). For the most part, the observed sporadic and isolated permafrost regions of Fig. 1b also lie on the cold side of the 0℃ 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 clearly needs more analysis.

Rather than focusing on mean annual temperature, we now 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 can be computed from daily-averaged air temperature, $T_{air}(t)$, and the freezing temperature, T_f (0^0C degree), as follows:

$$T_{cum} = \Sigma\, T_{pos} , \tag{1}$$

where

$$T_{pos} = \begin{cases} T_{air}(t) - T_f & if\ T_{air}(t) > T_f \\ 0 & if\ T_{air}(t) \le T_f \end{cases}, \tag{2}$$

The summation in Eq. (1) for year N is computed from 1 September of year (N-1) to 31 August of year N.

Figure 9a shows a map of the correlation coefficient between the 37-year time series (i.e., from September 1980 through August 2017) of 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 T_{cum}. Clearly, in these regions, air temperature exerts a dominant control on year-to-year ALT variability.

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 winter ground from becoming too cold, thereby facilitating the heating of the subsurface during non-winter months. Variable, but often thick, snowpack is in fact common in the areas of Fig. 9a that show a low (green) or negative (blue) correlation between ALT and T_{cum} – areas such as Central Siberia, the 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 snow water equivalent, SWE_{max}, during the preceding winter. A positive correlation is seen in many areas, most notably in areas with a poor or negative correlation between ALT and T_{cum} (Figure 9a) – for example, just west of Hudson Bay and along a zonal band at 60N in Russia. Apparently, in these areas, the impacts of snow physics on ALT outweigh the impacts of lumped energy input (T_{cum}).

In some other areas, however, ALT correlates positively with both T_{cum} and SWE_{max}. To quantify the joint contributions of T_{cum} and SWE_{max}, we performed a multiple linear regression analysis – we fit the equation

$$ALT = a_0 + a_1 T_{cum} + a_2 SWE_{max} , \tag{3}$$

to the available data. 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 variations in T_{cum} and 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 high mountains where more complex environmental controls might be playing a dominant role.

3.4 Evaluation of simulated permafrost extent and ALT climatology

We now evaluate, to the extent allowed by available CALM in-situ observations, and on a spatial scale much broader than that addressed in section 3.1, the accuracy of the permafrost fields simulated by the model. 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 the West Siberian Plain (labelled as A in Figure 1a). The reasons for this particular deficiency are unclear; perhaps the initial thermal conditions over the West Siberian Plain were too warm, or perhaps the soil column within the CLSM is not deep enough to allow a realistic permafrost table. The geographically thin disagreements between the simulated and observed permafrost extents toward the south in Figure 10a are not as much a concern, since the comparison in such areas is muddied by the interpretation of "isolated" permafrost in the observational map (Figure 1b). 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).

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 CALM network of in-situ ALT measurements allows a much more quantitative evaluation of the simulated ALTs, at least for the grid cells containing the measurement sites. Our comparisons here focus on multi-year averages; for a consistent comparison, we average the simulated ALTs over the same years for which observed ALTs are available. If multiple CALM sites lay within a single CLSM grid cell, a single "spatially-averaged" observed, multi-year value was computed for the cell.

The observed and simulated averages are compared in Figure 11. Generally, the CLSM-simulated ALT 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, observed ALTs tend to be much larger than the simulated ALTs (light purple dots in Figure 11a). The RMSE (and bias) associated with this scatterplot 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 11b). Given simplifications in the model, uncertainties in boundary conditions (e.g., soil properties), and representativeness issues stemming from the coarse-scale nature of the forcing data relative to the point-scale nature of the observations, 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 in isolation.

Comparing the observed and simulated spatial distributions of the ALT averages provides a further test of the accuracy of the simulation results. 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 (45^0N to 60^0N) and high latitudes (60^0N to 90^0N), respectively; in agreement with Figure 11a, the model clearly performs better in high-latitude regions, i.e., outside of Mongolia. Many of the sites north of 60^0N (Figure 12d) are coloured grey, indicating a small error in the simulation of ALT at these sites – the error at these sites ranges from only -0.10m to 0.10m.

The significant underestimation of ALT in Mongolia may very well 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 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. Another potential reason for the underestimation of Mongolian ALT involves the aforementioned representativeness error. The Mongolian CALM sites employ a single-point measurement method (section 2.4) and are thus not able to represent the presumed spatial heterogeneity of permafrost within an 81 km^2 CLSM grid cell.

3.5 Permafrost Variability as Indicated by ALT Trend from 1980 to 2017

The MERRA-2 dataset provides close to four decades of forcing data, and it is tempting to see if trends in this forcing imprint themselves on the simulated permafrost. We realize at the outset, however, that two difficulties hamper the accurate quantification of permafrost trends from our data. First, air temperature trends in MERRA-2 are known to be underestimated in high-latitude regions, especially in more recent years (Simmons et al., 2017). Second, our spin-up procedure essentially involved running our model over several cycles of MERRA-2 forcing prior to the final 1980-2017 simulation, meaning that

if a temperature trend does exist in the data, the initial conditions for 1980 already contain information from the warmer, later period. Note that both of these difficulties would contribute to a trend underestimation.

We keep these caveats in mind as we now compute permafrost trends through a linear regression of simulated yearly ALT
5 against the year of simulation, interpreting the slope obtained (in cm/yr) as the trend. We similarly compute trends in the mean annual air temperature (MAAT, ^{0}C/year), in T_{cum} (section 3.2), and in the number of days with air temperature above the freezing point (i.e., warm days). As in section 3.2, a given year's MAAT, accumulated air temperature, and number of warm days are computed for the year-long period ending on 31 August of that year.

10 The spatial distributions of the computed trends are shown in Figure 13. (Only trends that are statistically significant at the 0.05 level are shown in the Fig..) Figure 13a shows increasing/positive ALT trends (red colours) in many regions, particularly in northern Alaska and eastern Siberia but also in some parts of Canada and northern Mongolia. Such positive trends, which appear despite the aforementioned limitations in the analysis, can be interpreted as permafrost degradation. Particularly large positive trends (i.e., greater than or equal to 0.5 cm/year) are found along the coastline of the Okhotsk Sea
15 in eastern Siberia, in northern Mongolia, in northern Quebec, and in limited areas of central Canada and the north slope of Alaska. The fact that the highest trends are generally found at the margins of the simulated permafrost distributions is consistent with the findings of James et al. (2013).

The ALT trends in Figure 13a are, at least to some extent, consistent with the temperature trends shown in Figure 13b,c,d,
20 particularly in northern Alaska, northern Quebec, and (for MAAT) eastern Siberia. The widespread positive ALT trend in northern Mongolia is consistent with the positive trend in T_{cum} (Figure 13c). Some areas, however, show a large positive ALT trend but not an obvious corresponding trend in air temperature. Also, some locations near the regions marked B, C, and D in Figure 1a show negative ALT trends in the presence of a warming trend in T_{cum} (Figure 13c). It is possible that in such cases, the computed trends are strongly affected by snowpack variability, even though maximum SWE itself does not
25 tend to show a significant trend in these areas (not shown). It may be that in snow-dominated regions, joint variations and trends in temperature and snowpack complicate a simple interpretation of ALT trends in terms of trends in the meteorological forcing.

The ALT trend was similarly calculated from CALM observations at sites for which observed ALTs are available for at least
30 eight years (i.e., sample size ≥ 8). These observed trends are compared to the model-based trends in Figure 14. Here the model-based ALT trends were calculated using the same years for which observed ALTs are available. We exclude from the comparison the observational sites located in regions that, according to the CLSM, are permafrost-free (grey area in Figure 8a). Not counting the two Mongolian sites with trends at approximately 20-30 cm/yr, the simulated and observed trends at most of the sites roughly agree, though on balance the CLSM-based ALT trends are too low, likely due to the analysis

limitations noted at the beginning of the section. This particular comparison is, in any case, highly uncertain, given the limited number of data points available to compute the trends.

4 Conclusion and Discussion

Driving a tested model of permafrost dynamics (Tao et al., 2017) with a multi-decadal, reanalysis-based dataset of
5 meteorological forcing (Gelaro et al., 2017) allows a global-scale characterization of permafrost. Indeed, in the course of this work we have produced a dataset (effectively a derivative of MERRA-2) of permafrost variations in space and time at the 81 km^2 scale across middle-to-high latitudes. Analysis of these data, along with data produced in idealized experiments with specific homogenized controls, show how the statistics of permafrost variability in space and time 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 statistical analysis focusing on the global-scale fields reveals that yearly ALT strongly correlates with an accumulated air temperature diagnostic in most permafrost regions. In regions where they do not correlate well, yearly ALT does correlate with maximum SWE. Indeed, 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
20 CLSM-identified permafrost regions.

The spatial distribution of CLSM-simulated permafrost shows good general agreement with the observation-based permafrost map of Brown et al. (2002). The model-based product does seem to miss a large area of permafrost in the West Siberian Plain, but it captures correctly, for the most part, the southward extent of permafrost. Apparent errors along the
25 southern edge are, in any case, subject to significant uncertainty in this comparison given the presence of "sporadic" and "isolated" permafrost in the observational map, types that do not have direct analogues in the model-based (binary) product.

The CLSM-simulated ALT climatology was also compared to that derived from in-situ measurements collected through the CALM network. The simulated ALTs agree well with the in-situ observations for shallow permafrost in high-latitude
30 regions (above 60^{0}N latitude), but they generally underestimate ALTs in middle-latitude regions, especially in Mongolia. The RMSE of climatological ALT between model simulations and observations is 1.22m, and the mean bias is -0.48m. However, these reduce to 0.33m and -0.04m, respectively, when the Mongolia sites are excluded. In northern Alaska, ALT

retrievals from airborne remote sensing for 2015 and the corresponding simulated ALT exhibit reasonable accuracy vs. in-situ measurements. However, the remotely sensed ALT estimates generally demonstrate lower levels of spatial variability than the simulated ALT estimates. Moreover, the spatial patterns of the remotely sensed and simulated ALT differ considerably.

The simulated fields indicate permafrost degradation (as represented by a positive ALT trend) over the past 38 years in many areas, including large areas in Alaska and eastern Siberia as well as some limited patches in Canada and northern Mongolia. While concurrent trends in air temperature can reasonably explain many of the ALT trends, this connection appears to fall apart in regions dominated by snowpack variability. Modeled ALT trends agree with observed ALT trends (based on

10 analysis of data from the CALM network) within a reasonable range (i.e., -1cm/year to 1cm/year) at many sites. At other sites, however, modeled trends are biased low, and they grossly underestimate the observed trends at two Mongolian sites. We emphasize again that such trend analyses are in any case highly uncertain due to the relatively short temporal extent of the observational ALT record.

15 Spatial representativeness issues plague an evaluation of simulated ALT (representing averages across 81 km^2 grid cells) against site-based ALT measurements. That said, we fully expect that the discrepancies seen between the simulated and observational ALT estimates reflect problems on the modelling side. For example, the observed climatological ALTs at the Mongolia sites are all larger than 3m. This depth falls well within the 6th temperature 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.

20 Resolving this issue would require increasing the model's vertical resolution through the addition of soil layers. 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 soil carbon (not shown).

25 Another issue affecting our comparisons is the static representation of vegetation used in CLSM. 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 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; Kurylyk et al., 2014; Rowland et

30 al., 2011). Also relevant are potential errors in the MERRA-2 forcing, which has a particularly large impact on the trend analysis; as mentioned above, 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

5   MERRA-2 reanalysis, which will be avaiable 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.

Acknowledgments

Funding for this work was provided by the NASA Interdisciplinary Science program (NNX14AO23G). We thank Qing Liu

10   at GMAO/GSFC/NASA for providing us corrected MERRA-2 precipitation. We acknowledge the University of Maryland supercomputing resources (http://www.it.umd.edu/hpcc) made available for conducting the research reported in this paper.

References

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.

15   Barman, R., and Jain, A. K.: Comparison of effects of cold-region soil/snow processes and the uncertainties from model forcing 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.

Bosilovich, M. G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P.,

20   Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S., Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D., and Suarez, M.: MERRA-2: Initial Evaluation of the Climate, NASA Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2015-104606, Vol. 43, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 2015.

Bosilovich, M. G., Robertson, F. R., Takacs, L., Molod, A., and Mocko, D.: Atmospheric Water Balance and Variability in

25   the MERRA-2 Reanalysis, Journal of Climate, 30, 1177-1196, 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.

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.

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, 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., 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. (In Review), 2018.

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, 2016, 3672-3675,

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.

Cheng, L., and Zhu, J.: 2017 was the warmest year on record for the global ocean, Adv Atmos Sci, 35, 261-263, 2018.

Dankers, R., Burke, E. J., and Price, J.: Simulation of permafrost and seasonal thaw depth in the JULES land surface scheme, 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.

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.

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., 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.

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.

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.

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.

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.

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

Lawrence, D. M., and Slater, A. G.: A projection of severe near-surface permafrost degradation during the 21st century, Geophys Res Lett, 32, 2005.

Lawrence, D. M., and Slater, A. G.: Incorporating organic soil into a global climate model, Clim Dynam, 30, 145-160, 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-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.

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.

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 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.

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.

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

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

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

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., and De Lannoy, G. J. M.: Assessment
10 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.

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.

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

20 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.

25 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.

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
30 variations and trends from global reanalyses and monthly surface climatological datasets, Quarterly Journal of the Royal Meteorological Society, 143, 101-119, 2017.

Slater, A. G., and Lawrence, D. M.: Diagnosing Present and Future Permafrost from Climate Models, Journal of Climate, 26, 5608-5623, 2013.

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.

5 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.

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,
10   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.

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-
15   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 10m × 10m grid; 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).
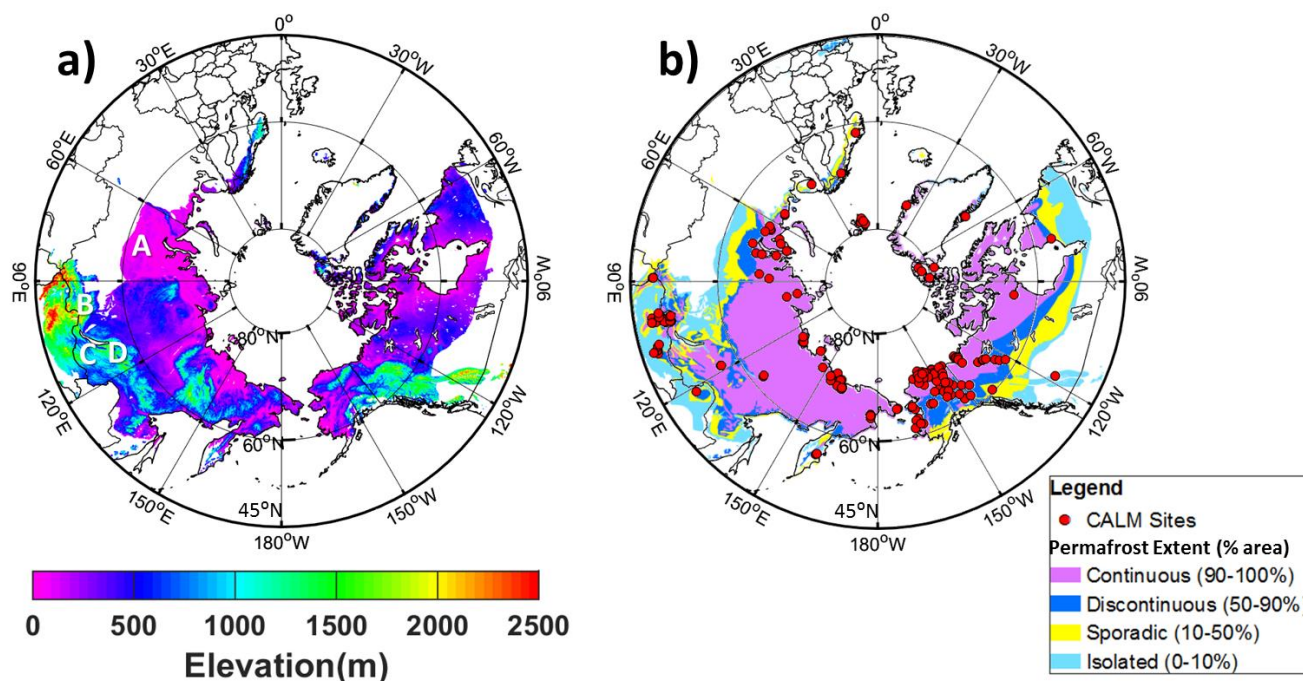
Figure 1: a) Elevation above mean sea level in the simulation domain, which is defined by the area for which NCSCDv2 data are available. Regions A, B, C, and D are discussed in the text. b) Permafrost and ground ice conditions adapted from (Brown et al., 2002). Red dots represent CALM sites.
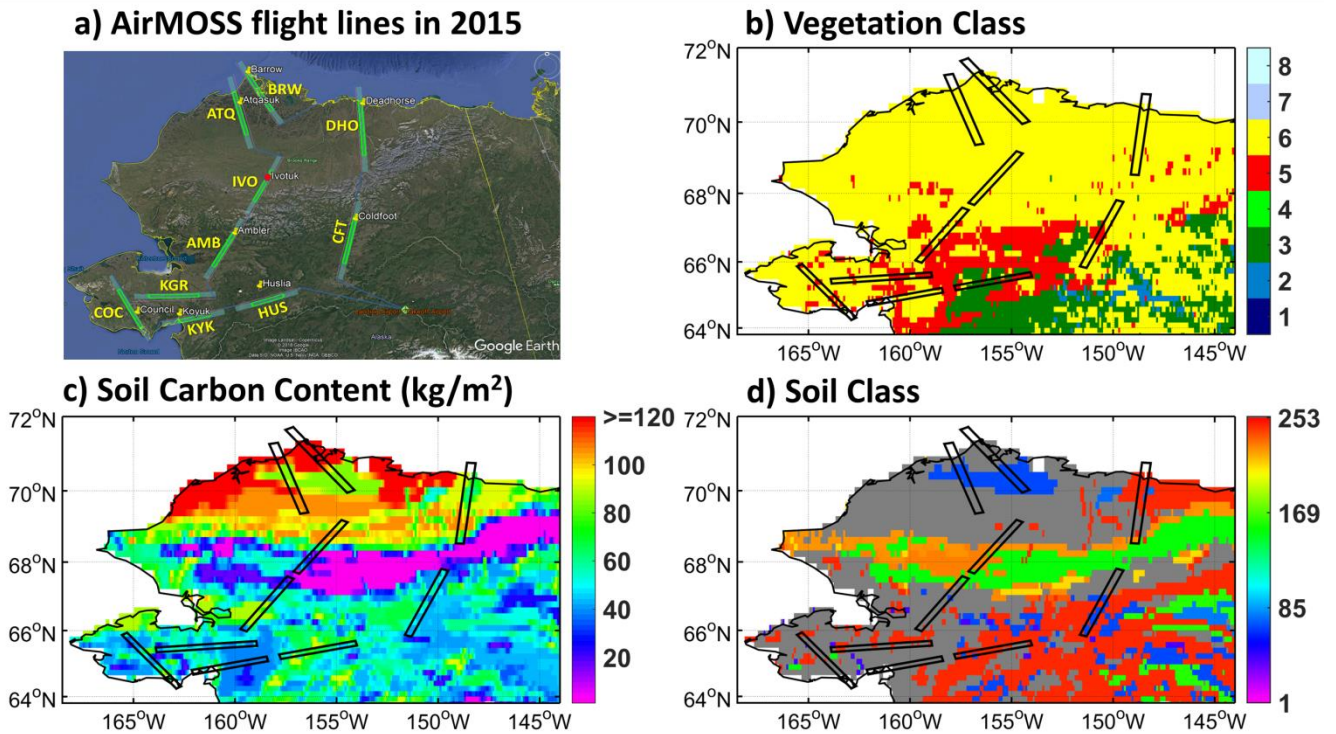
Figure 2: a) Ten transects of AirMOSS flights conducted in Alaska on 29 August 2015 and 1 October 2015, including HUS (Huslia), KYK (Koyuk), COC (Council), KGR (Kougarok), AMB (Ambler), IVO (Ivotuk), ATQ (Atqasuk), BRW (Barrow), DHO (Deadhorse), and CFT (Coldfoot). Each flight swath width is approximately 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) AirMOSS radar retrievals of ALT for 2015 at raw resolution (20 m x 60 m) derived from P-band radar observations on 29 August 2015 and 01 October 2015 for IVO, ATQ, BRW, and DHO. b) ALT radar retrievals aggregated to 81 km² model grid cells. c) CLSM-simulated ALT. Green squares in b) and c) represent CALM sites covered by the flight swath whereas black circles represent UAF sites.
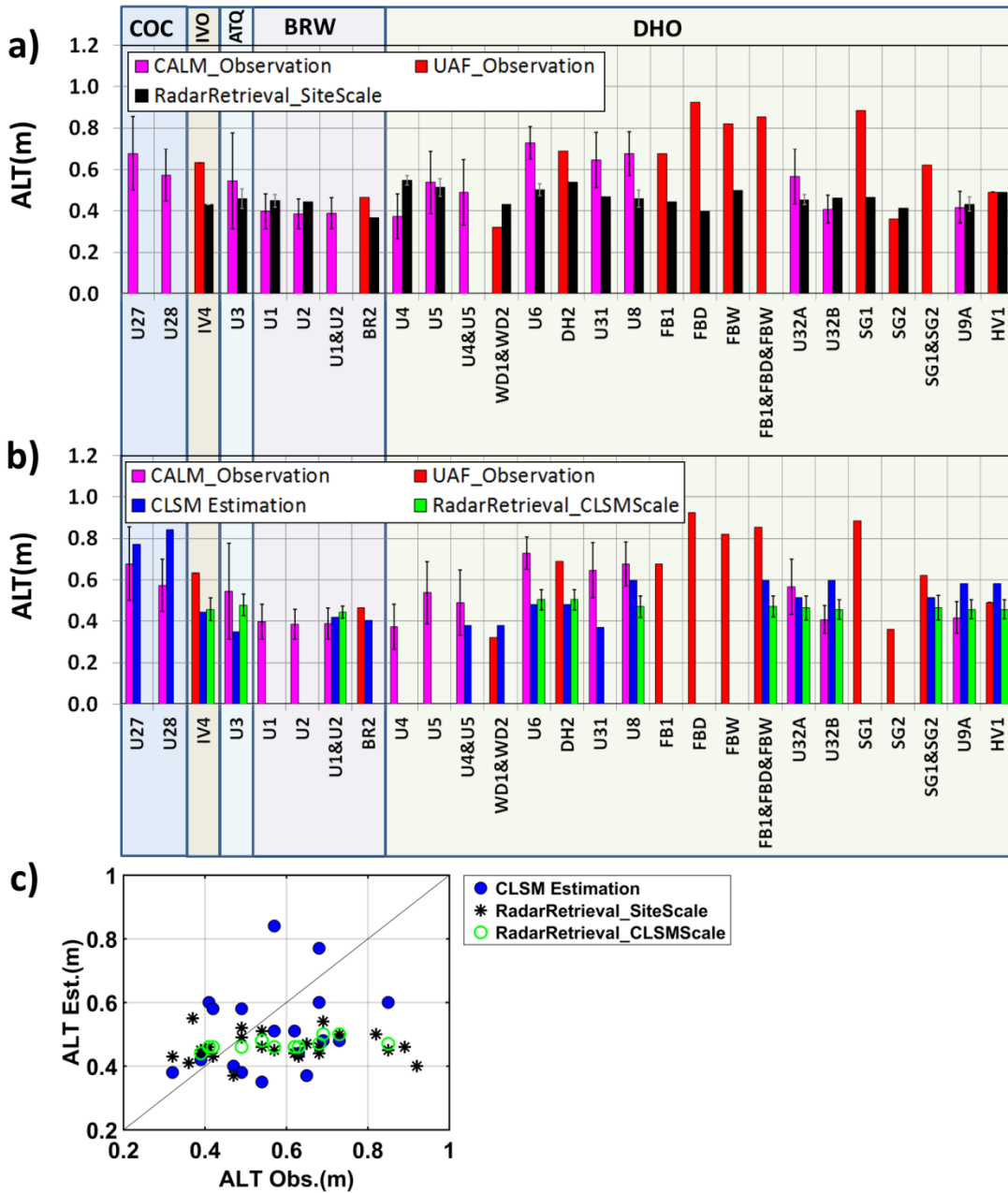
Figure 4: a) ALT observations for 2015 from CALM (magenta) and UAF (red) sites covered by AirMOSS swaths, and from radar retrievals at in-situ site scale (black). **b)** Same as a) but with ALT from radar retrievals aggregated to 81 km² grid cells (green), and from CLSM simulations at 81 km² (blue). The short name of the corresponding covering swath is shown on the top (see also Fig. 2a). Error bars represent the standard deviation for multiple observations at in-situ sites, and for radar retrievals within the CALM observing grid (at site scale), or within each 81 km² grid cell (at model scale). No standard deviations are provided for site-scale radar retrievals at CALM U2, U31, and U32B sites because sampling coordinates are not available nor at 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, FB1&FBD&FBW, and SG1&SG2) and compared with results at 81 km² grid cell. **c)** CLSM estimates and AirMOSS radar retrievals of ALT (at both site scale and model scale) for 2015 versus in-situ measurements.
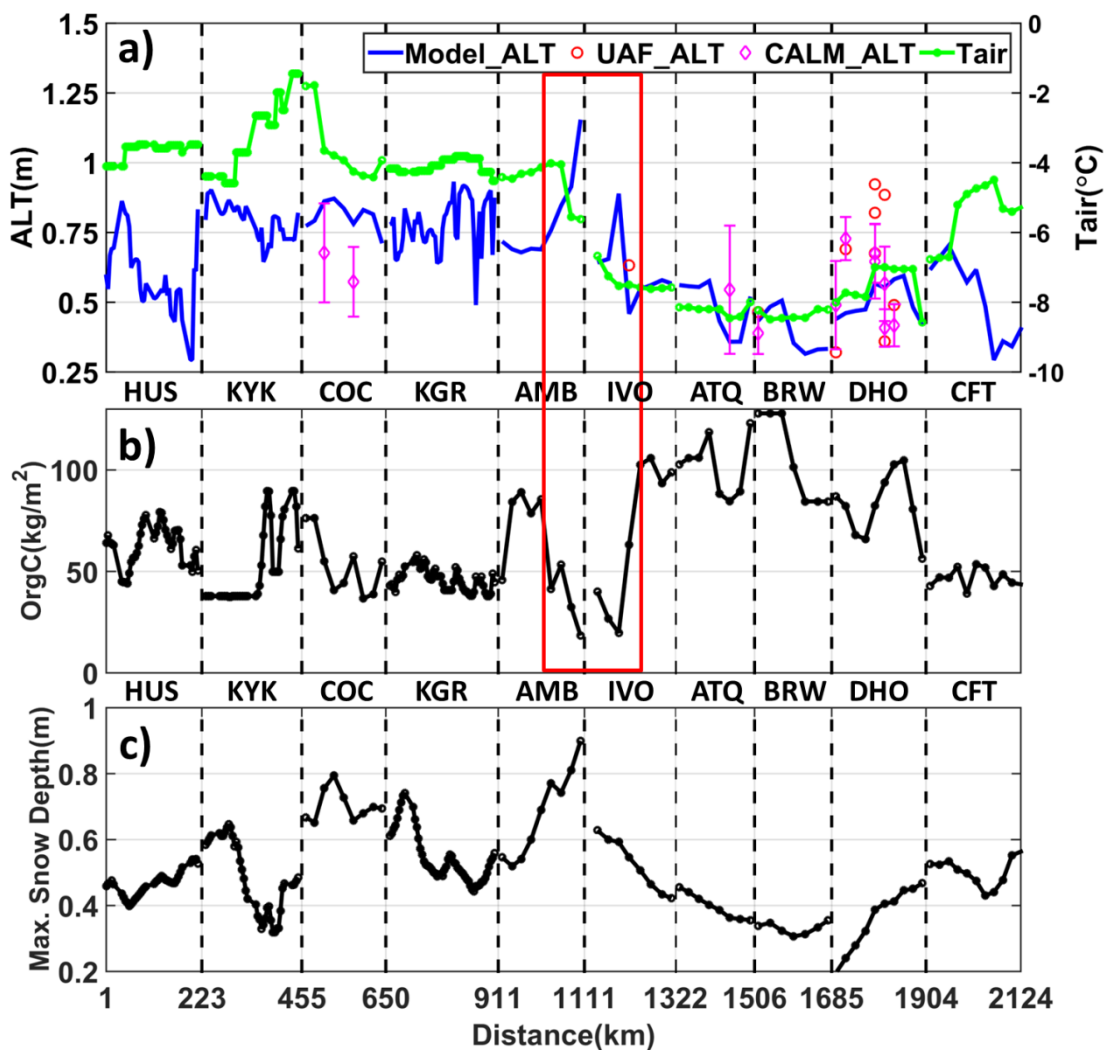
Figure 5: a) CLSM-simulated ALT (thawed-to-frozen depth) on 29 August 2015 along the AirMOSS flight transects. In-situ ALT observations from UAF and CALM are shown as red circles and magenta diamonds, respectively. Averaged air temperature at 2 m (Tair) from the preceding annual period (i.e., 01 September 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 abscissa in c) provides cumulative distances in units of km along the transects.
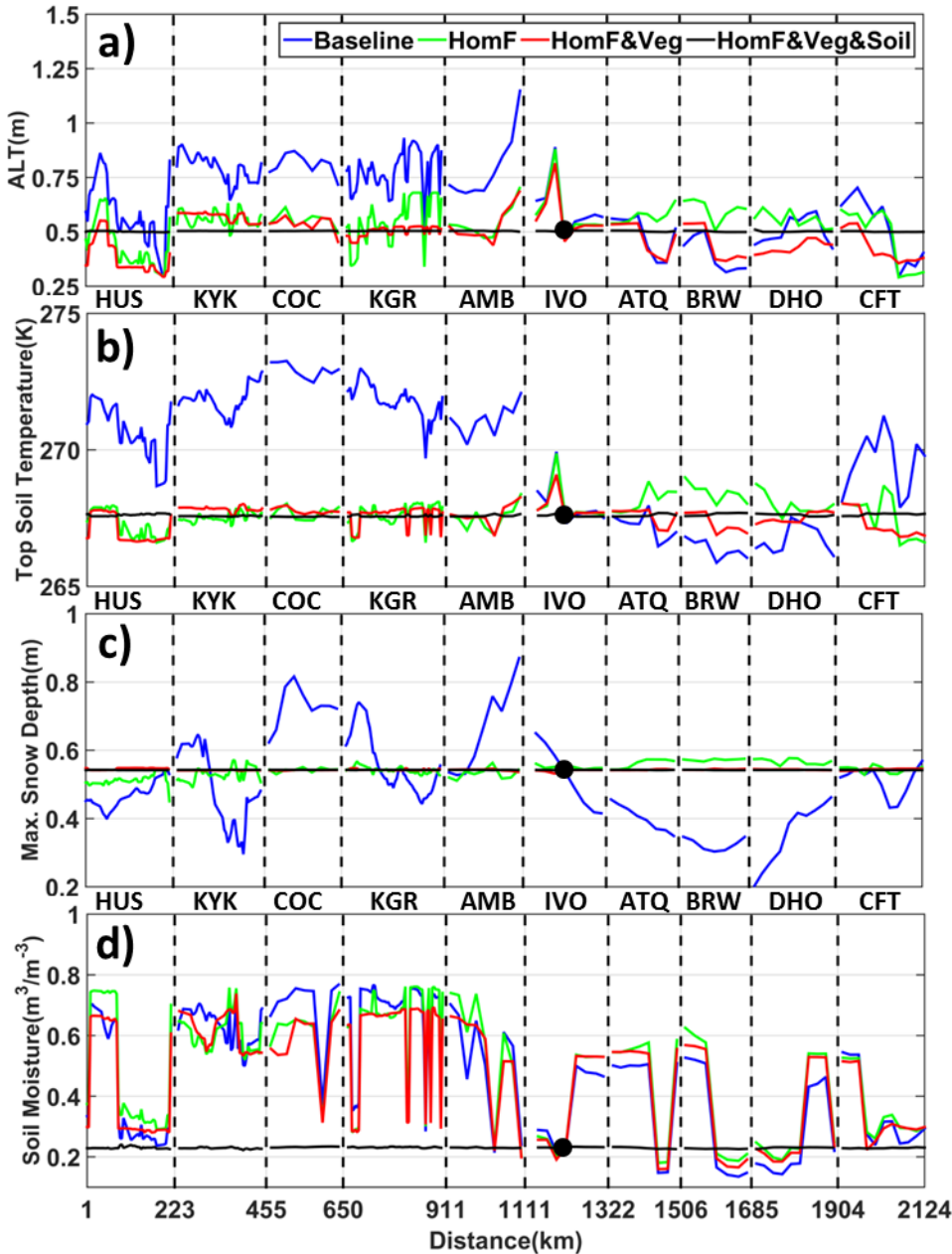
Figure 6: a) CLSM-simulated ALT (thawed-to-frozen depth) on the flight date (i.e., 29 August 2015) from the top four experiments listed in Table 2; b) simulated top layer soil temperature on the flight date, c) maximum snow depth the during the preceding annual period (i.e., from 01 September 2014 to 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.
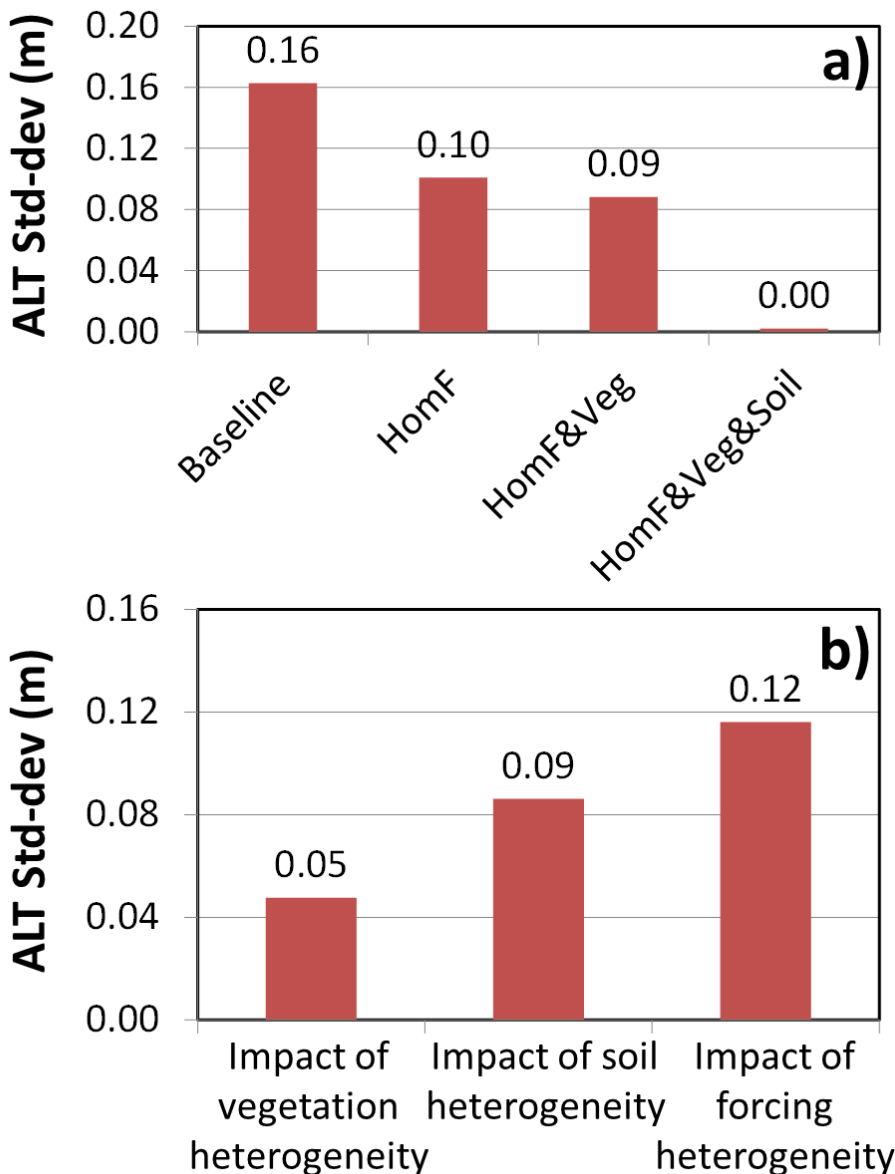
Figure 7: a) Standard deviation of ALT along the AirMOSS transects from the top four experiments listed in Table 2. Subplot b) shows the individual impact (or contribution) from the heterogeneous vegetation, soil type and meteorological forcing, respectively. For instance, the impact of vegetation (or soil, or forcing) heterogeneity is the ALT standard deviation along the transects from HomF&Soil (or HomF&Veg, or HomVeg&Soil).
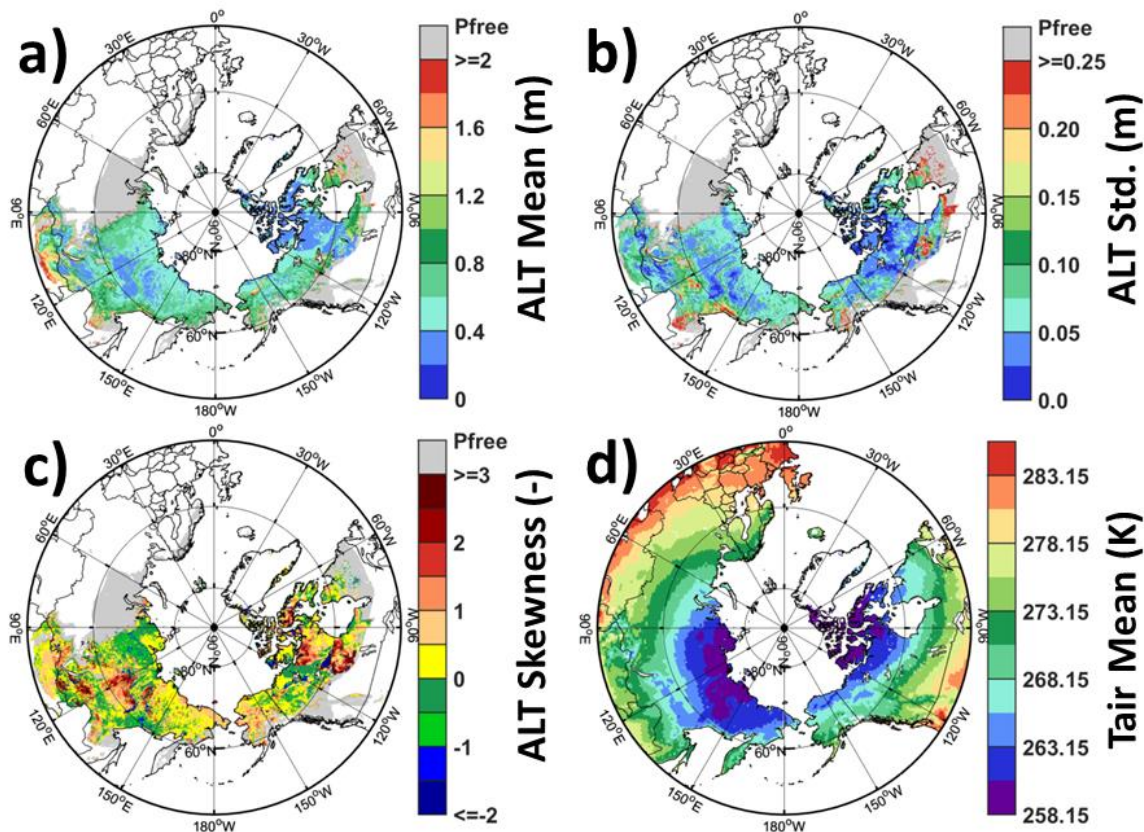
Figure 8: a) Mean, b) standard deviation, and c) skewness of CLSM-simulated ALT over the 38 years (1980 - 2017). Grey indicates permafrost-free (Pfree) areas in the simulation. d) 38-year averaged MERRA-2 annual atmospheric temperature at 2 m above displacement height (Tair).
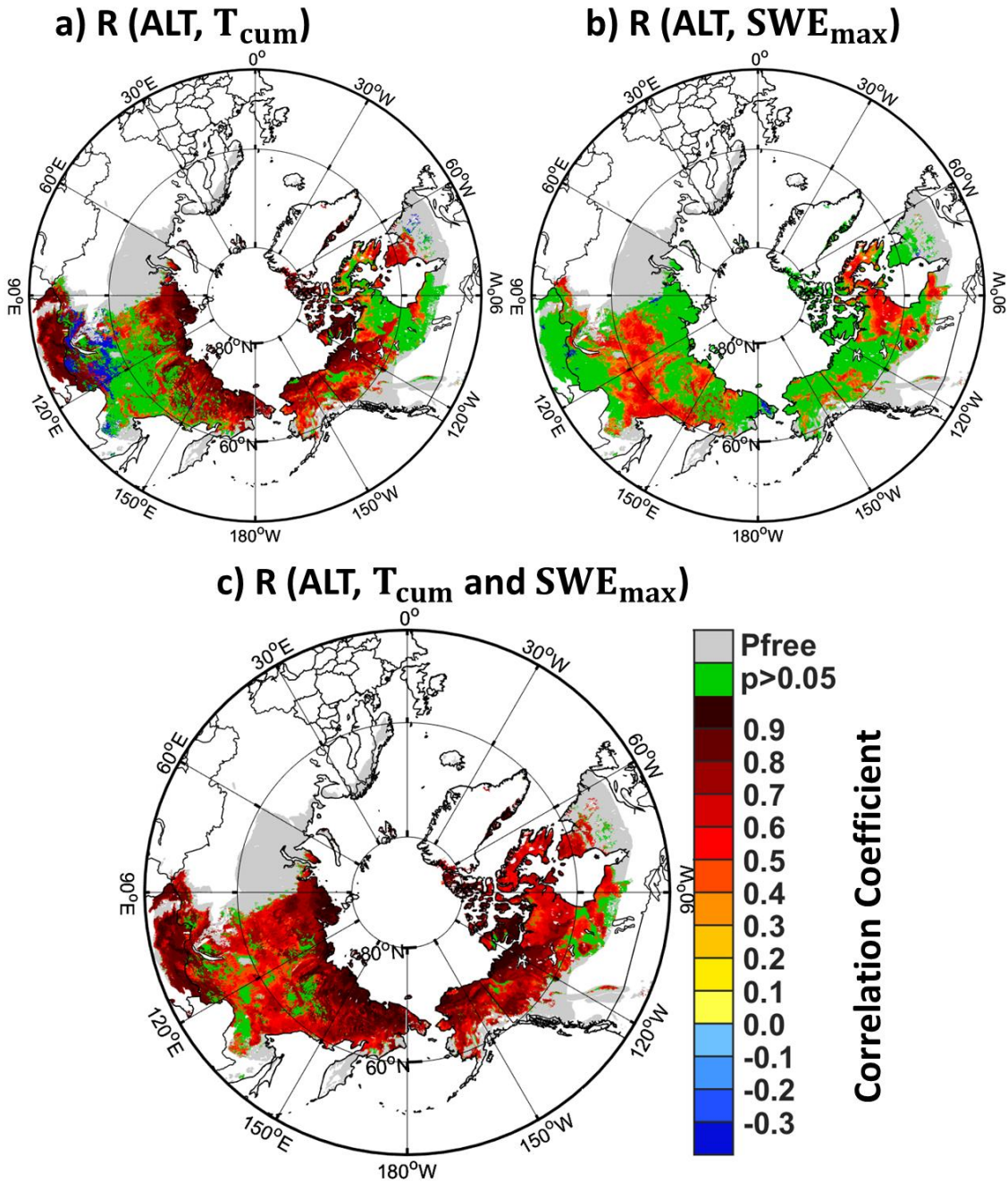
Figure 9: Correlation coefficient between a) ALT and effective accumulated air temperature (T_{cum}) and b) ALT and maximum SWE (SWE_{max}) from the preceding September to the present August over the period 1980-2017. c) Multi-variable coefficient of correlation for a fitted multiple linear regression model between ALT and the accumulated temperature and the maximum SWE. 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.
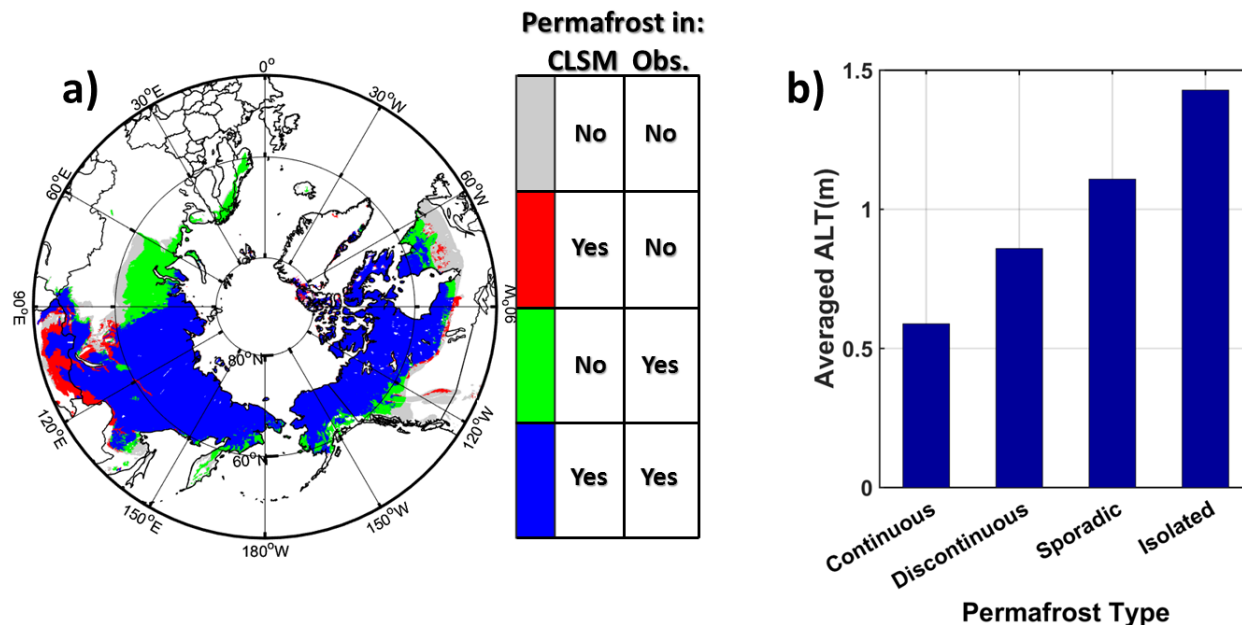
Figure 10: a) Four comparison categories include: 1) blue - CLSM collocates permafrost with the observation-based permafrost map of Brown et al. (2002) as either continuous, discontinuous, or sporadic permafrost; 2) green - CLSM has no permafrost, but the observation-based permafrost map does as either 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.
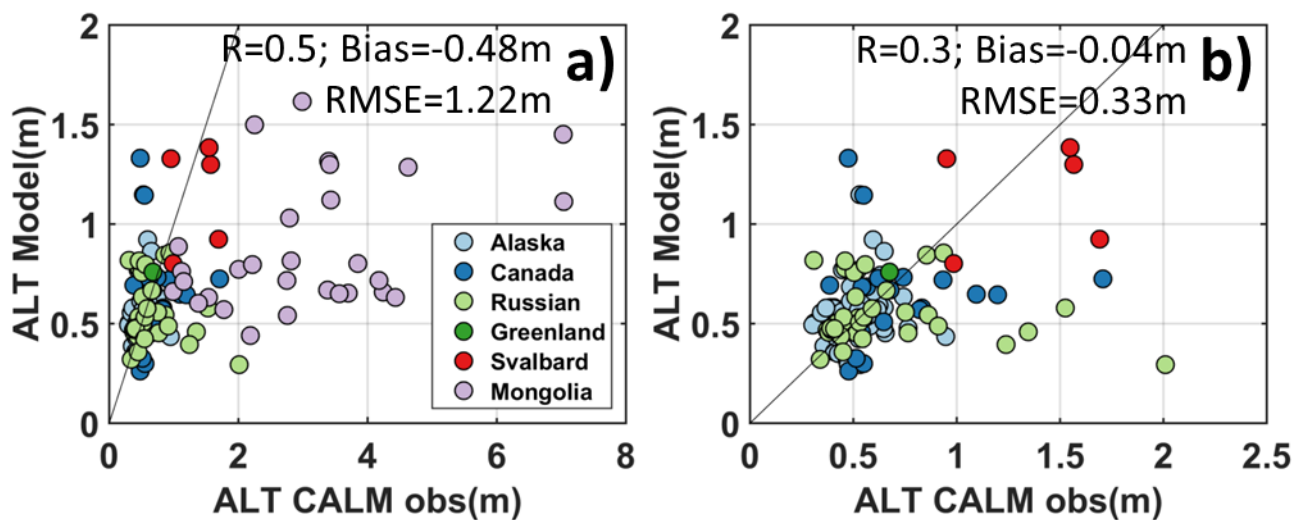
Figure 11: a) 38-yr average ALT for the period 1980-2017 from CLSM simulation vs. CALM observations. b) Same as a) but without the Mongolia sites. The correlation coefficient (R), bias, and root mean squared error (RMSE) are provided on each subplot.
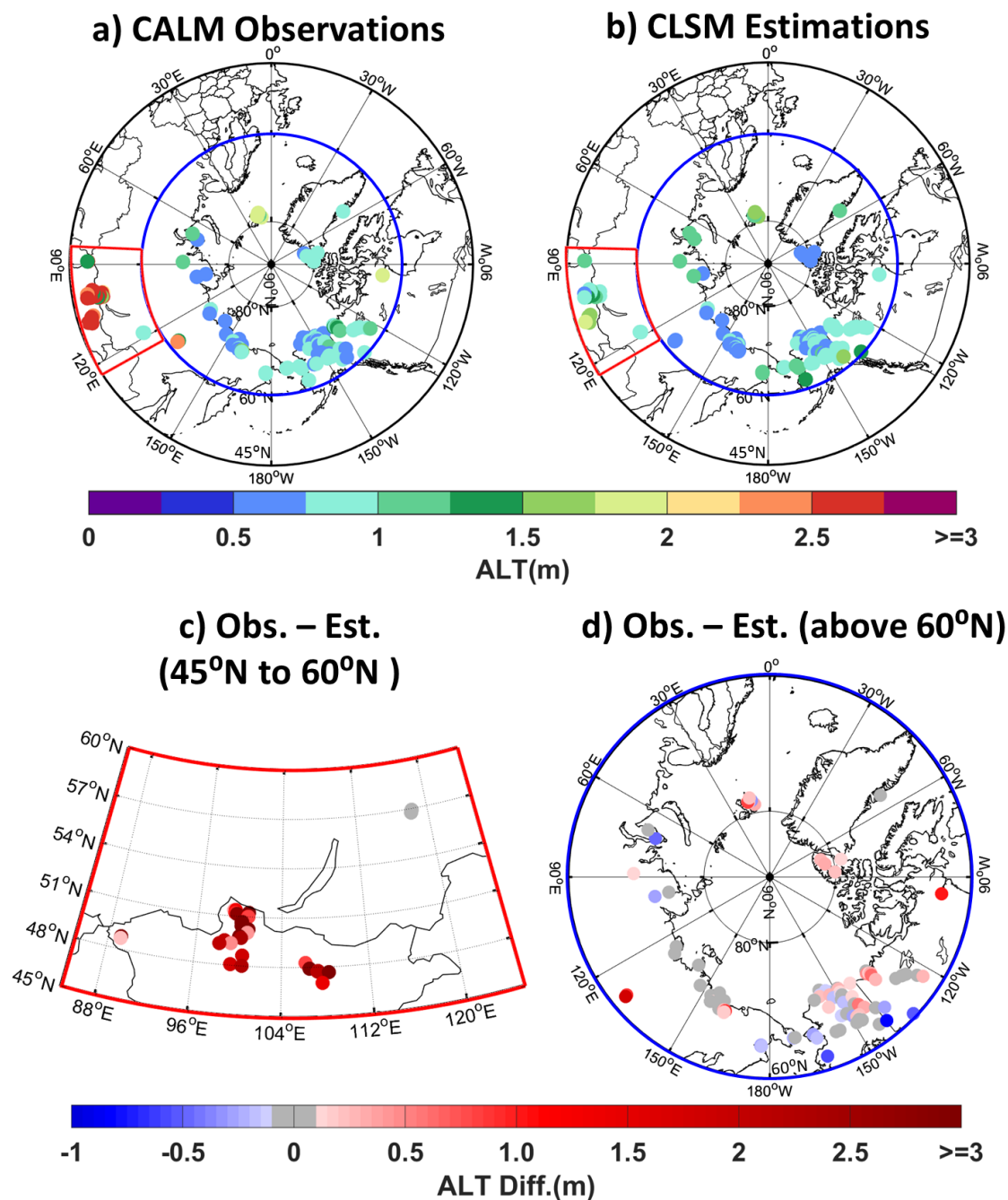
Figure 12: 38-yr average ALT for the period 1980-2017 at CALM site locations for a) CALM observations and b) CLSM results. c) ALT difference between observations and model results for locations within 45°N- 60°N latitude and 85°E-125°E longitude. d) Same as c) but for locations poleward of 60°N latitude. In c) and d) grey indicates absolute ALT differences less than 0.10 m.
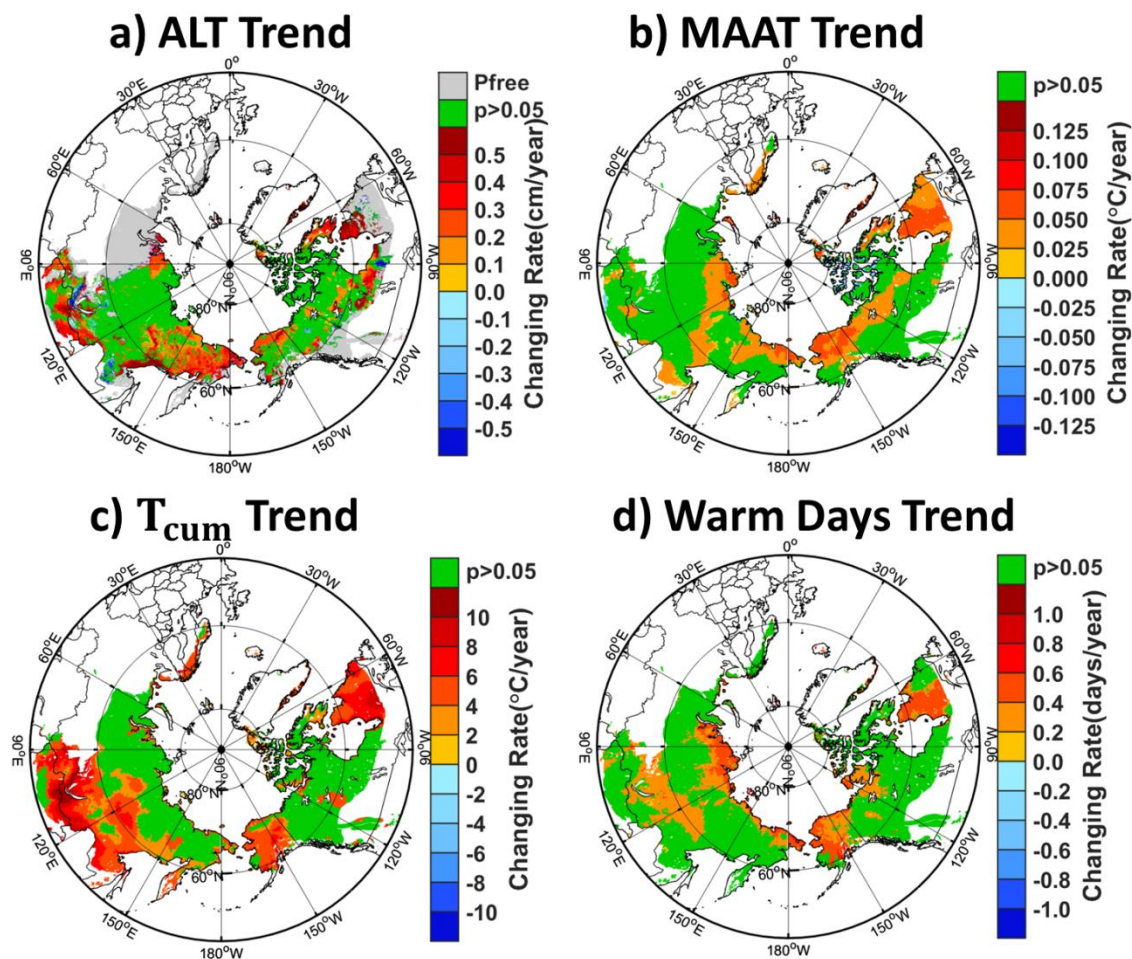
Figure 13: 1980-2017 trend in a) ALT, b) mean annual air temperature (MAAT), c) effective accumulated air temperature (T_{cum}), and d) warm days from CLSM simulations. Areas that have p-values larger than 0.05 (i.e., no statistically significant trend) are shown in green. In a), grey indicates permafrost-free (Pfree) areas in the simulation.
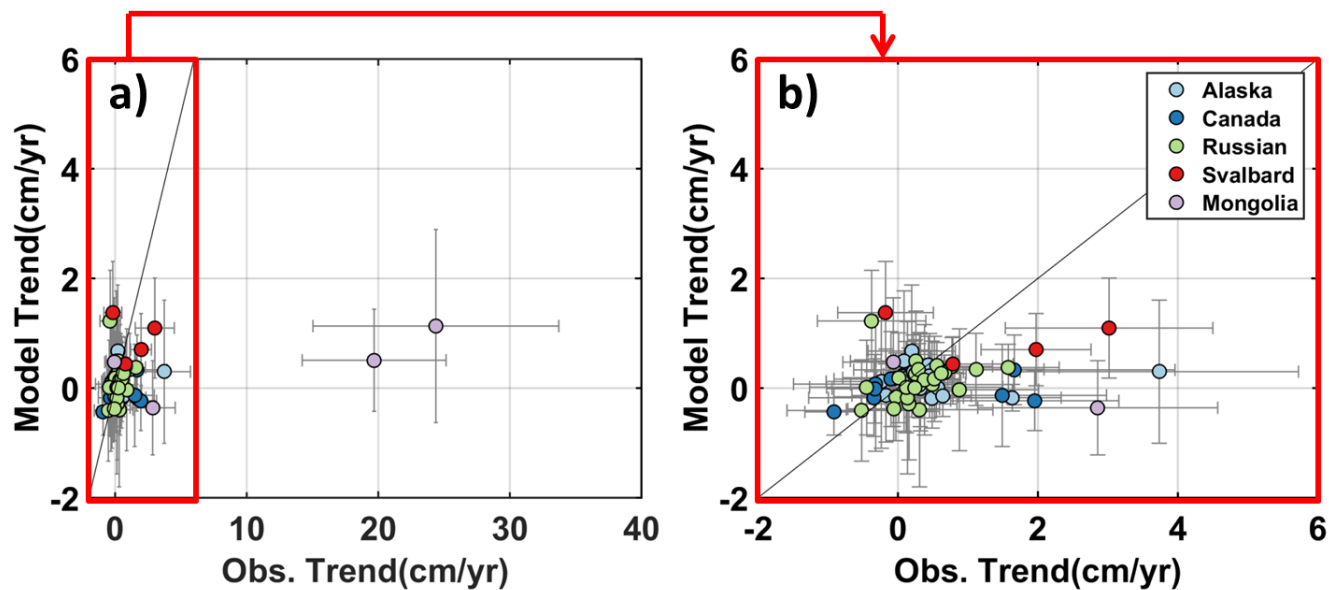
Figure 14: a) ALT trend from CLSM estimates vs. CALM observations, based on years common to both datasets within the period of 1990 to 2017. The horizontal and vertical error bars represent 95% confidence intervals of observed ALT trend (regression slope) and CLSM-simulated ALT trend, respectively. b) Same as a) but zoomed into observed ALT trends between -2 and 6 cm/yr.