Validation of Pan-Arctic Soil Temperatures in Modern Reanalysis and Data Assimilation Systems

Tyler C. Herrington [1], Christopher G. Fletcher [1], and Heather Kropp [2]

[1]Department of Geography and Environmental Management, University of Waterloo, 200 University Ave., Waterloo, Ontario, Canada, N2L 3G1
[2]Environmental Studies Program, Hamilton College, 198 College Hill Road, Clinton, 13323, New York, U.S.A.

Correspondence: Christopher G. Fletcher (chris.fletcher@uwaterloo.ca)

Abstract. Reanalysis products provide spatially homogeneous coverage for a variety of climate variables in regions where observational data are limited. However, very little validation of reanalysis soil temperatures in the Arctic has been performed to date, because widespread in situ reference observations have historically been unavailable there. Here we validate pan-Arctic soil temperatures from eight reanalysis and LDAS products, using a newly-assembled database of in situ data from diverse

5 measurement networks across Eurasia and North America. We find that most products have soil temperatures that are biased cold by 2-7 K across the Arctic, and that biases and RMSE are generally largest in the cold season. Monthly mean values from most products correlate well with in situ data (R > 0.9) in the warm season, but show lower correlations (r = 0.6 - 0.8), in many cases, over the cold season. Similarly, the magnitude of monthly variability in soil temperatures is well captured in summer, but overestimated by 20% to 50% for several products in winter. The suggestion is that soil temperatures in reanalysis products

10 are subject to much higher uncertainty when the soil is frozen and/or when the ground is snow-covered. We also validate the ensemble mean of all products, and find that when all seasons, and metrics are considered, the ensemble mean generally outperforms any individual product in terms of its correlation and variability, while maintaining relatively low biases. As such, we recommend the ensemble mean soil temperature product for a wide range of applications - such as the validation of soil temperatures in climate models, and to inform models that require soil temperature inputs, such as hydrological models, or for

15 permafrost simulations.

1 Introduction

Soil temperature is an important control of many physical, hydrological, and land surface processes, as soils act as a reservoir for energy and moisture underground. It provides an important initial condition for numerical weather prediction, as energy and water fluxes from the land are important for convective processes (Dirmeyer et al., 2006; Kim and Wang, 2007; Siqueira

20 et al., 2009). As soils react relatively slowly to variations in weather, soil temperature is also an important predictor of seasonal and mid-term weather forecasts (Xue et al., 2012). Soils over large portions of the Arctic are perennially frozen (permafrost soil). Roughly 800 gigatonnes of carbon (GtC) is estimated to be stored in permafrost soils across the Northern Hemisphere (Hugelius et al., 2014); about twice the amount of carbon currently residing in the atmosphere (Tarnocai et al., 2009). Continued

warming, and thawing of permafrost soils, and related decomposition of carbon could act as a positive feedback on warming,

25 by releasing more methane (CO_2) and carbon dioxide (CH_4) to the atmosphere (Koven et al., 2011; Schuur et al., 2015).

In situ based soil temperature monitoring networks using thermistor probes, particularly at high latitudes, are limited in terms of their spatial and temporal coverage (Yi et al., 2019), making it difficult to assess hemispheric scale changes in permafrost. Reanalysis products have been used in a variety of weather and climate applications to provide information on a regular spatial grid; particularly in regions where limited or no observational data is available (Koster et al., 2004; Zhang et al., 2008). Previous

30 studies validating reanalysis soil temperature have primarily focused on the middle latitudes, such as across China (Yang and Zhang, 2018; Xu et al., 2019; Zhan et al., 2020), the Qinqhai-Tibetan plateau (Hu et al., 2019; Qin et al., 2020; Wu et al., 2018), Europe (Albergel et al., 2015; Johannsen et al., 2019), and the continental United States (Albergel et al., 2015; Xia et al., 2013), with a couple of recent studies validating soil temperatures globally (Li et al., 2020c; Ma et al., 2021). Relative to in situ ground temperature probe networks, most reanalysis products are biased cold by about 2°C - 5°C, on average (Hu

35 et al., 2019; Qin et al., 2020; Yang and Zhang, 2018). Ma et al. (2021) found that most reanalysis products show larger cold biases over polar regions than they do over tropical and temperature regions, while a recent study by Cao et al. (2020) found that ERA5-Land soil temperatures were biased warm over the Arctic, particularly in winter.

Several explanations have been suggested for the biases in reanalysis soil temperatures, including model parameterizations (Albergel et al., 2015; Cao et al., 2020; Chen et al., 2015; Wu et al., 2018; Xiao et al., 2013), air temperature biases (Cao et al.,

40 2020; Hu et al., 2017), errors in topography and elevation, arising from the coarse resolution of reanalysis products (Yang and Zhang, 2018; Zhao et al., 2008; Ma et al., 2021), and errors in simulated snow cover and snow thermal insulation (Cao et al., 2020; Royer et al., 2021).

While soil temperature biases in individual reanalysis products may limit their utility, a consensus is emerging that multi-reanalysis ensemble, based on the same principle as ensemble weather prediction (World Meteorological Organization, 2012),

45 are an effective way to increase the signal-to-noise ratio for many important geophysical variables. Ensemble mean datasets based on combinations of in situ, model, satellite and reanalysis data have been used to reduce biases in estimates of snow water equivalent (Mudryk et al., 2015), soil moisture (Dorigo et al., 2017; Gruber and Scanlon, 2019), as well as precipitation (Beck et al., 2017, 2019), and Hu et al. (2019) suggest that a similar method could be used to reduce biases in reanalysis soil temperatures.

50 Reanalysis soil temperatures have been relatively well characterized over the middle latitudes. Studies validating Arctic soil temperatures in reanalysis products, however, have either focused on a singular product (Cao et al., 2020), or have only considered a limited spatial extent (Li et al., 2020c; Ma et al., 2021).

Here we perform a validation of pan-Arctic (and Boreal) soil temperatures from eight reanalysis and land data assimilation system (LDAS) products. The main objectives are to 1) validate the 8 reanalysis and LDAS products in terms of their bias,

55 RMSE, correlation and standard deviation of Arctic soil temperatures, and 2) investigate whether an ensemble mean soil temperature product outperforms the individual reanalysis products.

2 Data

2.1 Reanalysis and LDAS Data

Table 1 outlines the six reanalysis and two LDAS soil temperature products used in this study. For simplicity, the term
60 "reanalysis" will hereafter be used to describe both reanalysis and LDAS products. A summary of each product follows below. Products were remapped onto the GLDAS-CLSM grid for comparison, using three different methods: nearest neighbour, bilinear interpolation, and first-order conservative remapping. The choice of remapping method did not affect the overall conclusions of the study, and the analysis is based on data remapped using the conservative remapping method, as it facilitated the use of the largest number of validation sites and grid cells.

65 The reanalysis products investigated span a wide range of horizontal resolutions, ranging between 0.1°, in the case of ERA5-Land, to 1.0° for both GLDAS products. Most products (CFSR, ERA-Interim, ERA5, ERA5-Land, and GLDAS-Noah) simulate soil temperature across 4 vertical layers, while MERRA and GLDAS-CLSM include 6 vertical layers, and JRA-55 calculates soil temperature across a single layer. The topmost soil layer has the highest resolution (7 cm to 10 cm in most cases), while the bottom soil layer often averages soil properties over a metre or more (Table 1).

70 The Noah Land Surface Model (Noah-LSM) (Chen et al., 1996; Betts et al., 1997; Koren et al., 1999; Ek, 2003) is used by CFSR and GLDAS-Noah. CFSR uses the Noah-LSM in a fully coupled mode to obtain a first-guess land-atmosphere simulation, before operating in a semi-coupled mode with GLDAS to obtain information about the state of the land surface (Saha et al., 2010). GLDAS, however, is run in an offline mode, utilizing meteorological forcing from Princeton University between 1948 to 2000 (Sheffield et al., 2006), and a combination of model and observational data from 2000 - onwards (Rui
75 et al., 2018).

ERA-Interim, ERA5 and ERA5-Land use versions of the Tiled ECMWF Scheme for Surface Exchanges over Land (TES-SEL) land model (Viterbo, 1995; Viterbo and Betts, 1999). In the case of ERA-Interim, TESSEL is informed by empirical corrections from 2m (surface) air temperature and humidity (Dee et al., 2011). Meanwhile, ERA5 and ERA5-Land use an updated version of TESSEL, known as the Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL)
80 (Balsamo et al., 2015). In ERA5, a weak coupling exists between the land surface and atmosphere, while an advanced LDAS that incorporates information regarding the near-surface air temperature, relative humidity, as well as snow cover (de Rosnay et al., 2014), along with satellite estimates of soil moisture and soil temperature from the top 1m of soil (de Rosnay et al., 2013). ERA5-Land, unlike ERA5, does not directly assimilate observational data. Instead, the ERA5 meteorology (such as air temperature, humidity and atmospheric pressure) is used as forcing information for HTESSEL; allowing it to be run at
85 higher resolutions (Muñoz-Sabater et al., 2021). It includes an improved parameterization of soil thermal conductivity allowing for it to account for ice content in frozen soil; improvements to soil water balance conservation; and the ability to capture rain-on-snow events (Muñoz-Sabater et al., 2021).

Both GLDAS-CLSM and MERRA2 utilize the the Catchment Land Surface Model (Ducharne et al., 2000; Koster et al., 2000). Though MERRA2 does not include a land surface analysis (Gelaro et al., 2017), CLSM is informed using an updated
90 version of the Climate Prediction Center unified gauge-based analysis of global daily precipitation (CPCU) precipitation cor-

rection algorithm that originated in MERRA-Land. No corrections are available, however, for high latitude regions north of 62.5° N (Reichle et al., 2017). In the case of GLDAS-CLSM, CLSM is run in an offline mode, in a similar configuration to GLDAS-Noah. Finally, JRA-55 uses the Simple Biosphere Model (SiB) (Onogi et al., 2007; Sato et al., 1988; Sellers et al., 1986) in an offline mode, forced by atmospheric data and data from land surface analyses that incorporate microwave satellite

95   retrievals of snow cover (Kobayashi et al., 2015).

2.2 Observational Data

Owing to the lack of dense soil temperature monitoring networks in the Arctic, most of the observed soil temperature record is characterized by sparse measurements with inconsistencies in the temporal periods that they cover (Yi et al., 2019). Rather than limit our validation to a small geographic region in the permafrost zone, as several prior studies have done (Hu et al.,

100   2019; Qin et al., 2020; Wu et al., 2018; Ma et al., 2021; Li et al., 2020c), we choose to make use of data from a variety of sparse networks. Such an approach has been used to validate soil temperature and permafrost performance in ERA5-Land (Cao et al., 2020), and allows for the examination of larger geographic regions, as well as for the inclusion of a more diverse set of vegetation types across the continent Ma et al. (2021).

To the authors' knowledge, this is one of the first studies to compile a comprehensive set of in situ soil temperature

105   measurements across the Eurasian and North American Arctic, from multiple diverse sparse networks. We incorporate data from the Roshydromet Network in Russia (Sherstiukov, 2012), Nordicana series D (Nordicana) (Allard et al., 2020; CEN, 2020a, b, c, d, e, f, g), Global Terrestrial Network for Permafrost (GTN-P) (GTN-P, 2018), and Kropp et al. (2020) - in an attempt to provide a representative estimate of soil temperature across the circumpolar Arctic. Our validation data also includes sites from outside regions typically underlain by permafrost, in order to facilitate a comparison of the performance of reanal-

110   ysis soil temperatures at high latitudes with their performance in regions outside the permafrost zone. This provides a unique baseline upon which to perform a hemispheric wide assessment of soil temperature in reanalysis and LDAS systems.

The authors do acknowledge that there is a large discrepancy in the sampling of grid cells across North America and Eurasia, however we have done so in order to make use of all available data. Unfortunately, data availability is limited across much of the Canadian North, as it has been historically under-sampled relative to Eurasia (Metcalfe et al., 2018).

115   In order to compare with data from reanalysis and LDAS products, temperatures were averaged across two depth bins: a near surface layer (0 cm to 30 cm), and soil temperatures at depth (30 cm to 300 cm). For each site, temperatures from all depths residing within a layer were averaged, producing an estimated layer averaged temperature for every time-step. Data from 51 North American stations, and 247 Eurasian station were used to estimate estimate monthly average temperatures in the near surface layer, while 38 (256) stations from North America (Eurasia) contributed to estimates of monthly average temperatures

120   at depth.

Many of the in situ (station) sites reported measurements at hourly or daily frequency, however we chose to perform the analysis at monthly time scales, as soil temperatures do not vary much over shorter timescales. As such, once a layer average had been calculated for a site, a monthly average temperature was then estimated. Daily average temperatures (T_{day}) were calculated for each station with a timestep smaller than 1 day, using all available data, when there was at least one temperature

Product	Data Period	Resolution	Land Model	Soil Layers	References
CFSR1	1979 - 2010	0.31° x 0.31°	Noah LSM	4	Saha et al. (2010)
CFSR2	2011 - Present	0.2° x 0.2°	Noah LSM	4	Saha et al. (2014)
ERA-Interim	1979 - Aug 2018	0.7° x 0.7°	TESSEL	4	Dee et al. (2011)
ERA5	1979 - Present	0.27° x 0.27°	HTESSEL	4	Hersbach et al. (2020)
ERA5-Land	1981 - Present	0.1° x 0.1°	HTESSEL	4	Muñoz-Sabater et al. (2021)
GLDAS-CLSM	1948 - Present	1.0° x 1.0°	Catchment LSM	6	Rodell et al. (2004)
GLDAS-Noah	1948 - Present	1.0° x 1.0°	Noah LSM	4	Rodell et al. (2004)
JRA-55	1956 - Present	0.56° x 0.56°	Simple Biosphere Model	3*	Harada et al. (2016) Kobayashi et al. (2015)
MERRA2	1980 - Present	0.5° x 0.67°	Catchment LSM	6	Gelaro et al. (2017)

Table 1. Summary of the 8 reanalysis and LDAS products, their equatorial resolution, land model, and the number of soil layers included. *The JRA-55 Simple Biosphere Model contains up to three soil layers (depending on vegetation type), but for the purposes of soil temperature, it is calculated as one layer.

125 record available on a given date. T_{day} were then used to calculate a station's monthly average temperature (T_{soil}). Outlier detection was performed on in situ data before monthly averaging, by removing datapoints that fell outside 3.5σ from the mean soil temperature of the dataset.

Since the station data often included days with missing observations, the sensitivity of the monthly averages to missing data was tested, by computing monthly averages in five ways: using all months with at least one valid day in a month, using all

130 months with at least 25, 50, and 75 percent valid data, and finally using all months with no missing data in a month. It was found that T_{soil} was not substantially impacted by the inclusion or exclusion of months containing missing data.

In order to be considered as a validation location, the grid cell was required to include soil temperature data for all eight reanalysis/LDAS products, and be collocated with at least one in situ station. Duplicate stations across datasets were excluded. In situ locations were only included if there was at least 2 years worth of in situ data, in order to properly assess the station's

135 seasonal cycle. For grid cells containing multiple in situ stations, the value used in the comparison is a simple spatial average of the in situ stations in that grid cell on each calendar day.

Over Eurasia, the vast majority of grid cells contain a single in situ measurement location. In North America, however, several Alaskan grid cells contain three or more in situ stations. The near surface layer layer includes 271 validation grid cells, while at depth, there are 272 grid cells with minimal variation in their location relative to the near surface (Figure 1, panel

140 A). Panel B of Figure 1 shows the spatial standard deviation of monthly surface soil temperatures for grid cells with more than two stations included. The number of stations included in each grid cell ranges between two to 12. The median standard deviation is generally $\leq 2°$ C, though the temporal variation (spread) is quite large for several stations. This suggests that while differences in T_{soil} at in situ stations within a grid cell are generally within a few degrees of one another, there are times when the differences can be quite large.

145 ## 3   Methods

3.1   Validation Metrics

Reanalysis/LDAS and observational (station) soil temperature data were collocated with one another spatially and temporally. Grid-cell level soil temperatures from each product were compared against in situ soil temperatures using the following statistical metrics: bias (Eq. 1), root-mean-squared-error (RMSE) (Eq. 2), normalized standard deviation (SDV_{norm}) (Eq. 3

150 and Eq. 4), and the Pearson correlation (R) (Eq. 5). Statistical metrics were calculated as follows:

$$Bias = \frac{1}{N}\sum_{n=1}^{N}(T_p - T_i) \tag{1}$$
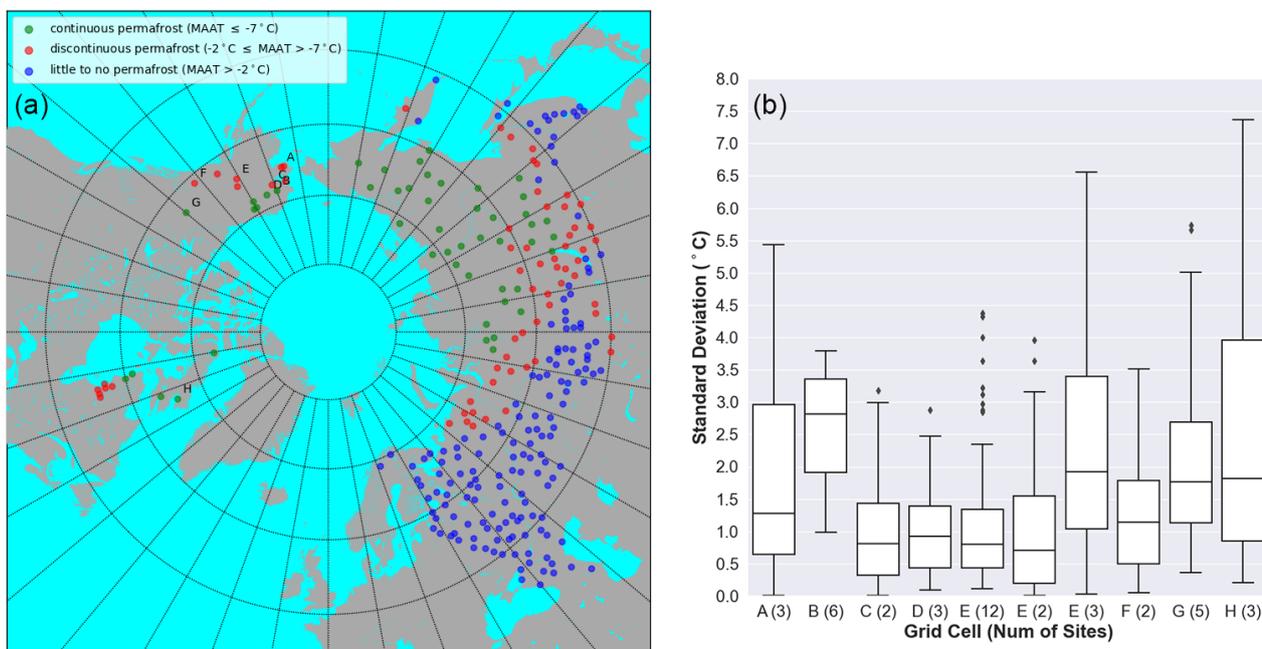
Figure 1. Panel A: location of the validation grid cells collocated with in situ stations in the near-surface layer. Grid cells are colour coded based on mean annual air temperature (MAAT); a proxy for permafrost zone. The locations of the validation grid cells differ minimally in the depth (30 cm to 300 cm) layer. Panel B: temporal spread in spatial standard deviation of monthly soil temperatures near the surface, for grid cells with multiple in situ stations (the number of stations is included in brackets beside the letter). Their location is shown in Panel A.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(T_p - T_i)^2} \tag{2}$$

155 $$SDV = \sqrt{\frac{\sum_{n=1}^{N}(x_n - \overline{x})}{N-1}} \tag{3}$$

$$SDV_{norm} = \frac{SDV_{T_p}}{SDV_{T_i}} \tag{4}$$

$$R = \frac{\frac{1}{N}\sum_{n=1}^{N}(T_p - \overline{T_p})(T_i - \overline{T_i})}{SDV_{T_p}SDV_{T_i}} \tag{5}$$

160   Where T_p is the T$_{soil}$ from the reanalysis product, T_i is the T$_{soil}$ of the in situ data, N is the number of monthly soil temperature values, SDV_{T_p} and SDV_{T_i} are the standard deviations of the reanalysis product soil temperatures and in situ soil temperatures, respectively, while x refers to the T$_{soil}$ (of a dataset).

Metrics were calculated separately for each individual grid cell, and then averaged to calculate a pan-Arctic estimate. Estimates for the permafrost zone and the zone with little to no permafrost were also calculated by averaging together metrics from
165   grid cells falling within a particular zone.

3.2   Binning of Datasets by Season and Mean Annual Air Temperature

Datasets were binned into a cold season and warm season using the Berkeley Earth Surface Temperature (BEST) 2m air temperature (T$_{air}$) for each grid cell. Cold season months are those where T$_{air}$ ≤ -2° C, while the warm season refers to months with T$_{air}$ > -2° C, where T$_{air}$ is the monthly mean air temperature. Sensitivity testing on the cold/warm season
170   revealed no substantive impact on our conclusions using a threshold of 0° C, -5° C, and -10° C. We also tested the impact of using a different temperature dataset to perform the binning; the ERA5 2m air temperature, which resulted in similar findings.

Soil temperatures were also also separated based on the 1981-2010 BEST mean annual air temperature, in order to estimate continuous (≥ 90% permafrost), discontinuous (50 - 89.9% permafrost), and regions with little to no permafrost (< 50% permafrost). Smith and Riseborough (2002) found that the continuous permafrost zone is roughly consistent with a MAAT
175   between -6° C to -8° C, while the discontinuous permafrost zone boundary varies somewhat depending on soil organic matter content. For regions with primarily inorganic (mineral) soil, the southern boundary of the discontinuous permafrost zone was roughly consistent with a MAAT of -2° C, whereas for areas underlain by soils with a heavy organic matter content, the boundary was closer to the +1° C MAAT isotherm. As soil organic matter content was not available for the reanalysis products, the following MAAT boundaries were used in this study as a general approximation of permafrost type across the
180   study domain:

1. locations with a MAAT ≤ -7° C are referred to as being in the 'continuous permafrost' zone

2. locations where -2° C ≤ MAAT > -7° C are referred to as having 'discontinuous permafrost'

3. locations with a MAAT > -2° C are referred to as falling within the 'little to no permafrost' zone

The resultant permafrost distribution (Figure 1) is roughly consistent with the Brown et al. (2002) permafrost map. To
185   maximize the sample size in each group, we merge the 'continuous' and 'discontinuous' permafrost zones into a single category called the 'permafrost zone', and compare against the zone with 'little to no permafrost'.

4 Validation of Reanalysis and LDAS Products

4.1 Warm Season

Most products show small to moderate negative (cold) biases over the warm season (Figure 2); a finding similar to previous
190   studies investigating reanalysis soil temperatures over the mid-latitudes and the Qinghai-Tibetan Plateau (Hu et al., 2017; Qin
et al., 2020; Yang et al., 2020). Warm season biases (Figure 2, Panels C and D) tend to be slightly larger at depth for most
products (by 1° C - 2° C). JRA-55, however, displays substantially more negative biases near the surface – suggesting that it
underestimates summer soil temperatures near the surface.

Over the warm season, most reanalysis products display correlations greater than 0.95 near the surface, and close to 0.9 at
195   depth (Figure 3, Panels C and D). JRA-55, however, displays a correlation of 0.4 near the surface; arising from the fact that its
T_{soil} display a large spread for a given station temperature (red scatter) (Figure 4, top row). For example, while the histogram
for the station data shows that T_{soil} rarely falls below 0° C near the surface, in the warm season, frozen soil occurs relatively
regularly in JRA-55 during the warm season (Figure 4); likely suggesting that JRA-55's single soil layer is more representative
of deeper soil layers (which are more likely to remain frozen year round). Most products generally capture the warm season
200   variability in T_{soil}, and are within 20% of the observed – both at depth and near the surface (Figure 3, Panels C and D).

4.2 Cold Season

Over the cold season, biases tend to be larger (more negative) near the surface, relative to the warm season; particularly in
the cases of GLDAS-Noah, GLDAS-CLSM, and ERA-Interim, where cold season biases are several degrees larger (Figure 2,
Panels A and B). ERA5-Land, however, shows a slight overall positive (warm) bias, because of substantial positive biases over
205   the coldest temperature ranges (Figure 5), and over Siberia and North America (Figure S1); qualitatively similar to the findings
of Cao et al. (2020). At depth, half of the products show smaller cold season biases (relative to the warm season), while the
biases are of a similar magnitude for the remaining products. Most products show a maximum bias when T_{soil} is between -2°
C to -10° C, and there is a tendency for biases to decrease or flip sign over the coldest temperatures (Figure 5).

Cold season correlations are generally lower; particularly near the surface, where correlations are often 25% to 35% smaller
210   than they are over the warm season (Figure 3, panel A). In addition, there is less agreement between products, with a spread of
0.26 between the product with the lowest correlation (JRA-55), and the product most correlated with in situ (Ensemble Mean).
At depth, seasonal differences in correlations are smaller, though cold season correlations for most products are 5% to 10%
lower than their warm season counterparts (Figure 3, panel B).

Several products overestimate the in situ standard deviation in the cold season - particularly over Eurasia, though ERA5-
215   Land underestimates the cold season standard deviation by approximately 25% to 50% (Figure 3, panel A and B) - arising in
part due to the wintertime positive biases over large portions of the northern hemisphere (see Figure S2 and Cao et al. (2020)
for further details). The increased variance in T_{soil} in the cold season suggests that reanalysis products have a greater degree
of temperature variability in the winter. Along the topmost row (near-surface) and leftmost column (depth) of Figure 4 are
scatterplots of soil temperatures from each of the eight reanalysis products, relative to in situ data. There is a larger range of
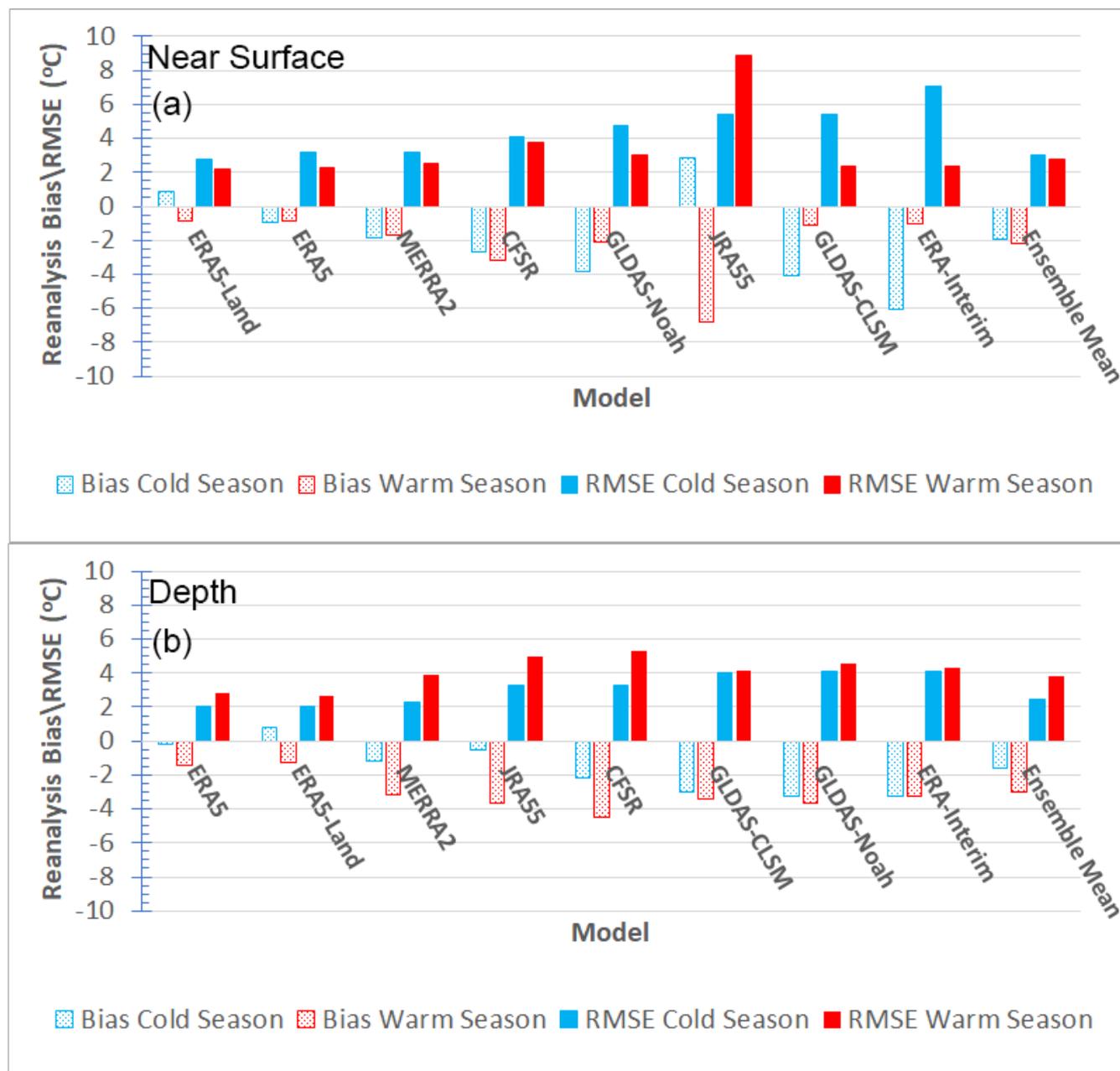
Figure 2. Bias (stippling) and Root Mean Square Error (RMSE) (solid colour) for the cold season (blue) (≤ -2° C) and the warm season (red) (> -2° C) performance of reanalysis products. Panel A displays the bias and RMSE for the near surface (0 cm to 30 cm) layer, while panel B displays the bias and RMSE at depth (30 cm to 300 cm). Models are ordered based on cold season RMSE (from the smallest to largest). The ensemble mean is shown beside for comparison.

220   temperatures for a given observed soil temperature in the cold season (blue scatter) than there is over the warm season (red scatter), both at depth, and near the surface. Second, there is a greater spread in standard deviations between products at colder
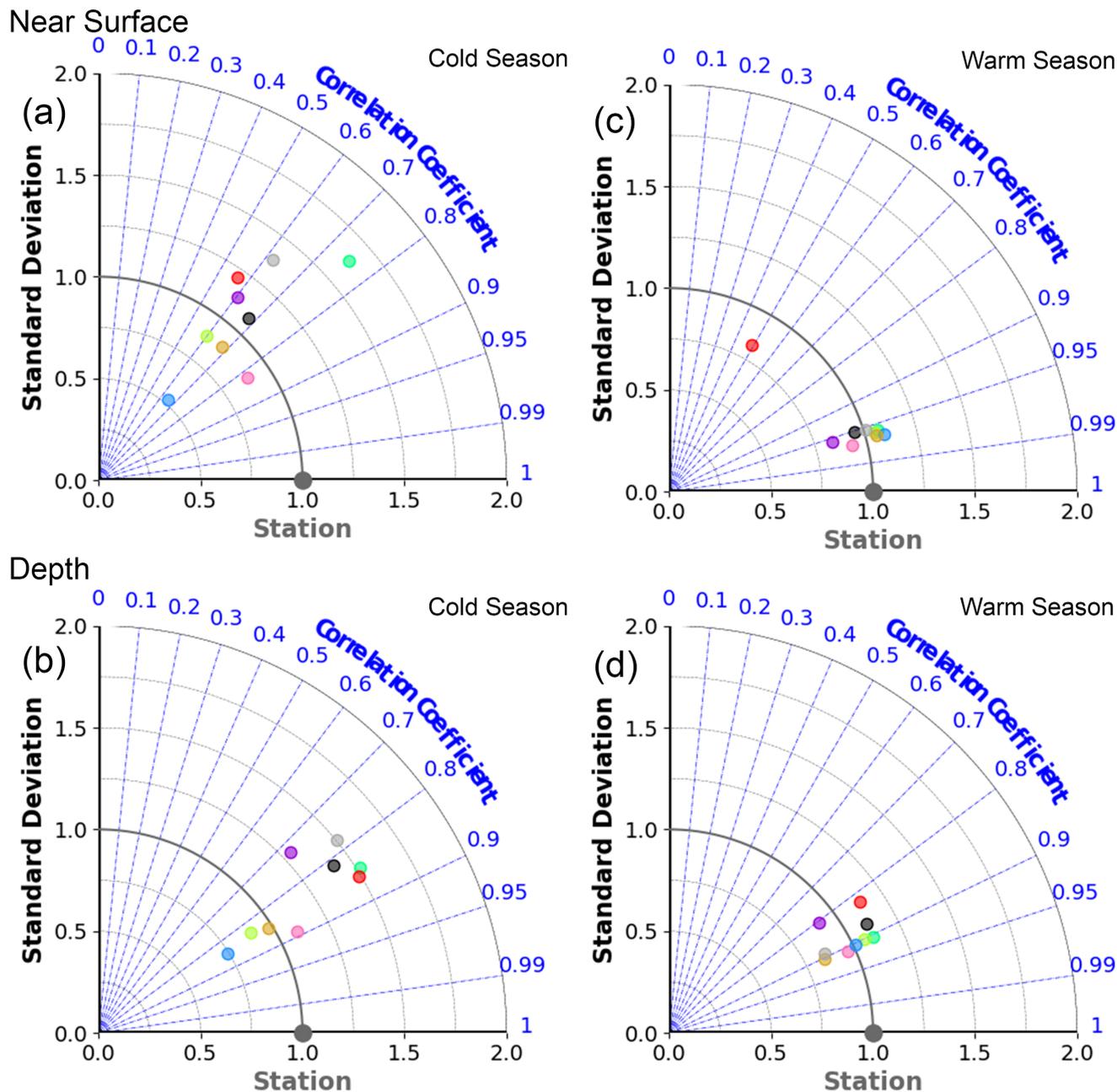
Figure 3. Taylor Diagram of the cold season (≤ -2° C) and the warm season (> -2° C) performance of reanalysis products. Panels A and B refer to the cold season, while panels C and D refer to the warm season. The top panels (panels A and B) are for the near surface (0 cm to 30 cm) while the bottom panels (panels C and D) refer to soil temperatures at depth (30 cm to 300 cm).
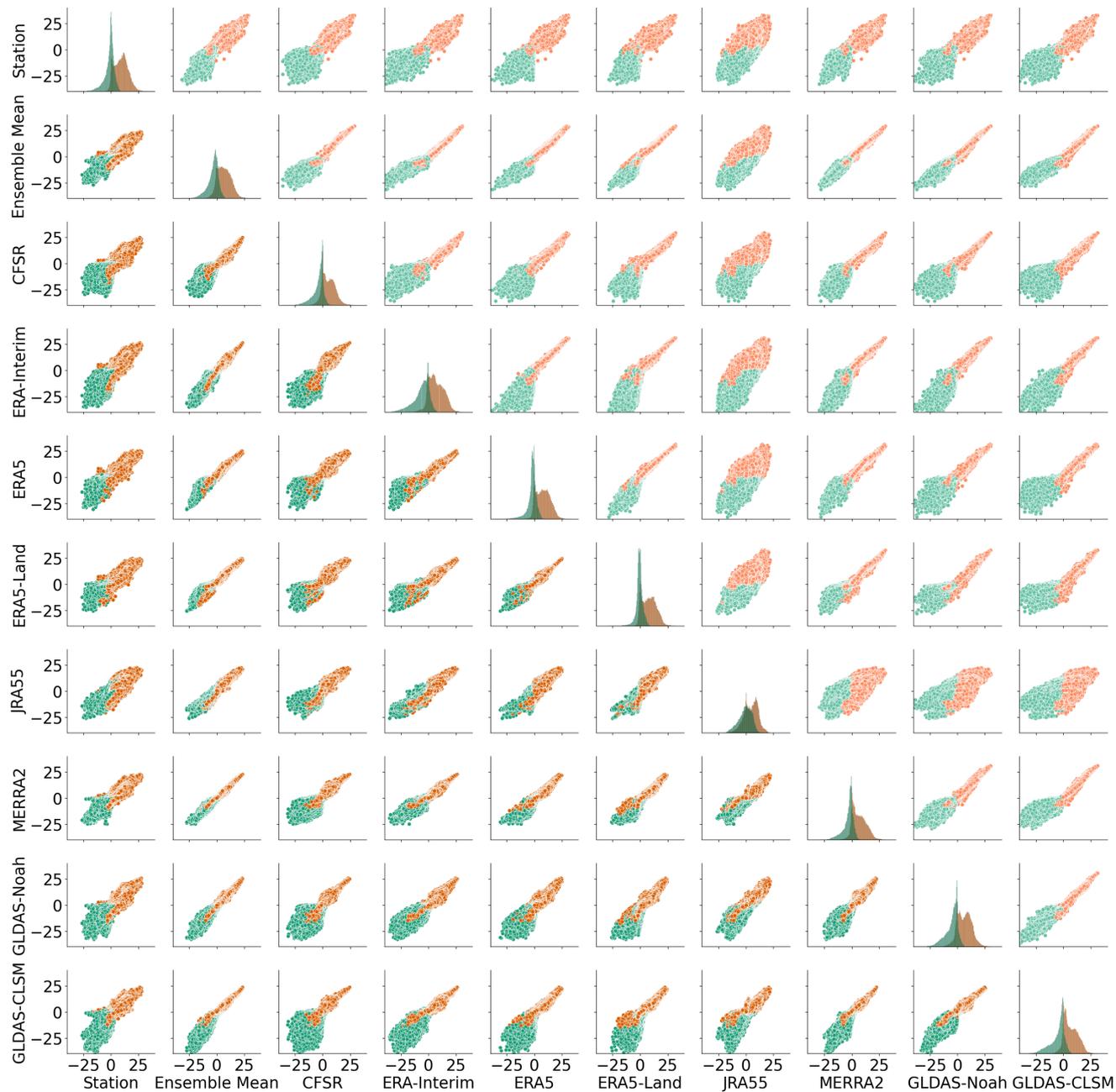
Figure 4. Scatterplot matrix of station and reanalysis soil temperatures. The top triangle (right of the histograms) refers to soil temperatures in the near-surface (0 cm to 30 cm) layer, while the bottom triangle represents soil temperatures at depth (30 cm to 300 cm). Seasons are stratified by the ERA5 air temperature, with the cold season ($\leq$ -2° C) in green and the warm season (>-2° C) in red.
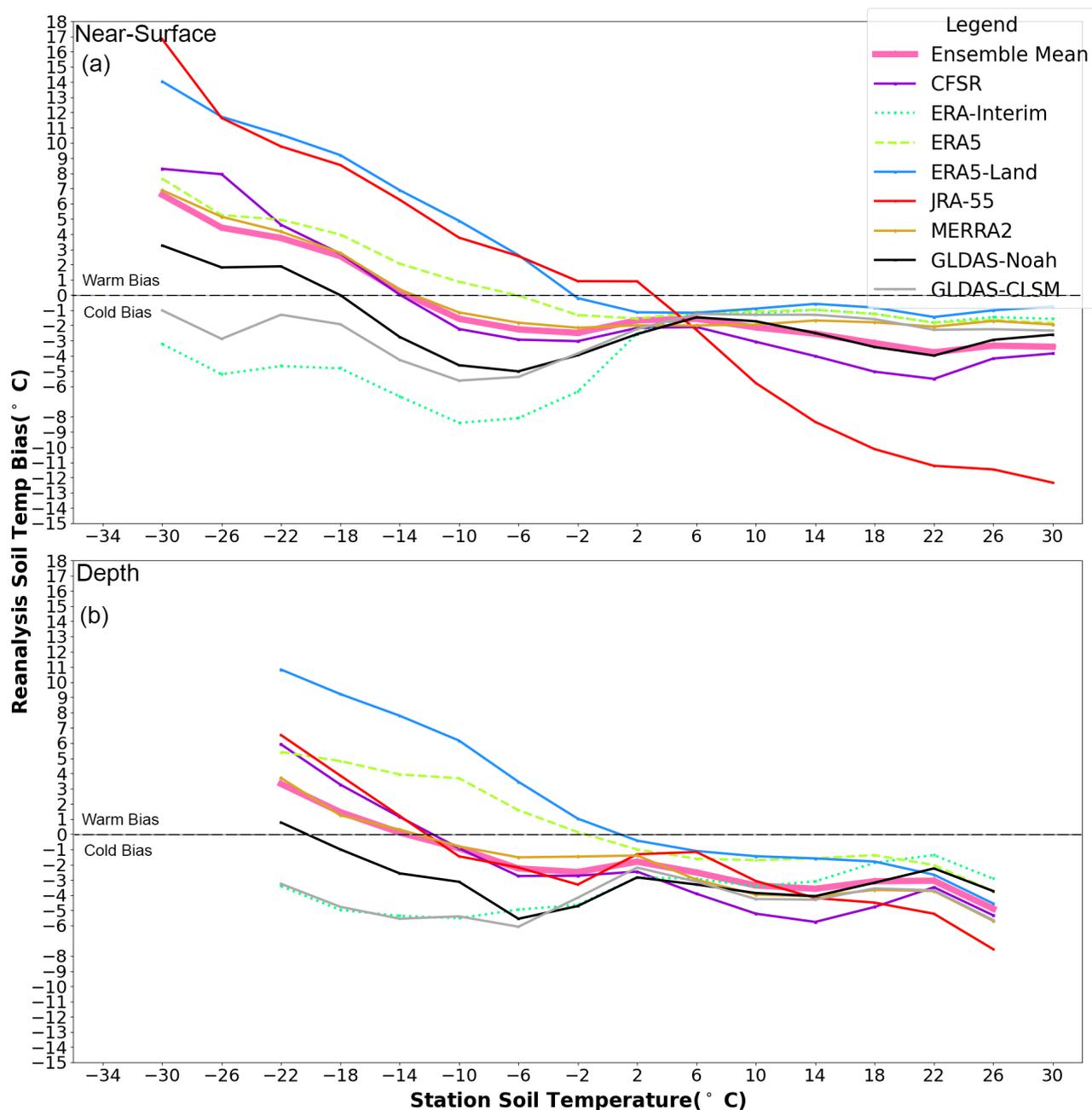
Figure 5. Reanalysis soil temperature bias as a function of station soil temperature for a) the near surface (0 cm to 30 cm) layer, and b) at depth (30 cm to 300 cm). Station temperatures are binned into 4° C intervals, beginning with the -32° C to -28° C bin, and ending with the 28° C to 32° C bin. The midpoint of each temperature bin is plotted along the x-axis.

temperatures. The spread is largest over the coldest temperature ranges (Figure 6), which suggests that colder climate regimes are likely an important controlling factor on reanalysis performance. Snow has a low thermal conductivity that insulates the soil, decoupling the air temperature above from the soil below the snow pack Zhang (2005). Cao et al. (2020) note that winter

225 warm biases in T_{soil} in ERA5-Land could be partially explained by HTESSEL having a snow density that was biased low (relative to observations), leading to a snowpack thermal conductivity that was too low, and an associated overestimate of the insulating effects of snow. Snow was also cited as a major controlling factor in soil temperature biases in ECMWF's Integrated Forecast System, which also uses the HTESSEL land surface model (Albergel et al., 2015).

Most products show a maximum standard deviation over the coldest T_{soil}, whereas JRA-55 shows a maximum standard

230 deviation when the observed T_{soil} are near freezing and shows a decline in standard deviation over the coldest temperatures (Figure 6), particularly near the surface (panel A).

4.3 Influence of Frozen Soils

Over our domain, there are grid cells over regions with substantial permafrost cover, as well as grid cells that fall outside the region typically covered by permafrost. As there are differences in the physical processes governing soil temperatures across

235 regions covered by permafrost, it is important to mention differences in performance between these two regions. Bias and RMSE are typically 2° C to 4° C larger over the permafrost zone (in both seasons) (Fig. S2). The mean bias and RMSE are typically 1° C to 3° C smaller over North America, relative to the permafrost zone in Eurasia (see Figure S3); however with fewer grid cells over North America, the uncertainty is also larger - as evidenced by the larger error bars.

Correlations between the permafrost region and the zone with little to no permafrost are generally quite similar, rarely

240 differing by more than 0.1 (not shown). Individual models are more likely to overestimate the near surface variance over the zone with little to no permafrost, while the opposite is true at depth (not shown). In the permafrost zone, however, the spread in standard deviation, at depth, is about 2.75 times larger than it is over the zone with little to no permafrost (not shown). Over North America, several products overestimate the observed standard deviation in the warm season, while there is good agreement between the reanalysis products and the observed standard deviation in Eurasia. Conversely, the cold season variance

245 is more likely to be overestimated over Eurasia (not shown). It is likely that sampling variability in T_{soil} is a large contributing factor to the differences in variance between North America and Europe.

Several factors may explain the increased variability in soil temperatures over permafrost regions. First - snow cover is present for longer, and may persist for longer on slopes with a northern aspect or where it has accumulated around vegetation. Munkhjargal et al. (2020) observed an average difference in soil temperatures of 3° C to 4° C across the Qinghai-Tibetan

250 plateau in response to aspect. Second, near surface soil temperature is particularly sensitive to vegetation cover; shrub height and vegetation type were found to account for roughly half the variability in late winter and spring soil temperatures in the low Arctic (Grünberg et al., 2020). Moreover, in the discontinuous permafrost zone, the presence of permafrost is often ecosystem protected or ecosystem driven. Variability in vegetation and drainage, for example, may lead to localized portions of the landscape where soil temperatures are colder than in surrounding areas (Jorgenson et al., 2010). Thirdly, latent heat interactions

255 in the active layer during spring can lead to long periods of time where the soil remains at or close to freezing (the zero-curtain
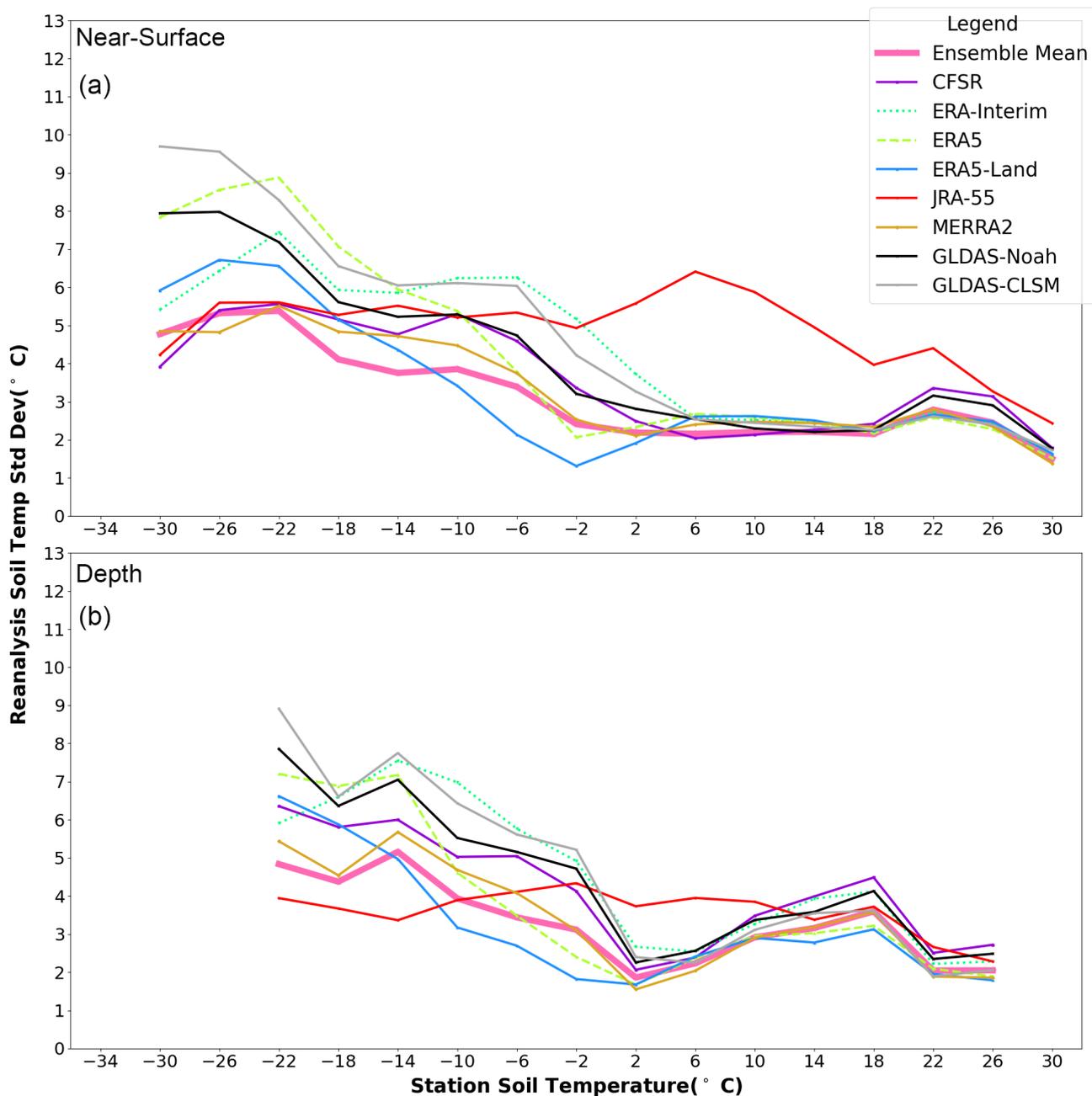
Figure 6. Reanalysis soil temperature standard deviation as a function of station soil temperature for a) the near surface (0 cm to 30 cm) layer, and b) at depth (30 cm to 300 cm). Station temperatures are binned into 4° C intervals, beginning with the -32° C to -28° C bin, and ending with the 28° C to 32° C bin. The midpoint of each temperature bin is plotted along the x-axis.

period). In regions where the active layer is deep (such as over the discontinuous permafrost zone), the zero curtain period is often longer and more pronounced (Chen et al., 2021) than it is over the continuous permafrost zone or regions with seasonally frozen soil. Many of the processes that control the zero-curtain effect, such as freeze-thaw parameterizations are relatively simplistic in many land models (Cao et al., 2020; Chen et al., 2015), and their coarse resolution would fail to capture local
260   scale variations in the zero-curtain period.

5   Ensemble Mean Product

5.1   Validation

Having validated the performance of the individual products, here we construct a soil temperature product based on the ensemble mean of the products, and investigate its performance. Taking the ensemble mean of all eight soil temperature prod-
265   ucts produces a soil temperature dataset that shows closer agreement with observed soil temperatures than most/all individual products. The ensemble mean bias and RMSE are generally quite close in magnitude to the best performing products over all regions, seasons, and depths (Figure 2). Moreover, by sequentially increasing the number of products included in the ensemble mean, we are able to demonstrate that greater number of products in the ensemble mean leads to a reduction in overall bias and RMSE (Figure 7, panels A and B). Interestingly, the number of products included appears to be of greater importance to the
270   performance of the ensemble mean than which products are included, suggesting that most products share similar uncertainties and signal-to-noise ratios, and resulting, to a certain degree, in the offsetting of biases or errors.

In addition, the ensemble mean maintains a high correlation (often the highest of all products) over all seasons and depths (Figure 3), and analogous improvements to correlation are seen as a greater number of models are included (Figure 7). Finally, the ensemble mean standard deviation is close to the observed value over all seasons and depths (Figure 3), suggesting that
275   the seasonal variability in T_{soil} is similar to the variability seen in the station data. Thus, the ensemble mean soil temperature dataset provides the best estimate of in situ temperatures for the broadest range of conditions.
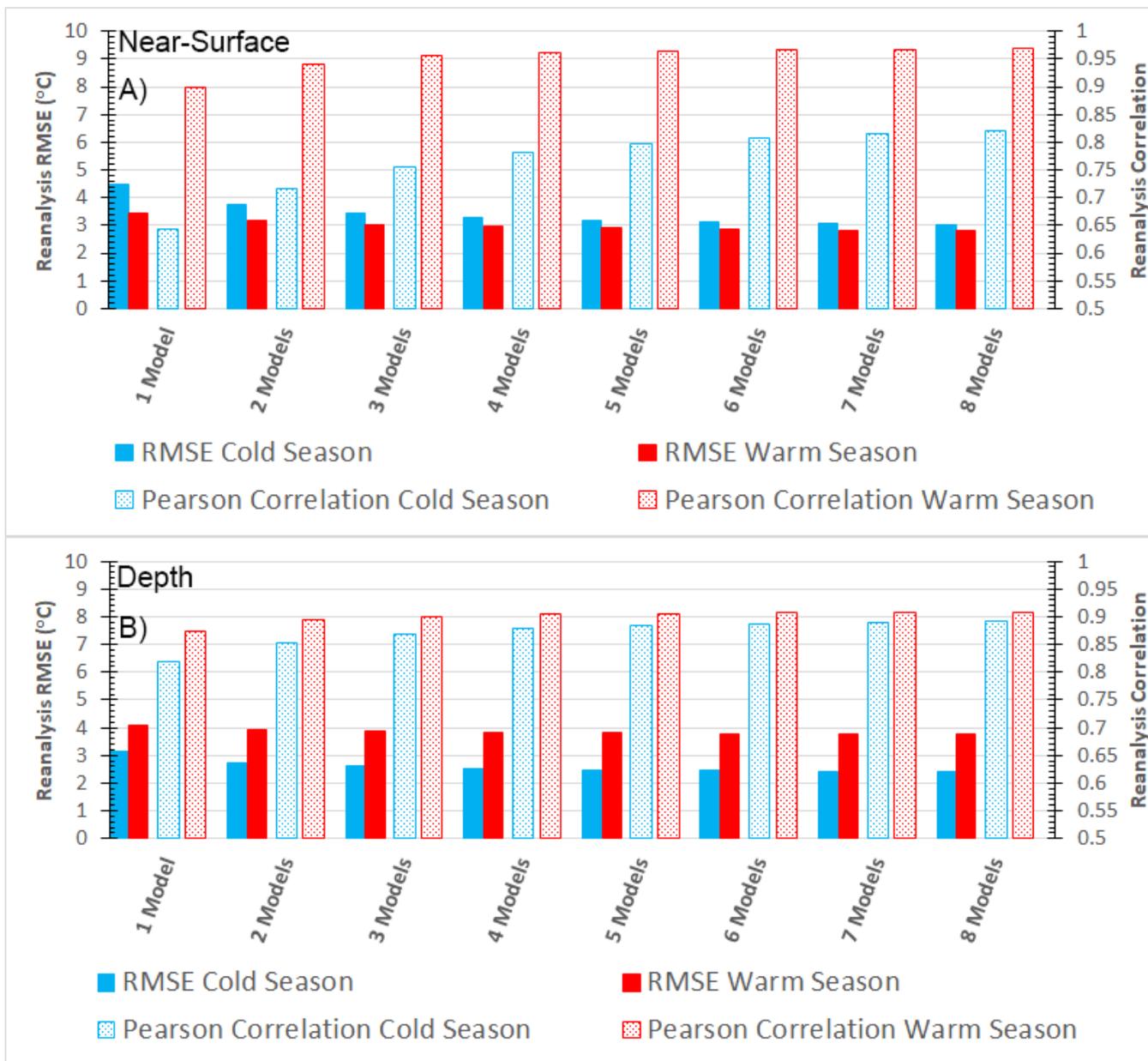
Figure 7. Root Mean Square Error (RMSE) (solid colour), and Pearson Correlation (stippling) for the cold season (blue) ($\leq$ -2° C) and the warm season (red) (> -2° C) averaged over all combinations of 1 model through to 8 model ensemble means. Panel A displays the RMSE and correlation for the near surface (0 cm to 30 cm) layer, while panel B displays the RMSE and correlation at depth (30 cm to 300 cm). Values are ordered based on cold season RMSE (from smallest to largest). Note that the y-axis scale is from -8° C to +10° C (rather than -10° C to +10° C).
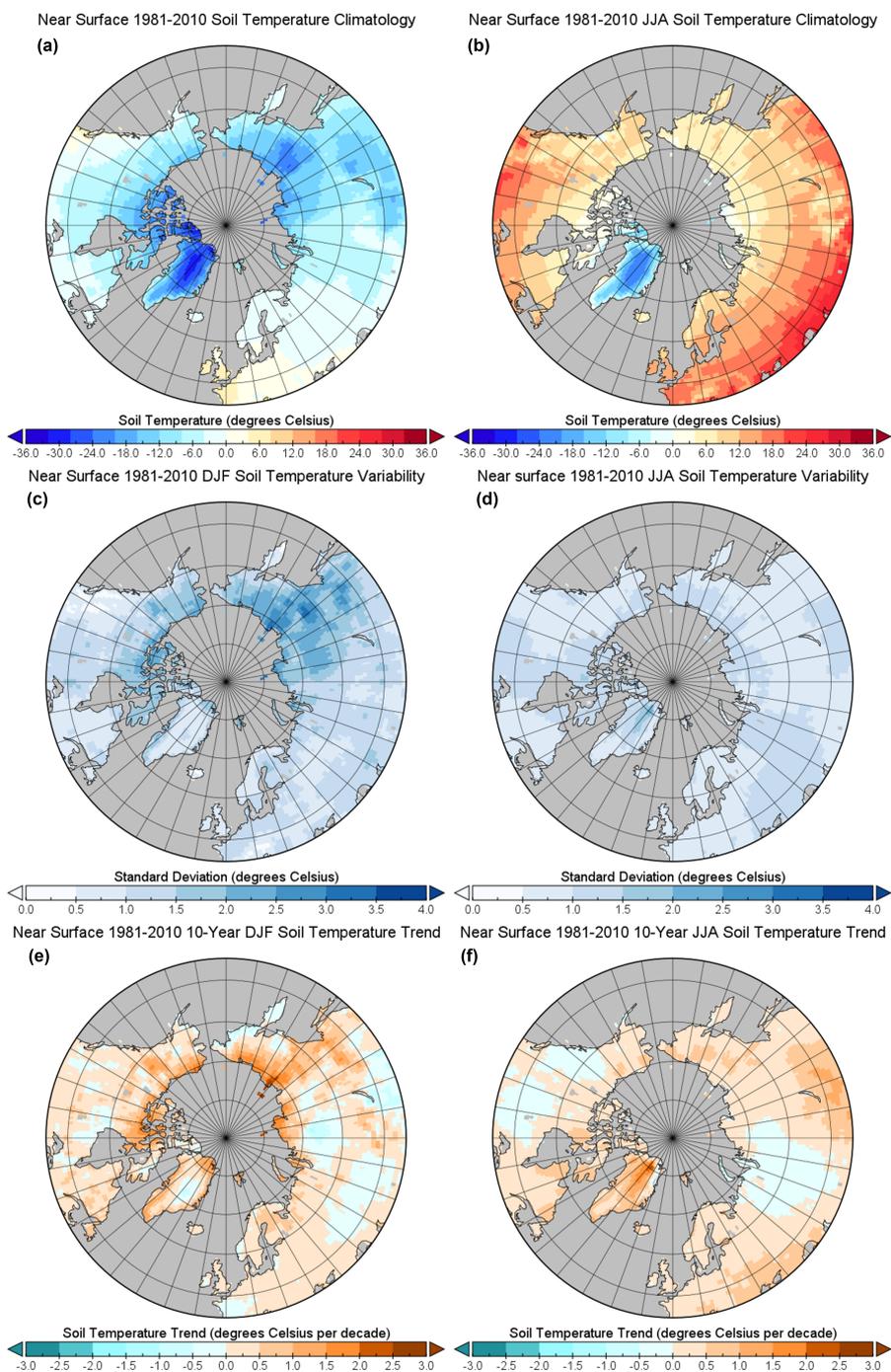
Figure 8. Ensemble mean 1981-2010 near-surface climatology (panels A and B), variability (panels C and D), and 30-year trend (panels E and F). Values for DJF are shown in the left hand panels, and those for JJA are shown in the right hand panels.

5.2 Climatology (1981-2010)

Having established that an ensemble mean of products provides the best choice for T_{soil} in most situations, in this Section we characterize the pan-Arctic mean, variability and trends of the ensemble mean T_{soil} product. Unlike in previous sections,
280 where the air temperature of a grid cell was used to separate the cold and warm season, this was not possible for a discussion of the climatology. Henceforth, in Sections 5.3 and 5.4, we use the DJF climatology to represent winter, and JJA climatology to represent summer.

Near surface soil temperatures generally range between -12° C and -30° C for most regions north of 60° N in DJF, with slightly warmer temperatures over the Yukon, Alaska, and Scandinavia (Figure 8, panel A). Temperatures are often a few
285 degrees warmer at depth, but the temperature pattern remains quite similar (DJF pattern correlation > 0.95 over the extratropical NH north of 45° N).

In JJA, most of the near surface soil temperatures across the Arctic rise above freezing, with the exception of a few locations along Baffin Island and Ellesmere Island (as the top 30 cm is representative of the active layer). Temperatures generally range between 6° C to 12° C, with cooler areas across Eastern North America, and along the northern coast of Siberia (Figure
290 8, panel B). At depth, (not shown), much of the Canadian Arctic Archipelago and the northern portions of Siberia remain below freezing, with most other regions showing soil temperatures between 0° C to 6° C (except for Scandinavia). JJA pattern correlations in near-surface and depth soil temperature climatologies are > 0.95.

Near Surface Cold Season Bias

(a)

Near Surface Warm Season Bias

(b)

Depth Cold Season Bias

(c)

Depth Warm Season Bias

(d)
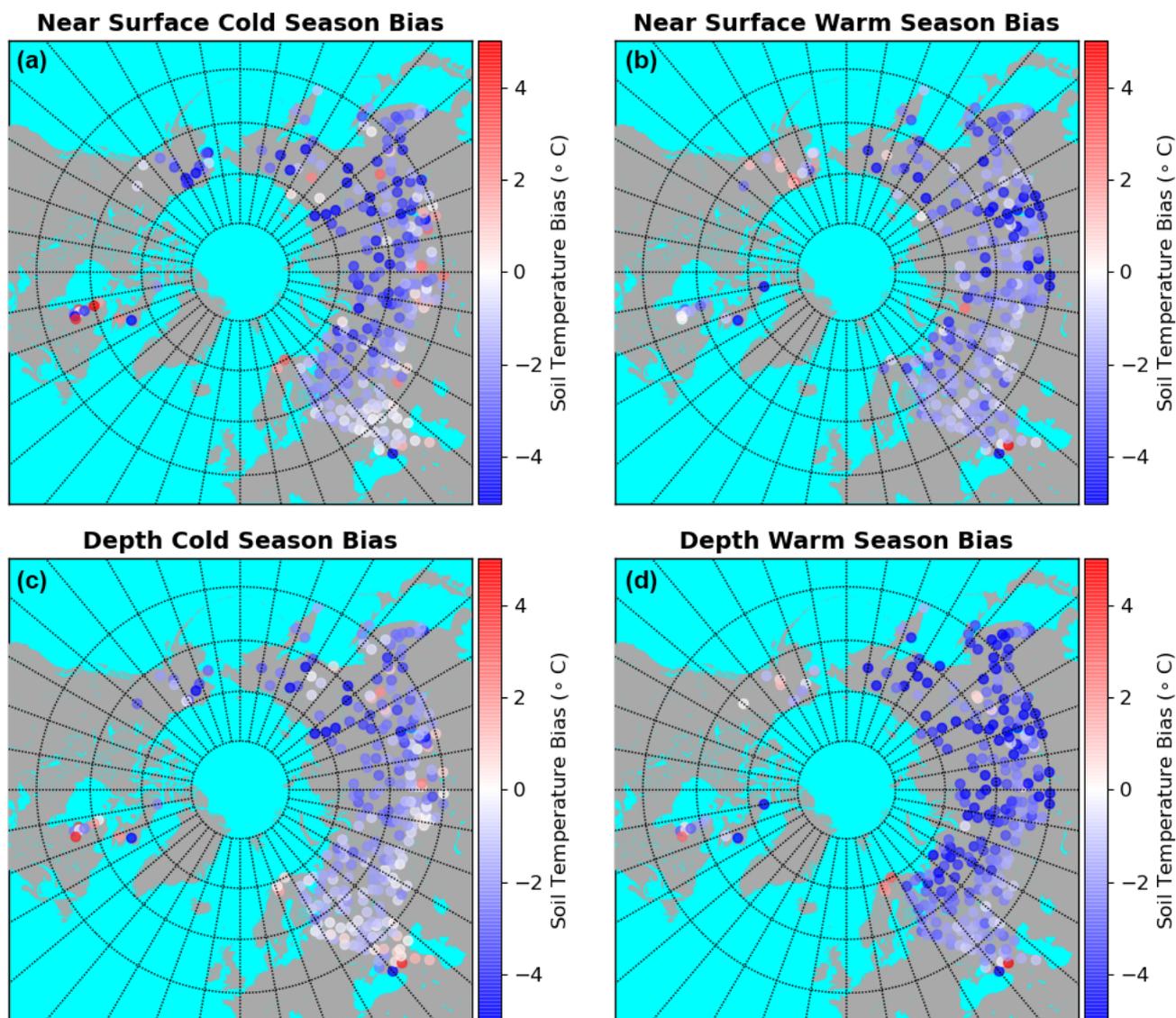
Figure 9. Spatial map of bias for the Ensemble Mean product. Values for the cold season are shown in the left hand panels, and those for the warm season are shown in the right hand panels. Panels A and B show the near surface bias, while biases at depth are shown in Panels C and D.

While, not directly comparable to the climatology figure, cold season biases in the Ensemble Mean product are typically -4° C to -5° C over the permafrost zone, and -2° C to -3° C over much of the zone with little to no permafrost (Figure 9, panel

295 A). There is little difference in the cold season biases at depth, except for being slightly smaller in western Eurasia (Figure 9, panel C).

In the warm season, the near surface biases are generally small (around -1° C over much of Eurasia), except for a few grid cells south of Siberia, where they are similar in magnitude to the cold season (around -4° C). Alaska and the Yukon, however,

appear to show a small to moderate warm bias on the order of $+1°$ C to $+3°$ (Figure 9, panel B). At depth, the biases are
300 somewhat larger than they are at the surface, particularly over Eurasia, but the pattern remains similar to the near surface
(Figure 9, panel D).

5.3   Variance and Trends (1981-2010)

Generally speaking, coastal regions show the greatest variability in T_{soil}, with inland regions showing substantially smaller
variation. Temperature variability is larger in DJF than it is for JJA (Figure 8, Panels C and D). A similar pattern is generally
305 seen at depth, but with reduced variability (pattern correlation greater than 0.9 in DJF). In JJA, the pattern in variability is
generally similar, though pattern correlations between surface and depth are closer to 0.58, owing to an increased sensitivity to
small changes standard deviation.

Decadal temperature trends (Figure 8) are the largest in DJF (panel E); reaching as large as $3°$ C per decade along the
northern coastlines. Inland regions typically see smaller DJF trends; closer to a degree per decade, at most. While most regions
310 show warming trends, portions of coastal Siberia and parts of western Eurasia see small to moderate cooling trends of $0.33°$
C per decade to $1°$ C per decade over the thirty year period. Pattern correlations between near-surface and depth temperature
trends are greater than 0.95 in DJF.

JJA temperature trends (Figure 8, panel F) are generally much smaller in magnitude; usually between $0.33°$ C per decade to
$0.66°$ C per decade, except along coastal regions where they range between $0.66°$ C per decade to $1°$ C per decade. Parts of
315 North America and central Eurasia show a slight cooling, while most of eastern Eurasia shows slight warming trends. Pattern
correlations between near-surface and depth temperature trends over JJA are 0.86.

6   Discussion and Conclusions

This study conducted a validation of pan-Arctic soil temperatures for eight reanalysis products, and validated a new ensemble
mean pan-Arctic soil temperature dataset. Most reanalysis products show a negative (cold) soil temperature bias of several
320 degrees across the Arctic; one that is most prominent during the cold season, and over the highest latitudes. There is also
greater disagreement between the reanalysis products' estimates of soil temperature variability and correlation during the cold
season. When all depths and seasons are considered, the ensemble mean product performs better than any individual product.

All estimates of T_{soil} are subject to uncertainty, and while we have not been able to present a complete quantitative assess-
ment of these sources of uncertainty, we have made some progress in their characterization. In particular, we identify four main
325 categories of uncertainty, and provide qualitative estimates of their importance:

1. Uncertainties in data assimilation and configurations of atmospheric model

   Many reanalysis products exhibit a positive air temperature bias near the surface over the Arctic (Beesley et al., 2000;
   Lindsay et al., 2014; Tjernström and Graversen, 2009); particularly in winter, in association with temperature inversions
   that are weaker than observed (Serreze et al., 2012; Tjernström and Graversen, 2009). Differences between products in

330   the parameterization of cloud cover and small scale turbulent mixing have been highlighted as key sources of uncertainty in atmospheric reanalyses (Boisvert et al., 2018; Taylor et al., 2018), along with the near surface Arctic energy budget (Graham et al., 2019). Reanalysis estimates of precipitation amount and phase are another important source of uncertainty, and are especially hard to constrain over the Arctic, due to a lack of observations (Boisvert et al., 2018).

2. Uncertainties in the land model configuration and/or parameterizations:

335   The methods used here did not allow for explicit study of the impacts of land model configuration on soil temperature. Nevertheless, it is important to acknowledge that most land models fail to simulate important aspects of permafrost soils, such as the influence of phase change on thermal conductivity (Cao et al., 2020), or include relatively crude parameterizations of freeze-thaw processes (Chen et al., 2015). Moreover, many current generation land models fail to capture feedbacks that exist between vegetation and permafrost (Chadburn et al., 2015) - a factor particularly important in the

340   discontinuous permafrost zone, where permafrost presence (or absence) is heavily influenced by vegetation (Jorgenson et al., 2010). Land surface models, such as the Simple Biosphere Model (used in JRA-55), that use the force restore method to estimate soil temperature, are prone to overestimating diurnal soil temperature range (Gao et al., 2004; Kahan et al., 2006), as well as the seasonal cycle of soil temperatures (Luo et al., 2003). This is because they underestimate heat capacity, and overestimate temporal variation in ground heat flux compared to more complex land models (Hong

345   and Kim, 2010). The force restore method assumes a strong diurnal forcing from above. In regions with substantial snow cover, such as the Arctic, this assumption is likely violated (Tilley and Lynch, 1998) because snow cover leads to a decoupling of the surface forcing from the soil below, and may, in part, explain why JRA-55 displays a reduced standard deviation in temperature over the cold season relative to most other products.

3. Instrumental uncertainties

350   The approach used here assumes that uncertainties in the in situ data are negligible, and hence are the 'true' soil temperature in a particular grid cell. In situ soil temperature measurements, however, are taken as a point measurement, and as such they may not be representative of the average soil temperature over a grid cell. As this is hard to explicitly account for in linear metrics such as bias, RMSE, and correlation, it can result in non-negligible errors (Crow et al., 2012; Loew and Schlenz, 2011); particularly when a limited number of stations are contributing to the spatially averaged station

355   temperature (as is the case over large parts of Eurasia). While the use of data from dense networks may improve the spatial representativeness of validation data (Zeng et al., 2015), their drawback is that they are often limited in terms of the land cover types they represent (Zeng et al., 2015; Ma et al., 2021).). This is particularly true in the Arctic where data from dense networks is extremely limited.

An estimate of the uncertainty of the validation data was calculated using grid cells where multiple stations were avail-

360   able. The uncertainty is represented by the standard deviation of soil temperatures between stations within a grid cell for a particular timestep. While the median standard deviation of soil temperature between stations is generally $\leq 2°$ C for eight of the nine grid cells, there are periods of time when the standard deviation is substantially larger than this (Figure 1, Panel B); consistent with the findings of in situ studies examining soil temperature variability at the site level

(Scharringa, 1976; Gubler et al., 2011; Gruber et al., 2018) and at regional scales (Kropp et al., 2021; Grünberg et al., 2020; Way and Lapalme, 2021; Zhang et al., 2021). Kropp et al. (2021) note that T_{soil} can vary widely over small spatial scales in response to vegetation canopy structure; on scales of tens of metres in many cases Gruber et al. (2018). This is particularly true in the cold season, as the presence or absence of vegetation has a strong control on snow accumulation. By trapping snow, trees and shrubs lead to substantially warmer soil temperatures in their vicinity (Way and Lapalme, 2021). Moreover, the impact of snow cover on soil temperature is generally more pronounced over permafrost regions (regions of seasonal frost) (Chen et al., 2021).

Uncertainties can also arise from processes that occur within the soil itself; particularly in seasonally frozen and permafrost soils. For example, rapid temperature fluctuations are often seen during the melt period, as melt-water can induce strong, localized latent heat fluxes in the active layer - factors not easily captured in monthly mean temperatures over large areas (Hinkel and Outcalt, 1995). Frost heave can lead to seasonal shifts in the position of the soil temperature probes, resulting in seasonal discontinuities in soil temperature datasets between summer and winter (Streletskiy et al., 2017; Urban and Clow, 2017).

4. Sampling uncertainties

Uncertainties exist due to inconsistencies in data availability, both temporally and spatially. Some sites (particularly those in Siberia, for example), included 30 years or more of data, while others were limited to a data record of a few years; resulting in inconsistencies in the time period over which metrics were calculated for each grid cell. The authors also acknowledge there are substantially more validation grid cells over Eurasia, than over North America. In order to test the impact of this, a Monte Carlo type simulation was constructed, whereby 25 (19) grid cells from the continuous and discontinuous permafrost zones of Eurasia were randomly selected, alongside the 25 (19) North American grid cells, and the mean bias, RMSE, standard deviation and Pearson Correlation were calculated for the 50 (38) grid cell subsample from the near surface (at depth). This process was repeated 10000 times, and the bias, RMSE, standard deviation, and Pearson Correlation from the subsamples were compared against those calculated over the entire continuous and discontinuous permafrost domain. Generally speaking, the bias, RMSE, standard deviation and correlation over the entire continuous and discontinuous permafrost domain fall within 1 quartile of the median value from the Monte Carlo Simulations; particularly in the top 30 cm. At depth, the median bias was 1° C smaller than the mean over the entire permafrost domain, and correlations were 0.1 lower in many cases. However, the sampling error was not large enough to alter the fundamental conclusions of the study.

It is possible that validation sites in the discontinuous permafrost zone were preferentially located in regions where permafrost is present. In parts of Alaska, and northern Quebec, for example, there are localized areas where warm biases exist (Figure 9). This may be expected in regions where permafrost presence is ecosystem protected, as reanalysis products are unable to account for local scale feedbacks of vegetation on permafrost presence. Given, however, that the discontinuous zone is a relatively small part of the validation dataset, and that warm biases in the discontinuous zone are

generally limited to a few localized areas, the authors do not believe that selection bias has fundamentally influenced the overall conclusions.

Time-series did not have a consistent start or end date - meaning that the metrics are calculated using different climatologies
400 across different grid cells. Thus, it is not possible to calculate a true uncertainty estimate. We've qualitatively acknowledged the main sources of uncertainty, and here we show an approximate estimate of the magnitude of uncertainty associated with each component, using the standard deviation in soil temperature across time and as a function of station temperature. The median standard deviation of in situ soil temperatures, for grid cells with more than 2 stations, is generally between 0.5° C and 2° C (Figure 1, Panel B) - an order of magnitude smaller than the standard deviation of reanalysis soil temperatures for a
405 given station soil temperature; particularly over the cold season (Figure 6). Reanalysis air temperatures maintain a relatively consistent standard deviation between 1.25° C to 1.75° C for most products, except over the coldest in situ temperatures (not shown). Unlike with soil temperature, the standard deviation of reanalysis 2m air temperature only increases modestly over the cold season; along with the spread in standard deviation between products (not shown). This would suggest that the largest degree of uncertainty in reanalysis soil temperatures over the Arctic is most likely contributed by differences in the land models
410 between products, rather than from uncertainties in observed soil temperatures, or from differences in product air temperatures.

Future work should aim to investigate how differences in snow cover and snow density between the reanalysis products may influence biases in the individual products. On a related note, future studies should emphasize how differences in the land models of reanalysis products may account for the spread in soil temperatures. The ensemble mean data product has potential applications in several fields, including the validation of model soil temperatures, species distribution models and
415 to inform models that require soil temperature inputs, such as hydrological models, or for permafrost simulations. A robust ensemble mean can be computed with four products (Figure 7), which means a higher resolution ensemble mean data product could be created using a subset of higher resolution reanalysis products. For example, ERA-Interim, MERRA2, and JRA-55 have resolutions lower than 0.5 degrees. A 0.31 degree ensemble mean product, based on ERA5, ERA5-Land, CFSR, and GLDAS-Noah is being explored, using a similar methodology.

420 *Data availability.* GTN-P data (GTN-P, 2018) is available from The Global Terrestrial Network for Permafrost, while the Kropp et al. (2020) dataset is available from Heather Kropp's Arctic Data Center page. Russian Hydromet (Sherstiukov, 2012) data are available from RIHMI-WDC, while Nordicana data can be obtained from Nordicana D. CFSR (Saha and Coauthors, 2010), CFSR2 (Saha and Coauthors, 2012), ERA-Interim (European Centre for Medium-Range Weather Forecasts, 2012), ERA5 (European Centre for Medium-Range Weather Forecasts, 2019) and JRA-55 (Japan Meteorological Agency, 2013) data were obtained from the National Center for Atmospheric Research
425 (NCAR)'s Research Data Archive (RDA). GLDAS-CLSM (Li et al., 2020a, b), GLDAS-Noah (Beaudoing et al., 2020a, b), and MERRA2 (Global Modeling and Assimilation Office, 2015) were obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC), while ERA5-Land data (Muñoz-Sabater, 2019) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS). The ensemble mean soil temperature dataset has been made available on the Arctic Data Center.

Author contributions. T. Herrington and C. Fletcher conceived of the study, T. Herrington gathered, and analyzed the data, and T. Herrington,
430 C. Fletcher interpreted the data. H. Kropp provided in situ data to the study, and T. Herrington, and C. Fletcher wrote the manuscript, with
contributions from H. Kropp.

Competing interests. The authors declare that they have no known conflicts - financial or personal, that could have appeared to influence this
work.

Acknowledgements. This work was funded under a National Science and Engineering Research Council (NSERC) PGS-D scholarship. The
435 authors would like to thank Dr. Hugh Brendan O'Neill as well as Dr. Andre Erler for their helpful comments.

References

Albergel, C., Dutra, E., Muñoz-Sabater, J., Haiden, T., Balsamo, G., Beljaars, A., Isaksen, L., de Rosnay, P., Sandu, I., and Wedi, N.: Soil Temperature at ECMWF: An Assessment Using Ground-Based Observations: Soil Temperature at ECMWF, Journal of Geophysical Research: Atmospheres, 120, 1361–1373, https://doi.org/10.1002/2014JD022505, 2015.

440 Allard, M., Sarrazin, D., and Hérault, L.: Borehole and Near-Surface Ground Temperatures in Northeastern Canada, v. 1.5 (1988-2019). Nordicana D8, 2020.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: A Global Land Surface Reanalysis Data Set, Hydrology and Earth System Sciences, 19, 389–407, https://doi.org/10.5194/hess-19-389-2015, 2015.

445 Beaudoing, H., Rodell, M., and NASA/GSFC/HSL: GLDAS Noah Land Surface Model L4 Monthly 1.0 x 1.0 Degree V2.0, , Greenbelt, Maryland, USA, 2020a.

Beaudoing, H., Rodell, M., and NASA/GSFC/HSL: GLDAS Noah Land Surface Model L4 Monthly 1.0 x 1.0 Degree V2.1, Greenbelt, Maryland, USA, 2020b.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and
450 Wood, E. F.: Global-Scale Evaluation of 22 Precipitation Datasets Using Gauge Observations and Hydrological Modeling, Hydrology and Earth System Sciences, 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily Evaluation of 26 Precipitation Datasets Using Stage-IV Gauge-Radar Data for the CONUS, Hydrology and Earth System Sciences, 23, 207–224, https://doi.org/10.5194/hess-23-207-2019, 2019.

455 Beesley, J. A., Bretherton, C. S., Jakob, C., Andreas, E. L., Intrieri, J. M., and Uttal, T. A.: A Comparison of Cloud and Boundary Layer Variables in the ECMWF Forecast Model with Observations at Surface Heat Budget of the Arctic Ocean (SHEBA) Ice Camp, Journal of Geophysical Research: Atmospheres, 105, 12 337–12 349, https://doi.org/10.1029/2000JD900079, 2000.

Betts, A., Chen, F., Mitchell, K. E., and Janjić, Z. I.: Assessment of the Land Surface and Boundary Layer Models in Two Operational Versions of the NCEP Eta Model Using FIFE Data, Monthly Weather Review, 125, 2896–2916, 1997.

460 Boisvert, L. N., Webster, M. A., Petty, A. A., Markus, T., Bromwich, D. H., and Cullather, R. I.: Intercomparison of Precipitation Estimates over the Arctic Ocean and Its Peripheral Seas from Reanalyses, Journal of Climate, 31, 8441–8462, https://doi.org/10.1175/JCLI-D-18-0125.1, 2018.

Brown, J., Ferrians, O., Heginbottom, J., and Melnikov, E.: Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2., Tech. rep., National Snow and Ice Data Center, Boulder, Colorado, USA, 2002.

465 Cao, B., Gruber, S., Zheng, D., and Li, X.: The ERA5-Land Soil Temperature Bias in Permafrost Regions, The Cryosphere, 14, 2581–2595, https://doi.org/10.5194/tc-14-2581-2020, 2020.

CEN: Climate Station Data from Northern Ellesmere Island in Nunavut, Canada, v. 1.7 (2002-2019). Nordicana D8, 2020a.

CEN: Climate Station Data from the Biscarat River Region in Nunavik, Quebec, Canada, v. 1.0 (2005-2019). Nordicana D62., 2020b.

CEN: Climate Station Data from the Clearwater Lake Region in Nunavik, Quebec, Canada, v. 1.1 (1986-2019). Nordicana D57, 2020c.

470 CEN: Climate Station Data from the Little Whale River Region in Nunavik, Quebec, Canada, v. 1.1 (1993-2019). Nordicana D58, 2020d.

CEN: Climate Station Data from the Sheldrake River Region in Nunavik, Quebec, Canada, v. 1.1 (1986-2019). Nordicana D61, 2020e.

CEN: Climate Station Data from Whapmagoostui-Kuujjuarapik Region in Nunavik, Quebec, Canada, v. 1.5 (1987-2019). Nordicana D4, 2020f.

CEN: Environmental Data from Boniface River Region in Nunavik, Quebec, Canada, v. 1.3 (1988-2019). Nordicana D7, 2020g.

475   Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An Improved Representation of Physical Permafrost Dynamics in the JULES Land-Surface Model, Geoscientific Model Development, 8, 1493–1508, https://doi.org/10.5194/gmd-8-1493-2015, 2015.

Chen, F., Mitchell, K., Schaake, Y., Xue, Y., Pan, H.-L., Koren, V., Duan, Q., Ek, M., and Betts, A.: Modeling of Land Surface Evaporation by Four Schemes and Comparison with FIFE Observations, Journal of Geophysical Research, 101, 7251–7268, 1996.

480   Chen, H., Nan, Z., Zhao, L., Ding, Y., Chen, J., and Pang, Q.: Noah Modelling of the Permafrost Distribution and Characteristics in the West Kunlun Area, Qinghai-Tibet Plateau, China: Noah Modelling of Permafrost, Permafrost and Periglacial Processes, 26, 160–174, https://doi.org/10.1002/ppp.1841, 2015.

Chen, L., Aalto, J., and Luoto, M.: Significant Shallow–Depth Soil Warming over Russia during the Past 40 Years, Global and Planetary Change, 197, 103 394, https://doi.org/10.1016/j.gloplacha.2020.103394, 2021.

485   Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., and Walker, J. P.: Upscaling Sparse Ground-Based Soil Moisture Observations for the Validation of Coarse-Resolution Satellite Soil Moisture Products: UPSCALING SOIL MOISTURE, Reviews of Geophysics, 50, https://doi.org/10.1029/2011RG000372, 2012.

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., and Isaksen, L.: A Simplified Extended Kalman Filter for the Global Operational Soil Moisture Analysis at ECMWF, Quarterly Journal of the Royal Meteorological Society, 139, 1199–1213,
490   https://doi.org/10.1002/qj.2023, 2013.

de Rosnay, P., Balsamo, G., Albergel, C., Muñoz-Sabater, J., and Isaksen, L.: Initialisation of Land Surface Variables for Numerical Weather Prediction, Surveys in Geophysics, 35, 607–621, 2014.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haim-
495   berger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dirmeyer, P. A., Koster, R. D., and Guo, Z.: Do Global Models Properly Represent the Feedback between Land and Atmosphere?, Journal
500   of Hydrometeorology, 7, 1177–1198, 2006.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for Improved Earth System Understanding: State-of-the Art and Future Directions, Remote Sensing of Environment, 203, 185–215,
505   https://doi.org/10.1016/j.rse.2017.07.001, 2017.

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, 24 823–24 838, https://doi.org/10.1029/2000JD900328, 2000.

Ek, M.: Implementation of Noah Land Surface Model Advances in the National Centers for Environmental Prediction Operational Mesoscale Eta Model, Journal of Geophysical Research, 108, https://doi.org/doi:10.1029/2002JD003296, 2003.

European Centre for Medium-Range Weather Forecasts: ERA-Interim Project, Monthly Means, 2012.

European Centre for Medium-Range Weather Forecasts: ERA5 Reanalysis (Monthly Mean 0.25 Degree Latitude-Longitude Grid), 2019.

Gao, Z., Chae, N., Kim, J., Hong, J., Choi, T., and Lee, H.: Modeling of Surface Energy Partitioning, Surface Temperature, and Soil Wetness in the Tibetan Prairie Using the Simple Biosphere Model 2 (SiB2): MODELING OF THE SURFACE PROCESSES, Journal of Geophysical Research: Atmospheres, 109, n/a–n/a, https://doi.org/10.1029/2003JD004089, 2004.

Gelaro, R., McCarty, W., Suárez, 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, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Global Modeling and Assimilation Office: MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, Greenbelt, MD, USA, 2015.

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., and Hudson, S. R.: Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer, Journal of Climate, 32, 4121–4143, https://doi.org/10.1175/JCLI-D-18-0643.1, 2019.

Gruber, A. and Scanlon, T.: Evolution of the ESA CCI Soil Moisture Climate Data Records and Their Underlying Merging Methodology, Earth System Science Data, 11, 717–739, https://doi.org/10.5194/essd-11-717-2019, 2019.

Gruber, S., Brown, N., Karunaratne, K., Kokelj, S. V., Stewart-Jones, E., and Subedi, R.: Air and Ground Temperature, Air Humidity and Site Characterization Data from the Canadian Shield Tundra near Lac de Gras, Northwest Territories, Canada, 2018.

Grünberg, I., Wilcox, E. J., Zwieback, S., Marsh, P., and Boike, J.: Linking Tundra Vegetation, Snow, Soil Temperature, and Permafrost, Biogeosciences, 17, 4261–4279, https://doi.org/10.5194/bg-17-4261-2020, 2020.

GTN-P: GTN-P Global Mean Annual Ground Temperature Data for Permafrost near the Depth of Zero Annual Amplitude (2007-2016), https://doi.org/10.1594/PANGAEA.884711, 2018.

Gubler, S., Fiddes, J., Keller, M., and Gruber, S.: Scale-Dependent Measurement and Analysis of Ground Surface Temperature Variability in Alpine Terrain, The Cryosphere, 5, 431–443, https://doi.org/10.5194/tc-5-431-2011, 2011.

Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., Onoda, H., Onogi, K., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: Representation of Atmospheric Circulation and Climate Variability, Journal of the Meteorological Society of Japan. Ser. II, 94, 269–302, https://doi.org/10.2151/jmsj.2016-015, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hinkel, K. M. and Outcalt, S. I.: Detection of Heat-Mass Transfer Regime Transitions in the Active Layer Using Fractal Geometric Parameters, Cold Regions Science and Technology, 23, 293–304, https://doi.org/10.1016/0165-232X(95)00003-T, 1995.

Hong, J. and Kim, J.: Numerical Study of Surface Energy Partitioning on the Tibetan Plateau: Comparative Analysis of Two Biosphere Models, Biogeosciences, 7, 557–568, 2010.

Hu, G., Zhao, L., Wu, X., Li, R., Wu, T., Xie, C., Pang, Q., and Zou, D.: Comparison of the Thermal Conductivity Parameterizations for a Freeze-Thaw Algorithm with a Multi-Layered Soil in Permafrost Regions, CATENA, 156, 244–251, https://doi.org/10.1016/j.catena.2017.04.011, 2017.

Hu, G., Zhao, L., Li, R., Wu, X., Wu, T., Xie, C., Zhu, X., and Su, Y.: Variations in Soil Temperature from 1980 to 2015 in Permafrost Regions on the Qinghai-Tibetan Plateau Based on Observed and Reanalysis Products, Geoderma, 337, 893–905, https://doi.org/10.1016/j.geoderma.2018.10.044, 2019.

Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., and Kuhry, P.: Estimated Stocks of Circumpolar Permafrost Carbon with Quantified Uncertainty Ranges and Identified Data Gaps, Biogeosciences, 11, 6573–6593, 2014.

Japan Meteorological Agency: JRA-55: Japanese 55-Year Reanalysis, Monthly Means and Variances., 2013.

Johannsen, Ermida, Martins, Trigo, Nogueira, and Dutra: Cold Bias of ERA5 Summertime Daily Maximum Land Surface Temperature over Iberian Peninsula, Remote Sensing, 11, 2570, https://doi.org/10.3390/rs11212570, 2019.

Jorgenson, M. T., Romanovsky, V., Harden, J., Shur, Y., O'Donnell, J., Schuur, E. A. G., Kanevskiy, M., and Marchenko, S.: Resilience and Vulnerability of Permafrost to Climate changeThis Article Is One of a Selection of Papers from The Dynamics of Change in Alaska's Boreal Forests: Resilience and Vulnerability in Response to Climate Warming., Canadian Journal of Forest Research, 40, 1219–1236, https://doi.org/10.1139/X10-060, 2010.

Kahan, D. S., Xue, Y., and Allen, S. J.: The Impact of Vegetation and Soil Parameters in Simulations of Surface Energy and Water Balance in the Semi-Arid Sahel: A Case Study Using SEBEX and HAPEX-Sahel Data, Journal of Hydrology, 320, 238–259, https://doi.org/10.1016/j.jhydrol.2005.07.011, 2006.

Kim, Y. and Wang, G.: Impact of Vegetation Feedback on the Response of Precipitation to Antecedent Soil Moisture Anomalies over North America, Journal of Hydrometeorology, 8, 534–550, https://doi.org/10.1175/JHM612.1, 2007.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Koren, V., Schaake, J., Mitchell, K., and Chen, F.: A Parameterization of Snowpack and Frozen Ground Intended for NCEP Weather and Climate Models, Journal of Geophysical Research: Atmospheres, 104, 19 569–19 585, 1999.

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, 24 809–24 822, https://doi.org/10.1029/2000JD900327, 2000.

Koster, R. D., Suarez, M. J., Liu, P., Jambor, U., Berg, A., Kistler, M., Reichle, R., Rodell, M., and Famiglietti, J.: Realistic Initialization of Land Surface States: Impacts on Subseasonal Forecast Skill, Journal of Hydrometeorology, 5, 1049–1063, https://doi.org/10.1175/JHM-387.1, 2004.

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: Permafrost Carbon-Climate Feedbacks Accelerate Global Warming, Proceedings of the National Academy of Sciences, 108, 14 769–14 774, https://doi.org/10.1073/pnas.1103910108, 2011.

Kropp, H., Loranty, M. M., Sannel, B., O'Donnell, J., and Blanc-Bates, E.: Synthesis of Soil-Air Temperature and Vegetation Measurements in the Pan-Arctic. 1990-2016., 2020.

585 Kropp, H., Loranty, M. M., Natali, S. M., Kholodov, A. L., Rocha, A. V., Myers-Smith, I., Abbot, B. W., Abermann, J., Blanc-Betes, E., Blok, D., Blume-Werry, G., Boike, J., Breen, A. L., Cahoon, S. M. P., Christiansen, C. T., Douglas, T. A., Epstein, H. E., Frost, G. V., Goeckede, M., Høye, T. T., Mamet, S. D., O'Donnell, J. A., Olefeldt, D., Phoenix, G. K., Salmon, V. G., Sannel, A. B. K., Smith, S. L., Sonnentag, O., Vaughn, L. S., Williams, M., Elberling, B., Gough, L., Hjort, J., Lafleur, P. M., Euskirchen, E. S., Heijmans, M. M., Humphreys, E. R., Iwata, H., Jones, B. M., Jorgenson, M. T., Grünberg, I., Kim, Y., Laundre, J., Mauritz, M., Michelsen, A., Schaepman-Strub, G., Tape,

590 K. D., Ueyama, M., Lee, B.-Y., Langley, K., and Lund, M.: Shallow Soils Are Warmer under Trees and Tall Shrubs across Arctic and Boreal Ecosystems, Environmental Research Letters, 16, 015 001, https://doi.org/10.1088/1748-9326/abc994, 2021.

Li, B., Beaudoing, H., Rodell, M., and NASA/GSFC/HSL: GLDAS Catchment Land Surface Model L4 Monthly 1.0 x 1.0 Degree V2.0, Greenbelt, Maryland, USA, 2020a.

Li, B., Beaudoing, H., Rodell, M., and NASA/GSFC/HSL: GLDAS Catchment Land Surface Model L4 Monthly 1.0 x 1.0 Degree V2.0,

595 Greenbelt, Maryland, USA, 2020b.

Li, M., Wu, P., and Ma, Z.: Comprehensive Evaluation of Soil Moisture and Soil Temperature from Third-generation Atmospheric and Land Reanalysis Datasets, International Journal of Climatology, p. joc.6549, https://doi.org/10.1002/joc.6549, 2020c.

Lindsay, R., Wensnahan, M., Schweiger, A., and Zhang, J.: Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic, Journal of Climate, 27, 2588–2606, https://doi.org/10.1175/JCLI-D-13-00014.1, 2014.

600 Loew, A. and Schlenz, F.: A Dynamic Approach for Evaluating Coarse Scale Satellite Soil Moisture Products, Hydrology and Earth System Sciences, 15, 75–90, https://doi.org/10.5194/hess-15-75-2011, 2011.

Luo, L., Robock, A., Vinnikov, K. Y., Schlosser, C. A., Slater, A. G., Boone, A., Etchevers, P., Habets, F., Noilhan, J., Braden, H., Cox, P., de Rosnay, P., Dickinson, R. E., Dai, Y., Zeng, Q.-C., Duan, Q., Schaake, J., Henderson-Sellers, A., Gedney, N., Gusev, Y. M., Nasonova, O. N., Kim, J., Kowalczyk, E., Mitchell, K., Pitman, A. J., Shmakin, A. B., Smirnova, T. G., Wetzel, P., Xue, Y., and Yang, Z.-L.: Effects

605 of Frozen Soil on Soil Temperature, Spring Infiltration, and Runoff: Results from the PILPS 2(d) Experiment at Valdai, Russia, Journal of Hydrometeorology, 4, 334–351, https://doi.org/10.1175/1525-7541(2003)4<334:EOFSOS>2.0.CO;2, 2003.

Ma, H., Zeng, J., Zhang, X., Fu, P., Zheng, D., Wigneron, J.-P., Chen, N., and Niyogi, D.: Evaluation of Six Satellite- and Model-Based Surface Soil Temperature Datasets Using Global Ground-Based Observations, Remote Sensing of Environment, 264, 112 605, https://doi.org/10.1016/j.rse.2021.112605, 2021.

610 Metcalfe, D. B., Hermans, T. D. G., Ahlstrand, J., Becker, M., Berggren, M., Björk, R. G., Björkman, M. P., Blok, D., Chaudhary, N., Chisholm, C., Classen, A. T., Hasselquist, N. J., Jonsson, M., Kristensen, J. A., Kumordzi, B. B., Lee, H., Mayor, J. R., Prevéy, J., Pantazatou, K., Rousk, J., Sponseller, R. A., Sundqvist, M. K., Tang, J., Uddling, J., Wallin, G., Zhang, W., Ahlström, A., Tenenbaum, D. E., and Abdi, A. M.: Patchy Field Sampling Biases Understanding of Climate Change Impacts across the Arctic, Nature Ecology & Evolution, 2, 1443–1448, https://doi.org/10.1038/s41559-018-0612-5, 2018.

615 Muñoz-Sabater, J.: ERA5-Land Monthly Averaged Data from 1981 to Present., 2019.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-82, 2021.

620 Mudryk, L. R., Derksen, C., Kushner, P. J., and Brown, R.: Characterization of Northern Hemisphere Snow Water Equivalent Datasets, 1981–2010, Journal of Climate, 28, 8037–8051, https://doi.org/10.1175/JCLI-D-15-0229.1, 2015.

Munkhjargal, M., Yadamsuren, G., Yamkhin, J., and Menzel, L.: Ground Surface Temperature Variability and Permafrost Distribution over Mountainous Terrain in Northern Mongolia, Arctic, Antarctic, and Alpine Research, 52, 13–26, https://doi.org/10.1080/15230430.2019.1704347, 2020.

625 Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 Reanalysis, Journal of the Meteorological Society of Japan. Ser. II, 85, 369–432, https://doi.org/10.2151/jmsj.85.369, 2007.

Qin, Y., Liu, W., Guo, Z., and Xue, S.: Spatial and Temporal Variations in Soil Temperatures over the Qinghai–Tibet Plateau from 1980 to 2017 Based on Reanalysis Products, Theoretical and Applied Climatology, 140, 1055–1069, https://doi.org/10.1007/s00704-020-03149-9, 630 2020.

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, https://doi.org/10.1175/JCLI-D-16-0570.1, 2017.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological 635 Society, 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Royer, A., Picard, G., Vargel, C., Langlois, A., Gouttevin, I., and Dumont, M.: Improved Simulation of Arctic Circumpolar Land Area Snow Properties and Soil Temperatures, Frontiers in Earth Science, 9, 685 140, https://doi.org/10.3389/feart.2021.685140, 2021.

Rui, H., Beaudoing, H., and Loeser, C.: README Document for NASA GLDAS Version 2 Data Products, README Document, National Aeronautics and Space Administration, Maryland, USA, 2018.

640 Saha, S. and Coauthors: NCEP Climate Forecast System Reanalysis (CFSR) Monthly Products, January 1979 to December 2010, 2010.

Saha, S. and Coauthors: NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products, 2012.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, 645 Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, 650 Journal of Climate, 27, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

Sato, N., Sellers, P., Randall, D., Schneider, E., Shukla, J., Kinter, III, J., Hou, Y.-T., and Albertazzi, E.: Effects of Implementing the Simple Biosphere Model in a General Circulation Model, Journal of the Atmospheric Sciences, 46, 2757–2782, https://doi.org/10.1175/1520-0469(1989)046<2757:EOITSB>2.0.CO;2, 1988.

Scharringa, M.: On the Representativeness of Soil Temperature Measurements, Agricultural Meteorology, 16, 263–276, 655 https://doi.org/10.1016/0002-1571(76)90046-7, 1976.

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

Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A.: A Simple Biosphere Model (SIB) for Use within General Circulation Models, Journal of the Atmospheric Sciences, 43, 505–531, https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2, 1986.

Serreze, M. C., Barrett, A. P., and Stroeve, J.: Recent Changes in Tropospheric Water Vapor over the Arctic as Assessed from Radiosondes and Atmospheric Reanalyses, Journal of Geophysical Research: Atmospheres, 117, n/a–n/a, https://doi.org/10.1029/2011JD017421, 2012.

Sheffield, J., Goteti, G., and Wood, E.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088–3111, 2006.

Sherstiukov, A.: Dataset of Daily Soil Temperature up to 320 Cm Depth Based on Meteorological Stations of Russian Federation [In Russian], Trudy VNIIGMI-MTsD, 176, 224–232, 2012.

Siqueira, M., Katul, G., and Porporato, A.: Soil Moisture Feedbacks on Convection Triggers: The Role of Soil–Plant Hydrodynamics, Journal of Hydrometeorology, 10, 96–112, https://doi.org/10.1175/2008JHM1027.1, 2009.

Smith, M. W. and Riseborough, D. W.: Climate and the Limits of Permafrost: A Zonal Analysis, Permafrost and Periglacial Processes, 13, 1–15, https://doi.org/10.1002/ppp.410, 2002.

Streletskiy, D., Biskaborn, B., Smith, S., Noetzli, J., Gonçalo, V., and Schoeneich, P.: Strategy and Implemantation Plan (2016-2020) For the Global Terrestrial Network for Permafrost (GTN-P), Tech. rep., The George Washington University, Washington, D.C., USA, 2017.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil Organic Carbon Pools in the Northern Circumpolar Permafrost Region: SOIL ORGANIC CARBON POOLS, Global Biogeochemical Cycles, 23, GB2023, https://doi.org/10.1029/2008GB003327, 2009.

Taylor, P., Hegyi, B., Boeke, R., and Boisvert, L.: On the Increasing Importance of Air-Sea Exchanges in a Thawing Arctic: A Review, Atmosphere, 9, 41, https://doi.org/10.3390/atmos9020041, 2018.

Tilley, J. S. and Lynch, A. H.: On the Applicability of Current Land Surface Schemes for Arctic Tundra: An Intercomparison Study, Journal of Geophysical Research, 103, 29 051–29 063, https://doi.org/doi:10.1029/1998JD200014, 1998.

Tjernström, M. and Graversen, R. G.: The Vertical Structure of the Lower Arctic Troposphere Analysed from Observations and the ERA-40 Reanalysis, Quarterly Journal of the Royal Meteorological Society, 135, 431–443, https://doi.org/10.1002/qj.380, 2009.

Urban, F. and Clow, G.: Climate and Active-Layer Data Acquired in the National Petroleum Reserve–Alaska and the Arctic National Wildlife Refuge, 1998–2015, Tech. Rep. Data Series, Department of the Interior. US Geological Survey, Reston, Virginia, USA, 2017.

Viterbo, P.: An Improved Land Surface Parametrization Scheme in the ECMWF Model and Its Validation, Technical Report 75, ECMWF, Reading, UK, 1995.

Viterbo, P. and Betts, A.: Impact on ECMWF Forecasts of Changes to the Albedo of the Boreal Forests in the Presence of Snow, Journal of Geophysical Research, 104, 27 803–27 810, 1999.

Way, R. G. and Lapalme, C. M.: Does Tall Vegetation Warm or Cool the Ground Surface? Constraining the Ground Thermal Impacts of Upright Vegetation in Northern Environments, Environmental Research Letters, 16, 054 077, https://doi.org/10.1088/1748-9326/abef31, 2021.

World Meteorological Organization: Guidelines on Ensemble Prediction Systems and Forecasting, Tech. Rep. WMO-No. 1091, World Meteorological Organization, Geneva, Switzerland, 2012.

Wu, X., Nan, Z., Zhao, S., Zhao, L., and Cheng, G.: Spatial Modeling of Permafrost Distribution and Properties on the Qinghai-Tibet Plateau, Permafrost and Periglacial Processes, 29, 86–99, https://doi.org/10.1002/ppp.1971, 2018.

695 Xia, Y., Ek, M., Sheffield, J., Livneh, B., Huang, M., Wei, H., Feng, S., Luo, L., Meng, J., and Wood, E.: Validation of Noah-Simulated Soil Temperature in the North American Land Data Assimilation System Phase 2, Journal of Applied Meteorology and Climatology, 52, 455–471, https://doi.org/10.1175/JAMC-D-12-033.1, 2013.

Xiao, Y., Zhao, L., Dai, Y., Li, R., Pang, Q., and Yao, J.: Representing Permafrost Properties in CoLM for the Qinghai-Xizang (Tibetan) Plateau, Cold Regions Science and Technology, 87, 68–77, 2013.

700 Xu, W., Sun, C., Zuo, J., Ma, Z., Li, W., and Yang, S.: Homogenization of Monthly Ground Surface Temperature in China during 1961–2016 and Performances of GLDAS Reanalysis Products, Journal of Climate, 32, 1121–1135, https://doi.org/10.1175/JCLI-D-18-0275.1, 2019.

Xue, Y., Vasic, R., Janjić, Z., Liu, Y., and Chu, P.: The Impact of Spring Subsurface Soil Temperature Anomaly in the Western U.S. on North American Summer Precipitation: A Case Study Using Regional Climate Model Downscaling, Journal of Geophysical Research: Atmospheres, 117, 2012.

705 Yang, K. and Zhang, J.: Evaluation of Reanalysis Datasets against Observational Soil Temperature Data over China, Climate Dynamics, 50, 317–337, https://doi.org/10.1007/s00382-017-3610-4, 2018.

Yang, S., Li, R., Wu, T., Hu, G., Xiao, Y., Du, Y., Zhu, X., Ni, J., Ma, J., Zhang, Y., Shi, J., and Qiao, Y.: Evaluation of Reanalysis Soil Temperature and Soil Moisture Products in Permafrost Regions on the Qinghai-Tibetan Plateau, Geoderma, 377, 114583, https://doi.org/10.1016/j.geoderma.2020.114583, 2020.

710 Yi, Y., Kimball, J. S., Chen, R. H., Moghaddam, M., and Miller, C. E.: Sensitivity of Active-Layer Freezing Process to Snow Cover in Arctic Alaska, The Cryosphere, 13, 197–218, https://doi.org/10.5194/tc-13-197-2019, 2019.

Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J., and Zou, P.: Evaluation of Remotely Sensed and Reanalysis Soil Moisture Products over the Tibetan Plateau Using In-Situ Observations, Remote Sensing of Environment, 163, 91–110, https://doi.org/10.1016/j.rse.2015.03.008, 2015.

Zhan, M.-j., Xia, L., Zhan, L., and Wang, Y.: Evaluation and Analysis of Soil Temperature Data over Poyang Lake Basin, China, Advances
715 in Meteorology, 2020, 1–11, https://doi.org/10.1155/2020/8839111, 2020.

Zhang, J., Wang, W.-C., and Wei, J.: Assessing Land-Atmosphere Coupling Using Soil Moisture from the Global Land Data Assimilation System and Observational Precipitation, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD009807, https://doi.org/10.1029/2008JD009807, 2008.

Zhang, T.: Influence of the Seasonal Snow Cover on the Ground Thermal Regime: An Overview, Reviews of Geophysics, 43, RG4002,
720 https://doi.org/10.1029/2004RG000157, 2005.

Zhang, Y., Touzi, R., Feng, W., Hong, G., Lantz, T. C., and Kokelj, S. V.: Landscape-scale Variations in Near-surface Soil Temperature and Active-layer Thickness: Implications for High-resolution Permafrost Mapping, Permafrost and Periglacial Processes, 32, 627–640, https://doi.org/10.1002/ppp.2104, 2021.

Zhao, T., Guo, W., and Fu, C.: Calibrating and Evaluating Reanalysis Surface Temperature Error by Topographic Correction, Journal of
725 Climate, 21, 1440–1446, 2008.