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 such as the Arctic where observational data are limited. Soil temperatures are an important control of many land-atmosphere exchanges and hydrological processes, and permafrost soils are thawing as the climate warms. 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 limited there. Here we validate pan-Arctic soil temperatures from eight reanalysis and land data assimilation system products, using a newly-assembled database of in situ observations from diverse measurement networks across Eurasia and North America. We examine product performance across the extratropical northern hemisphere between 1982 and 2018, and find that most products have soil temperatures that are biased cold by 1-5 K, with an RMSE of 2-9 K, 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 ($r = 0.55$ - 0.85) in 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 are subject to much higher uncertainty when the soil is frozen and/or when the ground is snow-covered, suggesting that the representation of processes controlling snow cover in reanalysis systems should be urgently studied. We also validate the ensemble mean of all available 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.

1 Introduction

Soil temperatures, both near the surface, and at depth, are an important control of many physical, hydrological, and land surface processes, as soils act as a reservoir for energy and moisture underground. They provide 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 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., 2011). Soils over large portions of

the Arctic are perennially frozen (permafrost soil). Roughly 1400 to 1600 gigatonnes of carbon (GtC) is estimated to be stored in soils in permafrost affected regions of the Northern Hemisphere (Hugelius et al., 2014). Continued warming, and thawing of permafrost soils, and related decomposition of carbon could act as a potential positive feedback on warming, by releasing more methane (CH_4) and carbon dioxide (CO_2) into the atmosphere (Koven et al., 2011). 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 studies validating reanalysis soil temperature have primarily focused on the middle latitudes, such as across China (Yang and Zhang, 2018; Zhan et al., 2020; Zhao et al., 2022), the Qinqhai-Tibetan Plateau (Hu et al., 2019; Jiao et al., 2023; 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., 2020). Relative to in situ ground temperature probe networks, most reanalysis products are biased cold by about 2° C - 5° C, on average (Hu 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., 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; Cao et al., 2022).

While soil temperature biases in individual reanalysis products may limit their utility, a consensus is emerging that multi-product ensembles, based on the same principle as ensemble weather prediction (World Meteorological Organization, 2012), 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 et al., 2019), precipitation (Beck et al., 2019), as well as for local scale permafrost simulations (Cao et al., 2019). Li et al. (2020) suggest that a similar method could be used to reduce biases in reanalysis soil temperatures.

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., 2020; 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 soil temperature products in terms of their bias, RMSE, correlation and standard deviation, 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 and 2 outline the six reanalysis and two LDAS soil temperature products used in this study. For simplicity, the term "reanalysis" will hereafter be used to describe both reanalysis and LDAS products. A summary of each product follows below. Products were remapped onto the European Reanalysis - Interim (ERA-Interim) 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.

The reanalysis products investigated span a wide range of horizontal resolutions, ranging between 0.1°, in the case of ERA5-Land and FLDAS, to 0.75° for ERA-Interim (Table 1). Most products - CFSR, ERA-Interim, ERA5, ERA5-Land, and the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) simulate soil temperature across 4 vertical layers, while MERRA2 includes 6 vertical layers, and JRA55 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 2).

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 FLDAS. 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). FLDAS, however, is run in an offline mode, utilizing meteorological forcing from MERRA2 (McNally et al., 2017), and rainfall information from NOAA's Global Data Assimilation (GDAS) (Derber et al., 1991), the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) (Funk et al., 2015), and the African Rainfall Estimation version 2.0 (RFE2) (Xie and Arkin, 1997).

ERA-Interim, ERA5 and ERA5-Land use versions of the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) 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) (Balsamo et al., 2009). In ERA5, a weak coupling exists between the land surface and atmosphere. It includes 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 1 metre 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 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).

Table 1. Summary of the 8 reanalysis and LDAS products, their equatorial resolution, land model, and relevant references.

Product	Data Period	Resolution	Land Model	References
CFSR	1979 - 2010	0.31° x 0.31°	Noah LSM	Saha et al. (2010)
CFSv2	2011 - Present	0.2° x 0.2°	Noah LSM	Saha et al. (2014)
ERA5	1940 - Present	0.25° x 0.25°	HTESSEL	Hersbach et al. (2020)
ERA5-Land	1950 - Present	0.1° x 0.1°	HTESSEL	Muñoz-Sabater et al. (2021)
ERA-Interim	1979 - Aug 2019	0.75° x 0.75°	TESSEL	Dee et al. (2011)
FLDAS	1982 - Present	0.1° x 0.1°	Noah LSM	McNally et al. (2017)
JRA55	1956 - Present	0.56° x 0.56°	Simple Biosphere Model	Harada et al. (2016) Kobayashi et al. (2015)
MERRA2	1980 - Present	0.5° x 0.625°	Catchment LSM	Gelaro et al. (2017)

Table 2. Summary of the 8 reanalysis and LDAS products and the number and depths of the soil layers included. *The JRA55 Simple Biosphere Model contains up to three soil layers (whose depths vary depending on vegetation type), but the soil temperature is averaged over all layers to produce a singular value at each grid cell.

Product	Soil Layers	Soil Depths (in cm)
CFSR	4	0 - 10, 10 - 40, 40 - 100, 100 - 200
CFSv2	4	0 - 10, 10 - 40, 40 - 100, 100 - 200
ERA5	4	0 - 7, 7 - 28, 28 - 100, 100 - 289
ERA5-Land	4	0 - 7, 7 - 28, 28 - 100, 100 - 289
ERA-Interim	4	0 - 7, 7 - 28, 28 - 100, 100 - 289
FLDAS	4	0 - 10, 10 - 40, 40 - 100, 100 - 200
JRA55	3*	temperature averaged over soil column
MERRA2	6	0 - 9.88, 9.88 - 29.4, 29.4 - 67.99, 67.99 - 144.25, 144.25 - 294.96, 294.96 - 1294.96

MERRA2 utilizes the Catchment Land Surface Model (CLSM) (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 version of the Climate Prediction Center unified gauge-based analysis of global daily precipitation (CPCU) precipitation correction algorithm that originated in MERRA-Land (Reichle et al., 2017b). No corrections are available, however, for high latitude regions north of 62.5° N (Reichle et al., 2017a). Finally, JRA55 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 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 a soil temperature record that is temporally and spatially sparse (Luo et al., 2020). While Russia has a more complete record of permafrost temperatures extending back to the 1980s (Sherstiukov, 2012), longer term permafrost records over North America are generally limited to the western Arctic (Smith et al., 2010). Portions of coastal Nunavik, in Northern Quebec, have records of permafrost temperatures from the 1990s onwards (CEN, 2020a, b, c, d, e, f, g), while soil temperature measurements in the Central Arctic are rather sparse (Smith et al., 2010). Rather than limit our validation to a small geographic region in the permafrost zone, as several prior studies have done (Hu et al., 2019; Qin et al., 2020; Wu et al., 2018; Ma et al., 2021; Li et al., 2020), we choose to combine data from a variety of sparse and dense 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).

The study compiles a comprehensive set of in situ soil temperature measurements, approximately 1700 stations in total, from across extratropical Eurasia and North America (Table 3, and Supplemental Metadata). Incorporating data from multiple diverse sparse networks, the dataset includes data from the Yukon Geological Survey (Yukon Geological Survey, 2021), the Northwest Territories (Cameron et al., 2019; Ensom et al., 2020; Gruber et al., 2019; Spence and Hedstrom, 2018a, b; Street et al., 2018), 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 reanalysis soil temperatures at high latitudes with their performance in regions outside the permafrost zone. These include stations from Kropp et al. (2020), Sherstiukov (2012), and GTN-P (2018), as well as locations from the Manitoba Mesonet network (RoTimi Ojo and Manaigre, 2021), the Michigan Enviro-weather Network (MAWN) (Enviro-weather, 2022), the North Dakota Mesonet Network (NDAWN) (North Dakota Mesonet Network, 2022), and the Alberta Climate Information Service (ACIS) network (Alberta Agriculture, Forestry and Rural Economic Development, 2022). Data is also sourced from a peatland ecosystem in Metro Vancouver (Lee et al., 2017), and several locations in central and Northern BC (Déry, 2017; Hernández-Henríquez et al., 2018; Morris et al., 2021). This provides a unique baseline upon which to perform a hemispheric wide assessment of soil temperature in reanalysis and LDAS systems, and to the authors' knowledge, presents the most comprehensive analysis to date of soil temperatures across Canada and the Great Lakes basin.

2.3 Collocation of Station and reanalysis Data

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. In order

Table 3. Summary of the observational data networks included in the study, including the dataset name, number of stations included, and their references. Note that Nordicana D references are listed by site in the supplemental metadata.

Dataset	Number of Sites	Reference
GTN-P	68	(GTN-P, 2018)
Heather Kropp	229	(Kropp et al., 2020)
Roshydromet Network	458	(Sherstiukov, 2012)
Nordicana D	34	See supplemental station metadata
NWT Open Report 2017-009	73	(Cameron et al., 2019)
NWT Open Report 2018-009	214	(Gruber et al., 2019)
NWT Open Report 2019-004	9	(Ensom et al., 2020)
NWT Open Report 2019-017	31	(Rudy et al., 2020)
Street and Wookey (2016)	5	(Street and Wookey, 2016)
Yukon Permafrost Database	112	(Yukon Geological Survey, 2021)
Baker Creek	6	(Spence and Hedstrom, 2018b)
Cariboo Alpine MesoNET	12	(Hernández-Henríquez et al., 2018) (Déry, 2017) (Morris et al., 2021)
Burns Bog	1	(Lee et al., 2017)
Manitoba Mesoscale Network	85	(RoTimi Ojo and Manaigre, 2021)
Enviro-Weather Network	75	(Enviro-weather, 2022)
ACIS	31	(Alberta Agriculture, Forestry and Rural Economic Development, 2022)
NDAWN	150	(North Dakota Mesonet Network, 2022)

to maximize the amount of observational data available, layer-averaged soil temperatures were calculated at each timestep with all available data. This tradeoff meant that layer averages often included a different number of depths at different timesteps, and as such, we needed to limit our analysis of soil temperature trends and variability to locations where layer averages had a consistent number of depths.

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, in order to focus on processes controlling the seasonal cycle of soil temperatures. As such, we use monthly averages of soil temperatures for validation purposes. Outlier observations with anomalies greater than $\pm$ 3.5σ were removed before monthly averaging.

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 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 increase sample size, we therefore included all months with at least 50 percent valid 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

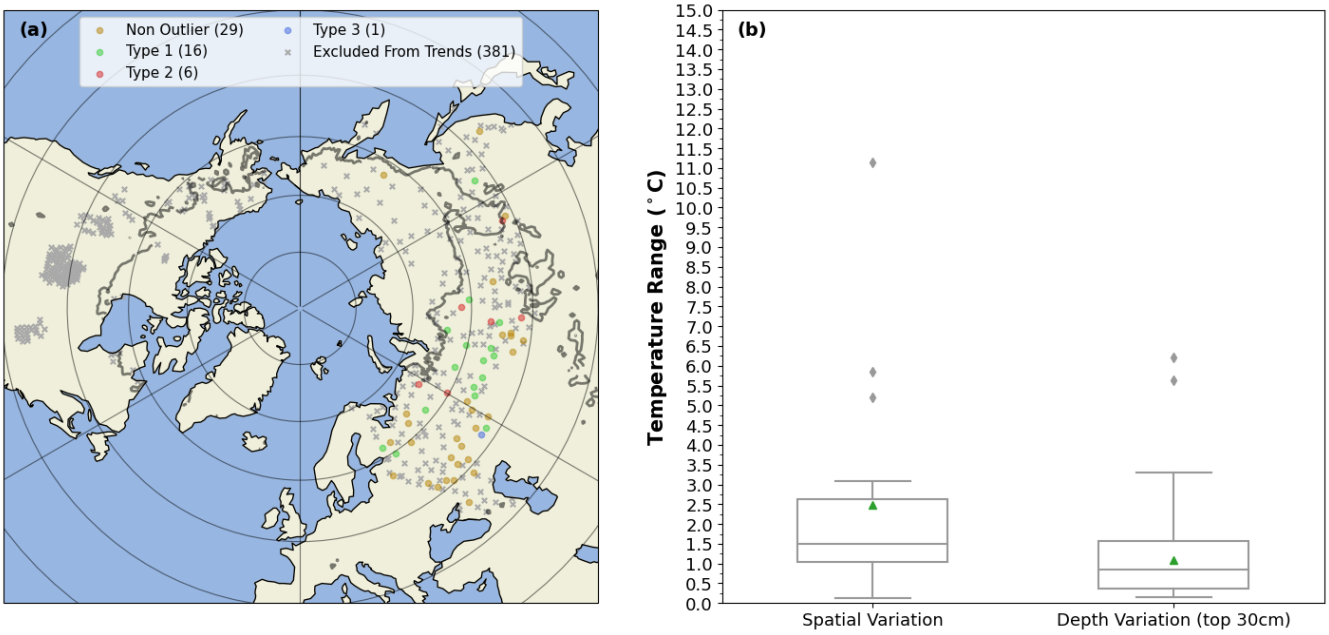

Figure 1. (a): location of the validation grid cells collocated with in situ stations in the near-surface layer. Grid cells excluded from the soil temperature trends analysis are shown as an "x". Type 1 refers to grid cells that where the ensemble mean simulates a winter minimum soil temperature that is too cold. Type 2 refers to grid cells where the ensemble mean simulates a summer maximum soil temperature that is too cold. Type 3 refers to grid cells where the ensemble mean underestimates the seasonal cycle of soil temperatures. The number in brackets beside each legend entry displays the number of grid cells in each category. The contour line encircles regions where the Obu et al. (2018) permafrost cover is at least 50 percent. (b): Impact of spatial variation and depth variation on the spread of soil temperatures in a grid cell. The mean is shown by a green triangle, and outliers are shown as grey diamonds.

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, grid cells contained a single in situ measurement location. In North America, however, a number of the grid cells contain two or more in situ stations. The near surface layer layer includes 430 validation grid cells (Figure 1, panel A), while at depth, there are 377 grid cells (not shown). A subset of stations with longer timeseries and a more complete data record are used to calculate soil temperature trends (Section 4.2). Stations included in the soil temperature trend and variability analysis are shown as circles of varying size and colour, while those excluded from the soil temperature trend and variability analysis are shown as an x (Figure 1, Panel A). The details of Figure 1 - Panel A will be described further in Section 5.3.

To calculate spatial averages, a simple average of (layer-averaged) soil temperatures from all stations within the bounds of a particular grid cell was calculated at each timestep, using all available stations. This meant that the number of stations included at each timestep was not always consistent, and the analysis of soil temperature trends was limited to a subset of 52 grid cells in Eurasia where the following conditions were met:

1. The timeseries included data between Jan 1985 and Dec 2010, with no missing data.

2. The number of stations included in the spatially averaged grid cell temperature was consistent over all timesteps.

3. The number of depths included in the layer averaged soil temperature of each contributing station remained consistent over all timesteps.

As a result North American grid cells were excluded from the soil temperature trends analysis, and the analysis is based on grid cells from Eurasia (where grid cells often only contained a single station) (Figure 1, Panel A). Using a subset of grid cells that incorporate multiple stations in the spatial average, and include a consistent number of stations and depths in the timeseries, we quantify the variability in soil temperatures between stations within a grid cell, and across depths within a layer average. It was found that the median temperature range between stations within a grid cell was approximately 1.5° C, roughly 1.75 times larger than the median temperature range across depths within the near surface layer of a station (Figure 1, Panel B), suggesting that temperature variability within a grid cell is substantially larger than variations in temperatures within the near surface layer of a particular station.

3 Methods

3.1 Validation Metrics

Reanalysis 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 (σ_{norm}) (Eq. 3 and Eq. 4), and the Pearson correlation (R) (Eq. 5). We also include an overall skill score for each model; a Thackeray et al. (2015) type formulation of the Taylor (2001) skill score (SS) (Eq. 6). Statistical metrics were calculated as follows:

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

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

$$\sigma = \sqrt{\frac{\sum_{n=1}^{N}(x_n - \overline{x})^2}{N-1}} \tag{3}$$

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

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

$$SS = \frac{2(1+R)}{(\sigma_{norm} + \frac{1}{\sigma_{norm}})^2} \tag{6}$$

Where T_p is the T_{soil} from the reanalysis product, and T_i is the T_{soil} of the in situ data. $\overline{T_p}$ and $\overline{T_i}$ refer to the mean T_{soil} of the reanalysis product and in situ data, respectively, while N is the number of monthly soil temperature values. σ_{T_p} and σ_{T_i} are the standard deviations of the reanalysis product soil temperatures and in situ soil temperatures, respectively. Finally, *x* refers to the T_{soil} (from a particular timestep in a dataset), and $\overline{x}$ is the mean T_{soil} of the dataset.

Metrics were calculated separately for each individual grid cell, and then averaged to obtain regional values. Estimates for the permafrost zone and the zone with little to no permafrost were also calculated by averaging together metrics from grid cells falling within a particular zone. Skill scores were calculated separately for the near surface, and depth, while the "overall" skill score represents an average of the near surface and depth skill scores.

3.2 Binning of Datasets by Season and Permafrost

Datasets were binned into a cold season and warm season using the University of East Anglia's Climatic Research Unit (CRU) TS version 4.07 2m air temperature (T_{air}) (Harris et al., 2020) for each grid cell. Cold season months are those where $T_{air} \leq$ -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 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.

Permafrost zonation was estimated using the Obu et al. (2018) permafrost map, which employs a temperature at the top of the permafrost (TTOP) model based on a 2000-2016 climatology, driven by a combination of remotely sensed land surface temperatures, downscaled atmospheric data from ERA-Interim, and landcover information from The European Space Agency (ESA) Climate Change Initiative (CCI) (Obu et al., 2019). To 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', which includes all regions with < 50% permafrost cover.

Table 4. Number of grid cells in each elevation bin for the near surface and at depth.

Elevation Range	Near Surface Grid Cells	Depth Grid Cells
Below 500 m	310	275
500 to 1000 m	105	87
1000+ m	15	15

3.3 Elevation Impacts

The authors examined the potential impacts of elevation differences between in situ datasets and reanalysis products by estimating the station elevation using the 90 m Copernicus Global Digital Elevation Model (GLO-90) (European Space Agency, 2021) and obtaining reanalysis elevations at their native resolution for the nearest grid cell to the station. For grid cells with more than one station, station elevations were averaged together to obtain a grid cell estimate of the "station" elevation.

Grid cells were grouped into three elevation bins based on the station elevation, and it was found that over 70% of grid cells are located in regions where the in situ station(s) are below 500 m (Table 4). Only 15 grid cells had station elevations above 1000 m, so the authors grouped all grid cells at or above 500 m together for the purposes of validation. While reanalysis products generally underestimated the elevation of higher elevation station, with an average RMSE of between 144 m and 589 m (not shown), this did not appear to have a major impact on soil temperature performance. Readers are referred to Section 4.3 (Spatial Variability) for more details.

3.4 Regridding of Reanalysis Products and Calculation of Ensemble Mean Soil Temperature

All products were first re-gridded to the ERA-Interim grid using a first-order conservative remapping technique (Jones, 1999). The near surface soil layers, which are representative of the top 30 cm of soil, were calculated as a simple average of the top 2 soil layers in each reanalysis product. For JRA55, the single soil layer was used as a near surface estimate. Soil temperatures at depth, for each product, were calculated as a simple average of all layers which fell between 30cm and 300cm, beginning with the third soil layer. For JRA55, the single averaged soil layer was used as an estimate of soil temperatures at depth. Readers are referred to Table 2 for further information about product soil layers. While the vertical discretization is coarser than that of the individual products, this approach allows the ensemble mean product to incorporate soil temperatures from products with different land models, whose vertical resolution is not constant.

After the near surface and deep soil layer average temperatures were calculated for each product, the ensemble mean soil temperature, for each layer, was calculated as the unweighted arithmetic mean from six individual soil temperature products (CFSR, ERA-Interim, ERA5, ERA5-Land, FLDAS and MERRA2). Owing to JRA55's simplified land model, which is unable to capture near surface soil temperatures, we decided to exclude it from the ensemble mean product, as its inclusion dramatically increased the bias and RMSE of the ensemble mean.

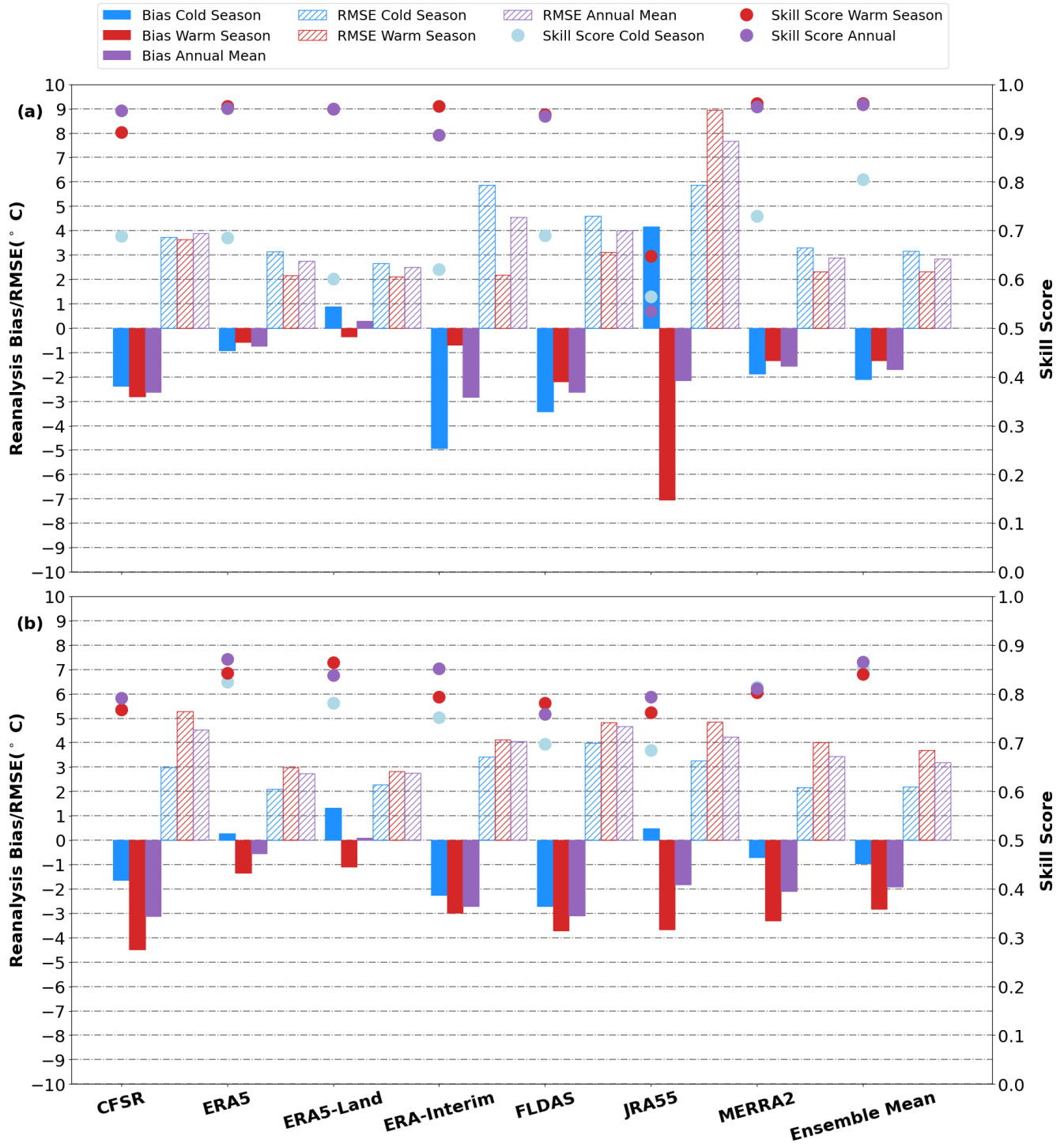

Figure 2. Bias (solid color), RMSE (hatching) and skill scores (circles) of each product the cold season (blue) ($\leq$ -2° C) and the warm season (red) (> -2° C) performance of reanalysis products. The skill score is also shown over the annual cycle (purple). (a) displays the bias, RMSE and skill score for the near surface (0 cm to 30 cm) layer, while (b) displays the bias, RMSE and skill score at depth (30 cm to 300 cm). The ensemble mean is shown beside for comparison.

4 Validation of Reanalysis Products

4.1 Annual Mean

The Taylor Skill score ranges between a minimum of zero and a theoretical maximum of one. A product with a skill score of 1.0 would display a perfect correlation of 1.0 relative to in-situ soil temperatures, and a soil temperature variance identical to that of the in-situ data. Near the surface, most products display relatively high annual mean skill scores of > 0.9, suggesting that they generally capture the overall seasonal cycle of soil temperatures. At depth, skill scores are typically about 0.1 lower (Figure 2, Table S1 and S2), arising from the lag between air temperatures and soil temperatures at depth. JRA55, however displays an annual mean skill score of 0.54 near the surface, and skill score of 0.79 at depth (Figure 2).This arises because JRA55 uses a simplified land model that uses just a single vertical layer; the soil temperatures used are computed as averages over the soil column that are, therefore, more similar to deeper soil layers than to the surface. Consequently, JRA55 underestimates the seasonal cycle of observed soil temperatures in the near- surface, and the timing of its annual maximum and minimum soil temperatures are offset by roughly a month relative to other products (not shown).

Most products are biased cold over the annual mean, and display soil temperature biases of between +0.3° C and -3.1° C, with RMSE values ranging between 2.5° C and 7.7° C over the extratropical NH (Figure 2, Table S1 and S2). ERA5-Land is a notable exception, as it shows a warm bias over the annual mean, driven by warm biases in winter; a factor that will be discussed further in Section 4.2 and Section 6.1.

4.2 Seasonal Cycle

As alluded to above, strong seasonal differences exist in reanalysis performance, with noticeably lower skill scores in the cold season. This is particularly true near the surface, where cold season skill scores are 0.08 to 0.35 lower than their warm season counterparts. Skill score declines at depth are reduced, but still show a decline of 0.02 to 0.08 relative to the warm season (Figure 2, Table S1 and S2).

The decline in skill score is mirrored by increases in near surface bias and RMSE for several products. This is particularly true for ERA-Interim, whose bias and RMSE are 4.1° C, and 3.7° C larger, respectively. Interestingly, biases for all products are somewhat larger in the warm season at depth, though seasonal differences are also generally smaller in the deeper soil layers (Figure 2). JRA55 and ERA5-Land are noticeable outliers to this trend, as they both exhibit positive (warm) biases in the cold season. In the case of JRA55, this is due to an underestimation of the seasonal cycle of temperatures, while for ERA5-Land, it is likely due to issues with snow cover properties (which will be discussed further in Section 6.1).

Near the surface, most products show a maximum cold (negative) bias when soil temperatures are between -2° C to -10° C, and there is a tendency for the biases of most products to decrease or flip sign over the coldest temperatures (Figure 3), suggesting that they may underestimate the coldest in-situ temperatures. At depth, however, most products display larger cold biases over warmer temperatures (Figure 3). JRA55 displays a maximum cold (negative) bias over the warmest temperatures, and the bias flips sign as temperatures approach freezing; linked to its reduced seasonal cycle (Figure 3). In the case of ERA-

Interim, the largest cold (negative) biases are found over the coldest temperatures; likely linked to issues with its snow cover representation (discussed in Section 6.1).

Most products generally capture the observed soil temperature variance during the warm season. This is evident in Figure 4, Panels B and D, as the normalized standard deviation is within 25% of the observed for all products. This is contrasted by the cold season, where several products overestimate soil temperature variability (particularly at depth), contributing to a decline in product skill (Figure 4, Panels A and C). For example, ERA5-Land's (blue diamond) cold season skill is impacted by its underestimation of cold season soil temperature variability (which is roughly half of the near surface observed variance), and arises in part because of its warm (positive) bias in winter (Figure 2).

The overestimation in soil temperature variance is also apparent in Figure 5 as products display larger soil temperature variances (as measured by their standard deviation) for a given in-situ soil temperature over colder temperatures. The spread in standard deviation between products (similar to their biases), is also generally largest over colder temperatures. The reduced standard deviation near the surface in the -32° C bin is likely a function of the small sample size (11).

Similar seasonal variation is seen in reanalysis soil temperature correlations (against station data), as most products show warm season correlations of greater than 0.93 near the surface, with reductions of between 0.20 and 0.39 for most products over the cold season. At depth, warm season correlations are generally 0.09 to 0.18 lower than near the surface, and seasonal differences are much smaller (Figure 4). The poor JRA55 correlation near the surface arises from its mismatched seasonal cycle.

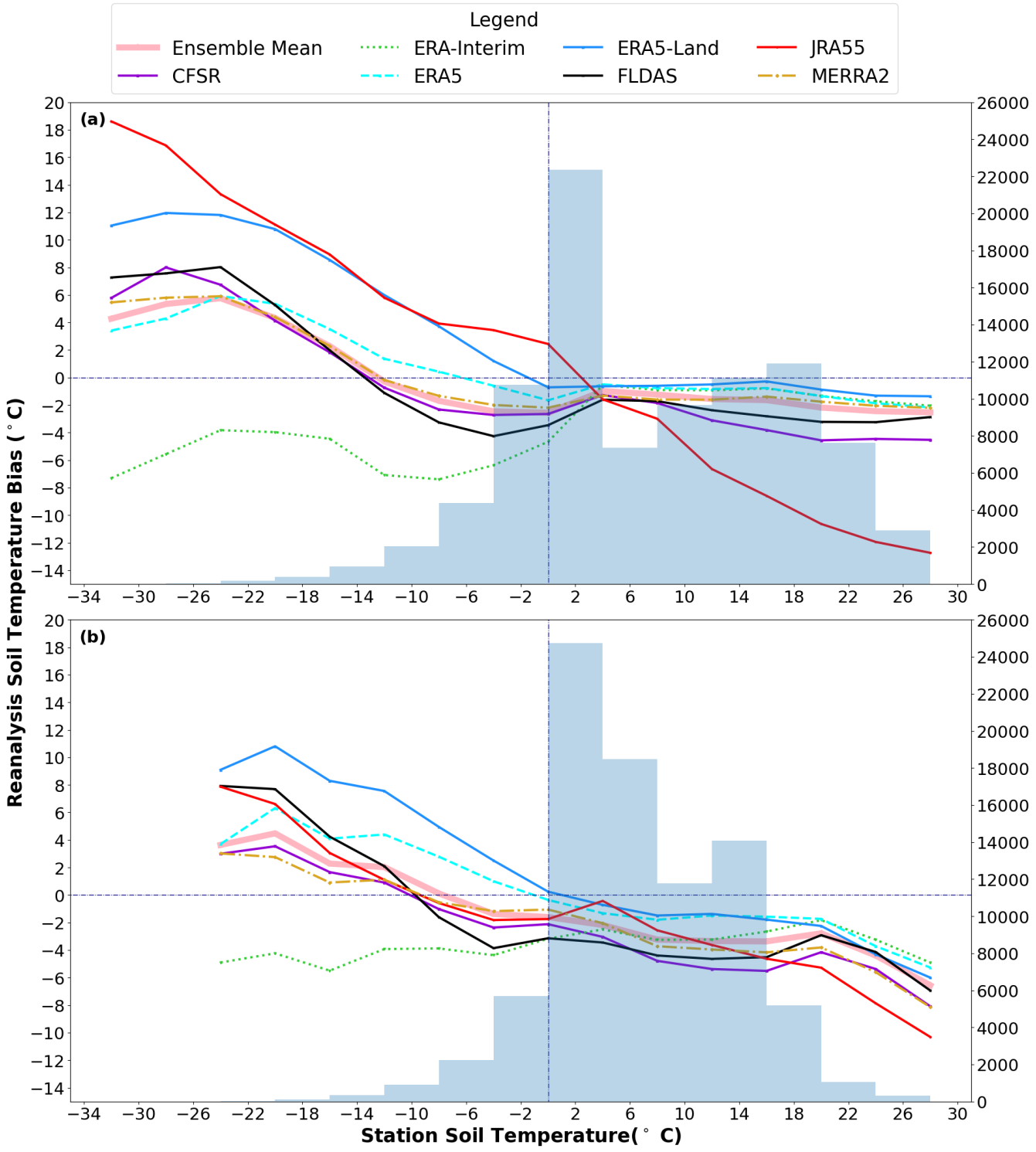

Figure 3. Reanalysis soil temperature bias as a function of station soil temperature for (a) the near surface (0 cm to 30 cm) layer, and (b) at depth (30 cm to 300 cm). Station temperatures are binned into 4° C intervals, beginning with the -32° C to -28° C bin, and ending with the 26° C to 30° C bin. The midpoint of each temperature bin is plotted along the x-axis. The secondary y-axis displays the number of datapoints in each bin (in conjunction with the histogram).

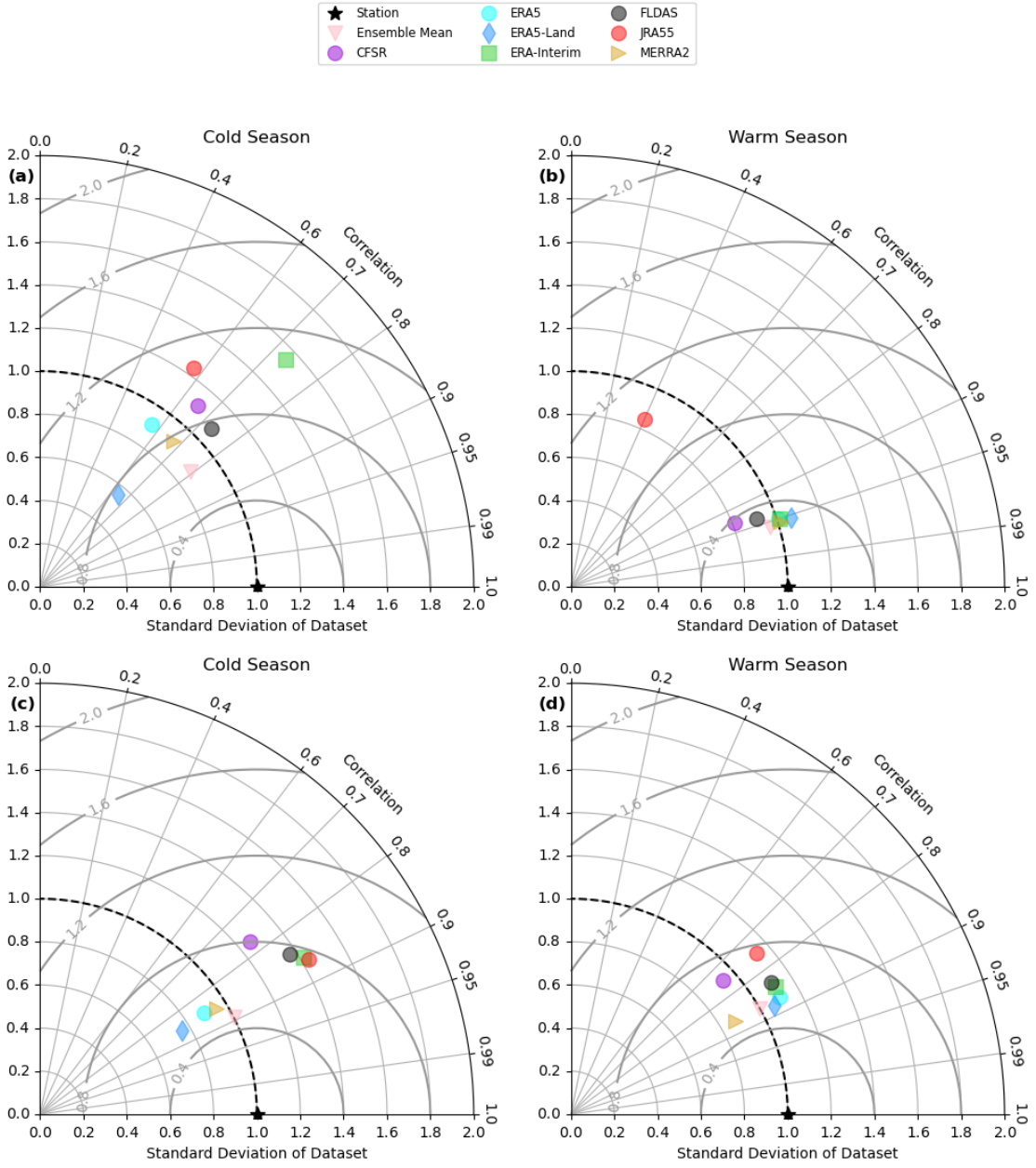

Figure 4. Taylor Diagram of the cold season and the warm season performance of reanalysis products. Panels A and B refer to the cold season, while panels C and D refer to the warm season. The top panels, (a) and (c) are for the near surface while the bottom panels, (b) and (d) refer to soil temperatures at depth. The concentric rings (solid grey lines) refer to the centralized root mean square error (CRMSE), and a product would have a CRMSE of zero if the timeseries of the reanalysis matched the station data perfectly; with a normalized standard deviation of one, and a correlation of one.

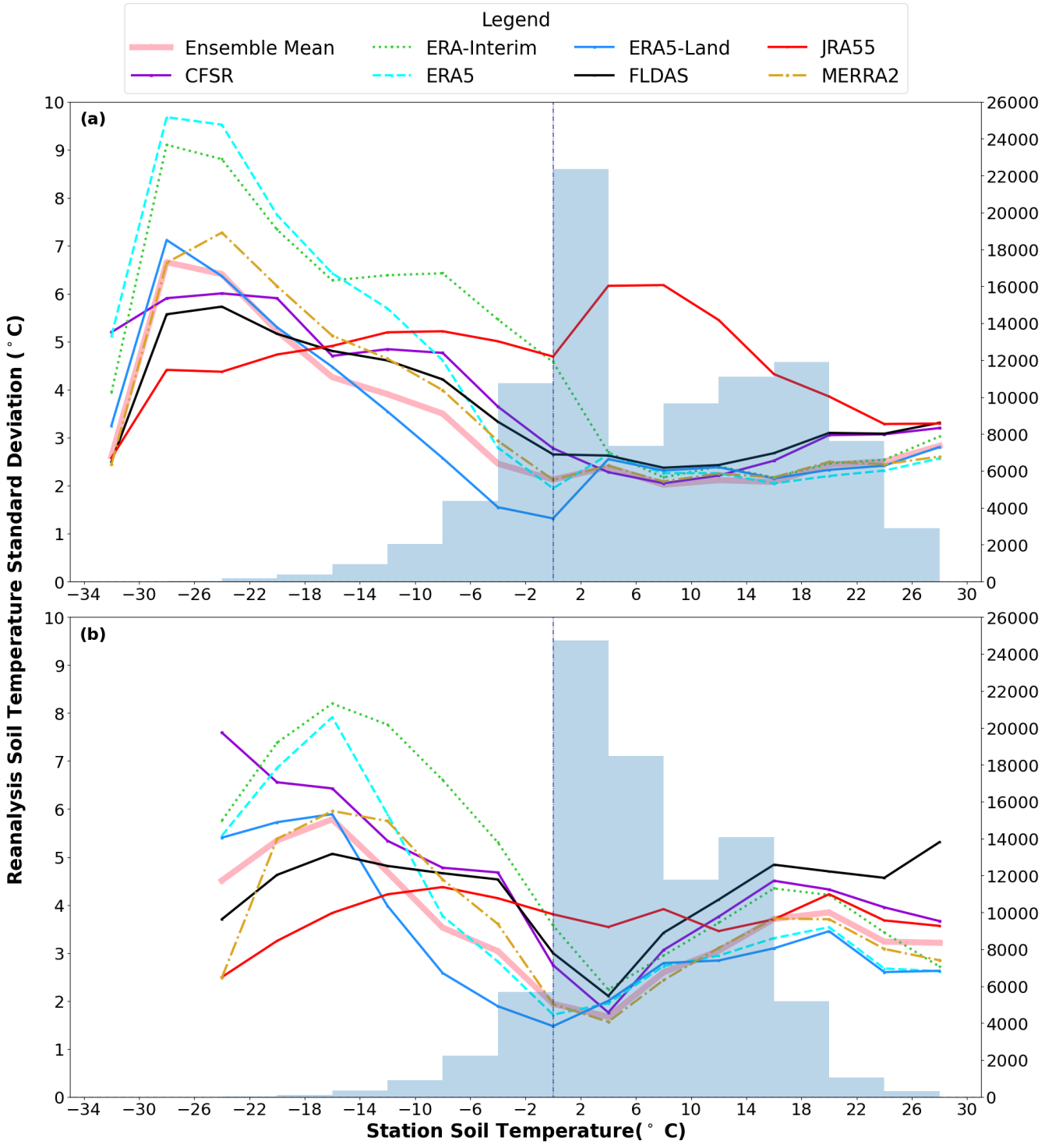

Figure 5. Reanalysis soil temperature standard deviation as a function of station soil temperature for (a) the near surface (0 cm to 30 cm) layer, and (b) at depth (30 cm to 300 cm). Station temperatures are binned into 4° C intervals, beginning with the -32° C to -28° C bin, and ending with the 26° C to 30° C bin. The midpoint of each temperature bin is plotted along the x-axis. The secondary y-axis displays the number of datapoints in each bin (in conjunction with the histogram).

4.3 Spatial Variability

Similar to the strong seasonal differences in product performance, we also see substantial spatial variability in performance that is strongly linked to latitude. Over the peramfrost regions, where snow cover lasts longer, soil temperature performance is typically worse than over regions further south. Near the surface, skill scores over permafrost regions are typically 0.05 - 0.1 lower over the annual mean than in regions with little to no permafrost, and at depth, they are generally between 0.07 and 0.26 lower (Figure 6). The one exception is JRA55, which actually sees a slightly higher skill score over permafrost regions, driven by slight declines in RMSE over the cold season relative to regions further south. It remains for future studies to determine whether these differences are due to permafrost regions being colder, or due to structural issues with the land models, as this is beyond the scope of this study.

Mirroring the lower skill scores in most products are larger RMSEs over the permafrost regions. This is particularly true during the cold season, when the RMSE for most products is typically 1.3° C to 4.5° C larger (Figure 6) than over the zone with little to no permafrost. Warm season RMSE is also larger for most products in the permafrost zone, though by a lesser margin (0.1° C to 2.1° C).

Bias is also typically larger over permafrost regions by 0.63° C to 3.9° C. The ERA5-Land warm (positive) bias in the cold season is largest over permafrost regions (Figure 6). In the case of JRA55, however, the warm (positive) biases over the cold season are largest further south. In fact, over many grid cells in the permafrost zone, JRA55 exhibits a cold (negative) bias during the cold season (not shown).

Reduced agreement between products is also apparent over permafrost regions, particularly at depth, where the spread in standard deviation between products is roughly 1.6 to 3.7 times larger over the permafrost zone (Figure S1), because of substantial differences in the variance of ERA5-Land, JRA55 and ERA-Interim. Interestingly, the differences in correlation and standard deviation between the permafrost zone, and the zone with little to no permafrost, in the near surface soil layers are less dramatic (Figure S2).

Generally speaking, the skill is higher over Eurasia than over North America (Figure 7). The lower skill in North America arises in part due to the underestimation of seasonal cycle over many grid cells in the Yukon, and an overestimation of variability of cold season temperatures over much of the Great Lakes Region (Figure S3). CFSR and JRA55 are an exception, however, as they greatly overestimate the cold season variability over much of western Eurasia (not shown), and consequently exhibit lower Eurasian skill scores. Soil temperature correlations (with in situ soil temperatures) are also lower by about 0.02 to 0.08 for most products in the warm season over North America, relative to Eurasia (not shown), which further contributes to reduced skill over North America.

As there are few stations above 1000m, elevation does not have a substantial impact on product performance (Figure 8). The slight improvement in near surface cold season performance in CFSR at higher elevations can be linked to a slightly higher correlation, while in ERA-Interim, the slightly higher skill score is due to a slight improvement in cold season temperature variance (not shown). Skill scores, and correlations (not shown), at depth, are lower by about 0.05 to 0.1 over higher elevation stations, however the overall conclusions are not altered.

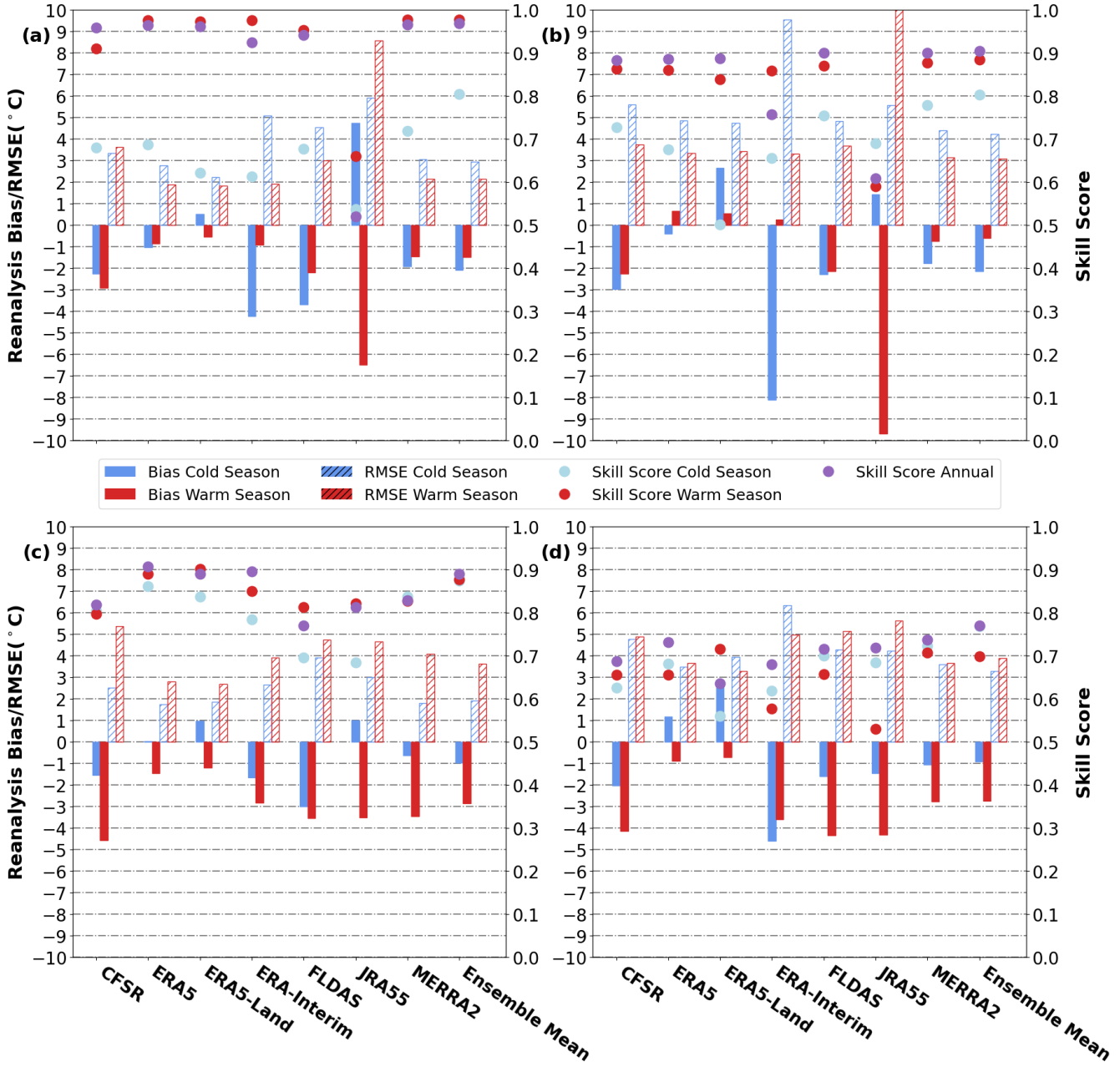

Figure 6. Bias (solid color), RMSE (hatching) and skill scores (circles) of each product the cold season (blue) and the warm season (red) performance of reanalysis products over the zone with little to no permafrost, (a) and (c), and the permafrost zone, (b) and (d). The skill score is also shown for the annual cycle (purple). (a) and (b) displays the bias, RMSE and skill score for the near surface layer, while (c) and (d) display the bias, RMSE and skill score at depth. The ensemble mean is shown beside for comparison.

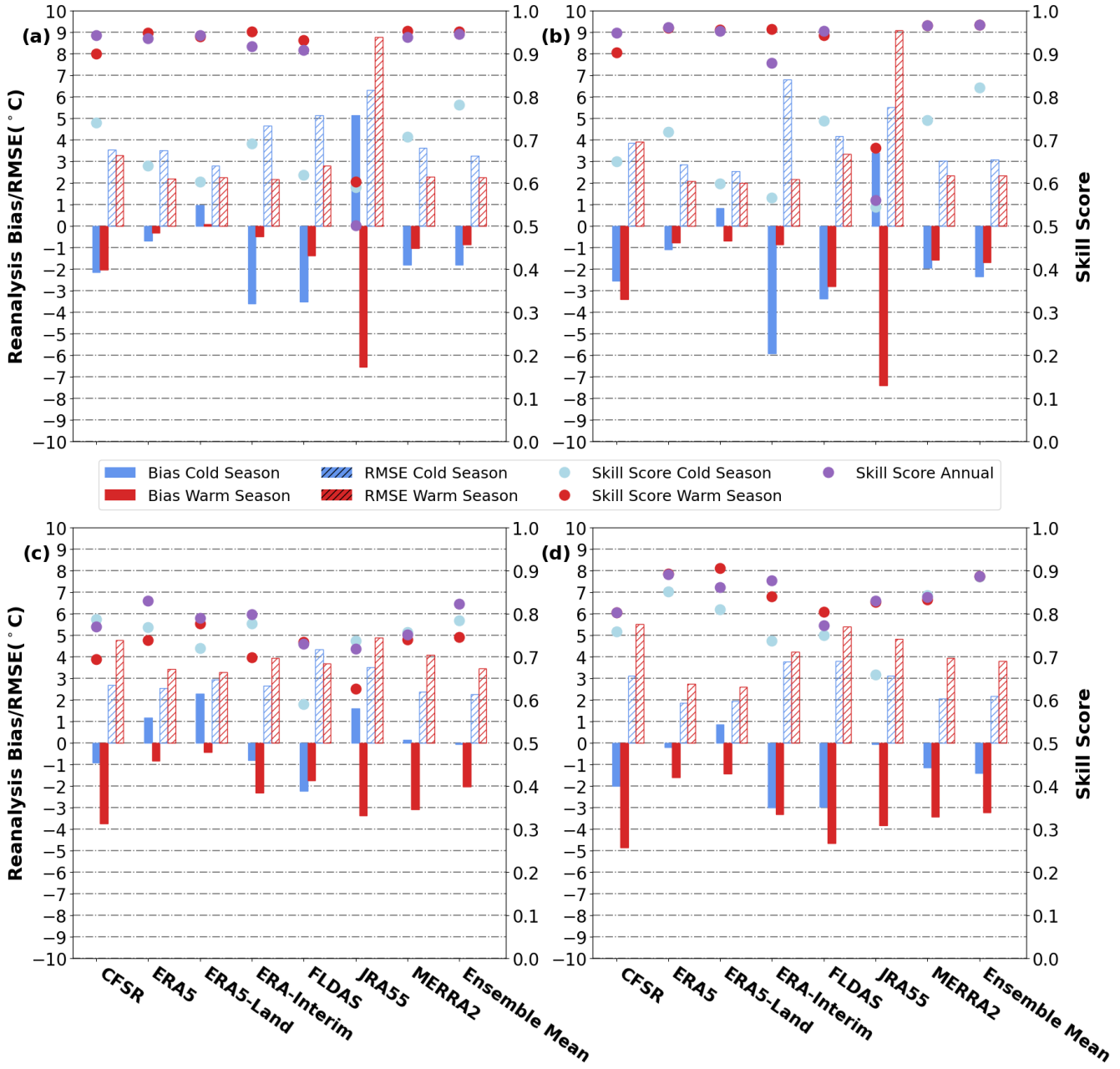

Figure 7. Bias (solid color), RMSE (hatching) and skill scores (circles) of each product the cold season (blue) and the warm season (red) performance of reanalysis products over North America, (a) and (c), and Eurasia, (b) and (d). The skill score is also shown for the annual cycle (purple). (a) and (b) displays the bias, RMSE and skill score for the near surface (0 cm to 30 cm) layer, while (c) and (d) display the bias, RMSE and skill score at depth (30 cm to 300 cm). The ensemble mean is shown beside for comparison

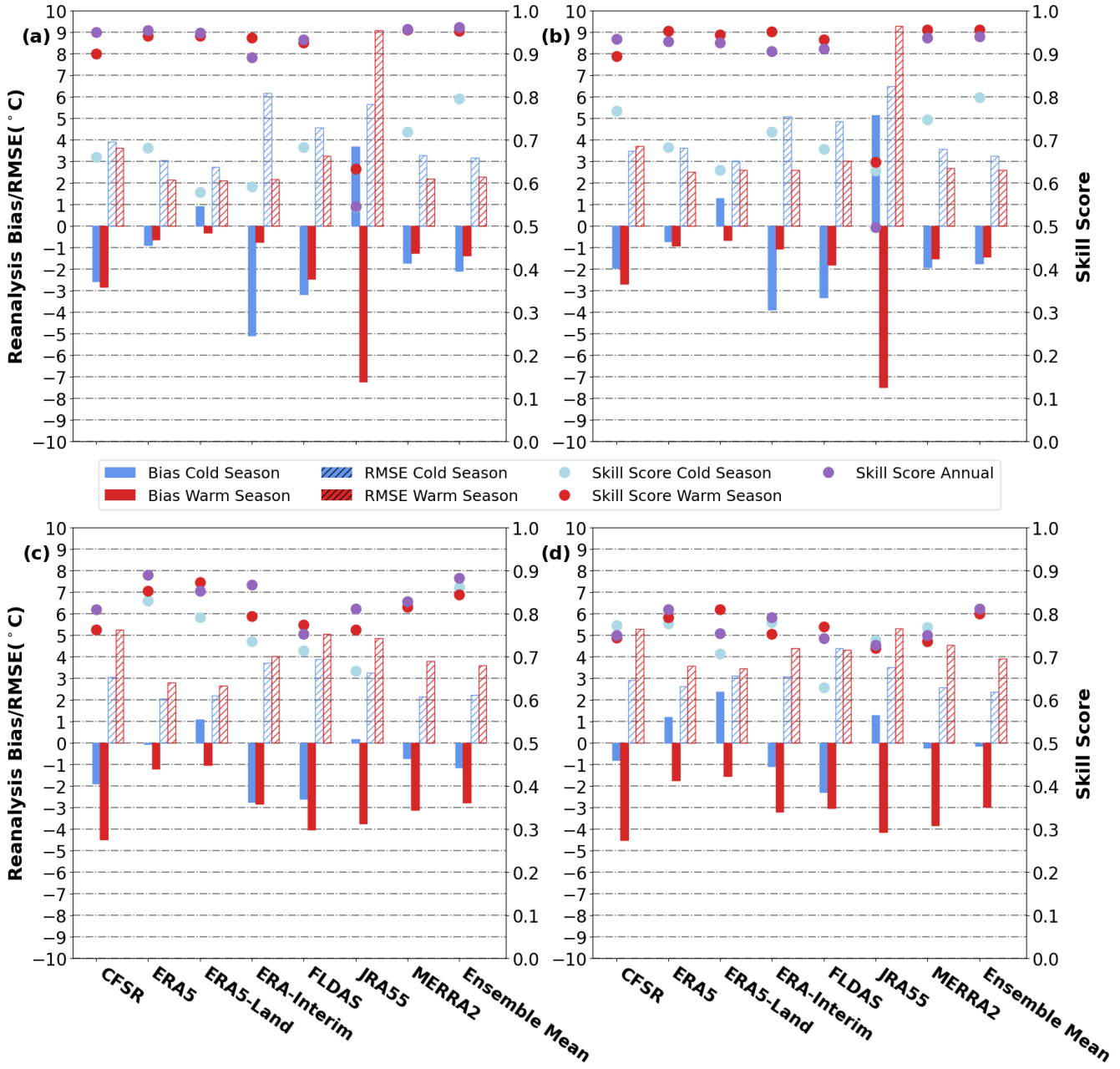

Figure 8. Bias (solid color), RMSE (hatching) and skill scores (circles) of each product the cold season (blue) and the warm season (red) performance of reanalysis products over low elevation grid cells (below 500m), (a) and (c), and grid cells at or above 500m elevation, (b) and (d). The skill score is also shown for the annual cycle (purple). (a) and (b) displays the bias, RMSE and skill score for the near surface (0 cm to 30 cm) layer, while (c) and (d) display the bias, RMSE and skill score at depth (30 cm to 300 cm). The ensemble mean is shown beside for comparison.

4.4 Soil Temperature Trends

We calculate product trends over the 1985-2010 period in order to be able to calculate a station estimate from a subset of 52 Eurasian grid cells that have a continuous timeseries, and a consistent number of sites and depths included over all dates and times (denoted as the Eurasian subset from hereon in). Trends at depth are very similar in magnitude and spatial pattern to the near surface, so we focus on the near surface results here. Trends in the Eurasian subset (hatched lines) are generally representative of the Eurasian average (blue), though are overestimated slightly in the case of CFSR.

Regionally averaged annual mean soil temperature trends show a small positive trend of $< 0.5^{\circ}$ C decade^{-1} in most products, over both Eurasia and North America, and trends in most products are generally consistent with the station estimate over the Eurasian subset (Figure 9, Panel C). In CFSR (purple), however, the trend is near zero over North America, and tends towards negative in Eurasia, arising because of anomalously cold years in 2009 and 2010 (see inset in Figure 9, Panel A), and anomalously warm periods in the 80s and early 90s at the beginning of the timeseries (Figure 9, Panels A and B). It is likely that the cold anomalies in 2009 and 2010 can be linked to issues with CFSR snow cover between January 2009 and January 2011 (Figure S4).

Similar to skill score, and RMSE, products show greater disagreement over higher latitudes (Figure S5), and during winter (Figure S6). ERA5 in particular, and ERA5-Land, and ERA-Interim to a lesser extent, show several pockets of cooling over Siberia and the western Arctic over North America (Figure S5), driven by strong cooling trends in DJF (Figure S6). Meanwhile, the cooling trends in DJF are not as apparent in FLDAS, JRA55 and MERRA2. Trends over JJA show good agreement between products, with most regions showing small warming trends of $< 1^{\circ}$ C decade^{-1}, and pockets of slight cooling over portions of Eurasia and Western North America (Figure S7).

Figure 9. Near surface soil temperature anomalies and trends for each of the reanalysis products. (a) displays the regionally averaged 1982-2018 annual mean soil temperature anomalies for each reanalysis product north of 40° N over Eurasia, while (b) displays the same, but over North America. (c) exhibits an estimate of the regionally averaged 1985-2010 annual mean decadal soil temperature trend for each of the individual products, and the ensemble mean for comparison (the error bars represent the 95% CI for the mean trend).

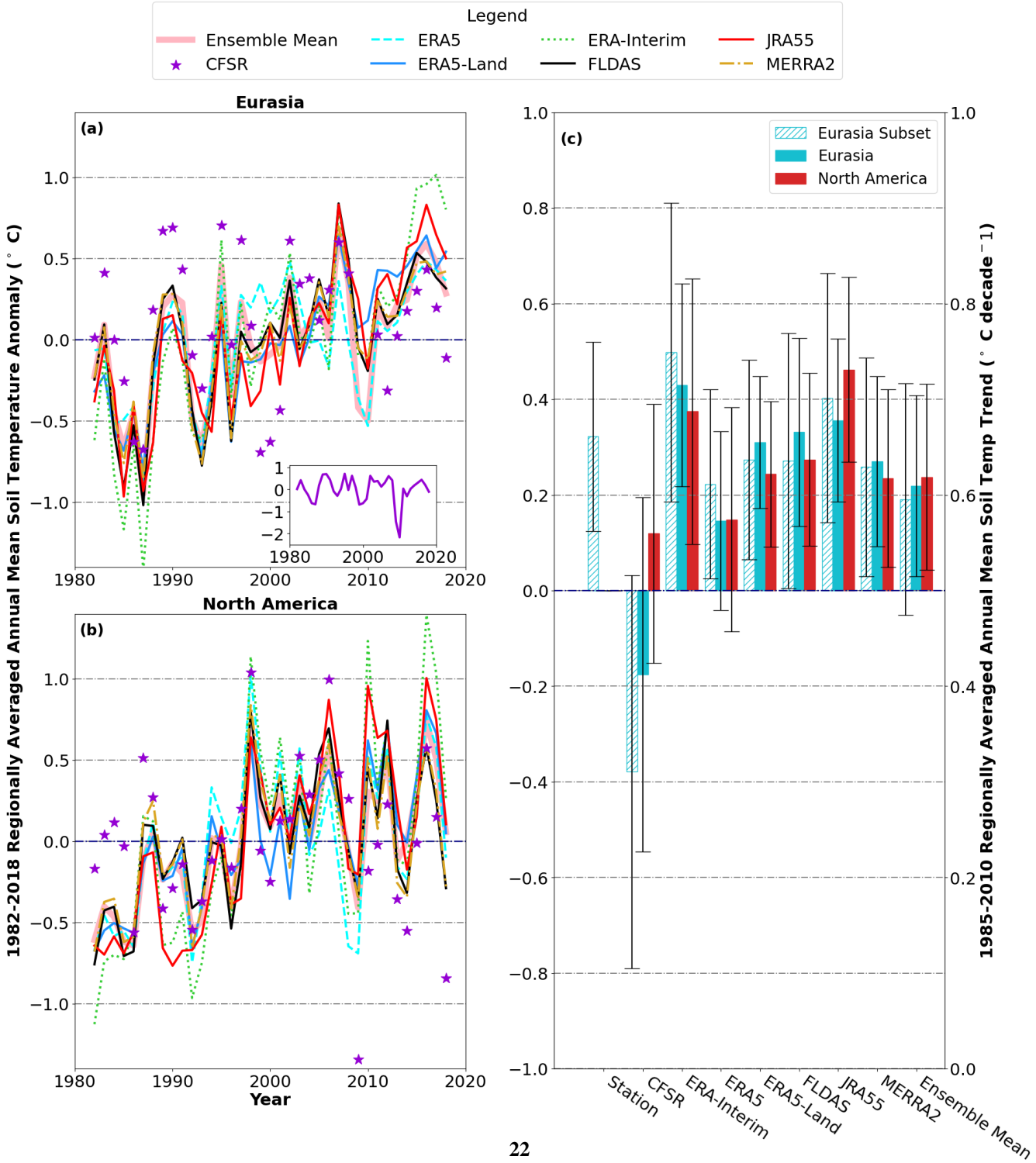

Table 5. Standard deviation (as a measure of spread between products) of the mean biases in winter minimum and summer maximum soil temperature, as a function of latitude and depth (from Figures 10 and S8, Panels C and D). Latitude bands are 10 degrees in width, such that the 40° N latitude band is an average between 40° N and 50° N, while the 60° N latitude band is an average between 60° N and 70° N, for example.

Latitude Band	Near Surface		Depth	
	Winter Minimum	Summer Maximum	Winter Minimum	Summer Maximum
40°N	2.80° C	2.32° C	1.51° C	1.25° C
50°N	2.80° C	2.41° C	1.56° C	1.43° C
60°N	3.81° C	2.95° C	2.41° C	1.90° C

4.5 Variability in Seasonal Extremes

The winter minimum and summer maximum show a cold bias over all latitude bands for most products (Figure 10 Panels C and D), similar to the mean biases in the warm and cold seasons (Figure 2). Similarly, the warm biases in ERA5-Land (sky-blue) and JRA55 (red) also extend to their winter maximum temperatures (Figure 10, Panel C). The conclusions regarding variability in soil temperature extremes, at depth, are generally similar to those near the surface (Figure S8, Panel C and D).

The spread between products in the bias of the summer maximum sees less latitudinal variation than the spread in the winter maximum over both depths, though the spread is largest near the surface (Table 5). Using the standard deviation as a measure of spread between product biases, the near surface standard deviation in winter minimum bias increases from 2.80° C over the 40° N latitude band, to 3.81° C north of 60° N (Table 5), in large part because of the large winter biases in ERA-Interim (green). Meanwhile, the spread in the summer maximum bias sees smaller increases (from 2.32° C at 40° N, to 2.95° C at 60° N) (Table 5).

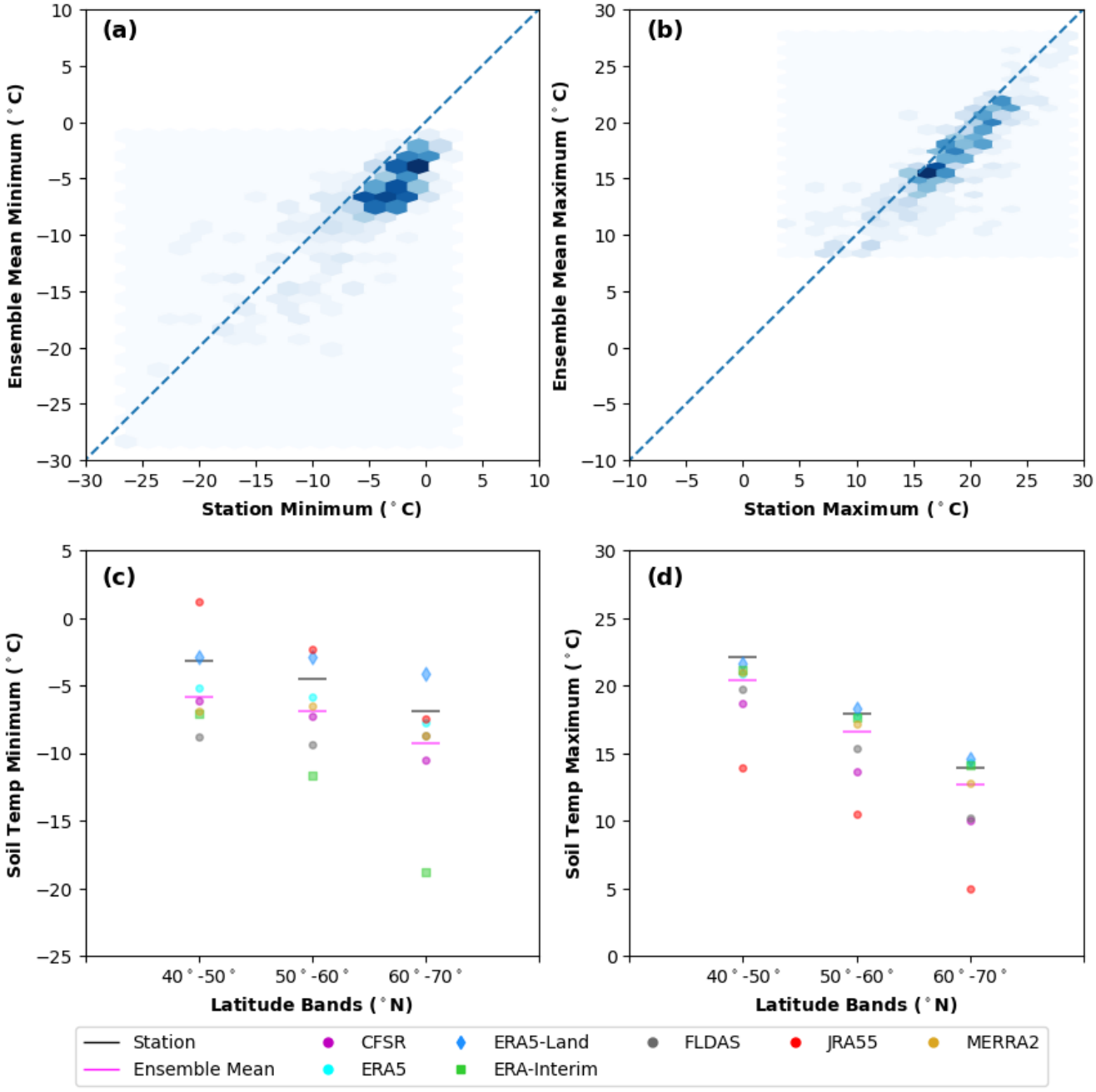

Figure 10. Performance of the near surface soil temperature variability in the Ensemble Mean. (a): Scatterplot of the station and ensemble mean winter minimum soil temperature. (b): Scatterplot of the station and ensemble mean summer maximum soil temperature. (c): latitudinal averages (from Eurasian grid cells) of near surface soil temperature winter minimum for the ensemble mean and contributing products. (d): latitudinal averages (from Eurasian grid cells) of near surface soil temperature summer maximum for the ensemble mean and contributing products.

5 Ensemble Mean Product

5.1 Ensemble Mean Validation

The ensemble mean soil temperature product shows closer agreement with in-situ soil temperatures than any of the individual product, when all depths, seasons and regions are considered as a whole. First, its annual mean and cold season skill scores are higher than in any individual product. Second, its bias and RMSE are generally close in magnitude to the product with the lowest bias and RMSE over all depths, and seasons (Figure 2). Third, the ensemble mean product (pink) displays a temporal variance within 20% of the observed variability over all depths and seasons, including in the cold season when many products fail to capture observed variance (Figure 4).

The value of using the ensemble mean soil temperature is particularly noticeable in the cold season when individual products see a decline in skill, and a larger spread in performance. This is particularly noticeable in that its near surface skill score in the cold season is nearly 10% higher than the next best product (Figure 2). Next, its correlation is roughly 0.05 larger than best individual product in cold season over both depths (Figure 4). Moreover, its bias and RMSE only see relatively small increases over permafrost regions 11), while products such as ERA5-Land, which have a small RMSE over more southern regions, see more substantial increases in bias and RMSE over the permafrost zone (Figure S1). Thus, the ensemble mean soil temperature dataset provides the best estimate of in situ temperatures for the broadest range of conditions.

5.2 Ensemble Mean Soil Temperature Trends

We focus our analysis of trends on the near surface data, as the spatial pattern of soil temperature trends near the surface and at depth show a pattern correlation of greater than 0.95 (not shown), and the conclusions regarding performance are generally similar.

The ensemble mean shows a small positive soil temperature trend of 0.23° C $\pm$ 0.09° C decade^{-1} over Eurasia, and 0.20° C $\pm$ 0.109° C decade^{-1} over North America between 1985 and 2010 (Figure 9, Panel C). Most regions show small positive mean annual trends of $< 0.5^\circ$ C decade^{-1}, though portions of N America and Siberia exhibit slight cooling trends of $< 0.5^\circ$ C decade^{-1} (Figure 12, Panel A).

Annual mean soil temperature trends in the Ensemble Mean over Eurasia show a strong correlation of 0.82 with observations (Figure 12 Panel B). The ensemble mean also generally captures the correct sign of the observed trends, though it has a slight tendency to underestimate the magnitude of the trend (Figure 12 Panel B). Trends at depth show a pattern correlation of 0.98 with the near surface (not shown) and the conclusions are generally the same.

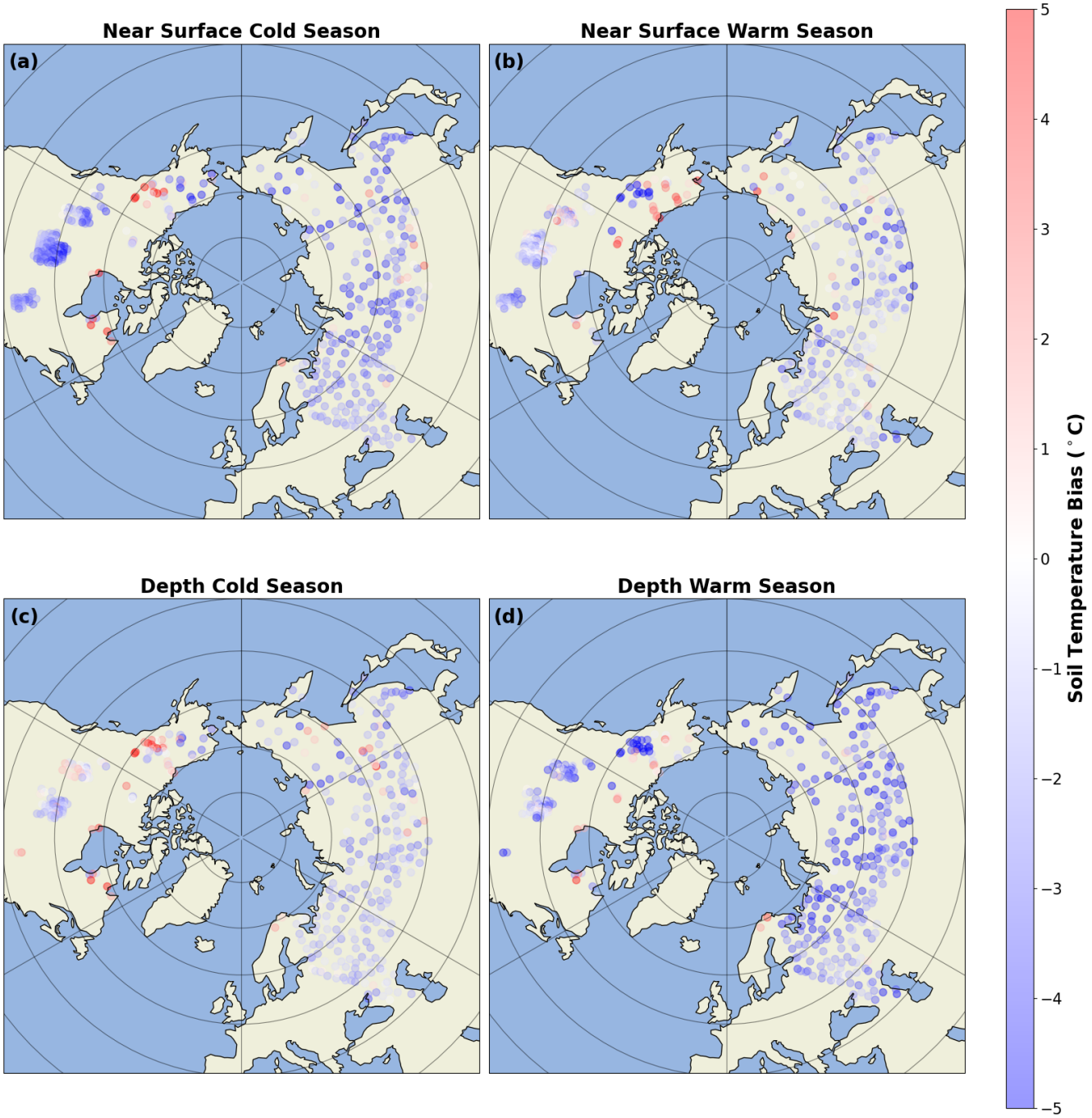

Figure 11. Spatial map of bias for the Ensemble Mean product. Values for the cold season are shown in the left hand panels, and those for the warm season are shown in the right hand panels. Panels A and B show the near surface bias, while biases at depth are shown in Panels C and D.

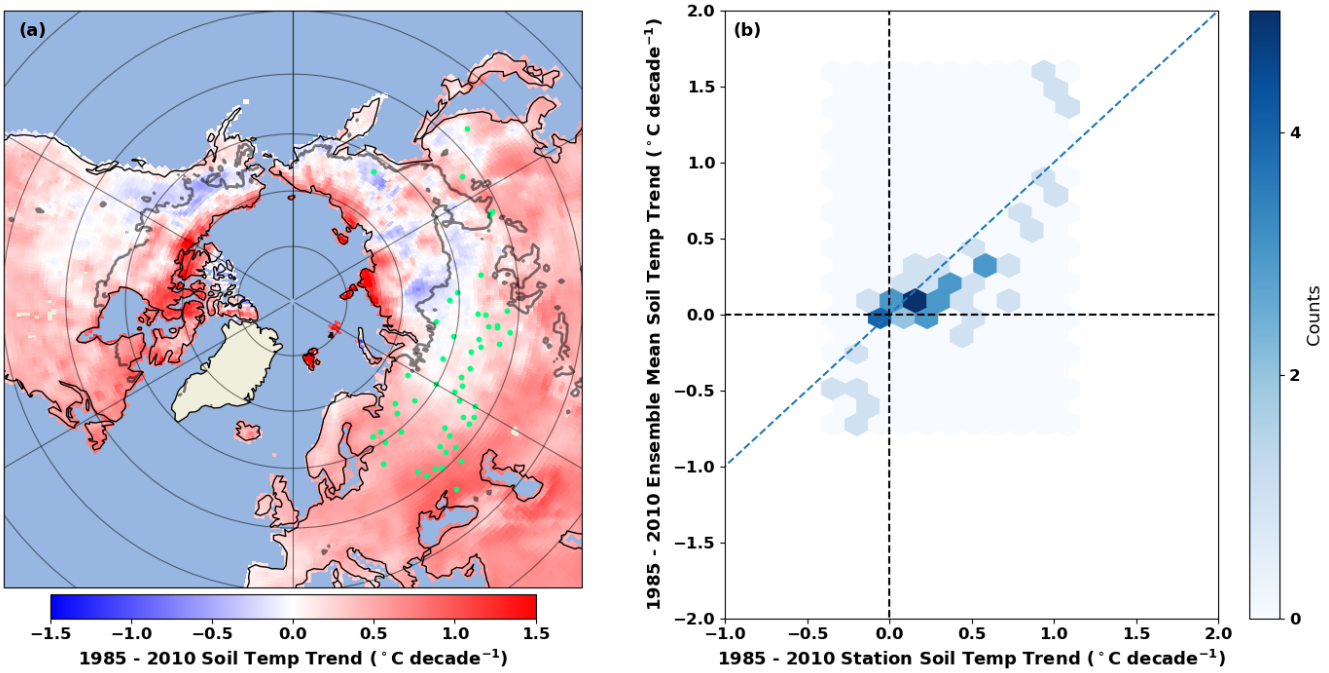

Figure 12. (a): 1985-2010 Ensemble Mean decadal soil temperature trends, near the surface, with the locations of validation grid cells included in the trend analysis shown as green dots. (b): Relationship between the near surface ensemble mean and station soil temperature trends (per decade). The black line represents the boundary of the permafrost zone (regions with at least 50% permafrost cover).

5.3 Ensemble Mean Variability in Seasonal Extremes

Similar to most products, the ensemble mean is biased cold in both the winter minimum (Figure 10, Panel A) and the summer maximum (Figure 10, Panel B) soil temperature. As Figure 10, Panels A and B show, however, there is a fair degree of variability in the behaviour of the ensemble mean seasonal extremes – making an assessment of the mean behaviour in seasonal extremes somewhat tricky. Therefore, in the following paragraphs, we will focus on the most robust findings.

Near the surface, biases in the winter minimum soil temperature (Figure 10, Panel A) are generally larger than in the summer maximum (Figure 10, Panel B). This can also be seen in the latitudinally averaged soil temperatures, where the ensemble mean (pink line) is further from the station (black line) in the winter minimum (Figure 10, Panel C) than in summer maximum (Figure 10, Panel D), which agrees with the findings that the near surface cold season bias is generally larger than the bias in warm season (Figure 2).

At depth, however, the summer maximum shows a larger bias (Figure S8, Panel C) than the winter maximum (Figure S8, Panel D), consistent with the finding that the extratropical mean bias is largest in the warm season at depth (Figure 2). From Figure S8, Panel B, we also see that the greatest disagreement in summer maximum occurs over the coldest regions.

Referring to Figure 1, Panel A, several different types of grid cells are denoted. The first group - Type 1 (16 occurrences) grid cells are characterized by a strong cold bias in (underestimate) the winter minimum soil temperature (Figure 13, Panel A). A second grouping of grid cells, which we refer to as Type 2 (6 occurrences) grid cells are defined as those which have a strong warm bias in (overestimate) the summer maximum temperature (Figure 13, Panel B).

A common feature of the third group, Type 3, is that they underestimate the observed seasonal cycle of soil temperatures (Figure 13, Panel C). While only 1 occurrence was found in the 52 Eurasian grid cells used for trend analysis, many grid cells in the Yukon would also show a similar underestimation of the seasonal cycle. This is evident as the ensemble mean normalized standard deviation (a measure of soil temperature variability) is substantially smaller than 1.0 in both seasons (Figure S3). Often the in situ stations located within Type 3 grid cells are located in areas devoid of vegetation, and it is likely that disagreements in the simulated vegetation cover in the contributing reanalysis products may partially account for the reduced seasonal cycle.

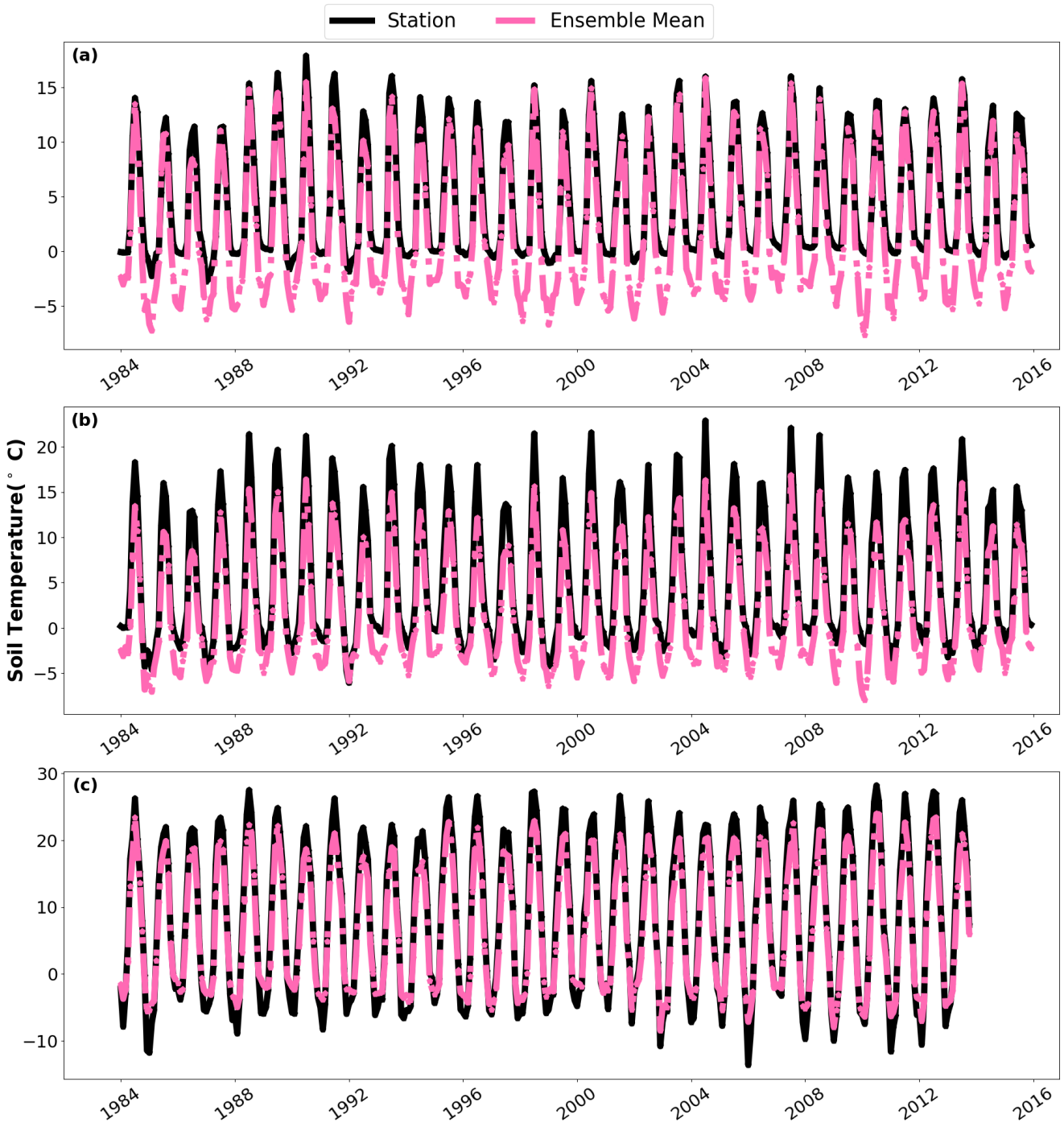

Figure 13. Timeseries from selected grid cells showing the ensemble mean (pink) and station (black) soil temperatures. (a): Timeseries where the ensemble mean simulates a winter minima that is too cold. (b): Timeseries where the ensemble mean simulates a summer maxima that is too cold. (c): Timeseries where the ensemble mean underestimates the seasonal cycle of soil temperatures.

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. The results are qualitatively similar to the findings of previous studies exploring reanalysis soil temperature performance in the extratropical northern hemisphere, which generally highlighted a cold bias in most products (Hu et al., 2019; Qin et al., 2020; Wu et al., 2018; Xu et al., 2019; Yang and Zhang, 2018; Zhan et al., 2020). Similar to Li et al. (2021), we note greater biases in cold season soil temperatures, and our results qualitatively reflect the findings of Cao et al. (2020), who found that ERA5-Land exhibited warm soil temperature biases - particularly over higher latitudes.

The soil temperature trends reported here of a similar magnitude to those reported by Biskaborn et al. (2019), who found that permafrost soil temperatures generally warmed at a rate of 0.29° C ± 0.12° C decade^{-1}, though ours differ in that they represent the mean extratropical northern hemisphere north of 40° N, whereas Biskaborn et al. (2019) predominantly focus on permafrost regions.

Other major differences here are that we develop an ensemble mean soil temperature product, and had a greater focus on higher latitude regions than most other studies. We also note a strong difference in seasonal performance. Relative to the warm season, the cold season is generally characterized by lower skill, larger near surface temperature biases, a larger spread in the reanalysis products' soil temperature variability and lower correlations with in situ soil temperatures. When all depths and seasons are considered, the ensemble mean product performs better than any individual product, exhibiting a consistently high skill, realistic soil temperature variability, and relatively small biases over all seasons.

Here we show an approximate estimate of the magnitude of soil temperature uncertainty associated with instrumental uncertainties, and those associated with structural differences and parameterizations in land models, using the standard deviation in soil temperature across time and as a function of station temperature. A complete quantitative assessment of the contributions of various sources of uncertainty is not possible using this dataset, as time-series did not have a consistent start or end date and consequently, the metrics are calculated using different climatologies across different grid cells. A more complete uncertainty analysis is beyond the scope here, but in the future could be achieved by limiting analysis to a subset of grid cells with consistent timeseries; for example by focusing on soil temperature networks such as the Michigan Enviro-weather Network, or the North Dakota Mesonet Network, or limiting the uncertainty analysis to a smaller portion of the permafrost region.

We find that the median spread in the spatially averaged soil temperature of stations in a grid cell is approximately 1.49° C (Figure 1, Panel B) – several degrees smaller than the standard deviation of reanalysis soil temperatures for a given station soil temperature; particularly over frozen soils (Figure 5). For example, when soil temperatures are below –20 ° C, soil temperature standard deviations increase to near 10° C in several products. Reanalysis two metre air temperatures maintain a relatively consistent standard deviation between 1.25° C to 1.75° C for most products, and only increase slightly to between 2.25° C and 3.5° C over the coldest station air temperatures (not shown). Unlike with soil temperature, the spread in reanalysis two metre air temperatures only increases modestly over the cold season (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 between products, rather than from uncertainties in observed soil temperatures, or from differences in product air temperatures.

6.1 Uncertainties Associated with Land Model Parameterizations and Structural Differences

Uncertainties in soil temperatures associated with structural differences and parameterizations in land models can be grouped into several categories. The first surrounds the simplified parameterizations controlling frozen soil processes. For example in the Noah LSM - utilized by CFSR and FLDAS, freeze-thaw processes are highly simplified, and unsuited for permafrost simulations (Hu et al., 2019) - and may have contributed to the relatively large soil temperature biases simulated in these products. Even the best performing products: ERA5 and ERA5-Land, are unsuited for simulation of permafrost soil temperatures, as they fail to simulate phase-dependent changes in soil thermal conductivity (Cao et al., 2020).

Yang et al. (2020) noted that larger soil temperature biases over the Qinghai-Tibetan Plateau in deeper soil layers were likely related to the fact that soil temperatures are less constrained by air temperature observations (and soil properties). This could also explain why soil temperature biases in the warm season are larger at depth than near the surface in this study. Moreover, the near surface soil layers tend to fall within the active layer (which undergoes seasonal freeze/thaw), while deeper soil layers are more likely to contain permafrost. Permafrost has a high degree of impermeability, which prevents soil water from infiltrating below the bottom of the active layer, and owing to latent heat considerations, leads to soil water freezing at the base of the active later (Zhao et al., 2000), however these processes are not well represented in reanalysis LSMs (Yang et al. (2020); Hu et al. (2019)).

LSMs such as the Simple Biosphere Model (used in JRA55), 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 and Kim, 2010). Moreover, the force restore method assumes a strong diurnal forcing from above, an assumption that is likely violated when snow cover is present (Tilley and Lynch, 1998; Slater et al., 2001), because snow cover leads to a decoupling of the surface forcing from the soil below. These factors may help explain why JRA55 is unable to simulate near surface soil temperatures as accurately as the other products explored in this study explicitly incorporate representations of soil heat flux between soil layers (Niu et al., 2011; Koster et al., 2000; van den Hurk et al., 2000; Balsamo et al., 2009), and hence they are able to simulate a dampening of seasonal variability of soil temperatures at depth (and greater variability near the surface).

Burke et al. (2020) note that differences in snow cover properties were important in explaining soil temperature biases of several Coupled Model Intercomparison Project 6 (CMIP6) models, and it is likely that differences in snow cover properties between the LSMs studied here could account for some of the observed spread - particularly in the cases of ERA-Interim, ERA5 and ERA5-Land, because during the warm season, these products have similar soil temperature biases, but their performance varies widely during the cold season (Figures 2 and 10), in large part because of snow density biases (Cao et al., 2020; Gao et al., 2022). In ERA-Interim, the large cold (negative) bias during the cold season is strongly related to the fact that it overestimates the observed snow density (Gao et al., 2022), and consequently also overestimates the thermal conductivity of the snow pack.

Conversely, snow density (and thermal conductivity) in ERA5-Land (and ERA5) are too low, and hence biases in snow density are a large contributing factor to the warm (positive) bias during the cold season (Cao et al., 2020).

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). In the case of the Noah LSM, which is included in CFSR/CFSv2 and FLDAS, Li et al. (2021) found that an overestimation of snow cover was a major contributor to larger soil temperature biases in winter over the Qinghai-Tibetan Plateau. Shukla et al. (2019) and Shukla and Huang (2020) found that overestimation of snow cover in CFSR during autumn and early winter leads to an overactive snow-albedo feedback, and excessive cooling of the near surface soil layers. This translates into a strong cold bias at depth over the spring and summer, and likely explains why CFSR's warm season bias and RMSE at depth are the largest of all seven products examined in this study (Figure 2).

6.2 Uncertainties Associated with Scale Effects

Here we evaluated soil temperatures at a relatively coarse resolution of 0.75°. As such, it is difficult for reanalysis products to capture local scale variability in soil temperature. The sub-grid scale variability in soil temperatures calculated in Figure 1, Panel B is of a similar magnitude to those calculated by previous studies exploring sub-grid scale variability in cryospheric soil temperatures (Gubler et al., 2011; Morse et al., 2012; Gisnås et al., 2014), though are generally smaller than those reported by Cao et al. (2019). We found that the spatial variability in soil temperatures in one high latitude grid cell is larger than 10° C at times (Figure 1) - of a similar magnitude to those reported by Cao et al. (2019).

Moreover, as many grid cells in Eurasia only included a single in situ station, there is a large chance that this single in situ station may not necessarily be reflective of conditions elsewhere in the grid cell (Gubler et al., 2011). When multiple in situ stations were available, we took the spatial mean of all stations, in an attempt to estimate the mean soil temperature over the grid cell.

6.3 Uncertainties Arising from Sampling Variability

As was described in Section 5.2, the presence of missing data created a challenge for calculating in situ soil temperature averages. While most grid cells in Eurasia had relatively consistent timeseries, and fewer issues with missing data, this was not the case over North America. Rather than limit our analysis to a small number of grid cells with little to no missing data (as we did for the calculation of soil temperature trends), we chose to make use of all available data at each timestep when calculating our validation metrics (bias, RMSE, standard deviation, correlation and skill score). Thus, the spatially averaged in situ soil temperature did not always contain a constant number of depths or grid cells at each timestep in many grid cells over North America.

From Figure 1, Panel B, it is apparent that the median variability of soil temperatures between stations within a grid cell (spatial variation), 1.49° C, is roughly 1.75 times larger than the median variability of soil temperatures at different depths, 0.84° C, for a particular station (depth variation). Thus, it would appear that fluctuations in the number of stations comprising the spatially averaged soil temperature are responsible for a greater proportion of the uncertainty than fluctuations in the

number of depths included. However, it is also apparent that the uncertainties arising from variations in the number of grid cells included in a station average are substantially smaller than the spread between reanalysis products. During the cold season, the uncertainty in soil temperatures associated with the spread between reanalysis products is often two to three times larger than the uncertainty arising from fluctuations in station availability.

6.4 Applications and Future Work

The ensemble mean data product provides gridded, monthly-averaged soil temperature estimates of near surface, and deeper soil temperatures at a 0.75° resolution. Therefore, it is most suitable to regional or hemispheric-scale analyses of soil temperature climatologies, or their seasonal cycle, or to explore recent trends in soil temperatures. The product could also be used to provide boundary conditions for models that require soil temperature inputs, such as hydrological models, and for the validation of model soil temperatures. While the ensemble mean product still exhibits substantial cold biases over permafrost regions, and therefore is likely unsuitable for permafrost modeling, the RMSE over the permafrost zone of the ensemble mean product outperforms the RMSE of the best performing product by 0.5° C, on average (Figure S1) in the cold season, and hence it may still provide some added value for estimation of high latitude soil temperatures relative to the individual products.

A robust ensemble mean can be computed with four products (not shown), which means a higher resolution ensemble mean data product could be created using a subset of the higher resolution reanalysis products. For example, ERA5, ERA5-Land, MERRA2, and CFSR have resolutions lower than 0.5°. Using a similar blending methodology, we have been investigating the performance of a 0.31° product (using a smaller subset of products that provide data at higher spatial resolution). We have also performed similar analyses with a 0.05° soil temperature product, using interpolated soil temperatures from the Arctic System Reanalysis version 2 (ASR), ERA5-Land, and FLDAS. The goal has been to assess the impact of spatial resolution on performance of the ensemble mean product. We are hoping to include these results in a follow-up paper. 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 also emphasize how differences in the land model structure and parameterization may account for the spread in soil temperatures.

Data availability. The CRU TS v 4.07 2m air temperature dataset can be found on the CRU TS dataset website. 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 et al., 2010), CFSv2 (Saha and Coauthors, 2012), ERA-Interim (European Centre for Medium-Range Weather Forecasts, 2012), and JRA55 (Japan Meteorological Agency, 2014) data were obtained from the National Center for Atmospheric Research (NCAR)'s Research Data Archive (RDA). FLDAS (McNally and NASA/GSFC/HSL, 2018), and MERRA2 (Global Modeling and Assimilation Office, 2015) were obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC). ERA5 (European Centre for Medium-Range Weather Forecasts, 2019), and 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. TH and CGF conceived of the study, TH gathered, and analyzed the data, and TH, CGF interpreted the data. HK provided in situ data to the study, and TH, and CGF wrote the manuscript, with contributions from HK.

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 authors would like to thank the editor and anonymous referees, as well as Dr. Hugh Brendan O'Neill and Dr. Andre Erler for their helpful comments towards improving the final manuscript.

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.

Alberta Agriculture, Forestry and Rural Economic Development: Current and Historical Alberta Weather Station Data., https://acis.alberta.ca, 2022.

Allard, M., Sarrazin, D., and Hérault, L.: Borehole and near-surface ground temperatures in northeastern Canada, v. 1.5 (1988-2019). Nordicana D8, doi:10.5885/45291SL-34F28A9491014AFD, 2020.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, Journal of Hydrometeorology, 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

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.

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.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A. B. K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., and Lantuit, H.: Permafrost is warming at a global scale, Nature Communications, 10, 264, https://doi.org/10.1038/s41467-018-08240-4, 2019.

Burke, E. J., Zhang, Y., and Krinner, G.: Evaluating permafrost physics in the CMIP6 models and their sensitivity to climate change, The Cryosphere, 14, 3155–3174, https://doi.org/https://doi.org/10.5194/tc-2019-309, 2020.

Cameron, E., Lantz, T., O'Neill, H., Gill, H., Kokelj, S., and Burn, C.: Permafrost Ground Temperature Report: Ground temperature variability among terrain types in the Peel Plateau region of the Northwest Territories (2011-2015), Tech. Rep. NWT 2017-009, Northwest Territories Geological Survey, Northwest Territories, Canada, 2019.

Cao, B., Quan, X., Brown, N., Stewart-Jones, E., and Gruber, S.: GlobSim (v1.0): deriving meteorological time series for point locations from multiple global reanalyses, Geoscientific Model Development, 12, 4661–4679, https://doi.org/10.5194/gmd-12-4661-2019, 2019.

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.

Cao, B., Arduini, G., and Zsoter, E.: Brief communication: Improving ERA5-Land soil temperature in permafrost regions using an optimized multi-layer snow scheme, The Cryosphere, 16, 2701–2708, https://doi.org/10.5194/tc-16-2701-2022, 2022.

CEN: Climate station data from Whapmagoostui-Kuujjuarapik Region in Nunavik, Quebec, Canada, v. 1.5 (1987-2019). Nordicana D4, doi:10.5885/45057SL-EADE4434146946A7, 2020a.

CEN: Climate station data from the Sheldrake river region in Nunavik, Quebec, Canada, v. 1.1 (1986-2019). Nordicana D61, doi:10.5885/45480SL-C89DEB92A4FE4536, 2020b.

CEN: Climate station data from the Clearwater lake region in Nunavik, Quebec, Canada, v. 1.1 (1986-2019). Nordicana D57, doi:10.5885/45475SL-5A33FE09B0494D92, 2020c.

CEN: Climate station data from the Little Whale River region in Nunavik, Quebec, Canada, v. 1.1 (1993-2019). Nordicana D58, doi:10.5885/45485SL-78F4F9C368364100, 2020d.

CEN: Climate station data from the Biscarat river region in Nunavik, Quebec, Canada, v. 1.0 (2005-2019). Nordicana D62., doi:10.5885/45495SL-78FA5A95C5FB4D, 2020e.

CEN: Climate station data from Northern Ellesmere Island in Nunavut, Canada, v. 1.7 (2002-2019). Nordicana D8, doi:10.5885/44985SL-8F203FD3ACCD4138, 2020f.

CEN: Environmental data from Boniface river region in Nunavik, Quebec, Canada, v. 1.3 (1988-2019). Nordicana D7, doi:10.5885/45129SL-DBDA2A77C0094963, 2020g.

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.

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.

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, 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., Haimberger, 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.

Derber, J., Parrish, D. F., and Lord, S. J.: The New Global Operational Analysis System at the National Meteorological Center, Weather and Forecasting, 6, 538–547, 1991.

Dirmeyer, P. A., Koster, R. D., and Guo, Z.: Do Global Models Properly Represent the Feedback between Land and Atmosphere?, Journal of Hydrometeorology, 7, 1177–1198, http://www.jstor.org/stable/24910939, publisher: American Meteorological Society, 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, 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.

Déry, S.: Cariboo Alpine Mesonet (CAMnet) Database, https://doi.org/10.5281/zenodo.1195043, 2017.

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.

Ensom, T., Kokelj, S., and McHugh, K.: Permafrost Ground Temperature Report: Inuvik to Tuktoyaktuk Highway stream crossing and alignment sites, Northwest Territories, Tech. Rep. NWT Open Report 2019-004, Northwest Territories Geological Survey, Northwest Territories, Canada, 2020.

Enviro-weather: Enviro-weather Automated Weather Station Network, https://mawn.geo.msu.edu/, 2022.

European Centre for Medium-Range Weather Forecasts: ERA-Interim Project, Monthly Means, https://doi.org/10.5065/D68050NT, 2012.

European Centre for Medium-Range Weather Forecasts: ERA5 Reanalysis (Monthly Mean 0.25 Degree Latitude-Longitude Grid), https://doi.org/10.5065/P8GT-0R61, 2019.

European Space Agency: Copernicus Global Digital Elevation Model GLO-90, 2021.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific Data, 2, 150 066, https://doi.org/10.1038/sdata.2015.66, 2015.

Gao, S., Li, Z., Zhang, P., Zeng, J., Chen, Q., Zhao, C., Liu, C., Wu, Z., and Qiao, H.: An Assessment of the Applicability of Three Reanalysis Snow Density Datasets Over China Using Ground Observations, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, https://doi.org/10.1109/LGRS.2022.3202897, 2022.

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.

Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isaksen, K., Boike, J., and Etzelmüller, B.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, The Cryosphere, 8, 2063–2074, https://doi.org/10.5194/tc-8-2063-2014, 2014.

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, https://doi.org/10.5067/RKPHT8KC1Y1T, 2015.

Gruber, S., Brown, N., Stewart-Jones, E., Karunaratne, K., Riddick, J., Peart, C., Subedi, R., and Kokelj, S. V.: Permafrost Ground Temperature Report: Ground temperature and site characterisation data from the Canadian Shield tundra near Lac de Gras, Northwest Territories, Canada, Tech. Rep. NWT Open Report 2018-009, Northwest Territories Geological Survey, Northwest Territories, Canada, 2019.

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, publication Title: Supplement to: Biskaborn, Boris K; Smith, Sharon L; Noetzli, Jeannette; Matthes, Heidrun; Vieira, Gonalo; Streletskiy, Dmitry A; Schoeneich, Philippe; Romanovsky, Vladimir E; Lewkowicz, Antoni G; Abramov, Andrey; Allard, Michel; Boike, Julia; Cable, William L; Christiansen, Hanne H; Delaloye, Reynald; Diekmann, Bernhard; Drozdov, Dimitry S; Etzelmller, Bernd; Grosse, Guido; Guglielmin, Mauro; Ingeman-Nielsen, Thomas; Isaksen, Ketil; Ishikawa,

Mamoru; Johansson, Margareta; Johannsson, Halldor; Joo, Anseok; Kaverin, Dmitry; Kholodov, Alexander L; Konstantinov, Pavel; Krger, Silke; Lambiel, Christophe; Lanckman, Jean-Pierre; Luo, Dongliang; Malkova, Galina; Meiklejohn, Ian; Moskalenko, Nataliya G; Oliva, Marc; Ramos, Miguel; Sannel, A Britta K; Sergeev, Dmitry V; Seybold, Cathy; Skryabin, Pavel; Vasiliev, Alexander; Wu, Qingbai; Yoshikawa, Kenji; Zheleznyak, Mikail; Lantuit, Hugues (2019): Permafrost is warming at a global scale. Nature Communications, 10(1), https://doi.org/10.1038/s41467-018-08240-4, 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.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Hernández-Henríquez, M. A., Sharma, A. R., Taylor, M., Thompson, H. D., and Déry, S. J.: The Cariboo Alpine Mesonet: sub-hourly hydrometeorological observations of British Columbia's Cariboo Mountains and surrounding area since 2006, Earth Syst. Sci. Data, pp. 1655–1672, https://doi.org/https://doi.org/10.5194/essd-10-1655-2018, 2018.

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.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

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 Product Users' Handbook, 2014.

Jiao, M., Zhao, L., Wang, C., Hu, G., Li, Y., Zhao, J., Zou, D., Xing, Z., Qiao, Y., Liu, G., Du, E., Xiao, M., and Hou, Y.: Spatiotemporal Variations of Soil Temperature at 10 and 50 cm Depths in Permafrost Regions along the Qinghai-Tibet Engineering Corridor, Remote Sensing, 15, 455, https://doi.org/10.3390/rs15020455, 2023.

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.

Jones, P. W.: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates, Monthly Weather Review, 127, 2204–2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

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, place: Boston MA, USA Publisher: American Meteorological Society, 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, place: Boston MA, USA Publisher: American Meteorological Society, 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. Arctic Data Center. doi:10.18739/A2736M31X, doi:10.18739/A2736M31X, 2020.

Lee, S.-C., Christen, A., Black, A. T., Johnson, M. S., Jassal, R. S., Ketler, R., Nesic, Z., and Merkens, M.: Annual greenhouse gas budget for a bog ecosystem undergoing restoration by rewetting, Biogeosciences, 14, 2799–2814, https://doi.org/10.5194/bg-14-2799-2017, 2017.

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, 2020.

Li, X., Wu, T., Wu, X., Chen, J., Zhu, X., Hu, G., Li, R., Qiao, Y., Yang, C., Hao, J., Ni, J., and Ma, W.: Assessing the simulated soil hydrothermal regime of the active layer from the Noah-MP land surface model (v1.1) in the permafrost regions of the Qinghai–Tibet Plateau, Geoscientific Model Development, 14, 1753–1771, https://doi.org/10.5194/gmd-14-1753-2021, publisher: Copernicus GmbH, 2021.

Luo, D., Liu, L., Jin, H., Wang, X., and Chen, F.: Characteristics of Ground Surface Temperature at Chalaping in the Source Area of the Yellow River, Northeastern Tibetan Plateau, Agricultural and Forest Meteorology, 281, 107 819, https://doi.org/10.1016/j.agrformet.2019.107819, 2020.

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

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.

McNally, A. and NASA/GSFC/HSL: FLDAS Noah Land Surface Model L4 Global Monthly 0.1 x 0.1 degree (MERRA-2 and CHIRPS), Tech. rep., Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA, https://doi.org/10.5067/5NHC22T9375G, 2018.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C. D., and Verdin, J. P.: A land data assimilation system for sub-Saharan Africa food and water security applications, Scientific Data, 4, 170 012, https://doi.org/10.1038/sdata.2017.12, 2017.

Morris, J., Hernández-Henríquez, M., and Déry, S.: Cariboo Alpine Mesonet meteorological data, 2017-2021, https://doi.org/10.5281/zenodo.6518969, 2021.

Morse, P., Burn, C., and Kokelj, S.: Influence of snow on near-surface ground temperatures in upland and alluvial environments of the outer Mackenzie Delta, Northwest Territories., Canadian Journal of Earth Sciences, 49, 895–913, https://doi.org/10.1139/e2012-012, 2012.

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.

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 System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, D12 109, https://doi.org/10.1029/2010JD015139, 2011.

North Dakota Mesonet Network: The North Dakota Mesonet Network, https://ndawn.ndsu.nodak.edu/, 2022.

Obu, J., Westermann, S., Kääb, A., and Bartsch, A.: Ground Temperature Map, 2000-2016, Northern Hemisphere Permafrost, https://doi.org/10.1594/PANGAEA.888600, publication Title: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, 2018.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J., and Zou, D.: Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale, Earth-Science Reviews, 193, 299–316, https://doi.org/10.1016/j.earscirev.2019.04.023, 2019.

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, 2020.

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P. P., Koster, R. D., and De Lannoy, G. J. M.: Assessment of MERRA-2 Land Surface Hydrology Estimates, Journal of Climate, 30, 2937–2960, https://doi.org/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, https://doi.org/10.1175/JCLI-D-16-0570.1, 2017b.

RoTimi Ojo, E. and Manaigre, L.: The Manitoba Agriculture Mesonet: Technical Overview, Bulletin of the American Meteorological Society, 102, E1786–E1804, https://doi.org/10.1175/BAMS-D-20-0306.1, 2021.

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.

Rudy, A., Kokelj, S., Morse, P., and Ensom, T.: Permafrost Ground Temperature Report: Inuvik to Tukyoyaktuk Highway Sentinel Sites, Northwest Territories, Tech. Rep. NWT Open Report 2019-017, Northwest Territories Geological Survey, Yellowknife, NWT, Canada, 2020.

Saha, S. and Coauthors: NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products, https://doi.org/10.5065/D69021ZF, 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, 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, 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.

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.

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.

Shukla, R. P. and Huang, B.: Cumulative Influence of Summer Subsurface Soil Temperature on North America Surface Temperature in the CFSv2, Journal of Geophysical Research: Atmospheres, 125, https://doi.org/10.1029/2019JD031899, 2020.

Shukla, R. P., Huang, B., Dirmeyer, P. A., and Kinter, J. L.: The Influence of Summer Deep Soil Temperature on Early Winter Snow Conditions in Eurasia in the NCEP CFSv2 Simulation, Journal of Geophysical Research: Atmospheres, 124, 9062–9077, https://doi.org/10.1029/2019JD030279, 2019.

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.

Slater, A. G., Schlosser, C. A., Desborough, C. E., Pitman, A. J., Henderson-Sellers, A., Robock, A., Vinnikov, K. Y., Entin, J., Mitchell, K., Chen, F., Boone, A., Etchevers, P., Habets, F., Noilhan, J., Braden, H., Cox, P. M., Rosnay, P. d., Dickinson, R. E., Yang, Z.-L., Dai, Y.-J., Zeng, Q., Duan, Q., Koren, V., Schaake, S., Gedney, N., Gusev, Y. M., Nasonova, O. N., Kim, J., Kowalczyk, E. A., Shmakin,

A. B., Smirnova, T. G., Verseghy, D., Wetzel, P., and Xue, Y.: The Representation of Snow in Land Surface Schemes: Results from PILPS 2(d), Journal of Hydrometeorology, 2, 7–25, https://doi.org/10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Hydrometeorology, 2001.

Smith, S., Romanovsky, V., Lewkowicz, A., Burn, C., Allard, M., Clow, G., Yoshikawa, K., and Throop, J.: Thermal State of Permafrost in North America: A Contribution to the International Polar Year, Permafrost and Periglacial Processes, 21, 117–135, https://doi.org/10.1002/ppp.690, 2010.

Spence, C. and Hedstrom, N.: Baker Creek Research Catchment Hydrometeorological and Hydrological Data, 10.20383/101.026, 2018a.

Spence, C. and Hedstrom, N.: Hydrometeorological data from Baker Creek Research Watershed, Northwest Territories, Canada, Earth Syst. Sci. Data, 10, 1753–1767, 2018b.

Street, L. and Wookey, P.: Soil temperature, soil moisture, air temperature and relative humidity for vegetation at Siksik Creek, North West Territories, Canada, 2016.

Street, L. E., Mielke, N., and Woodin, S. J.: Phosphorus Availability Determines the Response of Tundra Ecosystem Carbon Stocks to Nitrogen Enrichment, Ecosystems, 21, 1155–1167, https://doi.org/10.1007/s10021-017-0209-x, 2018.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Thackeray, C. W., Fletcher, C. G., and Derksen, C.: Quantifying the skill of CMIP5 models in simulating seasonal albedo and snow cover evolution: CMIP5-SIMULATED ALBEDO AND SCF SKILL, Journal of Geophysical Research: Atmospheres, 120, 5831–5849, https://doi.org/10.1002/2015JD023325, 2015.

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.

van den Hurk, B. J. J. M., Viterbo, P., Beljaars, A. C. M., and Betts, A. K.: Offline Validation of the ERA-40 Surface Scheme, Tech. Rep. Technical Memorandum 295, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom, https://www.ecmwf.int/node/12900, 2000.

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.

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.

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.

Xie, P. and Arkin, P. A.: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs, Bulletin of the American Meteorological Society, 78, 2539–2558, https://doi.org/10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2, 1997.

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., Huang, B., Hu, Z.-Z., Kumar, A., Wen, C., Behringer, D., and Nadiga, S.: An assessment of oceanic variability in the NCEP climate forecast system reanalysis, Climate Dynamics, 37, 2511–2539, https://doi.org/10.1007/s00382-010-0954-4, 2011.

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, 114 583, https://doi.org/10.1016/j.geoderma.2020.114583, 2020.

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.

Yukon Geological Survey: Yukon permafrost reports data. In: Yukon Permafrost Database. Government of Yukon, https://service.yukon.ca/permafrost, 2021.

Zhan, M.-j., Xia, L., Zhan, L., and Wang, Y.: Evaluation and Analysis of Soil Temperature Data over Poyang Lake Basin, China, Advances 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, publisher: John Wiley & Sons, Ltd, 2008.

Zhao, C., Gong, C., Duan, H., Yan, P., Liu, Y., and Zhou, G.: Evaluation of Three Reanalysis Soil Temperature Datasets with Observation Data over China, Earth, 3, 1042–1058, https://doi.org/10.3390/earth3040060, 2022.

Zhao, L., Cheng, G., Li, S., Zhao, X., and Wang, S.: Thawing and freezing processes of active layer in Wudaoliang region of Tibetan Plateau, Chinese Science Bulletin, 45, 2181–2187, https://doi.org/10.1007/BF02886326, 2000.

Zhao, T., Guo, W., and Fu, C.: Calibrating and evaluating reanalysis surface temperature error by topographic correction, Journal of Climate, 21, 1440–1446, 2008.