Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer

Brian Groenke[1,3], Moritz Langer[1,4], Jan Nitzbon[2,4], Sebastian Westermann[5], Guillermo Gallego[3,6], and Julia Boike[1,4]

[1]Permafrost Research Section, Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Potsdam, Germany.
[2]Paleoclimate Research Section, Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany.
[3]Department of Electrical Engineering and Computer Science, Technical University of Berlin, Germany.
[4]Department of Geography, Humboldt University, Berlin, Germany.
[5]Department of Geosciences, University of Oslo, Oslo, Norway
[6]Einstein Center Digital Future and Science of Intelligence Excellence Cluster, Berlin, Germany.

Correspondence: Brian Groenke (brian.groenke@awi.de)

Abstract. Long-term measurements of permafrost temperatures do not provide a complete picture of the Arctic subsurface thermal regime. Regions with warmer permafrost often show little to no long-term change in ground temperature due to the uptake and release of latent heat during freezing and thawing. Thus, regions where the least warming is observed may also be the most vulnerable to permafrost degradation. Since direct measurements of ice and liquid water contents in the permafrost layer are not widely available, thermal modeling of the subsurface plays a crucial role in understanding how permafrost responds to changes in the local energy balance. In this work, we first analyze trends in observed air and permafrost temperatures at four sites within the continuous permafrost zone, where we find substantial variation in the apparent relationship between long-term changes in permafrost temperatures ($0.02\,\mathrm{K\,yr^{-1}}$ to $0.16\,\mathrm{K\,yr^{-1}}$) and air temperature ($0.09\,\mathrm{K\,yr^{-1}}$ to $0.11\,\mathrm{K\,yr^{-1}}$). We then apply recently developed Bayesian inversion methods to link observed changes in borehole temperatures to unobserved changes in latent heat and thaw depth using a transient model of heat conduction with phase change. Our results suggest that the degree to which recent warming trends correlate with permafrost thaw and variations in latent heat is heavily dependent on both local soil properties as well as historical climatology. At the warmest site, a nine meter borehole near Ny-Ålesund, Svalbard, modeled annual maximum thaw depth increases by an average of $(12\pm1)\,\mathrm{cm\,K^{-1}}$ rise in mean annual ground temperature. In stark contrast, modeled thaw rates for a borehole on Samoylov Island in the Lena River Delta (northeastern Siberia) appear far less sensitive to temperature change, with an almost negligible increase of $(1\pm1)\,\mathrm{cm\,K^{-1}}$. Although our study is limited to just four sites, the results urge caution in the interpretation and comparison of warming trends in Arctic boreholes, indicating substantial uncertainty in their implications for the current and future thermal state of permafrost.

1 Introduction

It is well known that the climate is changing, and the impact of climate warming on polar regions is disproportionately severe due to Arctic amplification (Serreze and Francis, 2006). Arctic regions are warming 2-3 times faster than the global average

(Hinzman et al., 2005; Isaksen et al., 2022) which in turn is causing accelerated degradation of permafrost (Jorgenson et al., 2006). Degrading permafrost can have a substantial impact on local hydrology and ecological systems (Schuur and Mack, 2018) through the formation of thermokarst lakes, trough networks, and landscape deformation. In addition, rapid changes in landscape can also pose a threat to human populations in Arctic regions by destabilizing existing infrastructure (Nelson et al., 2002; Schneider von Deimling et al., 2021). Understanding the past, present, and future evolution of Arctic permafrost therefore plays a crucial role in assessing, predicting, and mitigating the impacts of climate change.

A significant challenge in monitoring permafrost thaw, however, is the general lack of high quality and accessible long-term observational data in the Arctic. Over the past few decades, circumpolar borehole temperature measurement sites have been established to monitor changes in permafrost environments. Many of these sites include borehole sensor arrays which provide measurements of ground temperature as deep as 50 meters or more below the surface. Recent studies have attempted to leverage these borehole data from the Global Terrestrial Network of Permafrost (GTN-P[1]), a database of shallow and deep ground temperature measurements taken from borehole sites across the Arctic (Burgess et al., 2000; Biskaborn et al., 2015), to quantify changes in permafrost temperatures at a global scale (Romanovsky et al., 2007; Biskaborn et al., 2019; Chen et al., 2021). These studies have consistently found that permafrost is warming globally ($0.01\,\mathrm{K\,yr^{-1}}$ to $0.06\,\mathrm{K\,yr^{-1}}$ in the continuous permafrost zone) in response to the rapidly changing Arctic climate.

Temperature, however, describes only one part of the thermal state of permafrost (Smith et al., 2022). A considerable fraction of energy can be consumed and released during thawing and freezing in the form of latent heat, thereby masking trends in the temperature signal within the critical range where phase change occurs (Riseborough, 1990; Romanovsky and Osterkamp, 2000; Romanovsky et al., 2010). This partitioning of energy can be expressed mathematically as the sum of two constituent energy contents:

$$H = H_S + H_L \tag{1}$$

where $H_S = TC(\theta(T))$ and $H_L = L\theta(T)$ [$\mathrm{J\,m^{-3}}$] are referred to as *sensible* and *latent* heat, respectively. Here T [K] is temperature, $\theta(T)$ [-] is (temperature dependent) volumetric liquid water content, $C(\theta)$ [$\mathrm{J\,K^{-1}\,m^{-3}}$] is volumetric heat capacity, and L [$\mathrm{J\,m^{-3}}$] is the latent heat of fusion of water. The relationship between sensible and latent heat, and thus consequently the rate of permafrost thaw, depends heavily on the non-linear freezing characteristic $\theta(T)$, often referred to as the *soil freezing characteristic curve* (SFCC) (Koopmans and Miller, 1966) for porous material. The SFCC governs the non-linear relationship between liquid water content and temperature during phase change, which arises due to the capillary pressure difference between frozen and unfrozen water present in the pore space. The functional form of the SFCC is empirically determined from soil properties, which typically requires measurements of soil moisture content over the critical temperature range. However, such measurements are not available at most borehole sites. The resulting lack of information on the distribution of energy in the ground represents a significant source of uncertainty in understanding what temperature trends observed in many GTN-P boreholes really tell us about the current thermal state of permafrost across regions. In order to understand how permafrost is changing, it is therefore vital that we find ways to quantify this uncertainty and develop physically coherent models to ac-

---

[1]https://gtnp.arcticportal.org/

curately link observed temperature trends to unobserved or sparsely observed quantities such as latent heat and active layer
55 thickness (ALT).

The problem of linking unobserved components of the thermal state to observed quantities (e.g., temperature) is fundamentally a problem of *inversion* under uncertainty. In other words, we wish to recover the full thermal state of permafrost (including latent heat) given observed borehole temperatures and some partial, *a priori* information about the local landscape and meteorological conditions. Bayesian inversion methods solve this problem probabilistically by combining information
60 from observations with prior knowledge in order to obtain a distribution over the unobserved variables. These methods are particularly attractive for geoscientific applications where observational data often come from systems that are neither stationary nor well-controlled (e.g., the local environment or the broader Earth system), thereby making traditional, frequency-based statistical inference difficult to apply and interpret correctly (Caers, 2018). As a result, Bayesian analyses have become increasingly common in the geosciences (Franks and Beven, 1997; Berliner, 2003; Qu et al., 2008; Caers, 2018) and, more recently,
65 also in cryospheric research (Wainwright et al., 2017; Gopalan et al., 2018; Verjans et al., 2020; Gregory et al., 2021). Unfortunately, full-fledged Bayesian inference using standard numerical sampling methods is generally infeasible for complex, simulation-based models where forward evaluation of the model is computationally expensive, which is the case for most transient models of dynamical processes (Cranmer et al., 2020). In this setting, theoretical and practical compromises must generally be made in order to obtain computationally feasible approximate inferences (Reich and Cotter, 2015). This applies
70 also to transient thermal models of permafrost, which typically require solving a discretized partial differential equation for heat diffusion in the forward evaluation (Riseborough et al., 2008). Finding computationally feasible methods for performing simulation-based inference is, therefore, a key methodological challenge in investigating the thermal state of permafrost with numerical models.

The primary objectives of this study are (i) to analyze the relationship between near surface air and permafrost temperature
75 trends observed at several research stations in the Arctic continuous permafrost zone, and (ii) to relate observed borehole temperature trends to plausible changes in latent heat over the last two decades via ensemble inversion of a transient heat conduction model with phase change. In both cases, we focus on the probabilistic quantification of uncertainty using Bayesian methods. To overcome the aforementioned computational challenges in solving the inverse problem for heat conduction, we apply a recently developed particle-based sampling method for approximate posterior inference (Garbuno-Inigo et al., 2020)
80 suitable for expensive numerical models. We apply our methods to four different sites (Fig. 1) located in different climate zones in northeastern Siberia, Svalbard, northern Alaska, and northwestern Canada, where long-term borehole temperature records (>10 years) as well as stratigraphic information are available.

2 Study sites

In order to cover a broad range of environmental and climatic conditions, we apply our inverse modeling method to borehole
85 and meteorological data obtained from sites in four different regions within the continuous permafrost zone. Fig. 1 shows the
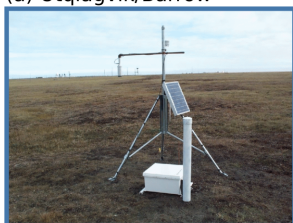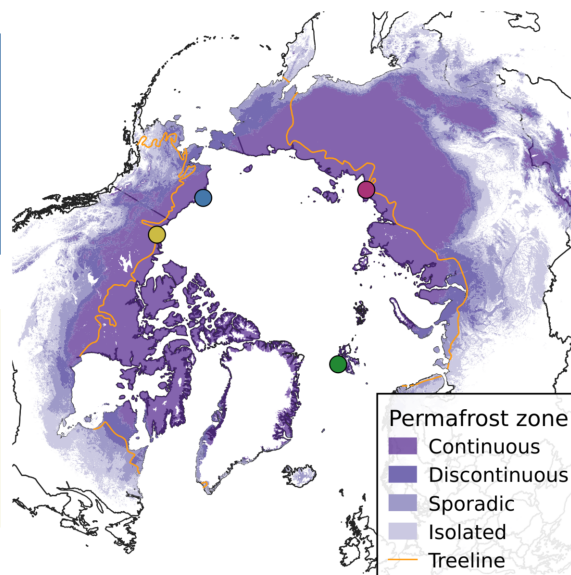
Figure 1. Map of the four study sites overlaid on the permafrost extent from Obu et al. (2019). Orange marks the tree line (Brown et al., 1997). Photos are reprinted/adapted with permission from published documentation for each site; Barrow (Bar, 2021), Samoylov Island (Boike et al., 2019), Bayelva (Boike et al., 2018a), and Parson's Lake (Smith et al., 2018) (permission granted by Open Government License - Canada: https://open.canada.ca/en/open-government-licence-canada).

geographical locations of each site along with photos of the local landscape. Table 1 provides a summary of available data on the long-term climatology of each site for a given reference period.

2.1 Samoylov

The Samoylov research station (Boike et al., 2019) lies on Samoylov Island (E 126° 28' 51.0" N 72° 22' 11.76") within the

90 Lena River Delta at the edge of the Laptev Sea in northeastern Siberia. The region is characterized by ice-wedge polygonal tundra, dense vegetation coverage, large water bodies, and sandy, ice-rich sediment. The site lies within the continuous permafrost zone and has permafrost reaching depths of $400\,\mathrm{m}$ to $600\,\mathrm{m}$ below the surface (Grigoriev, 1960) with active layer thickness ranging from $0.3\,\mathrm{m}$ to $0.7\,\mathrm{m}$. Recent surface temperature reconstructions show that the permafrost has been subjected to substantial warming over the last century (Kneier et al., 2018). The station provides automated monitoring of surface and climate

95 conditions in addition to an instrumented $27\,\mathrm{m}$ borehole drilled in 2006 from which hourly ground temperature measurements are available up until September 2021.

Table 1. Summary of climatological conditions for each of the four study sites included in this work. Air temperatures (for January and July) are monthly averages; rainfall refers to total annual unfrozen precipitation, and max snow depth refers to the maximum annual snow depth. Data are averaged over the reference period. All quantities were computed directly from published data for the indicated time period unless otherwise specified. Climatology data for Parson's Lake are taken from the nearest available long-term observation station at Trail Valley Creek which is about 37 km southeast of the borehole site.

Site name	Soil mixture	Period	Air temp. (Year, Jan, Jul)	Annual rainfall	Max annual snow depth	Reference
Samoylov	Peat, Sand	2002-2021	$-10.7\,°C$, $-29.8\,°C$, $9.3\,°C$	88 mm	37 cm	Boike et al. (2019)
Bayelva	Silt, Clay	1998-2021	$-4.3\,°C$, $-10.4\,°C$, $6.1\,°C$	197 mm	98 cm	Boike et al. (2018a)
Parson's Lake	Sand, Silt, Clay	1998-2019	$-7.9\,°C$, $-24.1\,°C$, $12.5\,°C$	145 mm[*]	40 cm[**]	ECCC (2011)
Utqiaġvik/Barrow	Peat, Silt	2004-2021	$-9.9\,°C$, $-23.7\,°C$, $4.4\,°C$	168 mm	50 cm[***]	Theisen (2022)

[*] Computed using ECCC data from the later time period of 2015-2021.

[**] Estimated visually from Figure 5 of Burn and Kokelj (2009) for the earlier time period of 2000-2007.

[***] Estimated visually from Figure 6 of Shiklomanov et al. (2010) for the earlier time period of 1994-2010.

2.2 Bayelva

The Bayelva station (Boike et al., 2018a) is located on the Brøgger peninsula of the island of Spitsbergen, Svalbard, approximately 3 km from the small research village of Ny-Ålesund (E 11° 50' 3.156" N 78° 55' 14.987"). The West Spitsbergen Current helps to maintain comparatively mild seasonal temperatures in the area around Ny-Ålesund. The unglaciated coastal areas are known to have continuous permafrost extending to depths as deep as 100 m (Humlum, 2005) with active layer depths of 1 m to 2 m. The site is situated on top of a hill, and the surrounding landscape is characterized by patterned permafrost ground and sparse (50-60%) vegetation coverage. The soils on the hill range from silty loam to silty clay with some marine deposits, while coarser grained silty-sands can be found in the surrounding area. The site includes an instrumented 9.0 m deep borehole with hourly measurements of ground temperatures dating back to 2009.

2.3 Parson's Lake

The Parson's Lake, "T5 Upland", borehole site (Wolfe et al., 2010; Smith et al., 2018) is situated east of the Mackenzie Delta, northeast of Inuvik in the Northwest Territories and within the continuous permafrost zone of northern Canada (W 133° 50' 17.159" N 68° 57' 29.879"). The region is classified as dwarf birch tundra with dense vegetation coverage and large water bodies (Burn and Kokelj, 2009). Soils in the area are mostly coarse-grained sand in the upper layers, with finer grained silty-clay soils in the deeper layers (> 1 m). Soil cores also show evidence of excess ground ice in moderate amounts (approximately 5-25% volumetric content) 1 m to 2 m below the surface. Permafrost extends down to 100 m or more in the outer delta area (Allen et al., 1988) and active layer thickness typically ranges from 0.35 m to 0.65 m. The site features a roughly 10 m deep borehole with average daily temperature measurements dating back to 2006.

2.4 Barrow

The Barrow North Meadow Lake 2 (NML2) borehole site (W 156° 38' 57.587" N 71° 19' 14.07") is situated near the village of Utqiaġvik (formerly known as Barrow) which lies at the northernmost point of Alaska along the Arctic coast. The surrounding landscape is characterized by drained thaw lakes and upland tundra consisting of gravel and silty soils with highly variable moisture and organic content in the active layer (Shiklomanov et al., 2010). The region lies within the continuous permafrost zone and is known to have permafrost extending down to $300\,\text{m}$ (Brewer, 1958) or more below the surface with highly variable active layer thickness ranging from $0.1\,\text{m}$ to $1.0\,\text{m}$. The upper permafrost layer is fairly ice rich with ground ice content as high as 50-75% by volume in some areas. In this work, we use data from a borehole at the North Meadow Lake 2 (NML2) site, where instrumented daily average soil temperature measurements in the permafrost layer down to $15\,\text{m}$ are sparsely available for the time frame 2012-2017 and annual measurements down to $47\,\text{m}$ are available from 2009-2021 (Romanovsky et al., 2002; Nelson et al., 2008; Bar, 2021).

3 Methods

3.1 Bayesian inference

The Bayesian approach to statistics provides a natural framework for inferring unobserved quantities of interest while simultaneously accounting for their associated uncertainties (Berliner, 2003). This is accomplished by applying Bayes rule to some observed and unobserved variables, Y and X, respectively:

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)} \quad \text{with} \quad p(Y) = \int\limits_{x \in X} p(Y|X = x)p(X = x)dx, \tag{2}$$

which can be seen a generic formula for obtaining the so-called *posterior* distribution of an unobserved quantity X *a posteriori* given observations Y from some sampling distribution or *likelihood* $p(Y|X)$. The prior distribution $p(X)$ encodes information about Y known *a priori* and plays a crucial role in the Bayesian inference workflow. Difficulties in the practical applications of Bayesian methods have historically arisen from the intractability of the integral in the denominator of (2), often referred to as the *marginal likelihood* $p(Y)$. However, advances in numerical sampling methods over the last few decades (Duane and Kogut, 1986; Bishop, 2006; Hoffman and Gelman, 2014) and the increase in available computing power have made Bayesian methods significantly more accessible.

3.2 Trend analysis

As a preliminary step of our study, we analyze trends in mean annual air and permafrost temperatures for each study site. For Samoylov, Bayelva, and Barrow, we use daily averaged in situ air temperature data where available, and bias-corrected ERA5-Land (Muñoz-Sabater et al., 2021) $2\,\text{m}$ air temperature in all years in the study period (2000-2020) where in situ observations are unavailable. The bias correction procedure for air temperature follows closely the empirical quantile mapping method of Piani et al. (2010); we deviate slightly by only applying the correction to values within the observed (in situ) temperature range

145   to avoid artificially clipping extreme temperatures in the reanalysis product. For the Parson's Lake borehole site, we use air
        temperatures from ERA5-Land without any correction since we were not able to find long-term in situ measurements near
        the station. We fit a two-stage, robust, Bayesian trend model to mean annual air and permafrost temperatures for all measured
        depths at each site. More details on the trend model and data preparation procedures can be found in Appendix B1.

3.3   Transient heat conduction model

150   In order to simulate the diffusion of heat in permafrost soils, we use a transient heat conduction model based on the same
        numerical procedures and parameterizations of CryoGrid (Westermann et al., 2016, 2022) which solves the 1-dimensional
        non-linear heat equation with phase change:

$$\frac{\partial H(z,t)}{\partial t} = \frac{\partial}{\partial z}\left[k_T(z,t)\frac{\partial T(z,t)}{\partial z}\right], \tag{3}$$

        where $H(z,t)$ [$\mathrm{J\,m^{-3}}$] is the discrete enthalpy function over d finite volumes $\Delta_i(z)$, $i = 1, \ldots, d$, defined on the depth domain

155     $z \in \mathcal{Z}$ with $\mathcal{Z} = [0\,\mathrm{m}, 1000\,\mathrm{m}] \subset \mathbb{R}$, and $k_T(z,t)$ [$\mathrm{W\,m^{-1}\,K^{-1}}$] is the (temperature dependent) thermal conductivity. Note that
        Eq. (3) is in so-called *mixed form*, i.e., the diffused variable (temperature) differs from the integrated variable (enthalpy). We
        can recover temperature from enthalpy via the following non-linear system:

$$T = \frac{H - L\theta(T)}{C(\theta(T))}, \tag{4}$$

        which can be solved using standard numerical methods, such as Newton iteration, or approximated using a pre-calculated

160   interpolant. Here $\theta(T)$ represents the SFCC for a given soil type. In this work, we use the SFCC specified by equations (21)-
        (23) in Dall'Amico et al. (2011) (see Appendix B2.2 for details), which is derived from the widely used van Genuchten (1980)
        soil-water retention equations.

        Since no general solution exists for Eq. (3) with arbitrary boundary conditions, numerical approximation is necessary in
        order to find a solution for $T(z,t)$. We discretize the spatial domain according to a fixed, non-uniform grid with the smallest

165   grid cell spacing of $0.05\,\mathrm{m}$ from the surface down to a depth of approximately $4\,\mathrm{m}$ and a pseudo-exponentially increasing grid
        spacing thereafter.

        We specify the soil stratigraphy for each site as stacked homogeneous layers, where the soil in each layer is parameterized
        according to the four characteristic fractions: excess ground ice (ξ_{xic}), natural porosity (ξ_{por}), saturation (ξ_{sat}), and organic
        content (ξ_{org}). From these independent fractions, the volumetric soil constituent fractions for excess ground ice (θ_x), pore

170   water/ice (θ_{wi}), air (θ_a), and mineral and organic soil fraction θ_m, θ_o can be computed via recursive partitioning of the volume,
        as shown in Fig. B1 and described in Appendix B2.2. Note that, while excess ground ice is represented through super-saturation
        of the soil volume (effectively just increasing the porosity), soil and water displacement from ground ice melt and hydraulic
        processes are neglected, and the total water/ice content $\theta_{wi} + \theta_x$ is assumed to remain constant throughout the simulation.

3.3.1 Initial and boundary conditions

175 Accurate simulation of the surface energy balance, snow dynamics, and surface hydrology is notoriously difficult and comes with many sources of uncertainty in landscapes with very heterogeneous surface characteristics (Langer et al., 2011a, b). An empirical alternative to approximate the bulk thermal effect of these complex surface processes is to use *n-factors* (Lunardini, 1978; Riseborough et al., 2008), which attempt to approximate the long-term deviation of surface temperature from air temperature via lumped scaling factors:

$$180 \quad n_F = \frac{\text{FDD}_\text{soil}}{\text{FDD}_\text{air}}, \ \ n_T = \frac{\text{TDD}_\text{soil}}{\text{TDD}_\text{air}}, \ \ \text{FDD} = \int_{t_1}^{t_2} T^-(t)dt, \ \ \text{TDD} = \int_{t_1}^{t_2} T^+(t)dt, \tag{5}$$

where $T^+(t) = \max(T(t), 0°C)$ and $T^-(t) = \min(T(t), 0°C)$. Note that FDD and TDD represent the commonly known freezing and thawing degree days. The resulting dimensionless factors $n_F, n_T \in [0, 1]$ can be used to approximate insulated surface temperatures by scaling the instantaneous air temperature:

$$T_\text{surf}(t) = \begin{cases} n_F(t)T_\text{air}(t) & T_\text{air}(t) \leq 0 \\ n_T(t)T_\text{air}(t) & T_\text{air}(t) > 0. \end{cases} \tag{6}$$

185 We use T_surf, with T_air derived from the same, bias-corrected ERA5-Land air temperature record described in Sect. 3.2, as the Dirichlet upper boundary condition for temperature in the transient heat conduction model. In order to account for possible decadal scale changes in surface conditions, the seasonal n-factors n_F and n_T are reparameterized as time-dependent, piecewise linear functions with knots at five pre-defined time points: 2000, 2005, 2010, 2015, and 2020. The values at each knot, e.g., $n_F^{(0)}, n_F^{(1)}, n_F^{(2)}, n_F^{(3)}, n_F^{(4)}$ (same for n_T), are varied as unknown parameters, thereby allowing T_surf, which is not generally

190 known or measured reliably, to be adapted according to the observed borehole temperatures. The geothermal heat flux at the lower boundary (1 km below the surface) is set to $0.053\,\text{W}\,\text{m}^{-2}$, and can generally be assumed to have a negligible effect on decadal time scales (Hermoso de Mendoza et al., 2020).

The initial temperature profile at each site is parameterized as a piecewise linear function between five knots: $(0\,\text{m}, \hat{T}_\text{air}(0))$, (z_ALT, T_1), (z_ZAA, T_2), $(z_\text{base}, 0°C)$, and $(1000\,\text{m}, T_3)$, where $\hat{T}_\text{air}(0)$ is the average annual air temperature for the first simu-

195 lation year, and T_1, T_2, T_3, and z_base (i.e. depth of the permafrost base) are varied as unknown parameters. z_ALT and z_ZAA nominally refer to the initial active layer thickness and depth of zero annual amplitude respectively; these values are set based on rough estimates from the relevant literature for each site. Each simulation additionally includes a 20-year "spin-up" period (1979-1999) excluded from analysis. The spin-up period allows the energy states of the upper subsurface layers to reach dynamic equilibrium with the forcing, while the parametric initial state broadly accounts for first-order uncertainty in the initial

200 temperature profile of the deeper layers.
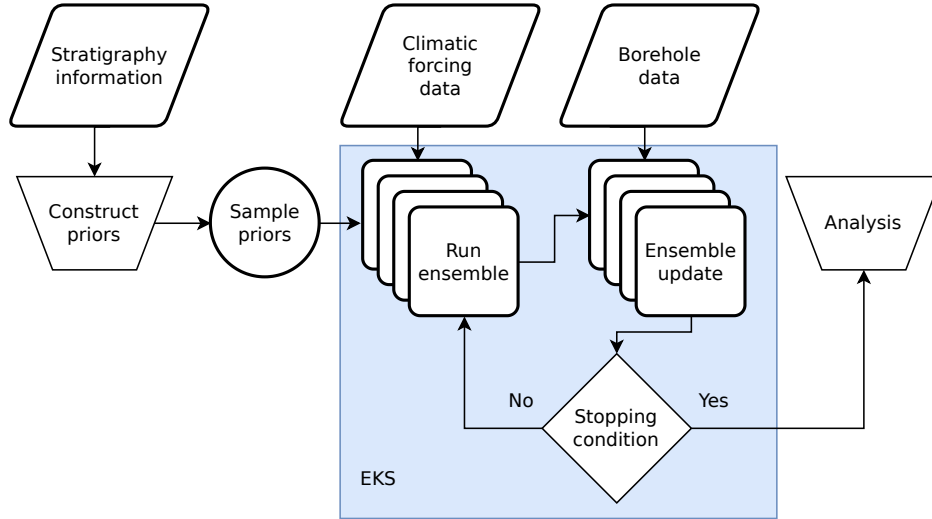
Figure 2. ISO standard flowchart showing a simplified overview of the inversion workflow using Ensemble Kalman Sampling (EKS).

3.4 Inversion method

3.5 Theoretical framework

The inversion problem of recovering some unobserved sequence of states $\mathbf{x}_{1:N} \in X^N$ from observations $\mathbf{y}_{1:N} \in Y^N$ over N temporal samples is commonly expressed via Bayes rule (Evensen et al., 2022):

205
$$p(\mathbf{x}_{1:N}, \boldsymbol{\phi}|\mathbf{y}_{1:N}, \mathcal{S}) \propto p(\mathbf{y}_{1:N}|\mathbf{x}_{1:N}, \boldsymbol{\phi}, \mathcal{S}) \, p(\mathbf{x}_{1:N}|\boldsymbol{\phi}, \mathcal{S}) \, p(\boldsymbol{\phi}) \, p(\mathcal{S}), \tag{7}$$

where $\phi \in \Phi$ represents the parameters characterizing the underlying hydrothermal processes in a dynamical model of the state variables, $\mathbf{x}$, and $\mathcal{S}$ represents any set of relevant exogenous variables (e.g., meteorological forcings, energy sources/sinks, etc), which define the boundary conditions of the system. We assume in Eq. (7) that the parameters, ϕ, are independent of the exogenous variables, $\mathcal{S}$, such that the joint prior density can be written as $p(\phi, \mathcal{S}) = p(\phi)p(\mathcal{S})$. Furthermore, we assume

210 the exogenous inputs, $\mathcal{S}$, to be constant, thereby allowing us to drop $p(\mathcal{S})$ from Eq. (7). The likelihood $p(\mathbf{y}_{1:N}|\mathbf{x}_{1:N}, \phi, \mathcal{S})$ determines how likely the observational data are, given a sequence of modeled states, while the distribution $p(\mathbf{x}_{1:N}|\phi, \mathcal{S})$, often referred to as the *process model*, relates a particular set of parameters ϕ to a sequence of states $\mathbf{x}_{1:N}$. In this work, we focus on probabilistically modeling parametric uncertainty induced by the prior $p(\phi) = \prod_{i=1}^{m} p_i(\phi_i)$ over m model parameters which encode *a priori* site-specific knowledge (e.g., soil characteristics and surface conditions) where available. The predictive

215 distribution over the model states can thus be expressed as:

$$p(\mathbf{x}_{1:N}|\mathbf{y}_{1:N}, \mathcal{S}) = \int_{\phi \in \Phi} p(\mathbf{x}_{1:N}|\phi, \mathcal{S}) \, p(\phi|\mathbf{y}_{1:N}, \mathcal{S}) d\phi \tag{8}$$

with the process model $p(\mathbf{x}_{1:N}|\phi,\mathcal{S}) = \delta\big(\mathbf{x}_{1:N} = f(\phi,\mathcal{S})\big)$, where f is the transient heat conduction model (described in Sect. 3.3) parameterized by ϕ, and δ is the Dirac delta distribution. This formulation of the problem departs somewhat from typical data assimilation procedures which tend to focus more on probabilistic modeling of the state, $\mathbf{x}$, with parametric uncertainty in ϕ often neglected (Caers, 2018; Geer, 2021). In contrast, our approach amounts to a kind of "perfect model" assumption where the dynamics are assumed to be error-free (though this is of course not really the case). This allows us to avoid the need to develop a stochastic description of $p(\mathbf{x}_{1:N}|\phi,\mathcal{S})$, while still being able to produce a distribution over plausible unobserved states via the parametric uncertainty induced by the prior. Estimating the predictive distribution in Eq. (8) requires integration over the parameter space Φ weighted by the posterior distribution $p(\phi|\mathbf{y}_{1:N},\mathcal{S})$ which is generally intractable. The integral can, however, be efficiently approximated via ensemble/particle-based sampling methods.

3.6 Approximate Inference via Ensemble Kalman Sampling

Due to the high computational cost of computing the likelihood, we employ a recently developed, probabilistic variant of ensemble Kalman inversion (EKI) (Iglesias et al., 2013) called *Ensemble Kalman Sampling* (EKS) (Garbuno-Inigo et al., 2020) to efficiently produce approximate samples from the posterior. The main advantage of EKS over more other commonly used numerical sampling methods, such as Markov Chain Monte Carlo (MCMC), is that it is a parallelizable, "ensemble" or particle-based method which requires only a handful of iterations to converge. In contrast, most MCMC methods require hundreds or even thousands of sequential (i.e., non-parallelizable) iterations. In contrast to EKI, which converges to a point estimate of the most likely parameters (i.e., a posterior mode), the ensemble produced by EKS can be interpreted as asymptotically unbiased samples from the stationary posterior distribution.

In order to apply EKS to our inversion problem, we assume that the observed borehole temperatures can be modeled as

$$T_{\mathrm{obs}} = (h_T \circ f)(\phi) + \eta, \tag{9}$$

where h_T is the forward map from the model states produced by f to comparable temperature observations, and $\eta \sim \mathcal{N}(0,\Sigma_T)$ is observation noise with zero mean and assumed covariance Σ_T. Additional inputs to f, such as forcing data, are assumed to be independent of ϕ and are here suppressed for brevity.

Since we aim to constrain the parameters ϕ of the heat conduction model f based on observed borehole temperatures, we choose h_T to be a simple aggregation function which computes mean annual ground temperatures (MAGT) from the daily temperature record produced by f. For each site, we compute the modeled MAGT at each depth where a borehole sensor is present, linearly interpolating between grid cells where necessary. The resulting predicted temperatures form a matrix $T_{\mathrm{pred}} = h_T \circ f(\phi) \in \mathbb{R}^{N \times \bar{d}}$, where N is the number of years and $\bar{d}$ the number of borehole sensor depths at which data are available. Note that, for the borehole data from Barrow, we were only able to make use of the annual measurements below the assumed depth of zero annual amplitude due to the sparse availability of instrumented measurements at depths closer to the surface. Model outputs at all points in time and space where no data are present are thus not explicitly included in the likelihood calculation, though it is important to note that they are still included implicitly through conditional dependence of the evaluation points on the time dynamics of Eq. 3.

250   The observed mean annual ground temperatures, T_{obs}, are used as the observation mean vector for the Ensemble Kalman Sampler described in Garbuno-Inigo et al. (2020). We set the observation covariance matrix Σ_T to the posterior predictive variances obtained from our trend analysis (see Sections 4.1 and B1) along the diagonal with a minimum variance of $0.01\,\mathrm{K}^2$ (to reflect an assumed baseline measurement uncertainty) and zero elsewhere, thereby assuming observation noise to be both independent and uncorrelated across time and space.

255   The prior distribution over model parameters, $p(\phi)$, is of crucial importance to our method. Some parameters in the heat transfer model, such as soil composition, will cause the resulting optimization problem on ϕ to be under-constrained, since there may be more than one possible combination of soil components which have similar thermal properties. Additionally, incorporating prior knowledge about plausible parameter values allows us to reduce the amount of computational effort wasted on physically implausible or incoherent model configurations that may arise from random sampling.

260   EKS assumes the m unconstrained parameters $\gamma(\phi) = \psi \in \Psi \subseteq \mathbb{R}^m$ to follow a multivariate Gaussian distribution, $\psi \sim \mathcal{N}(\mu_\psi, \Sigma_\psi)$, where $\gamma : \Phi \to \Psi$ is a bijective function which maps the m-dimensional possibly constrained parameters $\phi \in \Phi$ to their unconstrained values on the real line. We define our priors in the constrained parameter space Φ in order to more easily incorporate physically meaningful information about each site. We define suitable parameter priors for each site based on published field measurements and soil core analyses; full details on choices of priors for each site are in Appendix B2.3.

265 We construct a diagonal, multivariate Gaussian approximation of our priors in the unconstrained parameter space via Laplace approximation of the density induced by $\gamma(\psi)$. Note that this introduces some error into the EKS representation of the posterior since the resulting Gaussian distribution is only an approximation of the transformed density.

    We apply the implementation of EKS published by the authors (Constantinou et al., 2022) to our heat conduction model using borehole data obtained from each of the four sites described in Sect. 2. We use an ensemble of size $N_{ens} = 256$ (balancing between sample quality and computational cost), and we initialize the ensemble by sampling N_{ens} parameter vectors ϕ from

270 the prior. We iterate EKS until two units of ensemble/particle diffusion "time" are reached or for a maximum of 30 iterations. We refer the reader to Garbuno-Inigo et al. (2020) for further details about the EKS algorithm and iteration procedure.

4 Results

4.1 Permafrost and air temperature trends

275 The results of the trend analysis over the 20-year time period (2000-2020) described in Sect. 3.2 are shown in Fig. 3. The left panel shows the observed air and permafrost temperatures over time alongside the trend lines sampled from the posterior. The solid line represents the posterior mean while the shaded region shows the 95% central credible interval (CCI), which represents the interval between the 2.5% and 97.5% percentiles of the posterior samples. The box-plots in the right panel show summaries of the posterior distribution over slopes of the trend line, which can be more naturally interpreted as the average annual change

280 in temperature over the observed time period. The whiskers of the box-plot span the range $[q_1 - 1.5R, q_3 + 1.5R]$, where q_1 and q_3 are the first and third quartiles respectively and R is the interquartile range. All probabilities given in this section refer
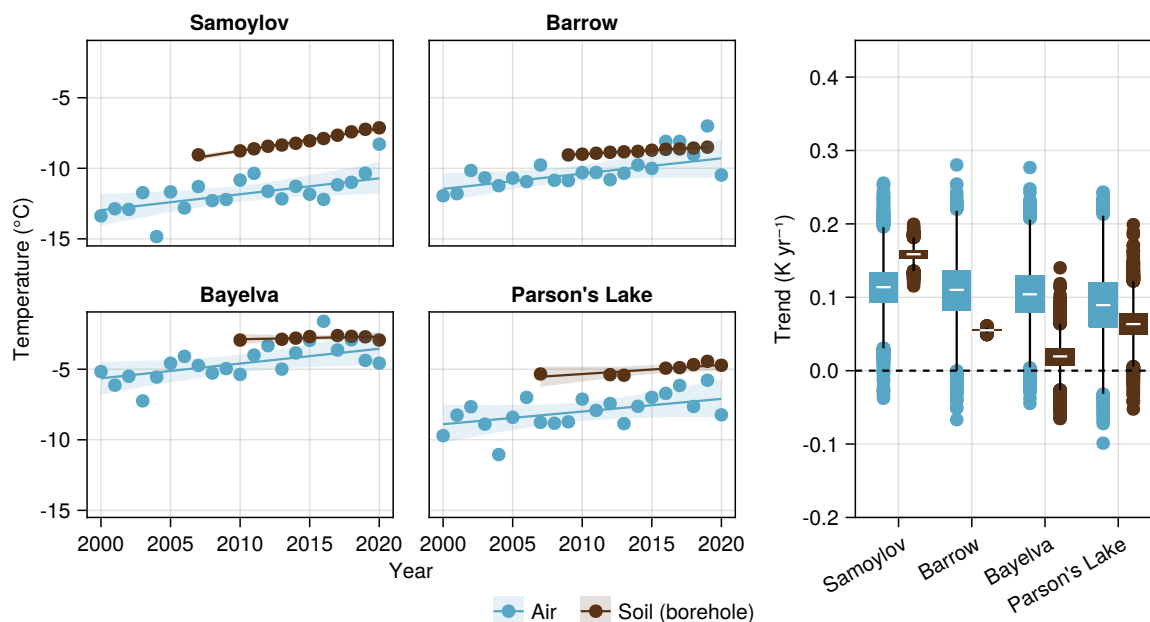
Figure 3. Trend Analysis: Observed mean annual temperature trends in air vs soil at all sites. Permafrost temperatures are from borehole measurements as close as possible to the depths of zero annual amplitude, 20.75 m (Samoylov), 20.0 m (Barrow), 9.0 m (Bayelva), 9.82 m (Parson's Lake). Dots represent observed mean annual temperatures at each location. For air temperature, the adjusted ERA-5 temperature reanalysis record is used where observations are not available. Trend lines and their corresponding 95% central credible intervals are obtained via numerical sampling from a robust trend model.

to posterior probabilities calculated from MCMC samples with Monte Carlo adjusted standard error (MCSE) strictly less than 2×10^{-3} in all cases.

Our analysis shows very strong evidence of increasing air temperatures (i.e. $\Pr(\mu_1 > 0) > 0.99$ for all sites) with the slope
285   of the trend line, μ_1, ranging from $0.09 \, \mathrm{K \, yr^{-1}}$ (95% CCI $0.00 \, \mathrm{K \, yr^{-1}}$ to $0.18 \, \mathrm{K \, yr^{-1}}$) at Parson's Lake to $0.11 \, \mathrm{K \, yr^{-1}}$ (95% CCI $0.04 \, \mathrm{K \, yr^{-1}}$ to $0.18 \, \mathrm{K \, yr^{-1}}$) at Samoylov Island. In comparison, mean annual permafrost temperatures near the depth of zero annual amplitude show similarly strong evidence of warming in three of the four sites; temperatures rose by approximately $0.16 \, \mathrm{K \, yr^{-1}}$ (95% CCI $0.14 \, \mathrm{K \, yr^{-1}}$ to $0.18 \, \mathrm{K \, yr^{-1}}$) on Samoylov Island, $0.06 \, \mathrm{K \, yr^{-1}}$ (95% CCI $0.05 \, \mathrm{K \, yr^{-1}}$ to $0.06 \, \mathrm{K \, yr^{-1}}$) at Barrow, and $0.06 \, \mathrm{K \, yr^{-1}}$ at Parson's Lake (95% CCI $0.02 \, \mathrm{K \, yr^{-1}}$ to $0.11 \, \mathrm{K \, yr^{-1}}$). For Bayelva, the observed deep permafrost
290   temperatures show a far milder warming trend of approximately $0.02 \, \mathrm{K \, yr^{-1}}$ (95% CCI $-0.02 \, \mathrm{K \, yr^{-1}}$ to $0.05 \, \mathrm{K \, yr^{-1}}$) and are considerably less conclusive ($\Pr(\mu_1 > 0) \approx 0.85$).

We next investigate the observed difference between soil and air temperature trends across sites. We find that soil temperature trends are generally lower than air temperature trends in both Barrow ($\Pr(\mu_1^{(\mathrm{soil})} < \mu_1^{(\mathrm{air})}) \approx 0.90$) and Bayelva ($\Pr(\mu_1^{(\mathrm{soil})} < \mu_1^{(\mathrm{air})}) \approx 0.98$). This relationship is less evident at the Parson's Lake borehole site ($\Pr(\mu_1^{(\mathrm{soil})} < \mu_1^{(\mathrm{air})}) \approx 0.68$) with ground
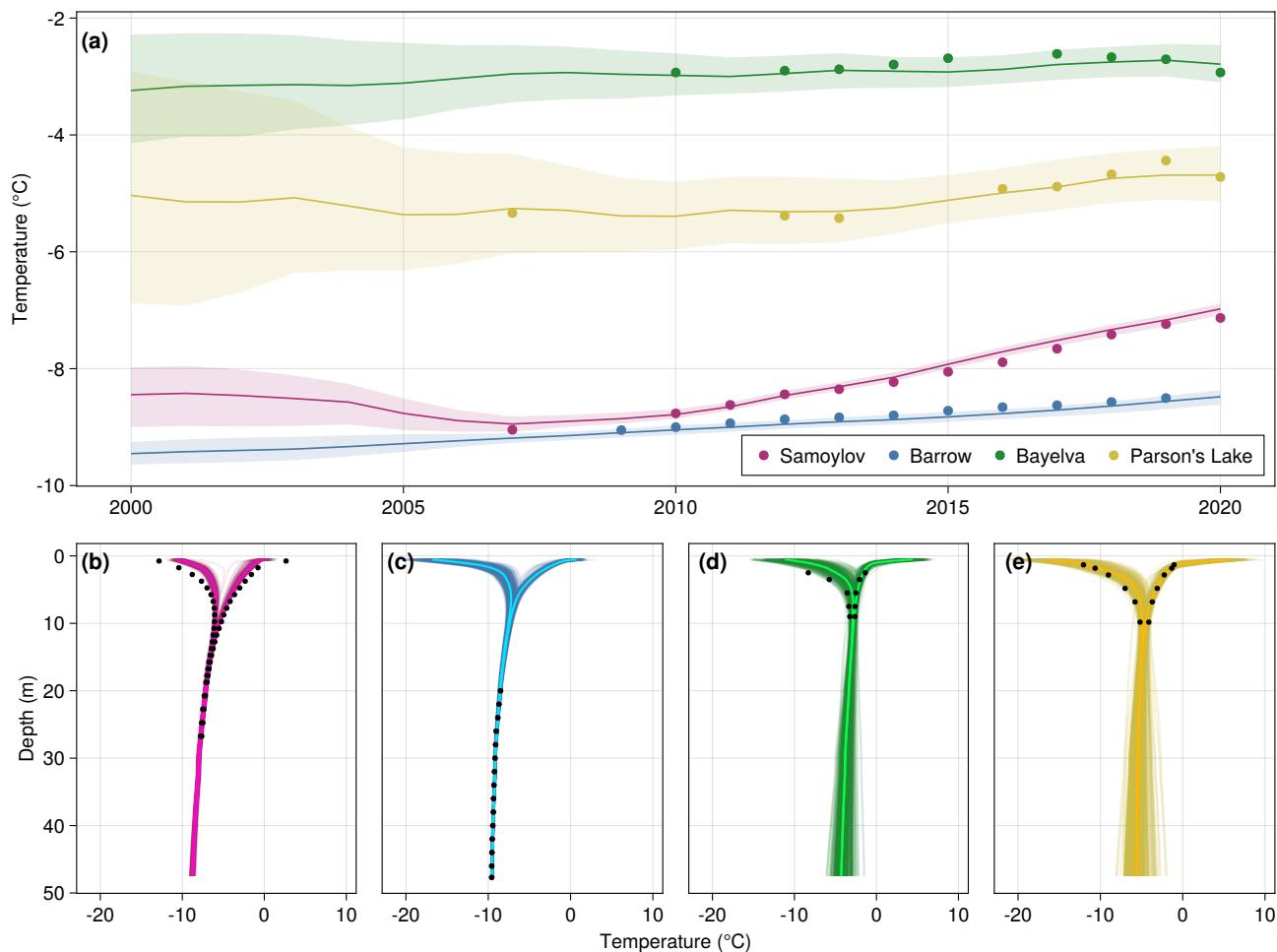
Figure 4. (a) Modeled vs observed mean annual ground temperatures at sensors which are as close as possible to the depth of zero annual amplitude; 20.75 m (Samoylov), 20.0 m (Barrow), 9.0 m (Bayelva), 9.82 m (Parson's Lake). Dots indicate observed annual borehole temperatures, and lines represent the mean of the ensemble predictions. Shaded regions cover 95% of the ensemble predictions. (b-e) Minimum and maximum annual temperature profiles at $2\,\mathrm{m}$ and below for the year 2020 (Samoylov, Bayelva, Parson's Lake) and 2019 (Barrow), observed (dots) vs model ensemble.

295   temperatures warming at a comparable rate to air temperatures. In stark contrast to the other sites, borehole measurements at Samoylov show deep soil temperatures warming on average $0.05\,\mathrm{K\,yr^{-1}}$ ($35\,\%$) faster than air temperatures with fairly high confidence ($\Pr(\mu_1^{(\mathrm{soil})} > \mu_1^{(\mathrm{air})}) \approx 0.92$).

4.2 Ensemble inversion

Modeled vs observed annual ground temperatures over the 20 year simulation period are shown in Fig. 4. The predicted temper-
300 atures from each ensemble member represents a (approximate) sample from the posterior predictive distribution, $p(T|T_{\text{obs}}, \mathcal{S})$.
The solid line in panel (a) is the ensemble mean temperature prediction while the shaded regions show the 95% CCI over
all ensemble members, i.e. the interval into which 95% of the predicted values for each ensemble member fall. Panels (b-e)
show the observed vs ensemble minimum and maximum annual temperatures for the last year in the simulation period where
observations are available. We can see in Fig. 4 that, for both Samoylov and Parson's Lake, the ensemble predicted tempera-
305 tures fit the observed borehole temperatures relatively well, although the 2020 annual min/max ground temperature curve for
Samoylov (Fig. 4b) indicates a slight bias in the min/max temperature range for the upper part of the profile affected by sea-
sonal variation. The ensemble temperatures for Bayelva also appear reasonable, though there is a clear warm bias in the winter
apparent in the temperature profile. The corresponding min/max temperature curves (Fig. 4d) reveal a substantial amount of
ensemble variation in the annual minimum and maximum temperature despite fairly modest variation in the ensemble mean at
310 9 m. This is also apparent in the 2020 ensemble temperature profile for Parson's Lake (Fig. 4e). There is much more variance
in the deeper part of the soil profile (i.e. below 10 m) for both of these sites in comparison to Samoylov and Barrow (Fig. 4b-e),
likely due to the lack of deeper borehole measurements.

Fig. 5 shows the joint and marginal distributions of average annual change in modeled latent heat (a) and thaw depth (b)
with average integrated ground temperature, i.e, the weighted average of grid cell temperatures, for each of the four sites. Here
315 average annual change is defined as the slope of the least squares trend line fitted over the entire simulation period (2000-2020)
for each ensemble member. We use integrated quantities in order to incorporate information from the entire upper 10 m of
the soil profile. The 10 m limit is somewhat arbitrary but is intended to balance the contributions of the active layer and the
permafrost layer to the aggregated quantities. Note that, since the model assumes constant hydrological conditions, changes
in latent heat can be directly attributed to phase change. The density plots on the top and right panels of Fig. 5 show the
320 ensemble marginal densities of temperature and latent heat change respectively. Each point in the plot represents the slope of
the least squares trend line for one ensemble member after fitting with EKS. The lines in the plot show the linear relationship
between changes in temperature and latent heat across all ensemble members. Table 2 shows the correlation and regression
coefficients associated with each linear fit. Stronger positive correlations between temperature and latent heat trends imply
higher sensitivity of permafrost thaw to changes in sensible heat uptake.
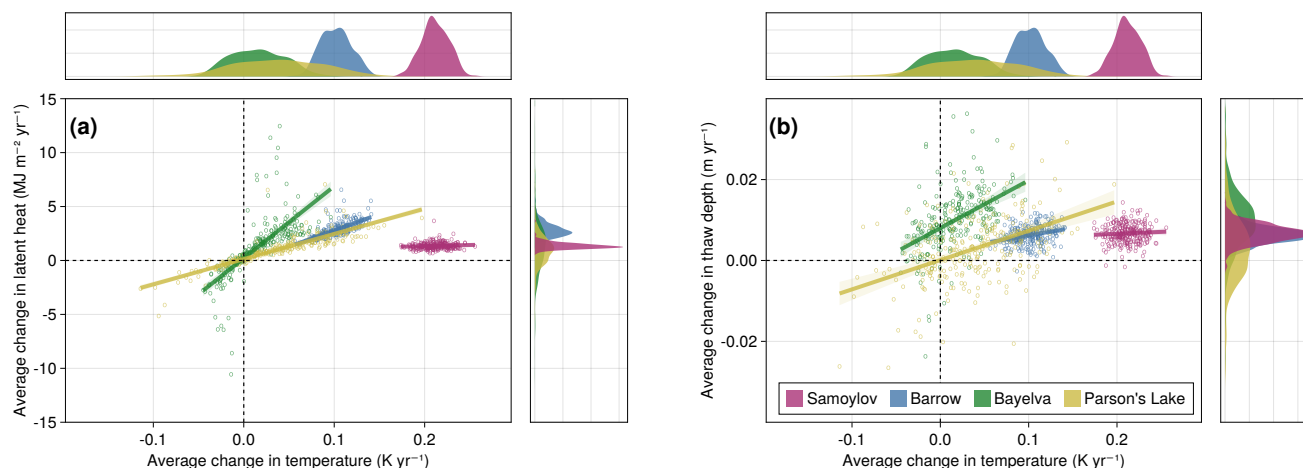
Figure 5. Joint densities of modeled mean annual change in latent heat (a) and thaw depth (b) vs mean annual change in ground temperature for all sites. Each point represents a single member of the fitted ensemble. Energy content and temperature are integrated over all grid cells in the upper 10 m of the soil profile. The density plots on the top and right panels of each plot show the marginal densities.

Table 2. Summary statistics (Pearson correlation and slope of the least squares line) for the linear relationships shown in Fig. 5 for both latent heat vs. temperature (a) and thaw depth vs. temperature (b). Means and standard errors are bootstrap estimates from 10,000 resamples of the posterior/ensemble trend values.

Site	Latent heat (a)			Thaw depth (b)		
	Correlation	Slope $(\mathrm{MJ\,m^{-2}\,K^{-1}})$	Std. dev. $(\mathrm{MJ\,m^{-2}\,yr^{-1}})$	Correlation	Slope $(\mathrm{m\,K^{-1}})$	Std. dev. $(\mathrm{m\,yr^{-1}})$
Samoylov	0.10 ± 0.05	2.0 ± 1.1	0.31 ± 0.03	0.06 ± 0.06	0.01 ± 0.01	$(2.0 \pm 0.1) \times 10^{-3}$
Barrow	0.71 ± 0.05	30.2 ± 1.6	0.71 ± 0.05	0.27 ± 0.06	0.04 ± 0.01	$(2.3 \pm 0.1) \times 10^{-3}$
Bayelva	0.73 ± 0.03	66.8 ± 3.9	2.71 ± 0.21	0.47 ± 0.05	0.12 ± 0.01	$(7.4 \pm 0.5) \times 10^{-3}$
Parson's Lake	0.86 ± 0.04	23.4 ± 1.1	1.38 ± 0.10	0.45 ± 0.06	0.07 ± 0.01	$(8.3 \pm 0.5) \times 10^{-3}$

325   We can see in Fig. 5a that the all three sites with silty or clay-like soils (i.e. Barrow, Bayelva, and Parson's Lake) exhibit a much stronger positive correlation between change in latent heat and temperature. Samoylov, in contrast, shows little to no correlation despite the comparatively rapid increase in ground temperature. Fig. 5b shows a clear difference in how changes in thaw depth correlate with changes in temperature between the two warmer sites, Parson's Lake and Bayelva, versus the two colder sites, Samoylov and Barrow. It is evident that the simulations for both Parson's Lake and Bayelva show a stronger

330 positive correlation between increasing thaw depths and increasing temperature, whereas for the colder sites this correlation is weaker (Barrow) or virtually non-existent (Samoylov). Both warmer sites also show more overall variability across the
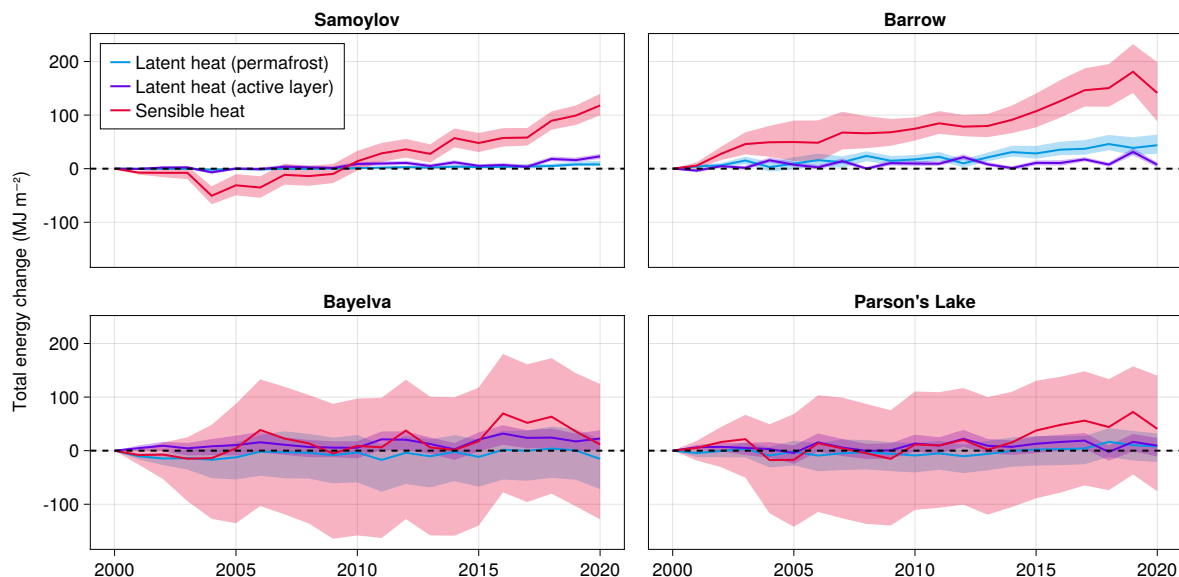
Figure 6. Total change in energy partitions (upper $10\,\mathrm{m}$ of soil profile) since the beginning of simulation period. Energy is partitioned into three categories: Latent heat in frozen grid cells (i.e. cells with maximum annual temperature $T_{\max} < 0\,^{\circ}\mathrm{C}$), latent heat in the active layer ($T_{\max} \geq 0\,^{\circ}\mathrm{C}$), and sensible heat. Solid lines show the median energy change while the shaded regions show the 95% CCI over the ensemble.

ensemble in both latent heat and thaw depth trends with standard deviations close to or sometimes more than double those of the two colder sites.

Lastly, Fig. 6 shows the long-term change in the mean annual energy contents of the upper $10\,\mathrm{m}$ of the soil profile over the
335   full simulation period (2000-2020) for each site. The total energy is partitioned into three parts: latent heat stored in "frozen" grid cells, i.e., those that have a maximum annual temperature below $0\,^{\circ}\mathrm{C}$, latent heat in the active layer, and sensible heat. The solid lines show the median energy change over all fitted ensemble members, and the shaded regions indicate the corresponding 95% CCI. Note that the categorization of a grid cell as "frozen" or "active" for the latent heat partitions is not constant; i.e., as the active layer grows or shrinks over time, the categorization of each grid cell based on maximum annual temperature changes
340   as well. This means that the change in latent heat for each partition is attributable not only to the thawing and freezing of water but also to the migration of grid cells between the two partitions.

5   Discussion

5.1   Estimating trends in observed permafrost and air temperatures

Numerous recent studies have highlighted warming trends observed in borehole sites all across the Arctic (Romanovsky et al.,
345   2007; Biskaborn et al., 2019; Chen et al., 2021; Smith et al., 2022). The results of our trend analysis are broadly consistent with

Biskaborn et al. (2019) who reported average decadal changes in permafrost temperatures of $(0.39 \pm 0.15)\,\mathrm{K\,dec^{-1}}$ across the continuous permafrost zone between 2007 and 2016, which translates into an average annual change of roughly $0.02\,\mathrm{K\,yr^{-1}}$ to $0.05\,\mathrm{K\,yr^{-1}}$. Observed warming trends at the Samoylov Island borehole site in the Lena River Delta (Siberia) over the longer period of 2007-2020 are unusually extreme (also noted by Biskaborn et al. (2019) for the same borehole) with a mean annual

350   change of $0.16\,\mathrm{K\,yr^{-1}}$, while observed changes at the Bayelva site near Ny-Ålesund (Svalbard) fall towards the lower end of the expected range with a mean annual change of $0.02\,\mathrm{K\,yr^{-1}}$. For $2\,\mathrm{m}$ air temperature, Biskaborn et al. (2019) report average changes of $(0.86 \pm 0.84)\,\mathrm{K\,dec^{-1}}$ ($0.02\,\mathrm{K\,yr^{-1}}$ to $0.17\,\mathrm{K\,yr^{-1}}$) across the continuous permafrost zone from 2007-2016. This is consistent with the results of our analysis which show mean air temperature trends ranging from $0.09\,\mathrm{K\,yr^{-1}}$ at Parson's Lake to $0.11\,\mathrm{K\,yr^{-1}}$ on Samoylov Island. The large difference in observed permafrost temperature trends between these two

355   sites, despite similar changes in air temperature, indicates considerable uncertainty in how permafrost is responding to the changing climate. Furthermore, the observation that deep permafrost on Samoylov Island is most likely warming faster than air temperature suggests that changes in air temperature alone cannot always fully explain permafrost warming.

These results motivate our inverse modeling study by demonstrating clear, localized differences and substantial uncertainty in how the permafrost thermal regime responds to long-term changes in air temperature. The discrepancies in the apparent

360   relationship between long-term changes in air and permafrost temperatures suggest that other factors are at play, such as surface conditions (e.g. snow cover) and variability in soil thermal properties. These factors significantly affect energy uptake in the subsurface, and ultimately, the current and future thermal state of permafrost in Arctic regions (Smith et al., 2022; Langer et al., 2022). We believe that this can be at least partially attributed to the latent heat effect, in addition to soil thermal properties, both of which are a major source of uncertainty in making inferences about the subsurface thermal regime (Riseborough, 1990;

365   Romanovsky and Osterkamp, 2000).

5.2   Variability and biases in the ensemble modeled temperature record

There is substantial variability in modeled temperatures over the simulation period (2000-2020) even after using EKS to calibrate the model ensemble to borehole measurements (Fig. 4). This variability reveals fundamental uncertainties about what observed changes in permafrost temperatures actually tell us. We see, for example, significant differences in the overall

370   spread of predicted temperatures between sites. This can be partially explained by differences in the depths of the observed temperatures; e.g., the deepest sensors available at the Bayelva and Parson's Lake sites are at $9\,\mathrm{m}$ and $9.82\,\mathrm{m}$ respectively, whereas both Barrow and Samoylov have much deeper measurements available near the depth of zero annual amplitude where there is little to no impact from seasonal variation. Important to note, however, is that the ensemble temperature spread in Bayelva is less than that of Parson's Lake, despite the sensor being roughly $82\,\mathrm{cm}$ closer to the surface, suggesting that depth

375   alone likely cannot explain these differences in variability. The wider spread in modeled temperatures at both Bayelva and Parson's Lake, particularly in the deeper parts of the soil profile, seems to indicate that there is more uncertainty in modeling the thermal dynamics of warmer permafrost. This may be due to higher sensitivity to soil properties, initial conditions, and changes in surface conditions and soil water content. We suspect that this sensitivity is in large part attributable to the nonlinear

effects of the freeze curve on the thermal dynamics, especially in deep permafrost (i.e. near the depth of zero annual amplitude)
380   where thermal gradients are smaller and heat diffuses more slowly.

The general agreement between the modeled and observed mean annual temperatures in Fig. 4 suggests that the fitted model
ensemble can provide a plausible estimate of the permafrost's thermal state at all four sites, subject to some uncertainties and
limitations which are discussed further in Sect. 5.6. There are, however, clear discrepancies between the model and obser-
vations, particularly in the temperature profiles for Samoylov and Bayelva, which should be considered. For Samoylov, the
385   ensemble median of the modeled annual temperature range of the permafrost layers is slightly too narrow, showing a warm
bias in the winter and a cold bias in the summer for the last simulation year of 2020 (Fig. 4b). This may be due the thermal
conductivities of the upper layers being too low, thereby over-insulating the lower. It is worth noting that, while some soil
properties are varied in the ensemble, the stratigraphy itself is not, which may be responsible for such biases. For Bayelva, we
suspect that the wintertime warm bias may be due to the model's assumption of static hydrology, which fails to capture the
390   effects of latent heat being lost due to drainage and evapotranspiration. We can see quite clearly from the volumetric water
content measurements shown in Figure 3 of Boike et al. (2018a) that soil saturation levels in the active layer vary substantially
both within and between years. It has also been well established by previous studies that hydrological conditions have a sub-
stantial effect on the subsurface thermal regime on Svalbard (Roth and Boike, 2001; Westermann et al., 2011). Nevertheless,
we believe the results of our inversion procedure agree well enough with temperature observations to provide insight into how
395   the thermal state of permafrost at each site is evolving.

5.3 What can temperature trends in deep boreholes tell us about the past?

It is worth noting that the spread in modeled temperatures across the ensemble is generally higher for earlier years (i.e. 2000-
2006), which is to be expected given the lack of observations to constrain the model. In many cases, the 95% CCI includes some
simulations where ground temperatures were actually warmer in the past, yet still produce temperature records comparable to
400   observations from the last two decades. This is particularly true for Samoylov, where nearly all of the ensemble members
show relatively stable permafrost temperatures for the first five years of the simulation period (2000-2005), which is the first
change point for the fitted non-stationary n-factors, which mimic changes in insulation effects at the surface (see Sect. 3.3.1
for details). This suggests that recent warming trends can be plausibly explained by changes in surface conditions rather than
longer term warming. This hypothesis is further supported also by the posterior distributions of the wintertime n-factors shown
405   in Fig. A1, which demonstrate clear temporal changes in the simulations for both Samoylov and Bayelva. It is important to
recognize, however, that conclusively verifying or falsifying this hypothesis would require the incorporation of snow depth
records, and ideally further historical borehole measurements (i.e., prior to 2005), into the inversion procedure, which we have
not attempted in this study. Previous studies of longer term borehole measurements dating back to the mid 20th century or
earlier (Romanovsky et al., 2007) as well as attempts to reconstruct historical surface temperatures in eastern Siberia (Kneier
410   et al., 2018) have shown that historical permafrost temperatures were substantially colder during the previous centuries. Our
results should nevertheless serve as a reminder that longer term, historical temperature trends should not be extrapolated from

recently observed temperature trends in deep boreholes without first considering other drivers of heat exchange at decadal and sub-decadal scales, such as the insulating effects of snow cover and vegetation.

5.4 Linking temperature trends to latent heat and thaw depth

415 The clear differences between the joint distributions for each site in Fig. 5 have substantial implications for the interpretation of permafrost temperature trends. Firstly, it is evident that the freezing characteristics of the soil, here represented through variation of the soil freezing characteristic curve parameters, have a substantial impact on the relationship between trends in temperature and latent heat. This is made most apparent in Fig. 5a by the presence of a strong linear relationship between latent heat change and temperature change in Barrow but not in Samoylov. Both sites are similar in many ways; they both have

420 historically very cold permafrost (below $-8\,°\mathrm{C}$ near the depth of annual amplitude), an insulating organic layer at the surface, and similar annual snow cover. The main difference is that the area around the Barrow borehole has silty soil (Romanovsky and Osterkamp, 2000) with freezing characteristics that allow for large amounts of unfrozen water at temperatures well below zero. The result is that, as permafrost close to the active layer warms, a larger amount of energy is consumed as latent heat, with less being diffused downward into the colder permafrost below. This is also reflected in Fig. 6, where the long-term latent

425 heat increase in the permafrost layer for the Barrow simulation clearly dwarfs that of Samoylov.

The joint density of thaw depth trends with temperature trends in Fig. 5b tell a similar but slightly different story. As with latent heat, simulations for the three sites with silty or clay-like soils in the permafrost layers show the strongest positive correlation while those for Samoylov show little to none. However, the correlation and effect size for the Barrow simulations are notably lower than those for the two warmer sites, Bayelva and Parson's Lake. We suspect that this is likely a result of the

430 colder, deep permafrost at Barrow acting as a counter balance to thaw.

We also note that the vast majority of ensemble members for both Samoylov and Barrow show clear warming trends whereas the marginal distributions of temperature change for Bayelva and Parson's lake are more diffuse and tend to be closer to zero. This is broadly consistent with the hypothesis of Riseborough (1990), i.e. that regions with warmer permafrost may show less detectable subsurface warming while actually being more at risk for thawing. The two warmer sites (particularly Bayelva) also

435 tend to show similar or larger changes in latent heat content and thaw depth even with relatively little change in temperature; this is likely due to a larger portion of energy being consumed as latent instead of sensible heat. This hypothesis is further supported by Fig. 6, which shows sensible heat as the clear dominant component of long-term energy change in Samoylov and Barrow. We discuss this further in Sect. 5.5.

A small number of ensemble members for both of these warmer sites lie in the upper left quadrants of Figs. 5a and 5b, i.e.,

440 where there is a positive trend in latent heat or thaw depth despite average ground temperatures actually declining over the simulated time period. Similarly, some ensemble members for these two sites fall into the lower two quadrants, indicating a decrease in latent heat or thaw depth with some parameter settings. This is in contrast to the two colder sites, Samoylov and Barrow, where nearly all ensemble members show positive trends in temperature, latent heat, and thaw depth. This underscores two features of warmer permafrost: (i) that observed temperature trends (or lack thereof) should be interpreted with caution, as

445 there is substantial uncertainty inherent in associated changes in latent heat, and (ii) that the sensitivity of warmer permafrost

to climatic changes can also imply rapid refreezing and contraction of the active layer under cooling conditions. We can expect the realistic impact of this effect, however, to be tempered by water drainage, which is not accounted for in our model.

5.5 Latent heat as a subsurface energy sink

While it is clear that Arctic permafrost is warming (Smith et al., 2010; Biskaborn et al., 2019; Chen et al., 2021; Smith et al., 450 2022), the latent heat effect acts on a larger scale as both an energy source and sink (Nitzbon et al., 2022), which can have the effect of masking long-term temperature changes in the climate signal (Riseborough, 1990), thereby introducing considerable uncertainty into the long-term implications of warming across the Arctic. Fig. 6 visualizes this uncertainty by comparing long-term changes in the partitioned subsurface energy budget across sites. There is a clear contrast between the two colder sites, Samoylov and Barrow, in the upper panels versus the two warmer sites, Bayelva and Paraon's Lake, in the lower panels. For 455 the colder sites, sensible heat change plays a dominant role with both the median and full 95% CCI of the ensemble showing a positive trend for most of the simulation period. The decrease in sensible heat at Barrow in the last simulation year, 2020, is likely due to the unusually large drop in air temperature which can be seen in Fig. 3.

Long-term change in latent heat uptake also differs between the colder and warmer sites. For both Samoylov and Barrow, a vast majority of ensemble members show an increase in latent heat contents both in the "frozen" partition (i.e., grid cells 460 where maximum annual temperatures are below zero) as well as in the active layer. This points to both a deepening active layer (consistent with Fig. 5) as well as a warming permafrost layer which is transitioning into a more "mushy" state as temperatures increase into the critical region of the freezing characteristic curve, where liquid water is present at temperatures below the nominal melting point. In contrast, both Bayelva and Parson's Lake show a much larger ensemble spread in latent heat changes for both partitions. Notably, however, the 95% credible ensemble spread for both sites (especially Bayelva) indicates a higher 465 probability of decreasing latent heat content in the frozen partition which is likely attributable to expansion of the active layer. It is also worth considering that both warmer sites also consist of more silty and clay-like soils which are often characterized by higher unfrozen water content at temperatures well below the melting point (Koopmans and Miller, 1966; Dall'Amico et al., 2011). The highly non-linear behavior of permafrost temperature changes due to the impacts of latent heat on the thermal dynamics thereby limits the usefulness of temperature records in evaluating the impact of climate warming on the thermal state 470 of permafrost.

5.6 Limitations

While we are confident that the ensemble inversion method and results presented in this study provide valuable insight into the relationship between air temperature trends, subsurface temperature trends, and permafrost thaw, there are several limitations of our methodology which are important to address.

475   1. **Neglected subsurface processes**. The transient heat conduction model described in Sect. 3.3 is capable of accurately simulating heat conduction with phase change at a relatively high spatial resolution. However, it neglects several processes such as water infiltration and percolation, excess ice melt and subsidence, as well as lateral exchange of heat and

water in the subsurface. In particular, the assumption that soil water content remains constant over the simulation period precludes factoring in the effects of long-term wetting or drying on the thermal properties (and therefore the thermal regime) of the ground closer to the surface. However, the assumption allows us to more objectively compare changes in the partitioning of latent vs. sensible heat in the subsurface (Fig. 6). We further note that for some sites, such as Samoylov which is typically waterlogged, no substantial wetting or drying has been reported during the simulation period (Boike et al., 2019).

2. **Simplified parameterization of surface processes**. We use only n-factor scaled air temperatures as the upper boundary condition for the heat equation. Our model does not explicitly represent surface processes such as radiative and turbulent heat fluxes, water bodies, vegetation and snow dynamics, but represents them in bulk via n-factors (Lunardini, 1978; Riseborough et al., 2008). Thus, inter-annual variability in surface conditions is not effectively represented, and meteorological forcings other than air temperature are not taken into account, thereby neglecting potentially valuable sources of information that would serve as additional constraints on the inverse problem.

3. **Under-constrained tuning of n-factors**. The parametric approach used for the n-factor scheme in our model (see Sect. 3.3.1), while powerful and flexible from a model calibration point of view, has the additional disadvantage of potentially creating non-physical or unrealistic surface conditions which are not supported or constrained by observational data. The optimization or sampling algorithm (in this work, EKS) has the freedom to adjust the half-decadal n-factors at the upper boundary in order to produce annual ground temperatures that better match the borehole observations, with the only constraint being the prior distributions. Since the focus of this work is not to generate realistic surface conditions but rather to understand the relationship between changes in sensible and latent heat changes at various sites, we consider this limitation to be acceptable. However, this is a problem that should be addressed in order to use our methodology to make inferences about likely causal drivers behind warming or long-term hydrothermal changes in the active layer.

4. **Selection and calibration of prior distributions**. Prior distributions provide a flexible tool for encoding domain-knowledge into inverse problems and, most importantly, can act as smooth regularizers in otherwise ill-posed or non-convex optimization problems. The model parameter priors employed in this work are loosely derived from both auxiliary data sources as well as field work and published values in the literature (see Appendix B2.3 for further discussion). However, the general lack of precise error and uncertainty estimates for some model parameters (in particular, soil properties) makes the selection of appropriate prior distributions difficult. We nevertheless emphasize that even imperfect (but plausible) prior distributions still do a better job of accounting for uncertainty than arbitrary point estimates of parameter values. Ideally, we would perform a sensitivity analysis to assess the impact of the priors on the results, but the large number of model parameters and computational cost of the dynamical model makes such comprehensive analyses prohibitively difficult.

5. **Other statistical considerations**. While EKS provides a computationally efficient alternative to MCMC for drawing samples from the posterior, it is nevertheless an approximation which provides a theoretical guarantee of convergence

to the posterior measure only over an infinite "time" horizon (i.e., iterations of the algorithm) (Garbuno-Inigo et al., 2020). Furthermore, the empirical results presented by the authors of EKS indicate that the method underestimates posterior variance (and thus uncertainty) on a finite time horizon unless the ensemble is very large. More recent work by Cleary et al. (2021) has attempted to circumvent this problem by performing exact posterior inference on emulated model outputs. Additionally, EKS requires specification of the observation noise covariance Σ_T, which is not, in general, known. Ideally, a fully Bayesian treatment of the inversion problem would include Σ_T as parameters to be inferred, which would have the benefit of producing an ensemble with a predictive distribution that is well calibrated on the training data. This is, however, not possible using EKS, and thus, we leave the problem of estimating noise parameters to future work.

These limitations provide a wide range of directions for future work; e.g. (1-3) could be addressed simply by incorporating additional physical processes into the model, such as those described in Westermann et al. (2022). However, this typically comes at a heavy computational cost, limiting the feasibility of running large ensembles over longer time periods spanning decades, centuries, or millenia. We are optimistic that recent advancements in simulation-based inference (Cranmer et al., 2020) and machine-learned based emulators may be able to address this challenge, as well as those described by (4) and (5). We also see a lot of potential in applying the inversion method outlined in this work at a larger, cirum-Arctic scale, perhaps by leveraging more efficient formulations of the heat conduction model such as those proposed by Tubini et al. (2020) or Langer et al. (2022). The greatest challenge here is the limited availability of high quality borehole data at a global scale, despite the ongoing efforts of collaborative data collection and publication projects such as GTN-P (Biskaborn et al., 2015).

6 Conclusions

In this work, we analyzed trends in annual air and permafrost temperatures at four different sites in the Arctic continuous permafrost zone. We found that long-term trends in permafrost temperatures over the study period (2000-2020) varied between the four sites ($0.02\,\mathrm{K\,yr^{-1}}$ to $0.16\,\mathrm{K\,yr^{-1}}$) more than trends in air temperatures ($0.09\,\mathrm{K\,yr^{-1}}$ to $0.11\,\mathrm{K\,yr^{-1}}$). In order to investigate what permafrost temperatures can tell us about the thermal state of permafrost, we then applied a recently developed method for ensemble inverse modeling, Ensemble Kalman Sampling (Garbuno-Inigo et al., 2020), to constrain a transient heat conduction model using long-term borehole temperature measurements from each site. We demonstrated that the response of the subsurface thermal regime to warming at the surface varies substantially between sites based on their historical climatology (i.e., colder versus warmer conditions) as well as the thermal properties of the soil. We found that the degree to which trends in permafrost temperatures correlate with changes in latent heat or thaw depth depend heavily on both historical climatology and the soil type. In particular, the sensitivity of modeled maximum annual thaw depths to changes in temperature trends were higher for the two sites with warmer permafrost, Parson's Lake in northwestern Canada ($0.07\,\mathrm{m\,K^{-1}}$) and Bayelva in Svalbard ($0.12\,\mathrm{m\,K^{-1}}$), versus the two colder sites, Barrow in northern Alaska ($0.04\,\mathrm{m\,K^{-1}}$) and Samoylov in northeastern Siberia ($0.01\,\mathrm{m\,K^{-1}}$). Furthermore, we showed that there is substantial uncertainty inherent in relating ground temperature trends to permafrost thaw in areas with warmer permafrost or silty/clay-like soils where the non-linear effects of latent heat play a much larger role. Our results suggest that currently observed trends at these sites can be plausibly mapped to a wide

range of different thermal states. This is largely due to the lack of information in the temperature signal of permafrost within the
critical temperature range where phase change of water occurs. The findings highlight the importance of considering regional
differences in soil characteristics and climatology when comparing permafrost temperature trends between sites. Our study also
highlights the importance of deep borehole measurements (i.e. deeper than $10\,\mathrm{m}$) in constraining permafrost thermal models
and providing insights into the past and present thermal state of permafrost in the Arctic. Future work will focus on the broader
application of probabilistic inversion methods, like those that used in this work, both to more sophisticated permafrost models
and at larger spatio-temporal scales.

Code and data availability. Preprocessed/collated data files (excluding those from Parson's Lake) in addition to the code for this study
are accessible online, at https://doi.org/10.5281/zenodo.6821348. Code for the transient heat conduction model is also available online, at
https://doi.org/10.5281/zenodo.6801740 (Groenke et al., 2022). Borehole data from the Parson's Lake borehole can be requested directly
from the Geological Survey of Canada/Natural Resources Canada (Smith et al., 2018). Original borehole data from the Samoylov Island and
Bayelva sites can be retrieved from the open-access Pangaea data store (Boike et al., 2017, 2018b). Orginal borehole data from the Barrow
site can be retrieved from the University of Alaska Fairbanks Permafrost Laboratory website (Bar, 2021).

Appendix A: Additional results

A1 Posterior distributions of wintertime n-factors indicate possible changes in snow cover

The posterior distributions of the wintertime n-factors are shown in Fig. A1. Each group of box-plots represents the fitted value
of the n-factor knot at that year. As described in Sect 3.3.1, the actual n-factor values are interpolated linearly between knots,
forming a piecewise-linear function over the simulation period. Values for the year 2000 are also applied for all years in the
spin-up period (1979–1999). The inter-annual differences between the knots indicate that, in addition to the long-term trend in
air temperature, changes in surface conditions are necessary in order to accurately reproduce temperature trends in the borehole
data.

Decreasing wintertime n-factors for Samoylov between 2005–2020 can be physically interpreted as an increase in surface
insulation, typically explained by an increase in snowfall or changes in the timing of the snow covered season. The extent
to which long-term snow depth measurements published in Boike et al. (2019) support this hypothesis is unclear, and further
validating this hypothesis will likely require further investigation with models that take into account snow dynamics. We briefly
discuss the unusually rapid warming observed on Samoylov Island in the following section (Sect. A3).

We observe the opposite effect for Bayelva, with the fitted wintertime n-factors clearly increasing from around 0.4 in 2000 to
roughly 0.6 in 2015. This can be interpreted as a reduction in surface insulation, suggesting decreasing snow cover or possibly
a longer snow-free season. Long-term snow observations presented by Boike et al. (2018a) verify this hypothesis. The fact that
fitted n-factor distributions recover this pattern suggests that changes in snow cover are necessary in order to reconcile observed
borehole temperatures with warming air temperatures. The higher posterior variance of the Bayelva n-factors vs. Samoylov
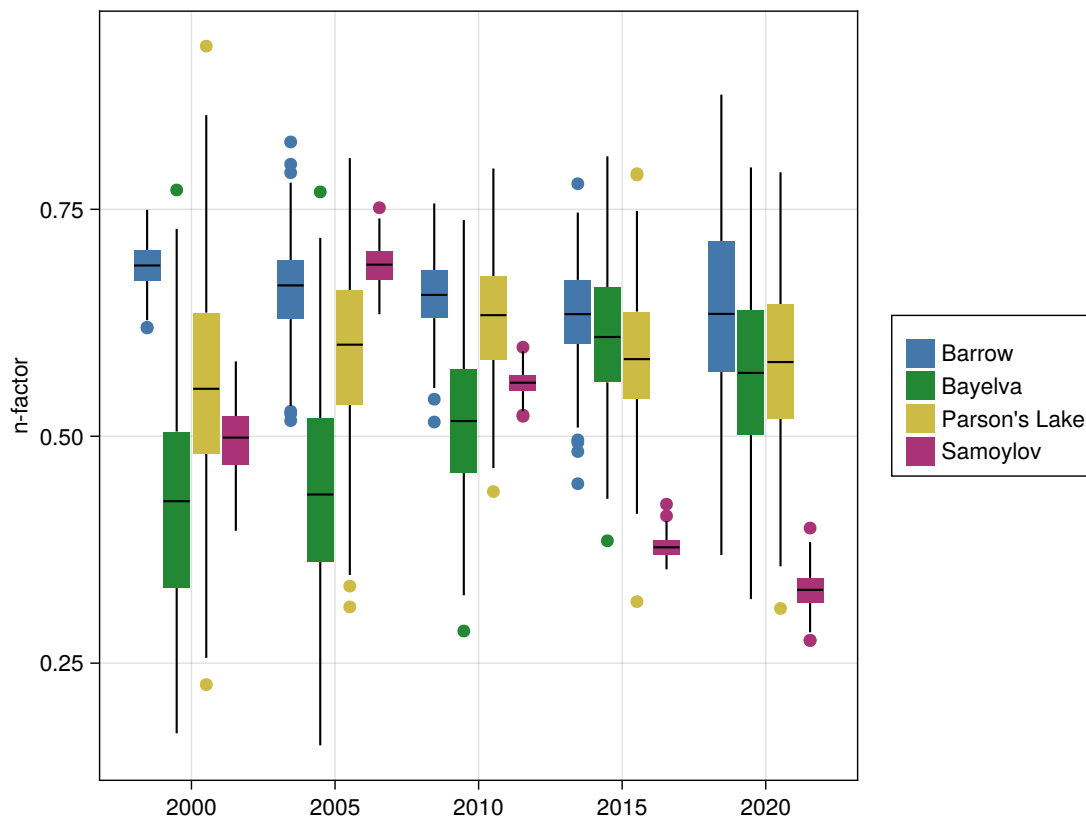
Figure A1. Box-plots showing posterior distributions of winter-time n-factors (i.e., applied when $T_{\mathrm{air}}(t) \leq 0$) after fitting with EKS.

575 also indicates more uncertainty in the strength of the effect. The n-factor distributions for Barrow and Parson's Lake show less of a clear temporal pattern, indicating that the observed borehole temperatures can be plausibly explained without changing insulating conditions at the surface.

A2 Posterior distributions of all parameters

Kernel densities for each parameter over the fitted ensembles for each site are presented in Figs. A2 to A5. Parameter names
580 follow the pattern LAYER.NAME[.PARA] where NAME refers to the model parameter being determined and PARA refers to the "sub-parameter" where applicable. The piecewise linear parameterization of the n-factors where INITIALVALUE refers to the knot at the year 2000, VALUE[1] to the knot at year 2005, VALUE[2] to the knot at year 2010, etc. Parameters named INIT.VALUE and INIT.DEPTH refer to the temperature and depth knots in the piecewise-linear initial temperature profile. Note that set of parameters varied differs slightly between sites due to differences in the stratigraphies; e.g. the excess ground ice
585 parameter is not considered in sites which are assumed (correctly or not) to have no excess ice within the soil column.
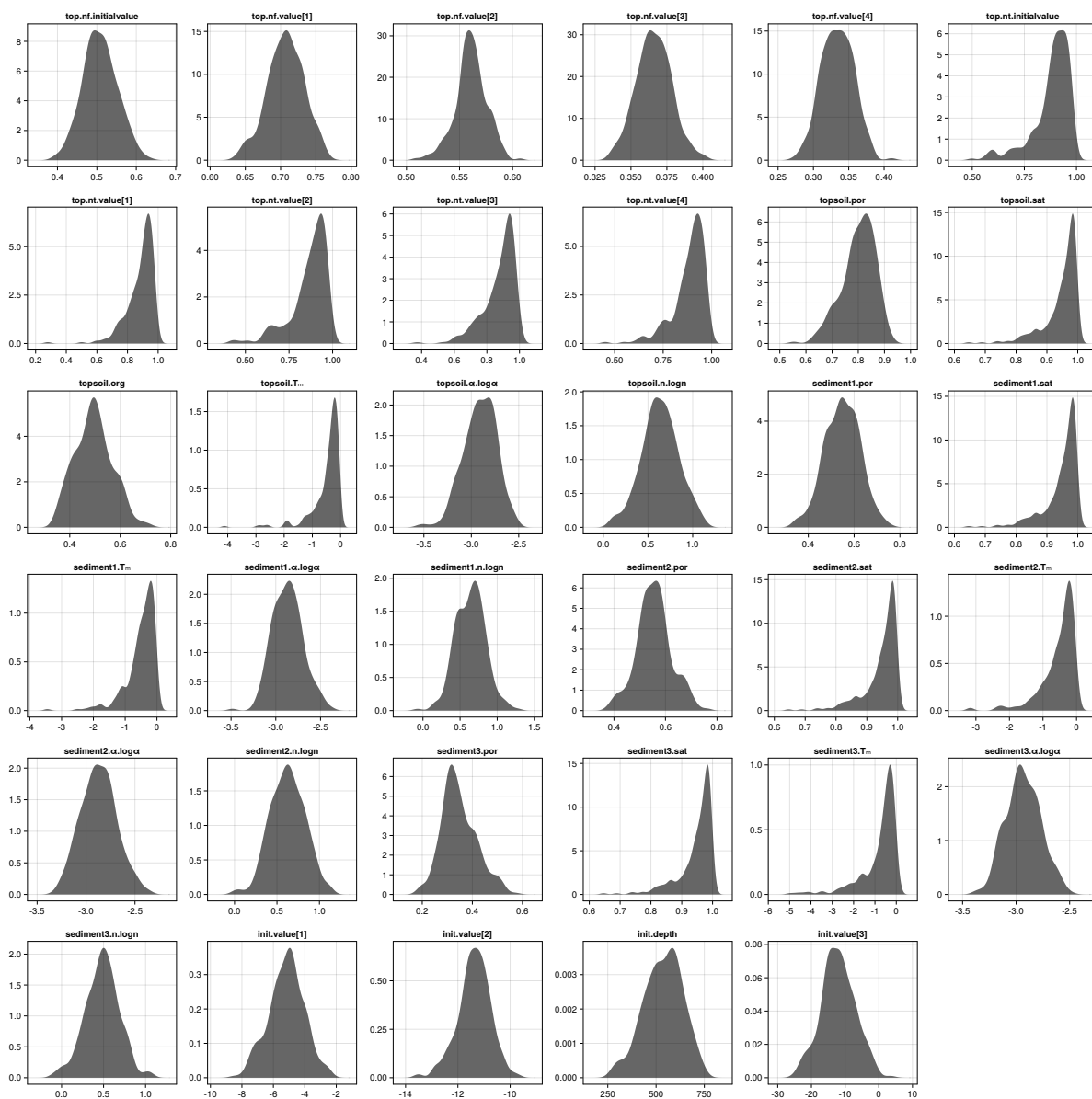
Figure A2. Fitted ensemble/posterior parameter distributions – Samoylov Island
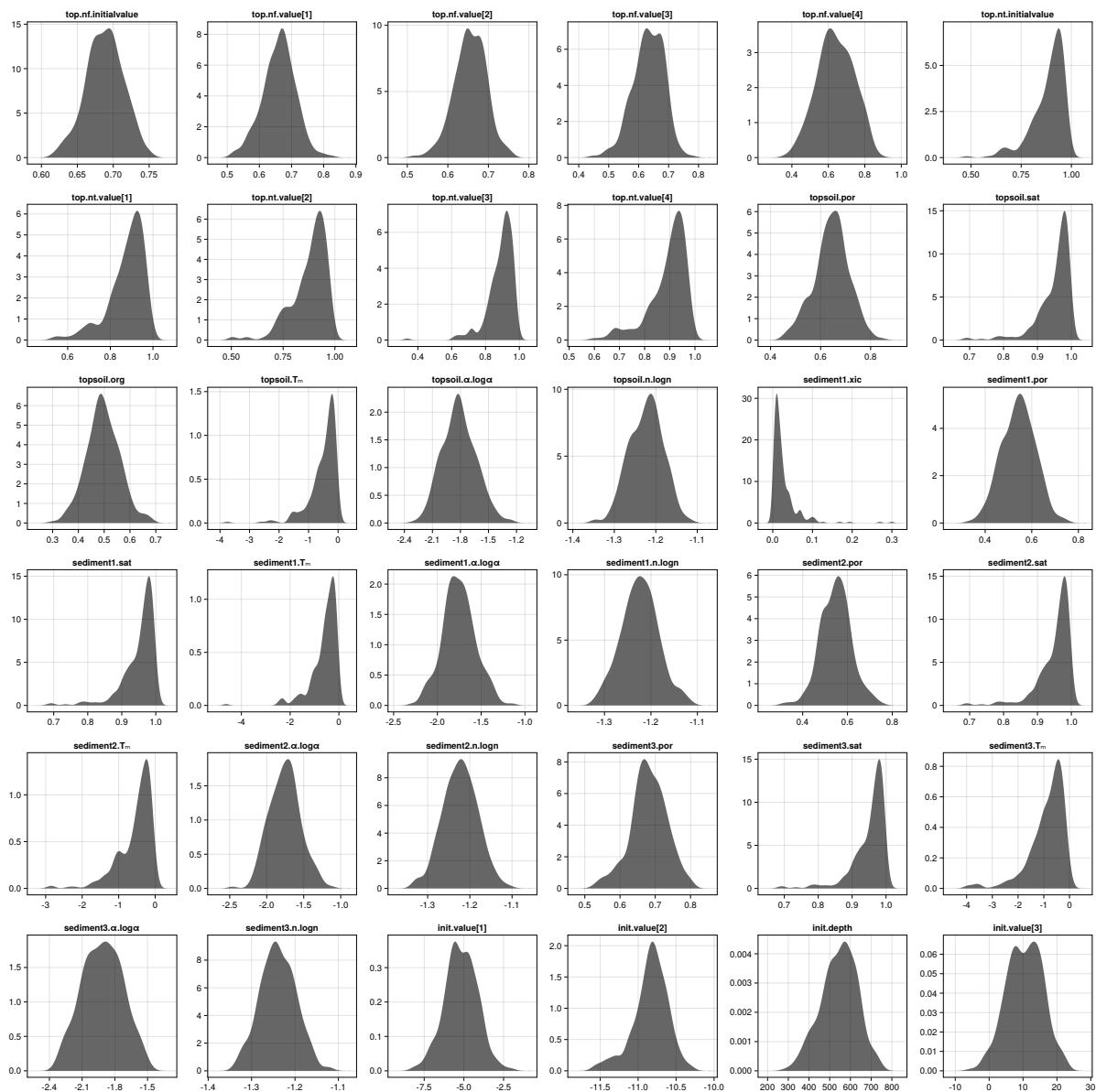
Figure A3. Fitted ensemble/posterior parameter distributions – Barrow
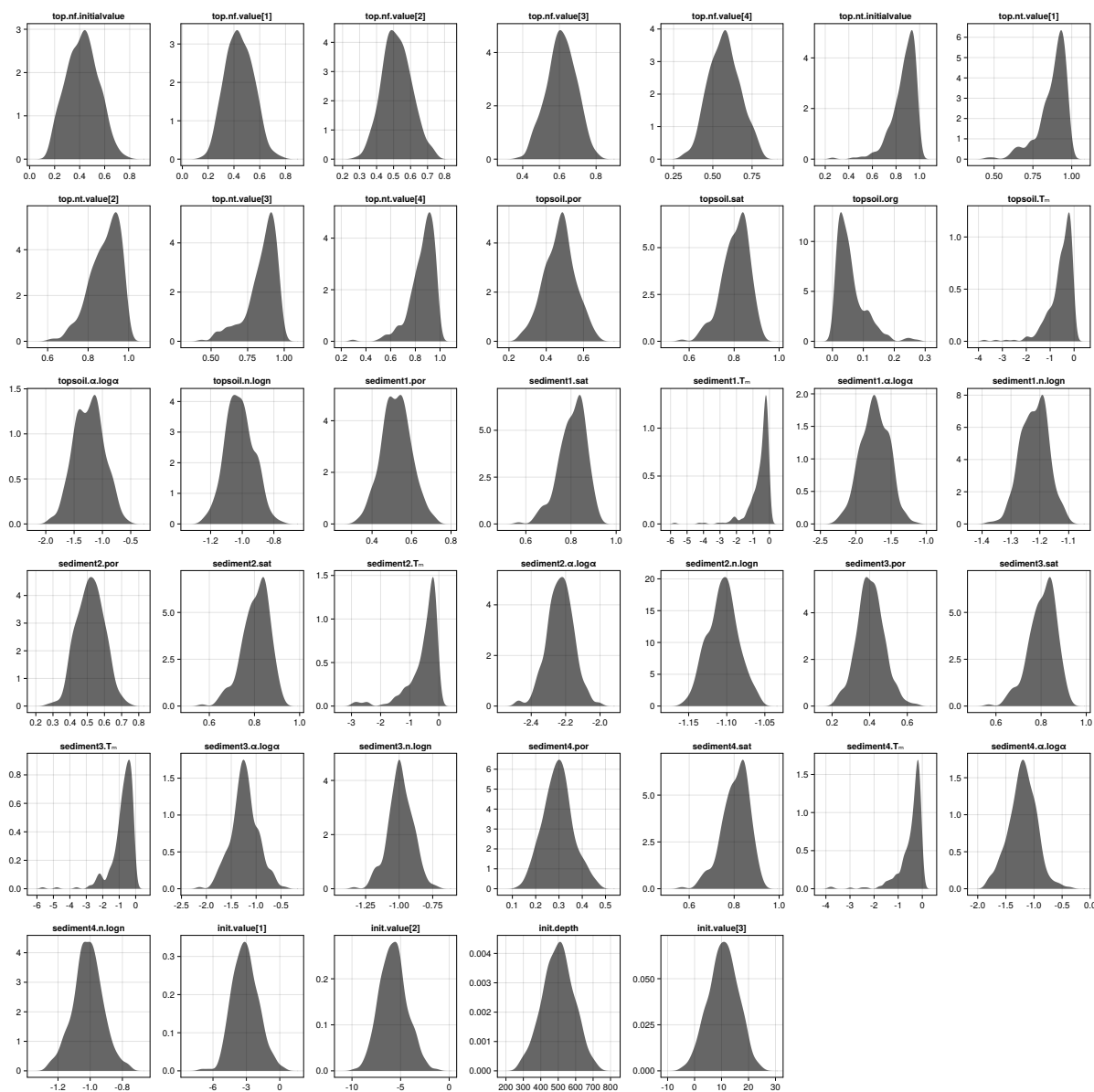
Figure A4. Fitted ensemble/posterior parameter distributions – Bayelva
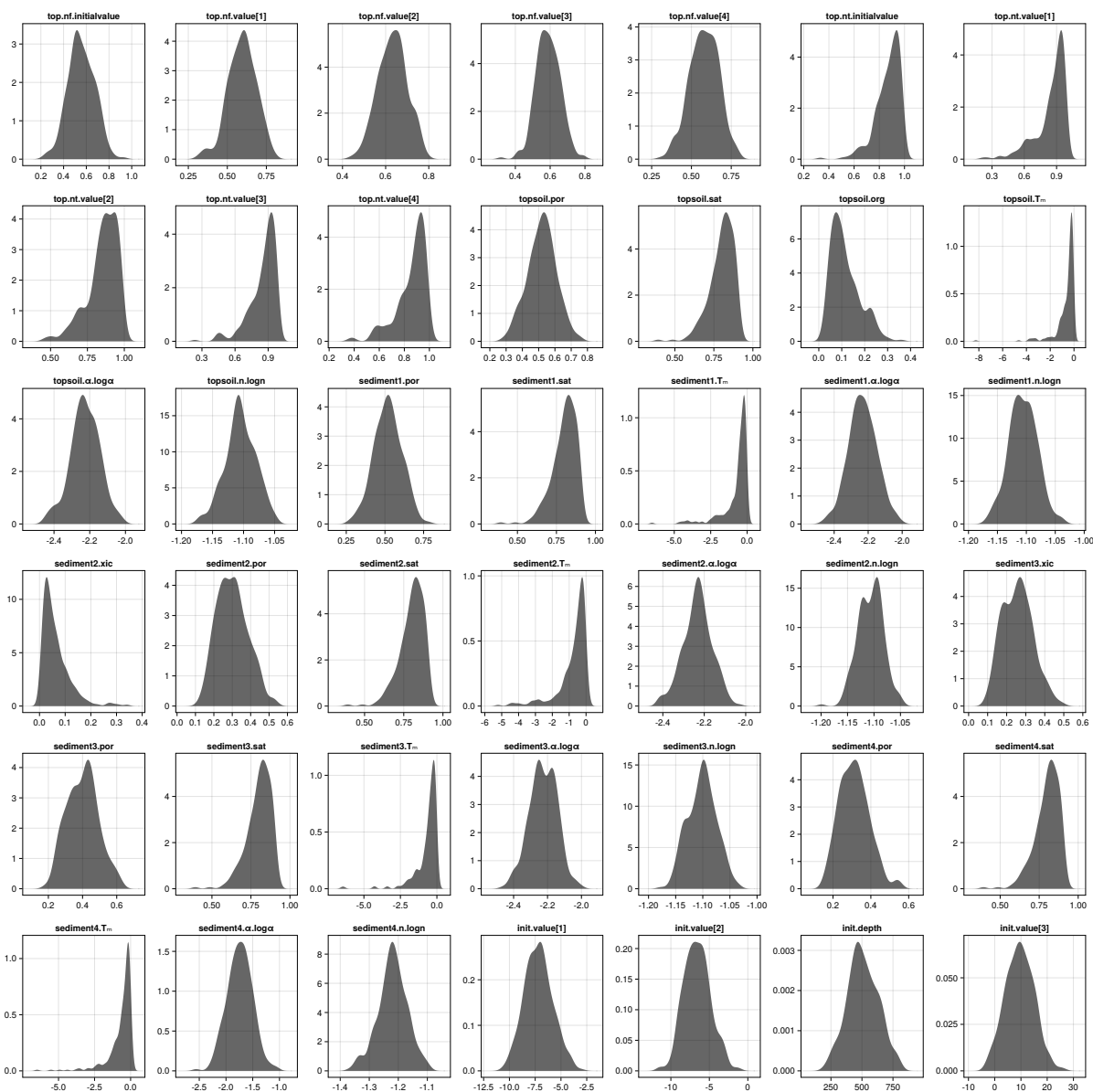
Figure A5. Fitted ensemble/posterior parameter distributions – Parson's Lake

A3   Rapid warming of deep permafrost temperatures on Samoylov Island

The rate at which deep ground temperatures (e.g. 20 m in Fig. 3) appear to be warming at the Samoylov Island site is remarkable, vastly exceeding trends observed at not only the three other borehole sites investigated in this study but also those reported by Biskaborn et al. (2019). As noted in Boike et al. (2019), a new research station was constructed within the immediate vicinity of the borehole site in 2013. The station includes a water delivery system which may act as a wind shield, thereby affecting seasonal snow cover patterns of the area around the borehole. Visual inspection in April of 2016 indicated that snow cover did indeed appear higher since construction of the water supply system (Boike et al., 2019).

Appendix B: Method details

B1   Trend analysis

The annual temperature trend model is specified as:

$$T_t \sim \mathcal{T}(\mu_1 t + \mu_0, \sigma, \nu)$$
$$\mu_1 \sim \mathcal{N}(0, 1)$$
$$\mu_0 \sim \mathcal{N}(0, 10)$$
$$\sigma \sim \mathrm{HalfCauchy}(0, 1)$$
$$\nu \sim \mathrm{Exponential}(10)$$

where $\mathcal{T}(\mu, \sigma, \nu)$ is the scaled and shifted Student's t-distribution with ν degrees of freedom, and t is the zero-centered year. We choose the Student's t-distribution as the likelihood due to the relatively small sample size (fewer than 20 years at all sites) and due to its longer tails allowing for more tolerance of unusually cold or warm years without severely affecting the overall trend (Lange et al., 1989). We use common, weakly-informative prior distributions for each parameter, with the exception of the slope μ_1 which follows a mildly informative unit Gaussian. This is justified by the fact that annual average change in temperature near and below the surface can be reasonably expected to fall well below $1\,°C$ per annum. Samples from the posterior distribution $p(\mu_1, \mu_0, \sigma, \nu | T_{\mathrm{obs}}, t)$ are obtained via the widely used "No U-turn" numerical sampler (NUTS) (Hoffman and Gelman, 2014). We combine 500 independent samples from eight different chains with 2500 discarded burn-in/adaptation samples to get a total of $n = 4000$ samples for each trend model.

B2   Transient heat conduction model

B2.1   Numerical scheme

The spatial gradients in Eq. (3) are approximated using the method of lines:

$$\nabla_t \cdot H_i(z, t) = \frac{1}{\Delta_i} \Big[\frac{k_i(T_{i+1} - T_i)}{\delta_i} - \frac{k_{i-1}(T_i - T_{i-1})}{\delta_{i-1}} \Big], \tag{B1}$$
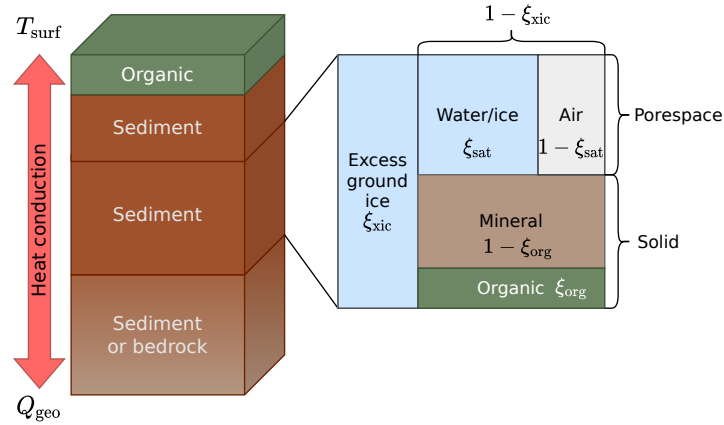
Figure B1. Simplified graphical representation of a four-layer stratigraphy configuration of the transient heat conduction model. The model uses stacked, homogeneous sediment layers, where each layer is represented as a partitioning of excess ice, pore space, and solid material. T_{surf} and Q_{geo} represent respectively the surface temperature and geothermal heat flux boundary conditions.

where Δ_i is the length of the i'th inner grid cell and δ_i is the distance between the midpoints of cells i and $i+1$.

615   For time integration, we use an implementation of the second-order, explicit, strong-stability-preserving time-stepping scheme of Shu and Osher (1988) provided by the Julia package, OrdinaryDiffEq.jl (Rackauckas and Nie, 2017). The time step size is set according to a suitable Courant-Fredrichs-Lewy (CFL) condition (Courant et al., 1967) for the nonlinear heat equation:

$$dt_{\max} = \mathcal{C} \cdot \min_i \left[\frac{\partial H}{\partial T} \right]_i k_i^{-1} \Delta_i^2, \tag{B2}$$

620   where $\frac{\partial H}{\partial T}_i$ $[\text{J K}^{-1}\,\text{m}^{-3}]$ is the apparent heat capacity at grid cell i, k_i $[\text{W m}^{-1}\,\text{K}^{-1}]$ is the thermal conductivity, Δ_i [m] is the vertical length of the grid cell, and $\mathcal{C} = \frac{1}{2}$ is the dimensionless Courant number which scales the step size to maintain stability.

B2.2   Soil parameterization

The soil constituent fractions by volume are parameterized using the four characteristic fractions: excess ice (ξ_{xic}), natural porosity (ξ_{por}), saturation (ξ_{sat}), and organic solid fraction (ξ_{org}). The volumetric soil constituent fractions for excess ice (θ_x),

625   pore water/ice (θ_w), air (θ_a), and mineral and organic content θ_m, θ_o can be computed according to the following linear system:

$$\theta_x = \xi_{\text{xic}},$$

$$\theta_w = (1 - \xi_{\text{xic}})\xi_{\text{por}}\xi_{\text{sat}},$$

$$\theta_a = (1 - \xi_{\text{xic}})\xi_{\text{por}}(1 - \xi_{\text{sat}}),$$

$$\theta_o = (1 - \xi_{\text{xic}})(1 - \xi_{\text{por}})\xi_{\text{org}},$$

630    $$\theta_m = (1 - \xi_{\text{xic}})(1 - \xi_{\text{por}})(1 - \xi_{\text{org}}),$$

Table B1. Stratigraphy configuration for Samoylov Island borehole simulations.

Layer	Depth (m)	Ground ice	Porosity	Saturation	Organic
topsoil	0.0	0.0	0.80	1.0	0.50
sediment1	0.1	0.0	0.80	1.0	0.25
sediment2	0.4	0.3	0.55	1.0	0.25
sediment3	3.0	0.0	0.50	1.0	0.0
sediment4	10.0	0.0	0.30	1.0	0.0

Table B2. Stratigraphy configuration for Bayelva borehole simulations.

Layer	Depth (m)	Ground ice	Porosity	Saturation	Organic
topsoil	0.0	0.0	0.45	0.8	0.05
sediment1	0.1	0.0	0.50	0.8	0.02
sediment2	0.5	0.0	0.50	0.8	0.02
sediment3	1.0	0.0	0.40	0.8	0.02
sediment4	10.0	0.0	0.30	1.0	0.0

where $\theta_x + \theta_w + \theta_a + \theta_o + \theta_m = 1$. Mean stratigraphy configurations (i.e. expected value of the prior) used for each site are given in Tables B1 to B4. Note that, in some cases, adjacent layers have identical mean values but are varied independently in the ensemble. Both Bayelva and Parson's Lake also have different freeze curve priors for some layers based on soil type; this is discussed further in Appendix B2.3.

635   We use the soil freezing characteristic of Dall'Amico et al. (2011), which is based on the soil water retention equation of van Genuchten (1980). The temperature dependent liquid water matric potential is expressed as:

$$\psi(T) = \begin{cases} \psi_0 + \frac{L_f}{gT^*}(T - T^*) & T < T^* \\ \psi_0 & T \geq T^*, \end{cases} \tag{B3}$$

where ψ_0 is the matric potential due to saturation, $L_f = 3.4 \times 10^5\,\mathrm{J\,kg^{-1}}$ is the specific latent heat of fusion of water, and $g = 9.80665\,\mathrm{m\,s^{-2}}$ is downward acceleration due to gravity. The critical temperature T^* is expressed as a function of the

640   melting temperature T_m and the matric potential:

$$T^* = T_m + \frac{gT_m}{L_f}\psi_0. \tag{B4}$$

The volumetric liquid water content then follows from the well known van Genuchten equation (van Genuchten, 1980):

$$\theta_w = \theta_r + (\theta_s - \theta_r)(1 + (-\alpha\psi(T))^n)^{-m}, \tag{B5}$$

where $\alpha \in (0, \infty)\,[\mathrm{m^{-1}}]$, $n \in [1, \infty)$, and $m = 1 - \frac{1}{n}$ are empirical fitting parameters of the van Genuchten curve which depend

645   on soil type, and θ_r and θ_s are the residual and saturated water contents respectively. We use the log-transform of the strictly

Table B3. Stratigraphy configuration for Barrow borehole simulations.

Layer	Depth (m)	Ground ice	Porosity	Saturation	Organic
topsoil	0.0	0.0	0.65	1.0	0.50
sediment1	0.2	0.0	0.55	1.0	0.0
sediment2	0.3	0.0	0.55	1.0	0.0
sediment3	1.0	0.0	0.55	1.0	0.0

Table B4. Stratigraphy configuration for Parson's Lake borehole simulations.

Layer	Depth (m)	Ground ice	Porosity	Saturation	Organic
topsoil	0.0	0.0	0.5	0.8	0.50
sediment1	0.5	0.0	0.5	0.8	0.0
sediment2	1.05	0.05	0.3	0.8	0.0
sediment3	1.40	0.25	0.40	0.8	0.0
sediment4	1.75	0.0	0.30	1.0	0.0

positive van Genuchten parameters in the inversion procedure to (i) allow for the use of Gaussian priors in unconstrained space and (ii) to improve the numerical properties of the likelihood which, as should be apparent from Eq. (B5), exhibits strong non-linearities in the derivatives with respect to α and n.

B2.3   Selection of prior distributions

650   There are broadly three categories of parameters which are considered in our analysis; soil parameters (see Appendix B2.2), n-factor parameters, and initial condition parameters. Separate soil parameters are defined for each stratigraphy layer. Soil freeze curve parameters (melting temperature T_m and van Genuchten parameters $\log \alpha$, $\log n$) as well as porosity are varied in all layers, while organic content is only varied in the uppermost soil layer and is generally assumed to be zero or negligible in all other layers. Ground ice content is also varied only in layers where it is suspected to be present. Saturation of pore-space

655   water/ice is varied as a single, global parameter used by each layer, excluding the bottom-most layer, which is held constant; since saturation is temporally constant throughout the simulation, this can be seen as effectively a bulk or average saturation level for the soil column.

   We based the stratigraphy priors for Samoylov, Bayelva, and Barrow on previous modeling studies, Nitzbon et al. (2020), Zweigel et al. (2021), and Romanovsky and Osterkamp (2000) respectively. For Parson's Lake, we designed the stratigraphy

660   based on the soil core analysis for the borehole site presented by Smith et al. (2018). We chose the priors for the van Genuchten parameters by fitting empirical freeze curves to temperature and volumetric water content measurements available for both the Samoylov Island and Bayelva sites (Boike et al., 2018a, 2019). The mean and standard deviation of the log-transformed

parameters are set according to the fitted posterior statistics, with the variance increased by a factor of 100 to roughly account
for uncertainty in extrapolation. We omit the details of this fitting procedure here for brevity, but the relevant code and results
665   are included in the supplementary material. For Barrow, we used the same van Genuchten parameter priors derived from the
$38\,\mathrm{cm}$ water content sensor at the Bayelva since the freeze curve visually agreed with the one presented in Romanovsky and
Osterkamp (2000), and both sites are known to have silty soil in the upper half meter (Boike et al., 2018a). We set these priors
for Parson's Lake based on the reported soil types in each layer (Smith et al., 2018), using the Samoylov priors in the upper
sandy layers, and the Bayelva clay/silt priors for the deeper layers. We stress that, in all cases, there is no methodical procedure
670   used to derive the prior distributions, and the lack of stratigraphy information for the deeper layers is a source of uncertainty in
our analysis. While the variability induced by sampling from the prior distributions accounts for some of this uncertainty, there
is a lot of room for improvement in the development of more rigorous and principled methods for deriving prior distributions
over soil stratigraphy parameters.

Tables B5 to B8 detail the prior distributions used for each parameter varied in the inversion procedure. We emphasize that
675   these stratigraphy values are generally designed to be rough, subjective estimates based on previous studies and characteristics
of the landscape; uncertainty in the exact values is accounted for through the corresponding prior distributions, which allow
for a modest amount of deviation (generally 10–20%) from the mean. Parameter names follow the same convention described
in Appendix A2.

Table B5: Parameter prior distributions for Samoylov Island simulations

parameter	prior	mean	stdev	units
top.nf.initialvalue	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[1]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[2]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[3]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[4]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nt.initialvalue	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[1]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[2]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[3]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[4]	Beta(α=18.00, β=2.00)	0.90	0.07	-
topsoil.por	Beta(α=40.00, β=10.00)	0.80	0.06	-
topsoil.org	Beta(α=25.00, β=25.00)	0.50	0.07	-
topsoil.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
topsoil.α.logα	Normal(μ=-2.88, σ=0.17)	-2.88	0.17	-
topsoil.n.logn	Normal(μ=0.64, σ=0.18)	0.64	0.18	-

parameter	prior	mean	stdev	units
sediment1.por	Beta(α=27.50, β=22.50)	0.55	0.07	-
sediment1.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment1.α.logα	Normal(μ=-2.88, σ=0.17)	-2.88	0.17	-
sediment1.n.logn	Normal(μ=0.64, σ=0.18)	0.64	0.18	-
sediment2.por	Beta(α=25.00, β=25.00)	0.50	0.07	-
sediment2.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment2.α.logα	Normal(μ=-2.88, σ=0.17)	-2.88	0.17	-
sediment2.n.logn	Normal(μ=0.64, σ=0.18)	0.64	0.18	-
sediment3.por	Beta(α=15.00, β=35.00)	0.30	0.06	-
sediment3.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment3.α.logα	Normal(μ=-2.88, σ=0.17)	-2.88	0.17	-
sediment3.n.logn	Normal(μ=0.64, σ=0.18)	0.64	0.18	-
init.depth	Truncated(Normal(μ=500.00, σ=100.00); lower=100.00, upper=900.00)	500.00	99.95	m
init.value[1]	Normal(μ=-5.00, σ=1.00)	-5.00	1.00	°C
init.value[2]	Normal(μ=-12.00, σ=3.00)	-12.00	3.00	°C
init.value[3]	Normal(μ=-12.00, σ=5.00)	-12.00	5.00	°C

Table B6: Parameter prior distributions for Bayelva simulations

parameter	prior	mean	stdev	units
top.nf.initialvalue	Beta(α=8.00, β=12.00)	0.40	0.11	-
top.nf.value[1]	Beta(α=8.00, β=12.00)	0.40	0.11	-
top.nf.value[2]	Beta(α=8.00, β=12.00)	0.40	0.11	-
top.nf.value[3]	Beta(α=8.00, β=12.00)	0.40	0.11	-
top.nf.value[4]	Beta(α=8.00, β=12.00)	0.40	0.11	-
top.nt.initialvalue	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[1]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[2]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[3]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[4]	Beta(α=18.00, β=2.00)	0.90	0.07	-
topsoil.por	Beta(α=22.50, β=27.50)	0.45	0.07	-

置

parameter	prior	mean	stdev	units
topsoil.org	Beta(α=2.50, β=47.50)	0.05	0.03	-
topsoil.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
topsoil.α.logα	Normal(μ=-1.24, σ=0.23)	-1.24	0.23	-
topsoil.n.logn	Normal(μ=-1.01, σ=0.08)	-1.01	0.08	-
sediment1.por	Beta(α=25.00, β=25.00)	0.50	0.07	-
sediment1.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment1.α.logα	Normal(μ=-1.76, σ=0.19)	-1.76	0.19	-
sediment1.n.logn	Normal(μ=-1.22, σ=0.04)	-1.22	0.04	-
sediment2.por	Beta(α=25.00, β=25.00)	0.50	0.07	-
sediment2.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment2.α.logα	Normal(μ=-2.23, σ=0.06)	-2.23	0.06	-
sediment2.n.logn	Normal(μ=-1.10, σ=0.02)	-1.10	0.02	-
sediment3.por	Beta(α=20.00, β=30.00)	0.40	0.07	-
sediment3.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment3.α.logα	Normal(μ=-1.24, σ=0.23)	-1.24	0.23	-
sediment3.n.logn	Normal(μ=-1.01, σ=0.08)	-1.01	0.08	-
sediment4.por	Beta(α=15.00, β=35.00)	0.30	0.06	-
sediment4.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment4.α.logα	Normal(μ=-1.24, σ=0.23)	-1.24	0.23	-
sediment4.n.logn	Normal(μ=-1.01, σ=0.08)	-1.01	0.08	-
init.depth	Truncated(Normal(μ=500.00, σ=100.00); lower=100.00, upper=900.00)	500.00	99.95	m
init.value[1]	Normal(μ=-3.00, σ=1.00)	-3.00	1.00	°C
init.value[2]	Normal(μ=-5.00, σ=2.00)	-5.00	2.00	°C
init.value[3]	Normal(μ=10.20, σ=5.00)	10.20	5.00	°C

Table B7: Parameter prior distributions for Barrow simulations

parameter	prior	mean	stdev	units
top.nf.initialvalue	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[1]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[2]	Beta(α=13.00, β=7.00)	0.65	0.10	-

parameter	prior	mean	stdev	units
top.nf.value[3]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nf.value[4]	Beta(α=13.00, β=7.00)	0.65	0.10	-
top.nt.initialvalue	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[1]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[2]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[3]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[4]	Beta(α=18.00, β=2.00)	0.90	0.07	-
topsoil.por	Beta(α=32.50, β=17.50)	0.65	0.07	-
topsoil.org	Beta(α=25.00, β=25.00)	0.50	0.07	-
topsoil.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
topsoil.α.logα	Normal(μ=-1.76, σ=0.19)	-1.76	0.19	-
topsoil.n.logn	Normal(μ=-1.22, σ=0.04)	-1.22	0.04	-
sediment1.xic	Beta(α=1.00, β=50.00)	0.02	0.02	-
sediment1.por	Beta(α=27.50, β=22.50)	0.55	0.07	-
sediment1.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment1.α.logα	Normal(μ=-1.76, σ=0.19)	-1.76	0.19	-
sediment1.n.logn	Normal(μ=-1.22, σ=0.04)	-1.22	0.04	-
sediment2.por	Beta(α=27.50, β=22.50)	0.55	0.07	-
sediment2.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment2.α.logα	Normal(μ=-1.76, σ=0.19)	-1.76	0.19	-
sediment2.n.logn	Normal(μ=-1.22, σ=0.04)	-1.22	0.04	-
sediment3.por	Beta(α=27.50, β=22.50)	0.55	0.07	-
sediment3.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment3.α.logα	Normal(μ=-1.76, σ=0.19)	-1.76	0.19	-
sediment3.n.logn	Normal(μ=-1.22, σ=0.04)	-1.22	0.04	-
init.depth	Truncated(Normal(μ=500.00, σ=100.00); lower=100.00, upper=900.00)	500.00	99.95	m
init.value[1]	Normal(μ=-5.00, σ=1.00)	-5.00	1.00	°C
init.value[2]	Normal(μ=-9.10, σ=3.00)	-9.10	3.00	°C
init.value[3]	Normal(μ=10.20, σ=5.00)	10.20	5.00	°C

Table B8: Parameter prior distributions for Parson's Lake simulations

parameter	prior	mean	stdev	units
top.nf.initialvalue	Beta(α=10.00, β=10.00)	0.50	0.11	-
top.nf.value[1]	Beta(α=10.00, β=10.00)	0.50	0.11	-
top.nf.value[2]	Beta(α=10.00, β=10.00)	0.50	0.11	-
top.nf.value[3]	Beta(α=10.00, β=10.00)	0.50	0.11	-
top.nf.value[4]	Beta(α=10.00, β=10.00)	0.50	0.11	-
top.nt.initialvalue	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[1]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[2]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[3]	Beta(α=18.00, β=2.00)	0.90	0.07	-
top.nt.value[4]	Beta(α=18.00, β=2.00)	0.90	0.07	-
topsoil.por	Beta(α=25.00, β=25.00)	0.50	0.07	-
topsoil.org	Beta(α=5.00, β=45.00)	0.10	0.04	-
topsoil.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
topsoil.α.logα	Normal(μ=-2.23, σ=0.06)	-2.23	0.06	-
topsoil.n.logn	Normal(μ=-1.10, σ=0.02)	-1.10	0.02	-
sediment1.por	Beta(α=25.00, β=25.00)	0.50	0.07	-
sediment1.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment1.α.logα	Normal(μ=-2.23, σ=0.06)	-2.23	0.06	-
sediment1.n.logn	Normal(μ=-1.10, σ=0.02)	-1.10	0.02	-
sediment2.xic	Beta(α=2.50, β=47.50)	0.05	0.03	-
sediment2.por	Beta(α=15.00, β=35.00)	0.30	0.06	-
sediment2.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment2.α.logα	Normal(μ=-2.23, σ=0.06)	-2.23	0.06	-
sediment2.n.logn	Normal(μ=-1.10, σ=0.02)	-1.10	0.02	-
sediment3.xic	Beta(α=12.50, β=37.50)	0.25	0.06	-
sediment3.por	Beta(α=20.00, β=30.00)	0.40	0.07	-
sediment3.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C
sediment3.α.logα	Normal(μ=-2.23, σ=0.06)	-2.23	0.06	-
sediment3.n.logn	Normal(μ=-1.10, σ=0.02)	-1.10	0.02	-
sediment4.por	Beta(α=15.00, β=35.00)	0.30	0.06	-
sediment4.Tm	Truncated(Normal(μ=0.00, σ=0.50); lower=-10.00, upper=0.00)	-0.40	0.30	°C

parameter	prior	mean	stdev	units
sediment4.α.logα	Normal(μ=-1.76, σ=0.19)	-1.76	0.19	-
sediment4.n.logn	Normal(μ=-1.22, σ=0.04)	-1.22	0.04	-
init.depth	Truncated(Normal(μ=500.00, σ=100.00); lower=100.00, upper=900.00)	500.00	99.95	m
init.value[1]	Normal(μ=-7.00, σ=1.00)	-7.00	1.00	°C
init.value[2]	Normal(μ=-5.10, σ=2.00)	-5.10	2.00	°C
init.value[3]	Normal(μ=10.20, σ=5.00)	10.20	5.00	°C

Author contributions. B.G., M.L., and S.W. designed the study and modeling scheme; B.G. designed and authored the model and analysis
680 code, carried out all numerical experiments, and wrote the manuscript with support from all co-authors. J.B.and G.G. supervised the project.
All authors discussed the results and contributed to the final version of the manuscript.

Competing interests. The authors declare no competing interest.

Acknowledgements. We are grateful to Caroline Duchense and Sharon Smith (Natural Resources Canada) for providing us with the borehole
temperature record from Parson's Lake as well as to Adam Theisen (Argonne National Laboratory) for providing climatology data from
685 the Northern Slope of Alaska. We also thank Inge Grünberg (Alfred Wegener Institute) for her valuable consultation and feedback. Brian
Groenke acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS).

References

Barrow 2 (N. Meadow Lake No.2 / NML-2), https://permafrost.gi.alaska.edu/site/br2, 2021.

Allen, D. M., Michel, F. A., and Judge, A. S.: The Permafrost Regime in the Mackenzie Delta, Beaufort Sea Region, N.W.T. and Its Signifi-
cance to the Reconstruction of the Palaeoclimatic History, Journal of Quaternary Science, 3, 3–13, https://doi.org/10.1002/jqs.3390030103,
1988.

Berliner, L. M.: Physical-Statistical Modeling in Geophysics: PHYSICAL-STATISTICAL MODELING IN GEOPHYSICS, Journal of Geo-
physical Research: Atmospheres, 108, n/a–n/a, https://doi.org/10.1029/2002JD002865, 2003.

Bishop, C.: Pattern Recognition and Machine Learning, vol. 4 of *Information Science and Statistics*, Springer-Verlag, New York, first edn.,
2006.

Biskaborn, B. K., Lanckman, J.-P., Lantuit, H., Elger, K., Streletskiy, D. A., Cable, W. L., and Romanovsky, V. E.: The New Database of the
Global Terrestrial Network for Permafrost (GTN-P), Earth System Science Data, 7, 245–259, https://doi.org/10.5194/essd-7-245-2015,
2015.

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.

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E. J., Overduin, P. P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I.,
Hauber, E., and Westermann, S.: Measurements in Soil and Air at Bayelva Station, https://doi.org/10.1594/PANGAEA.880120, 2017.

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Overduin, P. P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I.,
Hauber, E., and Westermann, S.: A 20-Year Record (1998–2017) of Permafrost, Active Layer and Meteorological Conditions at a
High Arctic Permafrost Research Site (Bayelva, Spitsbergen), Earth System Science Data, 10, 355–390, https://doi.org/10.5194/essd-10-
355-2018, 2018a.

Boike, J., Nitzbon, J., Anders, K., Grigoriev, M. N., Bolshiyanov, D. Y., Langer, M., Lange, S., Bornemann, N., Morgenstern, A.,
Schreiber, P., Wille, C., Chadburn, S., Gouttevin, I., and Kutzbach, L.: Measurements in Soil and Air at Samoylov Station (2002-2018),
https://doi.org/10.1594/PANGAEA.891142, 2018b.

Boike, J., Nitzbon, J., Anders, K., Grigoriev, M., Bolshiyanov, D., Langer, M., Lange, S., Bornemann, N., Morgenstern, A., Schreiber,
P., Wille, C., Chadburn, S., Gouttevin, I., Burke, E., and Kutzbach, L.: A 16-Year Record (2002–2017) of Permafrost, Active-Layer,
and Meteorological Conditions at the Samoylov Island Arctic Permafrost Research Site, Lena River Delta, Northern Siberia: An Op-
portunity to Validate Remote-Sensing Data and Land Surface, Snow, and Permafrost Models, Earth System Science Data, 11, 261–299,
https://doi.org/10.5194/essd-11-261-2019, 2019.

Brewer, M. C.: Some Results of Geothermal Investigations of Permafrost in Northern Alaska, Eos, Transactions American Geophysical
Union, 39, 19–26, https://doi.org/10.1029/TR039i001p00019, 1958.

Brown, J., Ferrians Jr, O., Heginbottom, J. A., and Melnikov, E.: Circum-Arctic Map of Permafrost and Ground Ice Conditions, 1997.

Burgess, M. M., Smith, S. L., Brown, J., Romanovsky, V., and Hinkel, K.: Global Terrestrial Network for Permafrost (GTNet-P): Permafrost
Monitoring Contributing to Global Climate Observations, Tech. Rep. 2000-E14, https://doi.org/10.4095/211621, 2000.

Burn, C. R. and Kokelj, S. V.: The Environment and Permafrost of the Mackenzie Delta Area, Permafrost and Periglacial Processes, 20,
725   83–105, https://doi.org/10.1002/ppp.655, 2009.

Caers, J.: Bayesianism in the Geosciences, in: Handbook of Mathematical Geosciences: Fifty Years of IAMG, edited by Daya Sagar, B.,
Cheng, Q., and Agterberg, F., pp. 527–566, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-78999-6, 2018.

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

730   Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., and Stuart, A. M.: Calibrate, Emulate, Sample, Journal of Computational Physics, 424,
109 716, https://doi.org/10.1016/j.jcp.2020.109716, 2021.

Constantinou, N. C., Lopez-Gomez, I., Iñigo, A. G., Bolewski, J., odunbar, Howland, M., Wu, J., Jackson, T., (Zhengyu)Huang,
D., Jaruga, A., Ervik, H. L. L., Monticone, P., Chor, T., and bielim: CliMA/EnsembleKalmanProcesses.Jl: V0.8.0, Zenodo,
https://doi.org/10.5281/zenodo.6516055, 2022.

735   Courant, R., Friedrichs, K., and Lewy, H.: On the Partial Difference Equations of Mathematical Physics, IBM journal of Research and
Development, 11, 215–234, 1967.

Cranmer, K., Brehmer, J., and Louppe, G.: The Frontier of Simulation-Based Inference, Proceedings of the National Academy of Sciences,
117, 30 055–30 062, https://doi.org/10.1073/pnas.1912789117, 2020.

Dall'Amico, M., Endrizzi, S., Gruber, S., and Rigon, R.: A Robust and Energy-Conserving Model of Freezing Variably-Saturated Soil, The
740   Cryosphere, 5, 469–484, https://doi.org/10.5194/tc-5-469-2011, 2011.

Duane, S. and Kogut, J. B.: The Theory of Hybrid Stochastic Algorithms, Nuclear Physics B, 275, 398–420, https://doi.org/10.1016/0550-
3213(86)90606-1, 1986.

ECCC: Historical Data - Climate - Environment and Climate Change Canada, https://climate.weather.gc.ca/historical_data/search_historic_data_e.html,
2011.

745   Evensen, G., Vossepoel, F. C., and van Leeuwen, P. J.: Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter
Estimation Problem, Springer Nature, https://doi.org/10.1007/978-3-030-96709-3, 2022.

Franks, S. W. and Beven, K. J.: Bayesian Estimation of Uncertainty in Land Surface-Atmosphere Flux Predictions, Journal of Geophysical
Research: Atmospheres, 102, 23 991–23 999, https://doi.org/10.1029/97JD02011, 1997.

Garbuno-Inigo, A., Hoffmann, F., Li, W., and Stuart, A. M.: Interacting Langevin Diffusions: Gradient Structure and Ensemble Kalman
750   Sampler, SIAM Journal on Applied Dynamical Systems, 19, 412–441, https://doi.org/10.1137/19M1251655, 2020.

Geer, A. J.: Learning Earth System Models from Observations: Machine Learning or Data Assimilation?, Philosophical Transactions of the
Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200 089, https://doi.org/10.1098/rsta.2020.0089, 2021.

Gopalan, G., Hrafnkelsson, B., Aðalgeirsdóttir, G., Jarosch, A. H., and Pálsson, F.: A Bayesian Hierarchical Model for Glacial Dy-
namics Based on the Shallow Ice Approximation and Its Evaluation Using Analytical Solutions, The Cryosphere, 12, 2229–2248,
755   https://doi.org/10.5194/tc-12-2229-2018, 2018.

Gregory, W., Lawrence, I. R., and Tsamados, M.: A Bayesian Approach towards Daily Pan-Arctic Sea Ice Freeboard Estimates from Com-
bined CryoSat-2 and Sentinel-3 Satellite Observations, The Cryosphere, 15, 2857–2871, https://doi.org/10.5194/tc-15-2857-2021, 2021.

Grigoriev, N.: The Temperature of Permafrost in the Lena Delta Basin–Deposit Conditions and Properties of the Permafrost in Yakutia,
Yakutsk, 2, 97–101, 1960.

760   Groenke, B., Nitzbon, J., and Langer, M.: CryoGrid.Jl: V0.10.3, Zenodo, https://doi.org/10.5281/zenodo.6801740, 2022.

Hermoso de Mendoza, I., Beltrami, H., MacDougall, A. H., and Mareschal, J.-C.: Lower Boundary Conditions in Land Surface Models – Effects on the Permafrost and the Carbon Pools: A Case Study with CLM4.5, Geoscientific Model Development, 13, 1663–1683, https://doi.org/10.5194/gmd-13-1663-2020, 2020.

Hinzman, L. D., Bettez, N. D., Bolton, W. R., Chapin, F. S., Dyurgerov, M. B., Fastie, C. L., Griffith, B., Hollister, R. D., Hope, A., Hunt-
765 ington, H. P., Jensen, A. M., Jia, G. J., Jorgenson, T., Kane, D. L., Klein, D. R., Kofinas, G., Lynch, A. H., Lloyd, A. H., McGuire, A. D., Nelson, F. E., Oechel, W. C., Osterkamp, T. E., Racine, C. H., Romanovsky, V. E., Stone, R. S., Stow, D. A., Sturm, M., Tweedie, C. E., Vourlitis, G. L., Walker, M. D., Walker, D. A., Webber, P. J., Welker, J. M., Winker, K. S., and Yoshikawa, K.: Evidence and Implications of Recent Climate Change in Northern Alaska and Other Arctic Regions, Climatic Change, 72, 251–298, https://doi.org/10.1007/s10584-005-5352-2, 2005.

770 Hoffman, M. D. and Gelman, A.: The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, p. 31, 2014.

Humlum, O.: Holocene Permafrost Aggradation in Svalbard, Geological Society, London, Special Publications, 242, 119–129, https://doi.org/10.1144/GSL.SP.2005.242.01.11, 2005.

Iglesias, M. A., Law, K. J. H., and Stuart, A. M.: Ensemble Kalman Methods for Inverse Problems, Inverse Problems, 29, 045 001, https://doi.org/10.1088/0266-5611/29/4/045001, 2013.

775 Isaksen, K., Nordli, t. O., Ivanov, B., Køltzow, M. A. t. O., Aaboe, S., Gjelten, H. M., Mezghani, A., Eastwood, S., Førland, E., Benestad, R. E., Hanssen-Bauer, I., Brækkan, R., Sviashchennikov, P., Demin, V., Revina, A., and Karandasheva, T.: Exceptional Warming over the Barents Area, Scientific Reports, 12, 9371, https://doi.org/10.1038/s41598-022-13568-5, 2022.

Jorgenson, M. T., Shur, Y. L., and Pullman, E. R.: Abrupt Increase in Permafrost Degradation in Arctic Alaska, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL024960, 2006.

780 Kneier, F., Overduin, P. P., Langer, M., Boike, J., and Grigoriev, M. N.: Borehole Temperature Reconstructions Reveal Differences in Past Surface Temperature Trends for the Permafrost in the Laptev Sea Region, Russian Arctic, arktos, 4, 1–17, https://doi.org/10.1007/s41063-018-0041-3, 2018.

Koopmans, R. W. R. and Miller, R.: Soil Freezing and Soil Water Characteristic Curves, Soil Science Society of America Journal, 30, 680–685, 1966.

785 Lange, K. L., Little, R. J., and Taylor, J. M.: Robust Statistical Modeling Using the t Distribution, Journal of the American Statistical Association, 84, 881–896, 1989.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The Surface Energy Balance of a Polygonal Tundra Site in Northern Siberia – Part 1: Spring to Fall, The Cryosphere, 5, 151–171, https://doi.org/10.5194/tc-5-151-2011, 2011a.

Langer, M., Westermann, S., Muster, S., Piel, K., and Boike, J.: The Surface Energy Balance of a Polygonal Tundra Site in Northern Siberia
790 – Part 2: Winter, The Cryosphere, 5, 509–524, https://doi.org/10.5194/tc-5-509-2011, 2011b.

Langer, M., Nitzbon, J., Groenke, B., Assmann, L.-M., Schneider von Deimling, T., Stuenzi, S. M., and Westermann, S.: The Evolution of Arctic Permafrost over the Last Three Centuries, EGUsphere, 2022, 1–27, https://doi.org/10.5194/egusphere-2022-473, 2022.

Lunardini, V.: A Correlation of N-Factors, in: Applied Techniques for Cold Environments, pp. 233–244, ASCE, 1978.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S.,
795 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.

Nelson, F. E., Anisimov, O. A., and Shiklomanov, N. I.: Climate Change and Hazard Zonation in the Circum-Arctic Permafrost Regions, Natural Hazards, 26, 203–225, https://doi.org/10.1023/A:1015612918401, 2002.

800 Nelson, F. E., Shiklomanov, N. I., Streletskiy, D. A., Romanovsky, V. E., Yoshikawa, K., Hinkel, K. M., and Brown, J.: A Permafrost Observatory at Barrow, Alaska: Long-Term Observations of Active-Layer Thickness and Permafrost Temperature, p. 6, 2008.

Nitzbon, J., Langer, M., Martin, L. C. P., Westermann, S., Schneider von Deimling, T., and Boike, J.: Effects of Multi-Scale Heterogeneity on the Simulated Evolution of Ice-Rich Permafrost Lowlands under a Warming Climate, The Cryosphere Discussions, pp. 1–29, https://doi.org/10.5194/tc-2020-137, 2020.

805 Nitzbon, J., Krinner, G., Schneider von Deimling, T., Werner, M., and Langer, M.: Quantifying the Permafrost Heat Sink in Earth's Climate System, Earth and Space Science Open Archive, p. 15, https://doi.org/10.1002/essoar.10511600.1, 2022.

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

810 Scale, Earth-Science Reviews, 193, 299–316, https://doi.org/10.1016/j.earscirev.2019.04.023, 2019.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical Bias Correction of Global Simulated Daily Precipitation and Temperature for the Application of Hydrological Models, Journal of Hydrology, 395, 199–215, https://doi.org/10.1016/j.jhydrol.2010.10.024, 2010.

Qu, Y., Wang, J., Wan, H., Li, X., and Zhou, G.: A Bayesian Network Algorithm for Retrieving the Characterization of Land Surface

815 Vegetation, Remote Sensing of Environment, 112, 613–622, https://doi.org/10.1016/j.rse.2007.03.031, 2008.

Rackauckas, C. and Nie, Q.: DifferentialEquations.Jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia, Journal of Open Research Software, 5, 15, https://doi.org/10.5334/jors.151, 2017.

Reich, S. and Cotter, C.: Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, 2015.

Riseborough, D.: Soil Latent Heat as a Filter of the Climate Signal in Permafrost, in: Proceedings of the Fifth Canadian Permafrost Confer-

820 ence, Collection Nordicana, vol. 54, pp. 199–205, Citeseer, 1990.

Riseborough, D., Shiklomanov, N., Etzelmüller, B., Gruber, S., and Marchenko, S.: Recent advances in permafrost modelling, Permafrost and Periglacial Processes, 19, 137–156, https://doi.org/10.1002/ppp.615, 2008.

Romanovsky, V., Burgess, M., Smith, S., Yoshikawa, K., and Brown, J.: Permafrost Temperature Records: Indicators of Climate Change, Eos, Transactions American Geophysical Union, 83, 589–594, https://doi.org/10.1029/2002EO000402, 2002.

825 Romanovsky, V. E. and Osterkamp, T. E.: Effects of Unfrozen Water on Heat and Mass Transport Processes in the Active Layer and Permafrost, Permafrost and Periglacial Processes, 11, 219–239, https://doi.org/10.1002/1099-1530(200007/09)11:3<219::AID-PPP352>3.0.CO;2-7, 2000.

Romanovsky, V. E., Sazonova, T. S., Balobaev, V. T., Shender, N. I., and Sergueev, D. O.: Past and Recent Changes in Air and Permafrost Temperatures in Eastern Siberia, Global and Planetary Change, 56, 399–413, https://doi.org/10.1016/j.gloplacha.2006.07.022, 2007.

830 Romanovsky, V. E., Smith, S. L., and Christiansen, H. H.: Permafrost Thermal State in the Polar Northern Hemisphere during the International Polar Year 2007–2009: A Synthesis, Permafrost and Periglacial Processes, 21, 106–116, https://doi.org/10.1002/ppp.689, 2010.

Roth, K. and Boike, J.: Quantifying the Thermal Dynamics of a Permafrost Site near Ny-Ålesund, Svalbard, Water Resources Research, 37, 2901–2914, https://doi.org/10.1029/2000WR000163, 2001.

Schneider von Deimling, T., Lee, H., Ingeman-Nielsen, T., Westermann, S., Romanovsky, V., Lamoureux, S., Walker, D. A., Chadburn,

835 S., Trochim, E., Cai, L., Nitzbon, J., Jacobi, S., and Langer, M.: Consequences of Permafrost Degradation for Arctic Infrastructure –

Bridging the Model Gap between Regional and Engineering Scales, The Cryosphere, 15, 2451–2471, https://doi.org/10.5194/tc-15-2451-2021, 2021.

Schuur, E. A. and Mack, M. C.: Ecological Response to Permafrost Thaw and Consequences for Local and Global Ecosystem Services, Annual Review of Ecology, Evolution, and Systematics, 49, 279–301, https://doi.org/10.1146/annurev-ecolsys-121415-032349, 2018.

840 Serreze, M. C. and Francis, J. A.: The Arctic Amplification Debate, Climatic Change, 76, 241–264, https://doi.org/10.1007/s10584-005-9017-y, 2006.

Shiklomanov, N. I., Streletskiy, D. A., Nelson, F. E., Hollister, R. D., Romanovsky, V. E., Tweedie, C. E., Bockheim, J. G., and Brown, J.: Decadal Variations of Active-Layer Thickness in Moisture-Controlled Landscapes, Barrow, Alaska, Journal of Geophysical Research: Biogeosciences, 115, https://doi.org/10.1029/2009JG001248, 2010.

845 Shu, C.-W. and Osher, S.: Efficient Implementation of Essentially Non-Oscillatory Shock-Capturing Schemes, Journal of computational physics, 77, 439–471, 1988.

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.

850 Smith, S., Chartrand, J., and Duchesne, C.: Report on 2017 Field Activities and Collection of Ground-Thermal and Active-Layer Data in the Mackenzie Corridor, Northwest Territories. Geological Survey of Canada Open File 8492, 109 Pp, Geological Survey of Canada, Ottawa, ON, 2018.

Smith, S. L., O'Neill, H. B., Isaksen, K., Noetzli, J., and Romanovsky, V. E.: The Changing Thermal State of Permafrost, Nature Reviews Earth & Environment, 3, 10–23, https://doi.org/10.1038/s43017-021-00240-1, 2022.

855 Theisen, A.: ARM-Climatologies North Slope Alaska, Zenodo, https://doi.org/10.5281/zenodo.6685908, 2022.

Tubini, N., Gruber, S., and Rigon, R.: A Method for Solving Heat Transfer with Phase Change in Ice or Soil That Allows for Large Time Steps While Guaranteeing Energy Conservation, Preprint, Frozen ground/Numerical Modelling, https://doi.org/10.5194/tc-2020-293, 2020.

van Genuchten, M. T.: A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils, Soil Science Society of America Journal, 44, 892–898, https://doi.org/10.2136/sssaj1980.03615995004400050002x, 1980.

860 Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., and van Wessem, J. M.: Bayesian Calibration of Firn Densification Models, The Cryosphere, 14, 3017–3032, https://doi.org/10.5194/tc-14-3017-2020, 2020.

Wainwright, H. M., Liljedahl, A. K., Dafflon, B., Ulrich, C., Peterson, J. E., Gusmeroli, A., and Hubbard, S. S.: Mapping Snow Depth within a Tundra Ecosystem Using Multiscale Observations and Bayesian Methods, The Cryosphere, 11, 857–875, https://doi.org/10.5194/tc-11-857-2017, 2017.

865 Westermann, S., Boike, J., Langer, M., Schuler, T. V., and Etzelmüller, B.: Modeling the Impact of Wintertime Rain Events on the Thermal Regime of Permafrost, The Cryosphere, 5, 945–959, https://doi.org/10.5194/tc-5-945-2011, 2011.

Westermann, S., Langer, M., Boike, J., Heikenfeld, M., Peter, M., Etzelmüller, B., and Krinner, G.: Simulating the Thermal Regime and Thaw Processes of Ice-Rich Permafrost Ground with the Land-Surface Model CryoGrid 3, Geosci. Model Dev., p. 24, 2016.

Westermann, S., Ingeman-Nielsen, T., Scheer, J., Aalstad, K., Aga, J., Chaudhary, N., Etzelmüller, B., Filhol, S., Kääb, A., Renette, C.,
870 Schmidt, L. S., Schuler, T. V., Zweigel, R. B., Martin, L., Morard, S., Ben-Asher, M., Angelopoulos, M., Boike, J., Groenke, B., Miesner, F., Nitzbon, J., Overduin, P., Stuenzi, S. M., and Langer, M.: The CryoGrid Community Model (Version 1.0) – a Multi-Physics Toolbox for Climate-Driven Simulations in the Terrestrial Cryosphere, Geoscientific Model Development Discussions, 2022, 1–61, https://doi.org/10.5194/gmd-2022-127, 2022.

Wolfe, S. A., Smith, S. L., Chartrand, J., Kokelj, S., Palmer, M., and Stevens, C. W.: Geotechnical Database and Descriptions of
Permafrost Monitoring Sites Established 2006-2010 in the Northern Mackenzie Corridor, Northwest Territories, Tech. Rep. 6677,
https://doi.org/10.4095/287167, 2010.

Zweigel, R. B., Westermann, S., Nitzbon, J., Langer, M., Boike, J., Etzelmüller, B., and Vikhamar Schuler, T.: Simulating Snow Redistribution
and Its Effect on Ground Surface Temperature at a High-Arctic Site on Svalbard, Journal of Geophysical Research: Earth Surface, 126,
e2020JF005 673, https://doi.org/10.1029/2020JF005673, 2021.