Replies to editor and referee comments on "Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer"

The Cryosphere, 10.5194/egusphere-2022-630

EC: ***Editor's Comment***, **RC:** ***Referee's Comment***, AR: Authors' Response, ☐ Manuscript Text

1. Response to editor comments

EC: ***Dear authors,***

I consider the manuscript a valuable contribution to TC, but I think the reviewers have raised questions about the approach that need to be addressed within the manuscript before being accepted for publication. I believe the comments should not be only addressed from the process-understanding point of view, but also the technical aspects of the approach proposed, as this study is for me more a technical paper than a case-study. I believe the comments from the reviewers are fair and your answer is in general adequate. Maybe it could be consider a sensitivity analysis to further understand the role from borehole data at different depths? The problem on the data uncertainty from sensors located at different depths was clearly a main issue for reviewer 1, which also was commented by reviewer 2. I understand that setting sensors at large depths is not always feasible; thus, this issue will come along in many other studies - also those using the approach you are proposed. Maybe, while preparing the revisions to your manuscript you could still test this in a synthetic study or a sensitivity analysis? I will ask the reviewers to read your revised version of the manuscript before considering it ready for publication.

Best regards,

Adrian Flores

AR: Dear Prof. Dr. Adrian Flores Orozco,

We thank you for your time and consideration as well as your positive feedback regarding our manuscript. We acknowledge your concerns about the referees' comments regarding the impact of available sensor depths on the inverse modeling procedure. Following your suggestion, we have conducted a sensitivity analysis for representative cold and warm sites (Samoylov and Bayelva) using synthetic pseudo-observations produced using model outputs from our previous simulations. The results of this analysis as well as a detailed description of the methodology are presented in Sect. S1.2 of the supplementary material. We found that the inclusion and exclusion of measurements at various depths had a fairly predictable impact on the resulting fitted ensemble. Predicted mean annual ground temperatures at the considered depths of $10\,\mathrm{m}$ and $20\,\mathrm{m}$ were in better agreement with the synthetic measurements when there were measurements available at these respective depths. The fitted ensemble tended to overestimate warming at “shallower” depths (above $10\,\mathrm{m}$) when only deeper measurements were available (below $20\,\mathrm{m}$) and tended to unerestimate warming at “deeper” depths when only shallow measurements were available. We have added a brief comment on these additional results in the main text.

We provide a brief summary of all major changes to the manuscript in Sect. 2 of this response letter. We include updated versions of our point-by-point responses to the referee comments in Sect. 3 and Sect. 4.

We would like to thank you again for your time, and we look forward to receiving your feedback on the revised manuscript.

Sincerely,
B. Groenke on behalf of all co-authors

2. Summary of major changes

We briefly summarize the most substantial changes to the manuscript below.

- As noted in our earlier responses to referee comments, we fixed a minor bug in our model code and re-ran inverse modeling simulations from scratch. There were no substantial changes to the results. We also used a larger ensembele size to improve the robustness of our results. All relevant figures and tables have been updated with these regenerated results in the revised manuscript.
- We have added a supplementary material document to cut down on the length of the primary manuscript and host some of the more technical details and additional results.
- In the process of addressing the editor and referee comments, we noticed an issue with Fig. 6 (now Fig. 7) in the original manuscript. The method by which we computed sensible heat for this figure previously, while technically correct, was slightly misleading in that it included changes in the heat capacity due to melting/freezing of pore ice/water. In order to more clearly isolate the relative contributions of temperature change and phase change, we have used an alternative formulation of sensible heat in the updated figure. Further explanation and discussion can be found in the revised manuscript (Sect. 4.2 and Sect. 5.4).
- Following the suggestion of the editor, we carried out additional experiments using synthetic data to assess the sensitivity of the inversion results to available sensor depths. The results are presented in the supplementary material (Sect. S1.2).
- Following the suggestion of referee 1, we have revised the discussion section to include more specific quantitative analysis.

– Following the suggestion of referee 1, we included the diagram describing the heat transfer model in the main text (Fig. 2). We also updated this diagram to make it more precise (e.g. referring to the layers as “soil” rather than “sediment” which is technically more correct). We additionally have added a new subfigure showing examples of the soil freezing characteristic curves used in our model. We felt this was necessary in order to provide a clearer picture of how phase change is actually represented in the model and why the soil freezing characteristics matter.

– Following the suggestion of referee 1, we improved the clarity and accessibility of the methods section, particularly the sections on Bayesian inference and the inverse modeling approach. Much of the inverse modeling detail has now moved to the Appendix. Furthermore, we have added some higher level explanation of EKS to provide more context to the reader.

– Following the suggestion of referee 2, we added further discussion on related work and methods in the introduction (Sect. 1).

– Following the suggestion of referee 2, we moved some of the technical description of the trend analysis method to the main text.

– Following the suggestion of referee 2, we have updated the parameter distribution plots in the Appendix to show both the prior and posterior.

– We have expanded the numerical implementation details in the Appendix. We further reorganized the Appendix for clarity.

– We have changed all mentions of “thaw depth” in the text with “active layer thickness” to be more precise; in this work, we are only ever discussing maximum annual thaw depth for which “active layer thickness” is a more common and more parsimonious term.

– We have revised and cleaned up the mathematical notation in the methods section to be more concise and consistent.

– We have condensed the “Limitations” section by combining points 2 and 3 from the original manuscript, the latter of which we assessed to be mostly redundant.

– We have corrected the parameter sensitivity analysis previously presented in the response to referee 2 to use only samples from the prior and not the posterior; this is necessary to avoid parameter correlations affecting the estimated indices (Li et al., 2010).

– We have made further minor corrections, clarifications, and improvements throughout the manuscript which we do not enumerate here.

3. Response to referee 1

The text in this section is adapted from Groenke (2023a) with updated excerpts from the revised manuscript included where relevant.

3.1. Issue 1: Research design and uncertainties

RC: ***The research design has some issues making it unclear if the main conclusions are attributed to the physical processes or the modeling uncertainties. First, the available depths of borehole data are not the same. At the two colder sites (Samoylov and Barrow, Fig 4b and 4c), both sites have deep borehole data although Barrow does not have shallow borehole data. In contrast, neither of the two warmer sites (Fig 4d and 4e) has deep borehole data. This could be the main reason causing the much larger temperature variability (Fig 4d and 4e), more scattered relationships in Fig 5, and more observed uncertainties in Fig 6. Therefore, the majority of conclusions made by comparing colder and warmer sites are not convincing. One or more warm sites with deep borehole data are needed to validate this study's conclusions. It is also worth performing the inverse modeling again on the Samoylov site excluding its deep borehole data to see if its thermal behavior stays the same or changes toward the warm sites.***

AR: We acknowledge that the disparity between available borehole temperature measurement depths is a concern. It is important to note, however, that this is not an intentional aspect of the study design but rather a limitation of the available data. It is common for researchers and practitioners to install automated temperature sensor instrumentation in the upper one meter of the ground since this is generally achievable without heavy drilling equipment. High quality, automated instrumentation of deep boreholes is unfortunately relatively rare and instead research teams typically collect manual measurements once per year (often in the summer when the borehole can be easily located). The data from the Barrow North Meadow Lake site featured in this work are an example of such measurements. As mentioned in the text, these annual measurements cannot be compared to mean annual temperatures recorded in instrumented boreholes (such as the other three sites) above the depth of zero annual amplitude (ZAA) due to the effects of seasonal variation. This is why we only use the manual measurements from Barrow at 20 m and below, as this is presumed to be deep enough that seasonal variation should be negligible.

Despite this limitation, we argue that our conclusions are justified for three primary reasons:

Firstly, the availability of deeper measurements should largely only affect uncertainty in the initial temperature profile at the beginning of the simulation period (i.e. at the year 2000); this is because deeper measurements help better constrain the range of plausible temperature profiles after the spin-up period (1979-1999). The impact on later years where observations are available in the upper $10\,\mathrm{m}$ will necessarily be less significant since, after the first 5 to 10 years, the climate signal will dominate the initial condition.

Secondly, one of the main reasons why we limit the analysis of energy contents to the upper $10\,\mathrm{m}$ is because this is the range in which all sites (excepting Barrow) have measurements available. While it is true that the temperature profile at the beginning of the simulation period would have some impact on the resulting distribution of observed trends, we would expect this effect to be mostly limited to temperature (i.e. a wider range of initial temperatures would spread out the distribution along the x-axis in Figure 6 from the revised manuscript). It should not affect the underlying relationship between temperature and latent heat, which is the central interest of this study.

Lastly, while it is true that the availability of deeper borehole measurements will affect the resulting spread of temperature predictions across the ensemble, we do not agree that this weakens the conclusions drawn from comparing the cold and warm sites. On the contrary, it is actually a strength of our method (and Bayesian methods more generally) that the posterior distribution meaningfully reflects uncertainty due to differences in data availability between sites and therefore allows us to make inferences despite these limitations of the available data.

To validate our arguments here, we followed the suggestion of the referee and ran an additional set of simulations for the Samoylov site with the measurements below $10\,\mathrm{m}$ omitted from the inference procedure. The results, along with an accompanying discussion, are presented in Sect. S1.1 of the supplementary material.

RC: ***Line 374 seems to demonstrate depth alone cannot explain the variability. However, the statement is not strong because 82 cm is too small on a 10 m scale. Also, the observations of Bayelva also have less variability than those of Parson's Lake, which likely explains the less variability in the modeled temperature at Bayelva.***

AR: We agree with the referee's assessment here and have removed this statement from the text.

RC: ***The authors do have a full section 5.6 to discuss the limitations. While these limitations are definitely important, the current research design is not strong enough to support the conclusions even neglecting other uncertainties.***

AR: We believe that the additional results presented in the supplementary material validate the study design and support the central arguments of our paper. The limitations detailed in section 5.6 are, as the referee states, important. However, as also argued in the main text, we believe that the model and study design are still strong enough to support the main conclusions.

RC: ***Secondly, section 5.3 discusses the role of surface conditions on ground warming based on the n-factor change before and after 2005. Again, uncertainties can be the main driver because no borehole observations are available to constrain the model before 2005. This is another key point made based on the comparison of two data not having the same conditions.***

AR: We have removed this paragraph from the text since it is, in hindsight, overly speculative given the limitations of the current study design and available data. We do not agree, however, that this is a particularly important point of the manuscript. The discussion here was largely tangential and was intended only to comment on the potential hazards of extrapolating recently observed trends in deep ground temperatures backwards in time.

3.2. Issue 2: Method description

RC: ***The manuscript has a large space describing the modeling method but most contents are too technical and not accessible to people who are in the cryosphere community but do not have expertise in numerical modeling, inversion, and Bayesian method, etc. The authors focus too much on the advanced topics of the method but completely missed the information on the basic idea of the applied method. Also, in many cases, the authors only cite some references without explicitly describing the methods, which makes the readers difficult to follow or understand.***

AR: We thank the referee for this valuable feedback. Although we already attempted in the original manuscript to keep technical language and details to a minimum, we recognize that this communication gap is one of the primary challenges of interdisciplinary research and that there are certainly further improvements that can be made to the existing text. We respond to each of the referee's specific concerns below.

RC: ***Section 3.1. The introduction of Bayesian inference involves too many technical terminologies. Please consider adding supporting sentences to make it easier for people not familiar with the Bayesian method to understand it.***

This section was intended to provide a basic introduction to the ideas of Bayesian inference for those not familiar with such methods, so it is of course important that it is accessible for this audience. We have revised the paragraph in question as follows:

> The Bayesian approach to statistics provides a natural framework for inferring unobserved quantities of interest while simultaneously accounting for their associated uncertainties [...]. This is accomplished ~~by applying Bayes ruleto some observed and unobserved variables, Y and X, respectively~~via Bayes rule:
>
> $$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)} \quad \text{with} \quad p(Y) = \int_{x \in X} p(Y|X=x)p(X=x)dx, \tag{1}$$
>
> 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~~. The *prior distribution* $p(X)$ ~~encodes information about Y known~~ reflects our pre-existing uncertainty about X *a priori* ~~and plays a crucial role in the Bayesian inference workflow.~~ (i.e. before observing Y) while the *likelihood* $p(Y|X)$ measures how well the model's predictions agree with the observations, Y. In this work, Y are temperature measurements, typically sampled over time and/or space, whereas X are unknown model parameters or unobserved physical quantities such as soil properties, thaw depth, or the ratio of sensible to latent heat. The overall objective is then to obtain the posterior distribution, $p(X|Y)$, of these unknown parameters given the temperature measurements which quantifies not only the best-fitting parameter settings but also the associated modeling uncertainties.

RC: ***Lines 140-144. Need to briefly explain the bias correction procedure.***

AR: We have added the following clarifying clause to briefly elaborate on the bias correction procedure of Piani et al. (2010):

The bias correction procedure for air temperature follows closely the empirical quantile mapping method of [...] in which the empirical quantiles of both the model (reanalysis) and observational data are computed over some reference period (in this study, we use the full time period for which observations are available); the model data are then mapped to the corresponding quantiles of the observations.

RC: ***Line 150. Need to briefly explain the numerical procedures and parameterizations of CryoGrid.***

AR: We have added two additional missing pieces of information to the Appendix section, namely the parameterization of the thermal conductivity, $k_T(z,t)$, and heat capacity, $C_T(z,t)$, functions.

The bulk, temperature-dependent thermal conductivity $k(T)$ is parameterized following (Cosenza et al., 2003) as the inverse quadratic mean of the constituents:

$$k(T) = (k_w * \sqrt{\theta_w(T)} + k_i * \sqrt{\theta_i(T)} + k_a * \sqrt{\theta_a} + k_m * \sqrt{\theta_m} + k_a * \sqrt{\theta_o})^2 \tag{2}$$

where k $[\mathrm{W\,m^{-1}\,K^{-1}}]$ refers to the thermal conductivity of the constituent material. The bulk heat capacity is similarly computed as a simple weighted average:

$$C(T) = \frac{1}{5}(c_w * \theta_w(T) + c_i * \theta_i(T) + c_a * \theta_a + c_m * \theta_m + c_o * \theta_o) \tag{3}$$

where c are the constituent heat capacities. Since we already vary the constituent fractions θ in our inversion procedure, we hold these values for the constituent thermal properties constant in all of our simulations. The values are shown in Table B5.

We have also added a table to the Appendix (Table B5) enumerating the relevant constituent conductivities and heat capacity parameters which are treated as constants.

RC: ***Section 3.6 This section introduces a key methodology EKS. It presents the advantages of EKS over MCMC and EKI without explaining the basic theory/idea of EKS in the first place. Again, this makes researchers not familiar with EKS very difficult to follow and understand it.***

AR: We have added some additional explanation of the basic idea behind the EKS algorithm:

EKS assumes that the observed borehole temperatures can be represented as:

$$T_{\text{obs}} = (g_T \circ f)(\phi) + \eta, \tag{4}$$

where g_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. EKS requires the prior distribution over parameters to be a multivariate Gaussian, so we must first construct a bijective function $\gamma : \Phi \to \Psi$ that maps the m-dimensional parameters $\boldsymbol{\phi} \in \Phi$ to values $\boldsymbol{\psi} \in \Psi$ with unbounded support on the real line. The approximate prior for EKS is then specified as $\boldsymbol{\psi} \sim p(\boldsymbol{\psi}) = \mathcal{N}(\mu_{\boldsymbol{\psi}}, \Sigma_{\boldsymbol{\psi}})$. The key insight behind the EKS algorithm is that an ensemble of parameter values $\{\boldsymbol{\psi}^{(i)}\}$ sampled from an initial density (i.e. the prior) can be transformed into samples from the (approximate) posterior distribution by treating them as a stochastic system of interacting particles. Their dynamics are then governed by the overdamped Langevin equation:

$$\frac{\partial \boldsymbol{\psi}^{(i)}}{\partial t} = -\nabla \mathcal{L}(\boldsymbol{\psi}^{(i)}) + \sqrt{2}R(t) \tag{5}$$

where $\mathcal{L}(\boldsymbol{\psi})$ is the log-likelihood of the unconstrained parameters $\boldsymbol{\psi}$ with respect to the observations and $R(t)$ is an m-dimensional Brownian motion. It can be shown that this interacting particle system then converges to the posterior density over an infinite time horizon, where "time" here refers to that of the particle diffusion rather than physical time in the forward model. We refer the reader to Garbuno-Inigo et al. 2020 for further details about the EKS algorithm and iteration procedure.

RC: ***Line 250. Need briefly explain what a mean vector from Garbuno-Inigo et al. 2020 is.***

AR: As discussed in the text, the observed temperatures are assumed to be generated according to equation (9), i.e:

$$T_{\text{obs}} = (h_T \circ f)(\boldsymbol{\theta}) + \eta$$

where f is the forward model evaluated at $\boldsymbol{\theta}$, h_T is the mapping function which extracts and aggregates the modeled temperatures, and $\eta \sim \mathcal{N}(0, \Sigma_T)$ is the observational noise. For the purposes of constructing the EKS algorithm, the model predictions can thus be equivalently seen as being sampled from a Gaussian distribution centered at T_{obs} (this follows from moving η to the left hand side of the above equation). Since we assume the observation noise to be independent across space and time, we can flatten the two-dimensional temperature field into a vector which thus constitutes the mean of this Gaussian distribution, hence the term "mean vector". However, we acknowledge that this term is non-standard and possibly confusing; in light of this, we have now rephrased this sentence in the revised manuscript:

~~The~~ We use the observed mean annual ground ~~temperatures,~~ temperatures from each borehole site as the observations, i.e. T_{obs} ~~, are used as the observation mean vector for the Ensemble Kalman Sampler described in [...]~~in Eq. 7, for the EKS algorithm.

This part of the text has also been moved to the Appendix in the interest of brevity.

3.3. Issue 3: Quantiative analysis in discussion

RC: ***The discussion needs more quantitative and specific analysis. When interpreting the results, the authors only briefly propose possible factors without explaining how would these factors impact the results.***

AR: We agree with the referee's assessment and have revised the discussion section accordingly. The diff is too extensive to include here, so we kindly refer the reader to the revised manuscript and its corresponding diff markup.

RC: ***Paragraph 255. I may miss something but I did not get the purpose of this paragraph. It states that the prior distribution over model parameters is important but does not explain what was done to improve performance.***

AR: The purpose of this paragraph was just to provide some basic motivation. We have rearranged this and the following paragraph to make this more clear:

> The prior distribution over model parameters, $p(\phi)$, ~~is of crucial importance to our method~~also plays a key role in the inversion procedure. 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.
>
> ~~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 [...]
>
> 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 (and possibly constrained) parameters $\phi \in \Phi$ to their unconstrained values on the real line.

This section has also been moved to the Appendix for brevity.

RC: ***Line 356. This sentence does not explain why Samoylov has deep soil temperature warming faster than the air temperature. Factors other than air temperature should be included here.***

AR: We agree that the other factors are also important to highlight. This was actually explained further in the following paragraph, but the connection was not necessarily obvious as written. We have revised these two paragraphs to make this point more clear:

> 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~~ This discrepancy in observed permafrost temperature trends between these two 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 andsubstantial uncertainty in how the permafrost thermal regime responds to long-term changes in air temperature. The discrepancies in the apparent 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~~ that is likely attributable to other factors. For example, thicker and/or less dense snow cover can accelerate permafrost warming by insulating the ground against rapid drops in air temperature characteristic to autumn and early winter, thereby delaying the refreezing of the active layer (Park et al., 2015). Additionally, soil thermal properties ~~. These factors~~such as the bulk conductivity and freezing characteristics due to soil texture can also play a significant role in modulating the effects of surface temperature changes [...]. Both of these factors, among others, can 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~~

The results of this trend analysis motivate our inverse modeling study by demonstrating clear, localized differences and substantial uncertainty in how the permafrost thermal regime at these four sites is responding to long-term changes in air temperature. We argue 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~~ factors affecting the uptake of latent heat in the subsurface such as soil freezing characteristics as well as historical climatology. We discuss in the following sections how our inverse modeling results suggest that both of these factors are major sources of uncertainty in making inferences about the ~~subsurface thermal regime (Riseborough, 1990; Romanovsky and Osterkamp, 2000)~~changing thermal state of permafrost.

RC: ***Paragraph 360. Besides only presenting the potential factors impacting the soil thermal states, I would include how they impact the thermal states. For example, how does the ground temperature change with air temperature giving increasing (or decreasing) snow thickness and soil thermal diffusivity?***

AR: We have added further discussion in the text to clarify the expected impacts on the thermal state (see the revised text above).

RC: ***Line 390. Please explain more about why latent heat is lost so that the temperature is warmer. Please also explain why drainage and evapotranspiration cause latent heat loss.***

AR: The revised manuscript now includes the following explanation:

~~For Bayelva, we suspect that the~~ The wintertime warm bias ~~may~~ for Bayelva ($(1.5 \pm 0.5)\,\mathrm{K}$) may also be due to ~~the model's~~ our assumption of static hydrology, which fails to capture the effects of ~~latent heat being lost~~ seasonal changes in soil moisture on the subsurface thermal regime. The release of latent heat during freezing slows the propagation of cold surface temperatures, thereby delaying the refreezing of the active layer in early winter (Romanovsky and Osterkamp, 2000). When unfrozen water is removed from the active layer due to drainage ~~and evapotranspiration.~~ or evapotranspiration, the latent heat stored in this water is removed as well. The result is that cold temperatures can propagate faster in the wintertime since a higher fraction of this energy is diffused as sensible heat. This would therefore serve as possible explanation for the warm bias in the model at this site.

RC: ***Section 5.6. It would be helpful if include some discussion about the expected changes after addressing each limitation.***

AR: We have condensed the limitations section and added some supporting sentences per the suggestion of the referee.

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. This could potentially alter the thermal dynamics of the ground in two ways: firstly, unfrozen water which is transported out of the pore space and not replaced can be seen as effectively removing energy from the system thereby altering the energy balance of the ground. Secondly, drier soils with higher air content will necessarily be less conductive, thereby becoming an insulator for deeper layers against further warming and cooling. However, the assumption of static hydrological conditions allows us to more ~~objectively~~ easily compare changes in the partitioning of latent vs. sensible heat ~~in the subsurface~~ between sites (Fig. 7). We further note that for ~~some sites, such as Samoylov~~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 ~~[...].~~ (see Sect. 3.3.2). 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~~ We might expect, for example, some of 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. inter-annual fluctuations to be explained by changes in snow or cloud cover. The inclusion of additional surface processes, however, also brings with it a whole range of additional uncertainties from both the forcings and relevant parameterizations.

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 [...] B3.3 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 construction of prior distributions difficult. Furthermore, it is very likely that many parameters are, in reality, highly correlated, which we are currently unable to account for in the prior. Prior distributions which are more informative or more realistic will help the inversion method converge faster and improve the reliability of the resulting posterior samples in making inferences about the real world system. 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. produced by traditional calibration methods.

5. **Other statistical considerationsStatistical limitations**. 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 outputspredictions produced by a surrogate model. Additionally, EKS requires a priori specification of the observation noise sampling distribution covariance Σ_T, which is not, in general, known. Ideally, a A fully Bayesian treatment of the inversion problem would include these Σ_T as parameters to be inferred, which (or equivalent) parameters in the posterior distribution. This would have the benefit of producing an ensemble with a predictive distribution that posterior samples where the predictive distribution is well calibratedon the training data. This is, however, not possible using EKS, and thus, we leave the problem of estimating noise parameters to future work, i.e. the 95% prediction interval should actually cover approximately

95% of the observations in the calibration period, therefore providing a built-in measure of the degree to which the model is capable of explaining the observations.

RC: ***Line 345 information about site location is needed for Biskaborn et al. 2019.***

AR: We have added a comment about the sites used by Biskaborn et al. 2019.

...who reported average decadal changes in permafrost temperatures of $(0.39 \pm 0.15)\,\mathrm{K\,dec^{-1}}$ for 53 GTN-P boreholes across the continuous permafrost zone...

RC: ***Line 386. This may be due 'to' the thermal…***

AR: Thank you. We have fixed this typo in the revised text.

RC: ***Line 445. The second point. Warm permafrost could have slow refreezing when warming due to the effects of latent heat.***

AR: Thank you for pointing this out. This sentence has been corrected in the revised manuscript.

This underscores two features of warmer permafrost: (i) that observed temperature trends (or lack thereof) should be interpreted with caution, as 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~~change is highly dependent on soil thermal properties, most notably the soil freezing characteristics. Soils which retain more unfrozen water at temperatures below $0\,°\mathrm{C}$ may delay thawing and refreezing due to the nonlinear effects of latent heat (Nicolsky and Romanovsky, 2018).

RC: ***Fig B1. In my opinion, this figure is important as it shows the basic settings of forward modeling. Please consider moving it into the main text.***

AR: The updated diagram is now featured in Fig. 2 of the revised manuscript.

RC: ***Line 170. The depth information of each layer is missing.***

AR: Depth information for the site stratigraphies can be found in the Appendix (Tables B1-B4).

4. Response to referee 2

The text in this section is adapted from Groenke (2023b) with updated excerpts from the revised manuscript included where relevant.

RC: ***I regret the absence of a more through sensitivity analysis for the parameters that were not included.***

AR: We agree that a parameter sensitivity analysis would be a valuable addition to our work. We are happy to report that we have re-run the simulations with a larger ensemble size ($N = 512$) and run a sensitivity analysis on the prior samples using the EASI method of Plischke (2010) as implemented by Dixit and Rackauckas (2022). We have included the results of this analysis in the suplementary material.

The only model parameters not included in our analysis are the geothermal heat flux (this is discussed further below), constituent material thermal properties (conductivities and heat capacities), and physical constants. We also exclude soil composition parameters in layers where there is strong a priori reason to do so (e.g. organic content is excluded from deeper layers). We have added a table summarizing these constant parameters and their values in the supplement of the revised manuscript.

Of these excluded parameters, the most influential are certainly the constituent thermal properties, and in particular, the constituent thermal conductivities. While, in principle, it would be better to include these parameters in the analysis, the current parameterization of the bulk thermal conductivity and heat capacity in terms of the constituents would lead to colinearity with the soil composition parameters (i.e. organic content, porosity, saturation, and excess ice) that are varied already in our analysis. This is both numerically problematic and functionally redundant since the bulk thermal properties can already be adjusted by varying the soil constituent fractions. Choosing only one of these two sets of parameters (i.e. fixing the constants and varying the composition, or vice versa) avoids this problem.

We should note, however, that this approach would not be appropriate if the primary goal of the inversion were to recover the correct thermal properties or composition parameters from temperature measurements. This would require a more careful analysis such as the hierarchical approach in Wang and Zabaras (2005) as well as higher resolution temperature measurements (we use only annual means in this study). In this work, however, we are primarily interested in the predictive densities and the corresponding relationship between temperature and latent heat rather than the exact parameter values.

RC: ***L65-73. I regret that other solutions to this issue are not discussed at all. If EKS is one possible avenue, others have been proposed, such as the combination of efficient multiple-chain McMC algorithm with reduced dimension representation of the parameter space (Laloy et al., 2018) , or bypassing the inverse problem by directly predicting the posterior distribution from simulation-based machine learning approaches (Thibaut et al., 2022 for example in hydrogeology dealing with the same type of prediction (temperature field)). Since this aspect is also included in the discussion, it could be interesting to expand the perspectives beyond the technique used in the paper (i.e. EKS).***

AR: We thank the referee for the additional references. In particular, we were not aware of the work of Thibaut et al. (2022) and Hermans et al. (2018) on "Bayesian Evidential Learning" which looks like a very promising alternative to the inversion method used in this work. We have added further discussion on alternative approaches to the Introduction:

> Unfortunately, full-fledged Bayesian inference using standard numerical sampling methods is ~~generally~~ often infeasible for complex, simulation-based models where forward evaluation of the model over longer time periods 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 also to transient thermal A wide range of strategies for overcoming this challenge have been proposed. One common family of methods known as "approximate Bayesian computation" (ABC) attempt to circumvent the issue of computing the likelihood by means of rejection sampling and summary statistics (Sisson et al., 2018; Rubin, 1984). Particle filter (also known as "sequential Monte Carlo") methods (Sisson et al., 2007; Doucet et al., 2001) have also been widely employed due to their efficiency in solving nonlinear filtering problems (Kantas et al., 2014; Noh et al., 2011; Moradkhani et al., 2005). Other more recently proposed methods include Bayesian Evidential Learning (BEL) (Thibaut et al., 2022; Hermans et al., 2018) and the "Calibrate, emulate, sample" algorithm of Cleary et al. (2021), both of which leverage the predictive power of machine learning algorithms to obtain an approximate form of the posterior distribution using a small number of initial forward simulations and a learned mapping to the relevant predictors.

These computational challenges in inverse modeling are naturally also relevant for transient models of permafrost , processes which typically require solving a discretized partial differential equation for heat diffusion one or more partial differential equations governing heat and water flow 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. However, relatively few studies to date have attempted to address this challenge. Romanovsky and Osterkamp (1997) used an analytical equilibrium model to invert air and soil temperatures measured at three sites in northern Alaska; Nicolsky et al. (2009) later used a traditional variational approach to invert measured soil temperatures with a numerical model of 1D heat conduction similar to the one used in our study. Harp et al. (2016) analyzed the effects of soil property uncertainties on permafrost thaw using the Null Space Monte Carlo (NSMC) method applied to the Advanced Terrestrial Simulator (ATS) of Coon et al. (2019) . More recently, Garnello et al. (2021) used Bayesian methods to calibrate the GIPL permafrost model (?) and make long-term permafrost thaw projections.

RC: ***L133-134. The prior does not encode information about Y, it encodes information known about the unobserved quantities (X), before the data Y is actually collected and thus correspond to what we know and don't know about X before the experience.***

AR: This was a typo in the text which we have now fixed.

RC: ***Section 3.2. It was not directly clear to me that a Bayesian approach was applied to the trend analysis. Maybe it could be more explicit.***

AR: We have updated this section to include more detail on the trend analysis method.

Due to the small sample size of our annual data (fewer than 20 years at all sites), we analyze air and permafrost temperature trends with a 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 linear trend model. We choose the Student's t-distribution as the likelihood due to its longer tails which allow for more tolerance of unusually cold or warm years without severely affecting the overall trend (Lange et al., 1989) . We use standard, weakly-informative prior distributions for each parameter, with the exception of the slope μ_1 to which we assign 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. The specification of the full probability model can be found in Appendix -B1.

RC: ***L134-136. Is it? Since this integral is a constant for a given problem, the comparison of the likelihood (ratio) is sufficient to sample the posterior distribution (see rejection sampling or Metropolis sampling) and the integral is not such a problem. The main problem lies in the computation of the likelihood p(Y|X) which requires to solve the forward problem, generally through numerical approximation of partial derivative equations.***

AR: In the original text, we were referring to the historical difficulties of applying Bayesian methods that predated the development of numerical sampling based inference algorithms. These methods were not, to our knowledge, widely used for Bayesian inference until the work of Geman and Geman (1984) and then were further developed by Gelman et al. (1995) and Neal (2011) among others.

As discussed later in Sec. 3.5, the primary impediment to the application of modern sampling methods to simulation-based problems like the one in this work is the cost of the likelihood as described by the referee. We have rephrased this section to make this more clear:

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 [...]and the in numerical sampling methods such as Markov Chain Monte Carlo (MCMC) [...]which sidestep the need to compute the marginal likelihood, as well as a general increase in available computing power, have made Bayesian methods significantly more accessible.

RC: ***L190-191. I am wondering about the effect of this fixed boundary. Given the effort for modelling the uncertainty on the upper BC, why not also considering uncertainty on the bottom one? This flux is certainly not known for sure and it could contribute to significant uncertainty at depth.***

AR: It is true that the geothermal heat flux at the lower boundary can have substantial effect in some cases when (i) it is very large, e.g. in a volcanic area, (ii) the simulation period is very long (centuries or longer), or (iii) when the total depth-wise thickness of the modeled volume is relatively shallow, i.e. $< 100\,\mathrm{m}$. As we state in the text, however, we can safely assume the effect of the lower boundary to be negligible in our simulations since our simulation period is relatively short (40 years) and the depth of the lower boundary condition is very deep ($1\,\mathrm{km}$) so there simply is not enough time for such a small energy flux to propagate all the way to the upper $50\,\mathrm{m}$. Hermoso de Mendoza et al. (2020) found that, even when applying the average continental heat flux (which we use here) at a relatively shallow depth of $42\,\mathrm{m}$, the effect on the ground temperature at $2.86\,\mathrm{m}$ was only $(0.04 \pm 0.01)\,\mathrm{K}$ per $20\,\mathrm{mW\,m^{-2}}$ change in the geothermal heat flux. Given the much deeper depth of the lower boundary in our model, we would expect this effect to be still 1-2 orders of magnitude smaller. Thus, we chose to simply neglect it in the interest of limiting the dimensionality of the parameter space under study.

RC: ***Section 3.4. is empty.***

AR: This was due to a LaTeX typo which we have now fixed in the revised version.

RC: ***L218-220. This is a weird formulation. Any Bayesian approach will include some prior uncertainty on model parameters, and if modelling error is often neglected, it is generally included in the observation error from the likelihood. Maybe what is specific to your approach is that the target X and the predictor Y are actually the same (temperature) and that you have first to estimate the distribution of parameter phi?***

AR: It is common for applied Bayesian methods to neglect certain variables or parameters in the model. A common example is in regression problems, where the posterior distribution $p(\phi|X,Y)$ is typically sought. A fully Bayesian treatment of this problem would assign priors to the predictors, i.e: $p(\phi|X,Y) \propto p(Y|X,\phi)p(X,\phi)$. This is often referred to as a "generative" model since integrating over ϕ yields the joint data distribution $p(Y,X)$. However, $p(X)$ is often neglected which simply corresponds to an implicit uniform prior in the resulting probability model.

Similarly, here we mean to emphasize that the model states, $\mathbf{x}_{1:\tau}$, on which temperature is functionally dependent, are a deterministic function of the parameters, ϕ, in the probability model described by Eq. 7. A fully Bayesian solution of the problem would account for uncertainty in the model states, $\mathbf{x}_{1:\tau}$, by treating them also as stochastic quantities. This is precisely the formulation used by Hidden Markov Models (HMM) and their continuous equivalents, state space models (SSM) and stochastic differential equations (SDE). Since our forward model is a deterministic function that does not include such stochasticity, this distribution is implicitly a Dirac density, as we describe in this paragraph.

It is true that model error is implicitly also included in the observation error from the likelihood. The limitations of EKS in this regard are also discussed in Sec. 5.6.

The paragraph discussed here has been moved to Appendix B3.1 in the revised manuscript. We have further added the following remark to the paragraph in this section:

> In principle, model error can be implicitly accounted for in the likelihood $p(\mathbf{y}_{1:\tau}|\mathbf{x}_{1:\tau},\phi,\mathbf{s}_{1:\tau})$. However, as discussed in Sect. 5.5, EKS assumes the noise scale of the liklihood to be fixed and known a priori, which is generally insufficient for nontrivial inverse problems. As a result, the predictive distribution described by Eq. (B8) cannot be guaranteed to be well calibrated.

RC: ***L224-225. It sounds like a classical McMC approach wouldn't work. Any method sampling the posterior can solve the problem, right?***

AR: MCMC is not feasible here due to the fact that the random walk is sequential, and a very large number of samples (typically thousands) is required in order to produce high quality posterior samples for nontrivial models. More advanced, gradient-based methods such as Hamiltonian Monte Carlo (HMC) are more efficient in this regard but require computation of the gradient which is equally (if not more) costly in a dynamical model such as the one used in this work.

But in principle, any method that can sample the posterior could work, notwithstanding computational limitations. Other particle-based methods such as sequential Monte Carlo (SMC) or even importance sampling would also, in principle, be viable.

It is worth noting that Garnello et al. (2021) used adaptive Metropolis-Hastings to solve a similar inverse problem (with a different end goal) for the GIPL permafrost model (Jafarov et al., 2012). However, their calibration period was much shorter (6 years) and limited to one site.

RC: ***Computational time for one forward model?***

AR: Each model run takes between 5 and 15 minutes for the full 40 year simulation period on a 5-node compute cluster, each with 24 x Intel® Xeon® Gold 6128 CPUs @ 3.40GHz and 200 GiB of RAM. The compute time for each individual simulation depends on the size of the calculated maximum timesteps which in turn depends on the parameter settings. Heavier tailed freeze curve configurations (e.g. for silty or clay-like soils) tend to be slower due to the strong nonlinearity induced by the freeze curve over a much wider range of subzero temperatures.

RC: ***L253-254. Assuming uncorrelated noise in time and space might be one of the unrealistic assumptions, depending of the type of sensors of course. Maybe mention it in the discussion?***

AR: We have revised the last point of Sect. 5.6:

> A fully Bayesian treatment of the inversion problem would include these Σ_T ~~as parameters to be inferred, which~~ (or equivalent) parameters in the posterior distribution. This would have the benefit of producing ~~an ensemble with a predictive distribution that~~ posterior samples where the predictive distribution 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~~, i.e. the 95% prediction interval should actually cover approximately 95% of the observations in the calibration period, therefore providing a built-in measure of the degree to which the model is capable of explaining the observations.

While this does not directly address the question about noise correlation, it addresses the broader point about the need to not make assumptions about the noise structure, in particular with regard to the problem of mdoel error.

For observation (i.e. measurement) error alone, assuming no correlation is actually fairly reasonable, notwithstanding long-term drift in the temperature sensor calibration. This is mitigated, however, by semi-regular maintenance (Boike et al., 2019)

RC: ***L258-259. It is also a requirement for any Bayesian inference. The posterior is directly related to the prior, so the prior should reflect the actual knowledge about the site. McMC is sampling from the prior distribution as well.***

AR: This paragraph has been moved to the Appendix B3.1. The line has also been revised to read:

> The prior distribution over model parameters, $p(\varphi)$, is of crucial importance to any Bayesian inference procedure.

RC: ***L284-291. I wonder about the validity of deducing trend with such short data sets. In the discussion, comparison with longer trend is introduced, but I feel this could be strengthened.***

AR: We use a Bayesian trend model with semi-informative priors and a robust (Student-t) likelihood for exactly this reason (see Appendix B1 for details). The wider tails of the Student-t likelihood makes the model much more "skeptical" of outliers and thus more conservative in estimating trends. The magnitude of the posterior slopes, despite this robust formulation, suggests that the observed trends are not spurious, although it is of course impossible to extrapolate them into the past without additional data.

Following the suggestion of the referee, we have revised the corresponding method section to make the trend analysis method more clear (see above).

RC: ***What is the reason? Is this biased visible consistenly throughout the year. If yes, all the models of the posterior must have a high misfit, what could indicate a lack of consistency between the prior and the data set (e.g., Lopez-Alvis et al., 2019).***

AR: The reasons for these biases are discussed in Sec. 5.2 of the original manuscript. Please also see our response to the related comment below along with the supplementary results presented in Fig. R1.

RC: ***Close to what? I am not sure the sentence makes sense. Please clarify.***

AR: We assume that the referee is referring to this sentence on line 331: "Both warmer sites also show more overall variability across the 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."

We meant to say that the standard deviations of the latent heat and thaw depth trend slopes computed across the ensemble are generally much larger than those for the two colder sites, i.e. nearly double or more than double. Fortunately, our updated results with a larger ensemble have made this distinction more clear:

> ... ~~close to or sometimes~~ of the trend slopes more than double those of the ~~two colder~~ cold sites.

RC: ***L374-375. Is this significant enough to state that the difference is not only due to the depth of the sensor? If both had some sensors deeper, this would largely reduce the uncertainty and has likely nothing to do with the fact that they are warmer, don't you think?***

AR: Following the suggestion of the other referee, we repeated the Samoylov simulations with all sensors below $10\,\mathrm{m}$ excluded from the inference procedure. The results showed that this does indeed increase the spread of the predicted temperature, as we might expect, but does not affect the uncertainty related to changes in latent heat and active layer thickness. We have revised the text and moved it to section 5.3:

> 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. 5). ~~This variability reveals fundamental uncertainties about what observed changes in permafrost temperatures actually tell us.~~ We see, for example, significant differences in the overall ~~spread~~ range of predicted temperatures between sites. This can be partially explained by ~~differences in the depths of the observed temperatures~~both differences in air temperature variability (Fig. 4) as well as in the depths at which temperature observations are available; 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 closer to the surface, suggesting that depth 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 (~~We present additional experiments exploring the sensitivity of our inversion method to available sensor depths in the supplementary material (Sect. S1). We find that the inclusion or exclusion of measurements at various depths has a fairly predictable effect on the results of the analysis; i.e. ~~near the depth of zero annual amplitude) where thermal gradients are smaller and heat diffuses more slowly.~~ the inclusion of data at additional sensor depths generally improves the fidelity of the predicted temperatures at these depths and reduces the associated uncertainty. However, the results suggest that constraining the model using only deep measurements, as we do for the Barrow site, may lead to overestimated warming closer to the surface.

RC: ***Why a reference to the median suddenly? The discrepancy is for all models, not only the median or the mean, all predictions are wrong.***

AR: We assume that the referee is referring to this sentence: "For Samoylov, the ensemble median of the modeled annual temperature range of the permafrost layers is slightly too narrow..."

This is actually a mistake; we have now corrected this line in the revised text:

> For Samoylov, the ensemble ~~median of the modeled~~ annual temperature range of the permafrost layers is slightly too narrow...

In general, we often refer to the median elsewhere in the text because, as a quantile, it is robust against distributional skew which does occur in some of the ensemble predictive distributions (most dramatically for Bayelva).

RC: ***L386-387. Have you tried enlarging the prior? Do you observe a similar bias for other years?***

AR: We generally tried to select priors for the n-factors and soil properties that were informative enough to be useful but not so informative where they would be restrictive. Ideally we would perform a robust sensitivity analysis with a wide range of priors, but given the complexity of the downstream modeling task, this is prohibitively difficult. While we did perform a number of ad-hoc experiments with wider or narrower priors, this did not affect this specific issue for Samoylov. The bias is also persistent across years (Fig. R1) with the exception of 2016 in which the observed temperatures are unusually warm due to early season snowfall.

It is very likely that the low temporal-resolution (bidecadal) of the n-factor parameterization is the source of this bias since it cannot capture interseasonal (or even interannual) variation in snow cover effects. It is only a very coarse approximation of snow cover effects to allow the model to capture long term trends.

RC: ***What do you mean by "smooth regularizers". One of the advantage of Bayesian inversion is to use realistic prior and avoid regularization (such as smoothing). Smoothing is normally only visible in the mean which is not necessarily a sample of the posterior.***

AR: The intent here was to relate the role of the prior in Bayesian inference to the role of regularization in classical optimization problems (Calvetti and Somersalo, 2018). It is a well known result that there is a correspondence between certain regularization schemes and families of prior distributions, e.g. ℓ_2-regularization in ridge regression corresponds to a multivariate Gaussian prior, and ℓ_1-regularization in LASSO corresponds to a multivariate Laplace prior (Hastie et al., 2009).

This sentence has now been removed from the revised manuscript since it is nonessential and, as the referee pointed out, potentially confusing.

RC: ***I think the term "appropriate" is badly chosen. The prior should reflect the uncertainty on model parameters before considering the data set, and what is mentioned should then result in larger prior. The main challenge is maybe to consider correlation between parameters?***

AR: We have improved the wording of this sentence and added a remark about correlation:

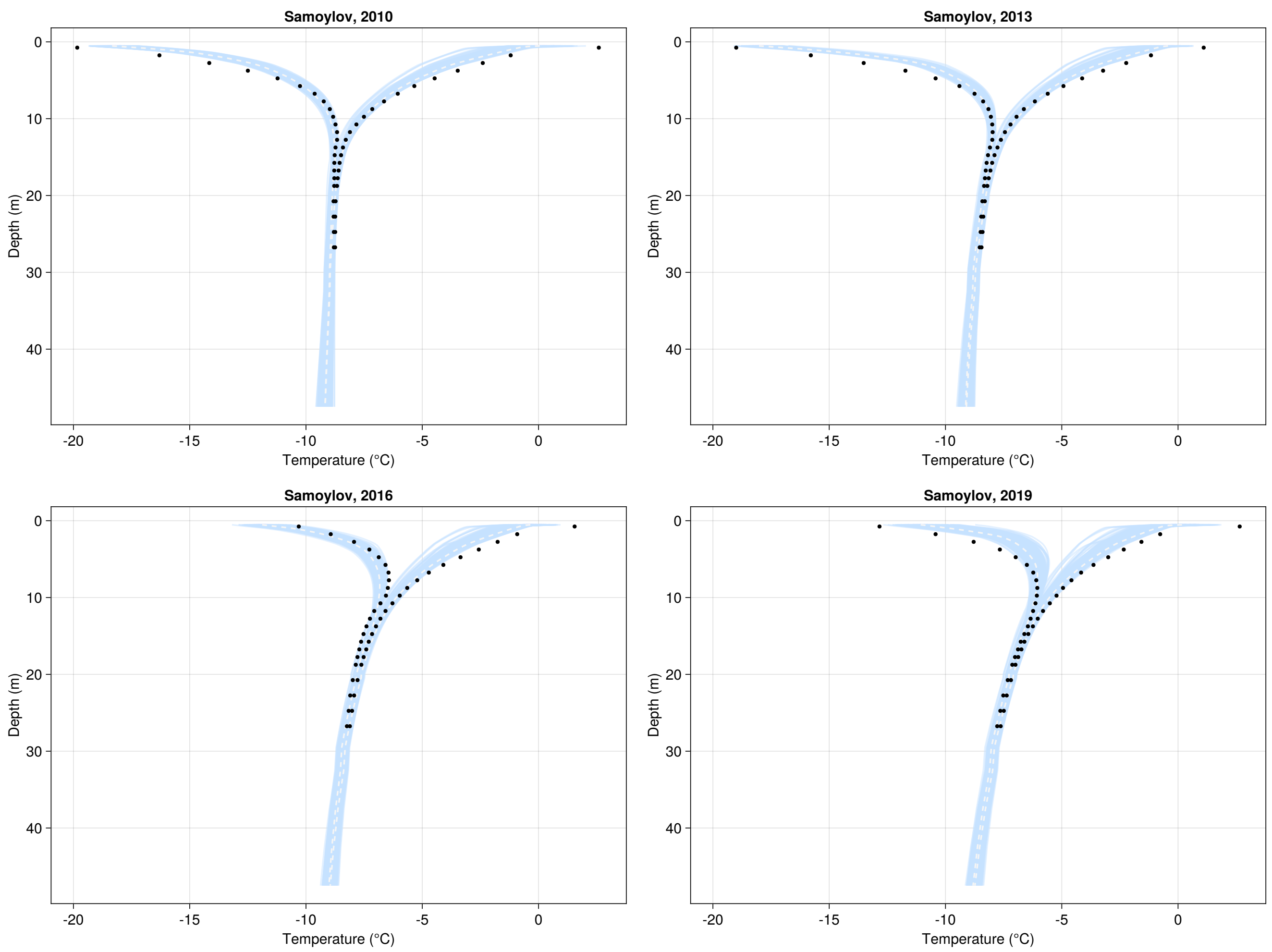

Figure R1. Samoylov ensemble trumpet curves (annual min vs max temperature) for four selected years with data available: 2010, 2013, 2016, 2019. The black dots are observed min/max temperatures for that year. The dashed white line shows the median annual temperature.

> However, the general lack of precise error and uncertainty estimates for some model parameters (in particular, soil properties) makes the ~~selection of appropriate~~ construction of prior distributions difficult. Furthermore, it is very likely that many parameters are, in reality, highly correlated, which we are unable to account for with our current approach.

RC: ***L515-516. This is also the case for most McMC approaches.***

AR: To clarify, we meant that the noise covariance must be fixed a priori, i.e. it is not treated as a parameter. Since MCMC is a more flexible method that can work on any target density, the noise parameters are typically included in the posterior distribution (Gelman et al., 1995) and sampled to fit the data. This is not possible with EKS. We have clarified this in the revised manuscript.

A fully Bayesian treatment of the inversion problem would include these Σ_T as parameters to be inferred, which (or equivalent) parameters in the posterior distribution. This would have the benefit of producing an ensemble with a predictive distribution that posterior samples where the predictive distribution is well calibratedon the training data. This is, however, not possible using EKS, and thus, we leave the problem of estimating noise parameters to future work, i.e. the 95% prediction interval should actually cover approximately 95% of the observations in the calibration period, therefore providing a built-in measure of the degree to which the model is capable of explaining the observations.

RC: ***L520-523. See comment 1.***

AR: We have now included further discussion on related methods (see above).

RC: ***Table B5. Display prior distributions behind posterior in figures A2 to A5 to immediately grasp the reduction in parameter uncertainty?***

AR: This is a great suggestion, and we have implemented it in the revised manuscript.

References

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 Opportunity 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.

Calvetti, D. and Somersalo, E.: Inverse Problems: From Regularization to Bayesian Inference, WIREs Computational Statistics, 10, https://doi.org/10.1002/wics.1427, 2018.

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.

Coon, E., Svyatsky, D., Jan, A., Kikinzon, E., Berndt, M., Atchley, A., Harp, D., Manzini, G., Shelef, E., Lipnikov, K., Garimella, R., Xu, C., Moulton, D., Karra, S., and Painter, S.: Advanced Terrestrial Simulator, USDOE Office of Science (SC), Biological and Environmental Research (BER) (SC-23), 2019.

Cosenza, P., Guérin, R., and Tabbagh, A.: Relationship between Thermal Conductivity and Water Content of Soils Using Numerical Modelling, European Journal of Soil Science, 54, 581–588, https://doi.org/10.1046/j.1365-2389.2003.00539.x, 2003.

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.

Dixit, V. K. and Rackauckas, C.: GlobalSensitivity. Jl: Performant and Parallel Global Sensitivity Analysis with Julia, Journal of Open Source Software, 7, 4561, 2022.

Doucet, A., de Freitas, N., and Gordon, N.: An Introduction to Sequential Monte Carlo Methods, in: Sequential Monte Carlo Methods in Practice, edited by Doucet, A., de Freitas, N., and Gordon, N., Statistics for Engineering and Information Science, pp. 3–14, Springer, New York, NY, https://doi.org/10.1007/978-1-4757-3437-9_1, 2001.

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

Garnello, A., Marchenko, S., Nicolsky, D., Romanovsky, V., Ledman, J., Celis, G., Schädel, C., Luo, Y., and Schuur, E. a. G.: Projecting Permafrost Thaw of Sub-Arctic Tundra With a Thermodynamic Model Calibrated to Site Measurements, Journal of Geophysical Research: Biogeosciences, 126, e2020JG006 218, https://doi.org/10.1029/2020JG006218, 2021.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, Chapman and Hall/CRC, 1995.

Geman, S. and Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, 721–741, https://doi.org/10.1109/TPAMI.1984.4767596, 1984.

Groenke, B.: Author Response to First Referee Comments on "Investigating the Thermal State of Permafrost with Bayesian Inverse Modeling of Heat Transfer", https://doi.org/10.5194/egusphere-2022-630-AC1, 2023a.

Groenke, B.: Author Response to Second Referee Comments on "Investigating the Thermal State of Permafrost with Bayesian Inverse Modeling of Heat Transfer", https://doi.org/10.5194/egusphere-2022-630-AC2, 2023b.

Harp, D. R., Atchley, A. L., Painter, S. L., Coon, E. T., Wilson, C. J., Romanovsky, V. E., and Rowland, J. C.: Effect of Soil Property Uncertainties on Permafrost Thaw Projections: A Calibration-Constrained Analysis, The Cryosphere, 10, 341–358, https://doi.org/10.5194/tc-10-341-2016, 2016.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2, Springer, 2009.

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A., and Caers, J.: Uncertainty Quantification of Medium-Term Heat Storage From Short-Term Geophysical Experiments Using Bayesian Evidential Learning, Water Resources Research, 54, 2931–2948, https://doi.org/10.1002/2017WR022135, 2018.

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.

Jafarov, E. E., Marchenko, S. S., and Romanovsky, V. E.: Numerical Modeling of Permafrost Dynamics in Alaska Using a High Spatial Resolution Dataset, The Cryosphere, 6, 613–624, https://doi.org/10.5194/tc-6-613-2012, 2012.

Kantas, N., Beskos, A., and Jasra, A.: Sequential Monte Carlo Methods for High-Dimensional Inverse Problems: A Case Study for the Navier–Stokes Equations, SIAM/ASA Journal on Uncertainty Quantification, 2, 464–489, https://doi.org/10.1137/130930364, 2014.

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., 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.

Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E., and Schoendorf, J.: Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs, The Journal of Physical Chemistry A, 114, 6022–6032, https://doi.org/10.1021/jp9096919, 2010.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S.: Uncertainty Assessment of Hydrologic Model States and Parameters: Sequential Data Assimilation Using the Particle Filter, Water Resources Research, 41, https://doi.org/10.1029/2004WR003604, 2005.

Neal, R. M.: MCMC Using Hamiltonian Dynamics, in: Handbook of Markov Chain Monte Carlo, pp. 139–188, Chapman and Hall/CRC, 2011.

Nicolsky, D. J. and Romanovsky, V. E.: Modeling Long-Term Permafrost Degradation, Journal of Geophysical Research: Earth Surface, 123, 1756–1771, https://doi.org/10.1029/2018JF004655, 2018.

Nicolsky, D. J., Romanovsky, V. E., and Panteleev, G. G.: Estimation of Soil Thermal Properties Using In-Situ Temperature Measurements in the Active Layer and Permafrost, Cold Regions Science and Technology, 55, 120–129, https://doi.org/10.1016/j.coldregions.2008.03.003, 2009.

Noh, S. J., Tachikawa, Y., Shiiba, M., and Kim, S.: Applying Sequential Monte Carlo Methods into a Distributed Hydrologic Model: Lagged Particle Filtering Approach with Regularization, Hydrology and Earth System Sciences, 15, 3237–3251, https://doi.org/10.5194/hess-15-3237-2011, 2011.

Park, H., Fedorov, A. N., Zheleznyak, M. N., Konstantinov, P. Y., and Walsh, J. E.: Effect of Snow Cover on Pan-Arctic Permafrost Thermal Regimes, Climate Dynamics, 44, 2873–2895, https://doi.org/10.1007/s00382-014-2356-5, 2015.

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.

Plischke, E.: An Effective Algorithm for Computing Global Sensitivity Indices (EASI), Reliability Engineering & System Safety, 95, 354–360, https://doi.org/10.1016/j.ress.2009.11.005, 2010.

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

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.

Riseborough, DW.: Soil Latent Heat as a Filter of the Climate Signal in Permafrost, in: Proceedings of the Fifth Canadian Permafrost Conference, Collection Nordicana, vol. 54, pp. 199–205, Citeseer, 1990.

Romanovsky, V. E. and Osterkamp, T. E.: Thawing of the Active Layer on the Coastal Plain of the Alaskan Arctic, Permafrost and Periglacial Processes, 8, 1–22, https://doi.org/10.1002/(SICI)1099-1530(199701)8:1<1::AID-PPP243>3.0.CO;2-U, 1997.

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.

Rubin, D. B.: Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician, The Annals of Statistics, 12, 1151–1172, 1984.

Sisson, S. A., Fan, Y., and Tanaka, M. M.: Sequential Monte Carlo without Likelihoods, Proceedings of the National Academy of Sciences, 104, 1760–1765, https://doi.org/10.1073/pnas.0607208104, 2007.

Sisson, S. A., Fan, Y., and Beaumont, M.: Handbook of Approximate Bayesian Computation, CRC Press, 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.

Thibaut, R., Compaire, N., Lesparre, N., Ramgraber, M., Laloy, E., and Hermans, T.: Comparing Well and Geophysical Data for Temperature Monitoring within a Bayesian Experimental Design Framework, Water Resources Research, 58, e2022WR033 045, 2022.

Wang, J. and Zabaras, N.: Hierarchical Bayesian Models for Inverse Problems in Heat Conduction, Inverse Problems, 21, 183–206, https://doi.org/10.1088/0266-5611/21/1/012, 2005.