Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik
Institut für Technische Informatik und Mikroelektronik
Robotic Interactive Perception Laboratory

Technische Universität Berlin

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

Master Thesis

Extension and Sensitivity Analysis of the EKS-Based Inversion Method for Reconstructing Historical Ground Surface Temperatures from Permafrost Borehole Data

Shorif Alamgir
Matriculation-Nr. 381305
Studiengang Computer Engineering

Berlin, April 15, 2026

Examiners: Prof. Dr. Guillermo Gallego, Prof. Dr. Julia Boike
Advisor: Dr. Frederieke Miesner

Selbständigkeitserklärung

Hiermit erkläre ich, dass die eingereichte Abschlussarbeit meine eigene, ohne fremde Hilfe erstellte Arbeit ist. Es wurden nur die aufgeführten Quellen und Hilfsmittel verwendet. Alle aus den Quellen und Hilfsmitteln entnommenen Stellen, sei es wörtlich oder sinngemäß, sind als solche kenntlich gemacht. Bezüglich generativer KI-Tools habe ich ***ChatGPT und DeepL*** zur Überprüfung meiner Texte hinsichtlich Grammatik, Rechtschreibung, Zeichensetzung und zur Verbesserung der Lesbarkeit verwendet. Ich bin vollständig verantwortlich für die Auswahl, Übernahme und alle Ergebnisse der von mir verwendeten KI-generierten Ausgaben. Ich habe die Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der TU Berlin vom 15. Februar 2023 zur Kenntnis genommen. https://www.static.tu.berlin/fileadmin/www/10000000/Forschen/Forschungsprofil/Dokumente/Implementation_Regulations_for_Good_Research_Practice.pdf Ich erkläre weiterhin, dass ich die Abschlussarbeit nicht in gleicher oder ähnlicher Form bei einer anderen Prüfungsbehörde eingereicht habe.

Statement of authorship

I hereby declare that the thesis submitted is my own, unaided work, completed without any external help. Only the sources and resources listed were used. All passages taken from the sources and aids used, either unchanged or paraphrased, have been marked as such. As for generative AI tools, I have used ***ChatGPT and DeepL*** for checking my writing regarding grammar, spelling, punctuation, and improving readability. I am fully responsible for the selection, adoption, and all results of the AI-generated output I use. I have taken note of the Principles for Ensuring Good Research Practice at TU Berlin dated 15 February 2023. https://www.static.tu.berlin/fileadmin/www/10000000/Forschen/Forschungsprofil/Dokumente/Implementation_Regulations_for_Good_Research_Practice.pdf I further declare that I have not submitted the thesis in the same or similar form to any other examination authority.

Berlin, April 15, 2026

Shorif Alamgir

. . . . . .

Abstract

Die Rekonstruktion historischer Bodenoberflächentemperaturen (GST) aus Permafrost-Bohrlochdaten stellt ein starkes inverses Problem dar, bei dem Unsicherheiten in den Messungen, den Untergrundeigenschaften und den Modellannahmen die Zuverlässigkeit der rekonstruierten Temperaturverläufe begrenzen. Die meisten bestehenden Inversionsstudien auf Basis von Bohrlochdaten verwenden lediglich ein einzelnes Temperatur-Tiefen-Profil pro Standort und nutzen daher die in wiederholten Messungen enthaltene Information nicht aus.
Aus numerischer und rechentechnischer Sicht handelt es sich um ein stark gestelltes Inversionsproblem, dessen Lösungsstabilität stark von den Priorannahmen, der Struktur der Likelihoods und der Modellierung des Beobachtungsrauschens abhängt. Die vorgeschlagene Multiple-Likelihood-Formulierung verändert das Inferenzproblem selbst, indem bedingt unabhängige Datenrestriktionen eingeführt werden, was sich direkt auf die posteriori Unsicherheit, die numerische Stabilität und die Sensitivität gegenüber Modell-Daten-Abweichungen auswirkt.
Aufbauend auf dem Ensemble-Kalman-Sampling-Framework (EKS) von Groenke et al. [1] adressiert diese Arbeit diese Lücke durch die Einführung einer Multiple-Likelihood-Formulierung, die mehrere, potenziell nicht-aufeinanderfolgende Bohrloch-Temperaturprofile desselben Standorts innerhalb einer einzelnen bayesianischen Inversion integriert. Zusätzlich wird ein datengetriebener GST-Prior konstruiert,der auf Lufttemperaturen (bspw. von ERA5) basiert, die mithilfe eines n-Faktor-Ansatzes zur Beschreibung der Luft-Boden-Kopplung transformiert werden.
Die Methode wird auf drei Permafrost-Bohrlochstandorte im Lena-Delta angewendet (Sardakh, Samoylov und Tiksi). Eine systematische Sensitivitätsanalyse untersucht den Einfluss von Annahmen zum Messrauschen, der Anzahl und zeitlichen Verteilung der Likelihood-Profile sowie der Bohrlochtiefe auf die Stabilität und Unsicherheit der Rekonstruktionen.
Die Ergebnisse zeigen, dass die Einbeziehung mehrerer Likelihoods die posteriore Variabilität verringern und die zeitliche Stabilität der rekonstruierten GST-Verläufe verbessern kann. Quantitative Fehlermetriken wie RMSE und normalisierte Log-Likelihood verbessern sich jedoch für realistische Rauschannahmen nicht konsistent, was darauf hindeutet, dass der Hauptnutzen in einer Stabilisierung der Rekonstruktion und nicht in systematischen Genauigkeitsgewinnen liegt. Der datengetriebene Prior verbessert zusätzlich die physikalische Plausibilität, indem er die Rekonstruktionen an lokale klimatologische Bedingungen koppelt.Des Weiteren Ergebnisse zeigen, dass wiederholte Bohrlochmessungen zeitliche Information enthalten, die in traditionellen Inversionsansätzen typischerweise ungenutzt bleibt.
Auch wenn der Ansatz weder die durch thermische Diffusion gesetzten physikalischen Grenzen noch die Einschränkungen vereinfachter Modelle überwindet, stellt er ein praktikables Framework dar, um stabilere und besser interpretierbare GST-Rekonstruktionen aus bestehenden Bohrlochdatensätzen zu gewinnen, und weist zugleich auf zentrale Ansatzpunkte für zukünftige Modellweiterentwicklungen und Validierungsstudien hin.

Abstract

Reconstructing historical ground surface temperatures (GST) from permafrost borehole data is a strongly ill-posed inverse problem, where uncertainties in measurements, subsurface properties, and model assumptions limit the reliability of inferred temperature histories. Most existing borehole inversion studies rely on a single temperature–depth profile per site and therefore do not exploit information contained in repeated measurements.

From a computational and numerical perspective, the problem constitutes a ill-posed inverse problem, where solution stability strongly depends on prior assumptions, likelihood structure, and noise modeling. The proposed multiple-likelihood formulation modifies the inference problem itself by introducing conditionally independent data constraints, which directly affects posterior uncertainty, numerical stability, and sensitivity to model–data mismatch.

Building upon the Ensemble Kalman Sampling (EKS) framework of Groenke et al. [1], this thesis addresses this gap by introducing a multiple-likelihood formulation that incorporates several, potentially non-consecutive borehole temperature profiles from the same site within a single Bayesian inversion. In addition, a data-driven GST prior is constructed using either in-situ GST observations or ERA5 air temperatures combined with an n-factor approach to represent air–ground coupling.

The method is applied to three permafrost borehole sites in the Lena River Delta (Sardakh, Samoylov, and Tiksi). A systematic sensitivity analysis investigates the influence of measurement noise assumptions, the number and temporal spacing of likelihood profiles, and borehole depth on reconstruction stability and uncertainty.

The results show that incorporating multiple likelihoods can reduce posterior variability and improve the temporal stability of reconstructed GST histories. However, quantitative error metrics such as RMSE and normalized log-likelihood do not improve consistently for realistic noise levels, indicating that the main benefit lies in stabilisation rather than systematic accuracy gains. The data-driven prior further improves physical plausibility by anchoring reconstructions to local climatological conditions. Furthermore, these findings demonstrate that repeated borehole measurements contain exploitable temporal information that is typically ignored in traditional inversions.

While the approach does not overcome the physical limits imposed by thermal diffusion or model simplifications, it provides a practical framework for extracting more stable and interpretable GST histories from existing borehole datasets and highlights key directions for future model development and validation.

Contents

1 Introduction

1.1 Motivation

In recent decades, the Arctic has experienced a strong and continuous rise in air and ground surface temperatures, leading to widespread permafrost thaw. Permafrost is defined as ground material that remains at or below 0 °C for at least two consecutive years[2–5]. This process is not only a clear indicator of climate change, but also acts as a powerful feedback mechanism, amplifying warming through the release of greenhouse gases such as carbon dioxide and methane. Understanding the long-term evolution of permafrost temperatures is therefore critical for climate research [6].

However, direct temperature measurements in permafrost regions are only available for the past few decades, which makes it difficult to assess the full extent of long-term warming trends [7]. Borehole temperature profiles offer a unique opportunity to reconstruct past ground surface temperatures (GST) over several centuries [7]. Yet, this reconstruction is mathematically challenging, as it constitutes an ill-posed inverse problem that is sensitive to noise and missing data [1, 8].

Early studies demonstrated that borehole temperature profiles can be used to infer past ground surface temperature changes using linear inversion techniques. For example, Lachenbruch and Marshall [9] successfully applied linear regularization methods to identify long-term warming trends in Alaskan permafrost. Clow [10] further refined this approach by analyzing the temporal smearing inherent in borehole-based reconstructions, highlighting the ability of these methods to capture low-frequency climate signals. Despite these early successes, traditional inversion techniques, such as Tikhonov regularization or deterministic optimizations, often provide only a single solution and do not properly capture the uncertainty in the reconstructed temperature history. In contrast, probabilistic methods based on Bayesian inference allow the reconstruction to be described using probability distributions [11]. One such method is Ensemble Kalman Sampling (EKS), which is used in this study.

This study explores the use of EKS for reconstructing historical GST from borehole data in the Lena Delta, one of the most rapidly warming permafrost regions in the Arctic. The goal is not only to estimate the past temperature evolution, but also to quantify the uncertainty of the reconstruction and assess the reliability of borehole data for long-term climate analysis.

1.2 Contribution

Reconstructing past ground surface temperatures (GST) from borehole data is an important method to understand how the climate and permafrost have changed over time. When the surface gets warmer or colder, thermal energy is transferred into or out of the ground through the process of heat conduction. These changes leave small signals in the temperature–depth profile of the borehole. By looking at these signals, it is possible to find out how the surface temperature changed in the past.
Over the years, various inversion techniques have been applied to reconstruct past ground surface temperatures from these temperature–depth signals measured in boreholes. One of the most popular technique for Ground Surface Temperatures (GST) reconstruction is the acsvd ([12], [13], [14], [15], [16]). However, these deterministic methods face problems in integrating uncertainties and are not designed to represent complex physical processes such as latent heat release or phase changes in permafrost environments [1]. In contrast, Bayesian inversion frameworks offer a probabilistic formulation that naturally accounts for uncertainty and allows the integration of prior knowledge and multiple sources of variability [17]. Among Bayesian approaches, Markov Chain Monte Carlo (MCMC) methods and Bayesian hierarchical models have been applied to reconstruct past ground surface temperatures from borehole data [8, 18]. These methods allow an explicit treatment of uncertainty but are often computationally expensive. Nonetheless, only a few Bayesian GST reconstruction studies have been applied to permafrost regions in the Arctic, despite the fact that these areas have experienced some of the strongest warming trends over the past decades [19]. Moreover, many of the inversion methods, which were applied in permafrost sites, neglect important physical processes such as the presence of an active layer or the role of latent heat. A notable exception is the recent work by Groenke et al. (2024) [1], who introduced a stochastic inversion framework based on Ensemble Kalman Sampling (EKS).
A key innovation of this work is that it considers the seasonal freezing and thawing of the active layer and the effects of latent heat in the inversion of ground surface temperature histories. The reconstructed GST are based on temperature profiles of boreholes. This represents a significant step toward more physically realistic GST reconstructions in permafrost environments. However, Groenke's method [1] was applied to a single temperature–depth profile from one borehole and did not investigate how integrating multiple temperature profiles measured at different times from the same borehole could affect the robustness and accuracy of the reconstruction. Other studies, such as Guglielmin et al. (2018) [20], have used multiple temperature profiles from consecutive years to reduce observational noise through averaging. Yet, these profiles were not treated as independent constraints within the inversion and were based on classical linear methods, without accounting for processes such as latent heat effects or active layer dynamics. To date, no study has combined physically detailed, stochastic inversion methods with multiple borehole temperature profiles measured at different times, whether these measurements are taken in consecutive years or not.
First, relying on a single temperature profile makes the reconstruction sensitive to year-specific disturbances or random measurement noise. Using more than one profile reduces this dependence on a single observation year. When several profiles are treated as separate likelihood terms, random effects may be partially averaged out, reducing the influence of individual outliers on the overall reconstruction.

Second, the proposed approach does not attempt to bypass the physical constraints imposed by thermal diffusion in the subsurface. Instead, it uses repeated measurements to make more consistent use of the available thermal information within the same physical model. Any extension of the recoverable past therefore remains limited by borehole depth and the governing heat conduction processes, while allowing the available data to be exploited more effectively within these physical bounds.
Finally, the approach is motivated by the practical characteristics of borehole data in Arctic permafrost regions. Long and continuous temperature records are often unavailable due to harsh environmental conditions and sensor failures. In practice, available data may therefore consist of only a few observation years with complete temperature–depth profiles, while other years may contain gaps or missing measurements. The proposed formulation does not require all measurement years to be complete or temporally consecutive, as long as some years contain fully observed borehole temperature profiles that can be used in the inversion.
In this thesis, I extend the EKS framework of Groenke et al. (2024) [1] by introducing a multiple-likelihood approach that makes it possible to use several temperature–depth profiles from the same borehole, measured in different years. The objective of this extension is to investigate whether integrating multiple, possibly non-consecutive, temperature–depth profiles from the same borehole can enhance GST reconstructions from permafrost boreholes.
I hypothesize that extending the EKS framework to use multiple temperature profiles from different years, possibly non-consecutive, leads to more robust and accurate GST reconstructions from permafrost boreholes. In particular:

1. Integrating multiple reference temperature profiles into the EKS framework reduces posterior uncertainty and decreases the Root Mean Square Error (RMSE) of reconstructed GST relative to observed GST, compared to a single-profile inversion.

2. Using multiple reference temperature profiles improves GST reconstructions from shallower borehole temperature profiles compared to a single-profile approach.

Based on this objective, the following research questions are formulated:

1. RQ1: How does integrating multiple, non-consecutive borehole temperature profiles into the EKS framework improve ground surface temperature reconstructions in permafrost regions, by reducing their variability (standard deviation) and error compared to measured GST data (RMSE)?

2. RQ2: How does the number and temporal distribution of available profiles affect the variability (standard deviation) and temporal resolution (ability to resolve annual to decadal trends) of the reconstruction?

3. RQ3: How does this method affect the depth-dependence of GST reconstructions, and how does it affect the performance for shallower borehole temperature profiles compared to that of deeper reference profiles?

4. RQ4: How sensitive are the multiple-likelihood EKS reconstruction results to core configuration parameters, such as the assumed measurement noise, in terms of variability and error compared to measured GST data?

2 Background and Theoretical Foundations

2.1 Permafrost and its Role in the Changing Climate

2.1.1 Definition and Global Distribution

The cryosphere includes all areas of the Earth where water remains frozen for long periods, such as glaciers, snow, sea ice, and permafrost. One of the main components of the cryosphere is permafrost, which was termed by S. W. Muller, a Professor of Geology at Stanford University [2]. Permafrost is defined on the basis of temperature and time, that is, when the ground remains at or below 0°C for at least two consecutive years [2]. According to current estimates, it covers around 11% of the earth surface [4]. The majority of permafrost is found in the Northern Hemisphere, where it underlies nearly 15% of the exposed land area [4]. In addition to polar regions, smaller areas of permafrost are found in high mountain sites such as the Alps and the Himalayas.
Over the past decades, global temperatures have risen at an unprecedented rate, with the Arctic and subarctic regions warming nearly four times as fast as the global average [19]. As a result, permafrost has been gradually thawing due to rising air temperatures, which in turn leads to increased ground surface temperatures. This ongoing temperature increase not only drives further permafrost thaw but also amplifies feedback mechanisms, such as the release of methane and carbon dioxide from thawing organic material, that significantly accelerate global warming and further enhance permafrost degradation [4].

2.1.2 Permafrost Warming and Observed Temperature Trends

Recent studies provide strong evidence for a persistent increase in temperature in permafrost sites [6, 9, 21]. For example, Biskaborn et al. [6] identified a global average increase of $(0.29 \pm 0.12)\,^{\circ}\mathrm{C}$ in near-surface permafrost temperatures between 2007 and 2016, with the strongest warming observed in the cold, continuous permafrost zones [6]. If this process continues, ongoing permafrost degradation and the release of carbon dioxide and methane could amplify global warming by an additional $0.13\,^{\circ}\mathrm{C}$ to $0.27\,^{\circ}\mathrm{C}$ by 2100 [6].
To better understand spatial variability, permafrost regions are commonly categorized according to the areal extent of frozen ground. According to Obu et al. [3], continuous permafrost refers to regions where 90–100% of the ground is underlain by permafrost. In contrast, discontinuous permafrost is present in 50–90% of the area, often interrupted by unfrozen zones due to local heat sources or vegetation. Sporadic permafrost covers only 10–50% of the landscape. Lastly, isolated patches of permafrost are found in less than 10% of the area and exist only under very specific conditions.
These different permafrost zones also respond differently to climate warming. Previous studies have reported a pronounced warming trend in Arctic permafrost regions, with continuous

permafrost zones showing particularly strong increases in mean annual ground temperatures over recent decades [6]. This is especially relevant for sites such as Samoylov and Sardakh in the Lena River Delta, which are located in continuous permafrost regions and are therefore highly sensitive to atmospheric warming.
Reconstructions of past ground surface temperatures (GST) from boreholes play a key role in permafrost research, as they provide the long-term perspective required to assess permafrost–climate interactions beyond the instrumental period [7]. Several methods and models have been developed to infer historical GST from permafrost boreholes, each with different approaches to handling borehole data [10, 20, 22].
However, traditional approaches often struggle to account for latent heat effects caused by seasonal freeze–thaw processes in the active layer, which can introduce systematic biases in the reconstructions. To address this, Groenke et al. [1] proposed a novel framework that couples a two-phase heat transport model with ensemble-based Bayesian inference. This approach explicitly incorporates the thermal dynamics of the active layer, resulting in more robust reconstructions.
The methodological framework introduced by Groenke et al. forms the basis of the forward model and inversion strategy employed in this thesis. Its components, including the two-phase heat transport formulation and the ensemble-based Bayesian inference scheme, are described in detail in the following sections.

2.2 Heat Conduction in Frozen Soils

Heat conduction in frozen soils controls how surface temperature variations propagate into the subsurface, making it a fundamental component of GST reconstruction from borehole data. In permafrost, heat transfer is strongly influenced by the presence of ice and thin layers of unfrozen water surrounding soil particles, which modify the thermal properties of the ground [23].
Differences in thermal conductivity between frozen and unfrozen soil lead to distinct heat transfer behaviour, while latent heat effects during freezing and thawing further slow down or accelerate temperature changes. Since temperature variations at the surface diffuse gradually into the ground, the thermal characteristics of frozen soil determine how past climate signals are preserved and interpreted.
This process is commonly described by the heat conduction equation (Eqn. 2.2), which forms the basis of forward models used to simulate vertical temperature propagation in the subsurface.

2.2.1 Heat Conduction Equation

Heat conduction in soil transfers thermal energy from warmer to cooler regions, driven by temperature gradients [23]. The general mathematical formulation for heat conduction follows *Fourier's Law* and in three dimensions, the equation is written in vector notation as follows[24, 25]:

$$q = -k \cdot \nabla T \tag{2.1}$$

where q ($\mathrm{W\,m^{-2}}$) is the heat flux vector and $T = T(x, y, z)$ is the temperature field. The coefficient k with the unit $\mathrm{W\,m^{-1}\,K^{-1}}$ is called thermal conductivity and it determines how efficiently heat is conducted through the medium. The equation 2.1 describes a vector quantity that is proportional to and opposite in direction to the temperature gradient ∇T [24].
To describe how temperature evolves over time within the subsurface, Fourier's law is combined with conservation of energy, leading to the transient heat conduction equation. The equation takes the form [26]:

$$\nabla \cdot (k \nabla T) = c \frac{\partial T}{\partial t}, \tag{2.2}$$

where c ($\mathrm{J\,m^{-3}K^{-1}}$) is the volumetric heat capacity, and t is time. Equation (2.2) expresses that temporal changes in temperature are governed by spatial gradients in conductive heat flow. Both k and c depend strongly on whether the soil is frozen or unfrozen. Frozen soil generally exhibits higher thermal conductivity due to the presence of ice, whereas unfrozen soil contains liquid water, which has lower conductivity but increases the effective heat capacity. This difference leads to pronounced thermal asymmetry between freezing and thawing conditions—an essential feature of permafrost ground.
In permafrost regions, phase change processes further modify the thermal response of the subsurface. To account for latent heat effects associated with freezing and thawing of pore water, Equation (2.2) is often reformulated in terms of volumetric enthalpy (e.g., [1, 27]), which incorporates both sensible and latent heat contributions.

One-Dimensional Heat Conduction in Permafrost

For GST reconstructions based on boreholes, heat transfer is commonly approximated as one-dimensional and purely vertical. This assumption is justified for long-term reconstructions, where lateral temperature gradients are expected to be small compared to the dominant surface–depth gradient. Under this assumption, Fourier's law reduces to the one-dimensional form [23]:

$$q = -k \frac{dT}{dz} \tag{2.3}$$

Here $\frac{dT}{dz}$ is the temperature gradient in the z-direction, and it describes how temperature changes with depth.
The thermal conductivity of frozen soil depends strongly on its frozen or unfrozen state. Ice increases thermal conductivity, whereas liquid water reduces it, leading to a strong contrast between frozen and thawed conditions. This contrast plays an important role in how temperature signals propagate downward in permafrost. This assumption underlies the forward model used throughout this thesis.

2.2.2 Thermal Diffusivity

Another key factor in the reconstruction of GST from permafrost is the thermal diffusivity α. Thermal diffusivity describes the speed at which a material responds to changes in temperature.

It is defined as the ratio of thermal conductivity, which is explained in the previous chapter, to volumetric heat capacity [23]:

$$\alpha = \frac{k}{c} = \frac{k}{c_p \rho} \tag{2.4}$$

where:

- k is the thermal conductivity ($\mathrm{W\,m^{-1}\,K^{-1}}$),
- ρ is the density of the material ($\mathrm{kg\,m^{-3}}$),
- c_p is the specific heat capacity ($\mathrm{J\,kg^{-1}\,K^{-1}}$).

A material with high thermal diffusivity allows temperature changes to spread quickly, while a material with low thermal diffusivity slows down temperature changes.
In frozen soils, thermal diffusivity depends on temperature, moisture, and phase changes. Ice has a higher diffusivity ($\alpha = 1.12\mathrm{mm^2\,s^{-1}}$) than water ($\alpha = 0.14\mathrm{mm^2\,s^{-1}}$), so frozen soil reacts faster to temperature changes than thawed soil [23]. However, when ice melts or water freezes, latent heat effects can slow down heat transfer, causing a delay in temperature changes.
These properties influence how quickly surface temperature signals penetrate permafrost, which is critical for interpreting borehole data in GST models.

2.2.3 Latent Heat in Frozen Soil

Latent heat plays a significant role in the thermal response of frozen ground. During freezing and thawing, energy is absorbed or released without a change in temperature, which modifies how heat propagates through permafrost. In frozen soils, the total amount of energy in the phase change process depends on the total amount of water in the given soil volume and the fraction of this water that changes phase. In the following, we can see the mathematical expression for the latent heat of fusion in frozen soil [23]:

$$L = \rho_d L'(w - w_u) \tag{2.5}$$

In equation 2.5, L denotes the soil volumetric latent heat of fusion with units of $\mathrm{kJ\,m^{-3}}$, $L' = 333.7\,\mathrm{kJ\,kg^{-1}}$ is the specific latent heat of fusion of water, ρ_d with units of $\mathrm{kg\,m^{-3}}$ is the dry soil density, w is the total water content, and w_u is the unfrozen water content of the frozen soil.
The release or absorption of latent heat leads to a delay in temperature propagation within the soil [28]. This effect must be considered in ground surface temperature (GST) reconstructions, as it can cause deviations between actual and modeled temperature profiles.

2.2.4 Two-Phase Heat Transport and Latent Heat in Permafrost

This study uses the two-phase heat transport model, which was also employed by Groenke et al. [1], to reconstruct GST from permafrost borehole data.

In this approach, the temperature field $T(z,t)$ is represented as a deviation from a quasi-linear steady-state solution, which separates the long-term geothermal background from transient climate signals:

$$T(z,t) = T_0 + \frac{Q_{\text{geo}}}{k(z)} z + \Delta T(z,t) \tag{2.6}$$

Here, T_0 is a reference temperature at the surface, Q_{geo} is the geothermal heat flux, $k(z)$ is the depth-dependent thermal conductivity, and $\Delta T(z,t)$ captures the time-dependent deviations due to past climate variations.
Furthermore, to model the latent heat, we use the enthalpy method. It combines sensible heat (from temperature) and latent heat of fusion (from phase change) in one formula:

$$\frac{\partial}{\partial z}\left(k(T)\frac{\partial T}{\partial z}\right) + \frac{\partial h(T)}{\partial t} = S(z,t) \tag{2.7}$$

Here, $h(T)$ is the enthalpy, and $S(z,t)$ is a possible heat source or sink. The thermal conductivity $k(T)$ depends on temperature and soil type.
The enthalpy function combines sensible and latent heat:

$$h(T) = c(T)(T - T_{\text{ref}}) + L\theta_w(T)$$

where $c(T)$ is the volumetric heat capacity, $T_{\text{ref}} = 0^\circ C$ the freezing point, L the latent heat of fusion, and $\theta_w(T)$ the temperature-dependent unfrozen water content. This model does not include groundwater flow or lateral heat transport. It only models vertical heat conduction. Boreholes with strong groundwater effects should not be used with this method [1].
We calculate the thermal conductivity $k(T)$ and heat capacity $c(T)$ of the soil as a mixture of four components: unfrozen water θ_w, ice θ_i, organic matter θ_o, and mineral particles θ_m. Their sum must be equal to 1:

$$\theta_w + \theta_i + \theta_o + \theta_m = 1 \tag{2.8}$$

The total water content is defined as $\theta_{\text{tot}} = \theta_w + \theta_i$. The organic fraction ω defines how much of the dry part of the soil is organic. Based on this, we use:

- $\theta_i = \theta_{\text{tot}} - \theta_w$ (volumetric ice content)
- $\theta_o = (1 - \theta_{\text{tot}}) \cdot \omega$ (organic content)
- $\theta_m = (1 - \theta_{\text{tot}}) \cdot (1 - \omega)$ (mineral content)

We assume saturated conditions, so the total water content is equal to the porosity of the soil. This way, the thermal properties can be calculated based on soil composition and temperature. The final representation of the GST history combines both seasonal variations and long-term changes. It consists of a sinusoidal annual cycle and a piecewise linear trend function that models past temperature deviations. The complete formula is given by:

$$T(0,t) = T_0 + A \cdot \sin\left(\frac{2\pi}{P}(t - t_0)\right) + \sum_{i=1}^{N} \psi_i(t) \cdot \left[\tau_i + \frac{\tau_{i+1} - \tau_i}{t_{i+1} - t_i}(t - t_i)\right] \tag{2.9}$$

In this equation, T_0 is the initial surface temperature at the start of the simulation period. The term $A \cdot \sin\left(\frac{2\pi}{P}(t - t_0)\right)$ represents the annual temperature cycle, where A is the amplitude and

P is the period, typically set to one year. The second term models the long-term temperature trend using a piecewise linear function. Here, τ_i are the surface temperature values defined at specific time points t_i, and the function $\psi_i(t)$ is equal to 1 within the interval $[t_i, t_{i+1})$, and 0 elsewhere. If the annual cycle is not considered, the amplitude A is set to zero
To account for the limited temporal resolution of borehole data caused by diffusive heat transport, the time intervals $[t_i, t_{i+1})$ are chosen such that each interval spans a time range of approximately $\pm\frac{1}{3}t'$, where t' is the time before the borehole measurement, following the approach by [22].

2.3 Inversion Techniques for Reconstructing Temperature Profiles

The reconstruction of past GST from borehole temperature measurements is a classic inverse problem in geophysics. In particular, instrumental temperature records have not been available for more than the past 150 years, making the reconstruction of past ground surface temperature crucial [7]. Furthermore, reconstructing past ground surface temperatures provides important information on long-term climate trends and can help assess whether recent warming is primarily driven by industrialization or influenced by natural climate variability.
Since this study employs an inverse method, it is necessary to first clarify what an inverse method is. In general, an inverse problem can be described mathematically as follows:

$$y = G(u) + \eta$$

where:

- G: denotes the forward model
- η is the observational noise
- y is the observed data

The goal of the inverse problem is to recover the unknown u from observation y. However, such problems are often ill-posed, meaning that small variations in y can lead to large uncertainties in u, requiring additional constraints or regularization techniques to obtain stable solutions [29–31].
Borehole temperature profiles are valuable records of past climate changes because thermal diffusion in the subsurface naturally smooths temperature variations over time. This property allows for the reconstruction of past surface temperatures, where deeper borehole measurements correspond to older climatic conditions. In the classical sense of inverse problems, a well-posed problem requires three conditions [32]:

1. Existence: A solution exists for all given data.
2. Uniqueness: The solution is unique.
3. Stability: Small errors in the data lead only to small changes in the solution.

If any of these conditions fail, the problem is then ill-posed. Because thermal diffusion acts as a low-pass filter, high-frequency variations in past temperatures are smoothed out, making direct inversion highly sensitive to noise and missing data, which is common in borehole measurements.
To overcome these challenges, various regularisation-based inversion methods have been developed.

2.3.1 Overview of GST Inversion Methods

In classical approaches, to overcome challenges from inverse problems is achieved through explicit regularisation methods such as Tikhonov regularisation. Within a Bayesian inference framework, regularisation is introduced implicitly through prior assumptions, and the solution is described probabilistically. In this study, inference is performed using Ensemble Kalman Sampling (EKS) as a computational method to approximate the Bayesian posterior distribution.
Classical GST reconstruction approaches often rely on linear regularization techniques, such as Tikhonov regularization, to stabilize the inversion of ill-posed borehole temperature problems. These methods enforce smoothness constraints on the reconstructed temperature history and are well suited for recovering long-term, low-frequency climate signals from borehole data.
However, linear regularization methods typically yield a single deterministic solution and are sensitive to measurement noise as well as to inaccuracies in the assumed forward model. In particular, model discrepancies arising from simplified physical assumptions in practical implementations, such as neglecting latent heat effects or active layer dynamics, are not explicitly represented within the inversion framework. As a result, uncertainties related to data noise, modeling errors, and non-uniqueness of the solution cannot be quantified in a systematic way, which may lead to overly confident reconstructions [33]. These limitations motivate the use of probabilistic inversion approaches, which are discussed as follows.
Bayesian inversion provides a probabilistic way to reconstruct past ground surface temperatures (GST) from borehole measurements. Unlike classical methods that produce only one "best" solution, Bayesian methods describe a distribution of possible GST histories. They combine a prior (knowledge before seeing the data), a likelihood (how well the model fits the data), and a posterior (updated knowledge after including the data). A widely used Bayesian approach is Markov Chain Monte Carlo (MCMC). MCMC generates samples from the posterior distribution by constructing a Markov chain that explores the space of possible GST histories. Some studies have used advanced versions such as reversible-jump Markov Chain Monte Carlo (rjMCMC), which can automatically adjust the complexity of the GST model during sampling [18].
Other Bayesian approaches include empirical Bayes and hierarchical Bayes methods. These techniques estimate model parameters—such as noise levels or smoothing strength—directly from the data. For example, Woodbury and Ferguson used information-based criteria (ABIC) to select prior settings in a Bayesian framework for GST reconstruction [34]. Overall, Bayesian inversion methods provide a powerful framework for uncertainty quantification, but classical Bayesian inference algorithms such as Markov Chain Monte Carlo (MCMC) can become computationally expensive when the forward model is costly to evaluate, since each step typically requires solving the heat conduction model [35].

2.3.2 Ensemble Kalman Sampling

To address the high computational cost of classical Bayesian inference methods, this thesis employs Ensemble Kalman Sampling (EKS) as the primary inversion technique for reconstructing past GST from borehole temperature profiles [1].
Ensemble Kalman Sampling is an ensemble-based Bayesian inference method that builds on Bayes' theorem and extends the classical Kalman filter, originally developed for linear state

estimation. A key development was the Ensemble Kalman Filter (EnKF), which introduced Monte Carlo techniques to handle nonlinear and high-dimensional systems. EKS further extends the EnKF by incorporating stochastic Langevin dynamics, enabling the sampling of the full posterior distribution rather than producing a single state estimate.
In the following sections, the theoretical foundations of EKS are outlined to provide a clear understanding of the method and its application to GST reconstruction [36].

Bayes' Theorem

Bayes' theorem forms the mathematical foundation for probabilistic state estimation and inverse problems, including the reconstruction of GST from borehole data. In this work, the notation $p(\cdot)$ denotes probability density functions. The prior $p(X)$ describes the uncertainty about the unknown parameters before including data, the likelihood $p(Y \mid X)$ expresses how probable the observed data Y are under a given parameter set X, and the posterior $p(X \mid Y)$ combines prior knowledge with observational information.

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}, \tag{2.10}$$

$$p(Y) = \int_{x \in \mathcal{X}} p(Y \mid X = x)\, p(X = x)\, \mathrm{d}x, \tag{2.11}$$

The term $p(Y)$ is the marginal likelihood (or evidence) and ensures that the posterior density integrates to one. In the context of Bayesian inversion, X represents the unknown parameters (e.g., the ground surface temperature history), and Y represents the evidence (borehole temperature measurements) [37]. This principle underpins recursive methods like the Bayes filter and its extended versions, such as the Kalman filter and Ensemble Kalman Sampler (EKS), by combining prior knowledge with observational data [38].

Kalman Filter

The Kalman filter is the classical solution for linear-Gaussian state estimation, based on the bayes filter, and can be viewed as a recursive application of Bayes' theorem. It provides an optimal estimate of a hidden state when the system dynamics and the observation model are both linear and the uncertainties are Gaussian [39]. In its prediction step, the state is propagated through a linear model

$$\bar{x}_t = A_t x_{t-1} + B_t u_t, \qquad \bar{P}_t = A_t P_{t-1} A_t^\top + Q_t,$$

where Q_t is the process noise covariance. The correction step incorporates a new measurement

$$z_t = C_t x_t + \delta_t, \qquad \delta_t \sim \mathcal{N}(0, R),$$

using the Kalman gain

$$K_t = \bar{P}_t C_t^\top (C_t \bar{P}_t C_t^\top + R)^{-1}.$$

The updated state estimate becomes

$$x_t = \bar{x}_t + K_t(y_t - C_t \bar{x}_t), \qquad P_t = (I - K_t C_t)\bar{P}_t.$$

Although Kalman Filter (KF) itself is not used directly in this study, it forms an important conceptual foundation for ensemble-based methods. The Ensemble Kalman Filter (EnKF) generalizes the KF to nonlinear and high-dimensional systems by replacing covariance matrices with ensemble statistics. The Ensemble Kalman Sampling (EKS) method applied in this work builds on this extension and adapts it to inverse problems.

Ensemble Kalman Filter

While the Kalman filter works really well for linear systems with Gaussian noise, it tends to struggle when models become nonlinear or when the state space is very large, which leads to huge computational demands [40][41]. In the case of reconstructing ground surface temperatures, the underlying processes, like heat conduction in permafrost, are often nonlinear and affected by many uncertain parameters. This makes it difficult to apply the classical Kalman filter directly.
To avoid these issues, the Ensemble Kalman Filter was developed, which is also the foundation for the Ensemble Kalman Sampling approach used in this study. Instead of representing uncertainty through a big covariance matrix, it uses an ensemble of model states, which means you run several model simulations and use their spread to represent the uncertainty. This makes it more flexible, especially when dealing with non-linear models or complex systems where computing analytical derivatives is not really feasible.
Furthermore, it should be mentioned that the EnKF is an approximation of the standard Kalman Filter based on Monte Carlo sampling [40]. As the number of ensemble members N increases, the sampling error decreases proportionally to $\frac{1}{\sqrt{N}}$. In the limit $N \to \infty$, the Ensemble Kalman Filter converges to the exact Kalman filter solution [40].
The main idea behind the EnKF is to propagate each of these ensembles individually during the analysis step instead of the n-dimensional estimate $x_{k|k}$ and the $n \times n$ covariance $P_{k|k}$ of the KF [41].

$$\overline{x}_{t|t} = \frac{1}{N} \sum_{j=1}^{N} x_t^{(j)} \tag{2.12}$$

The equation 2.12 approximates the true posterior estimate $\hat{x}_{k|k}$ in the limit $N \to \infty$.

$$\overline{P}_{t|t} = \frac{1}{N-1} \sum_{j=1}^{N} \left(x_t^{(j)} - \overline{x}_{t|t}\right) \left(x_t^{(j)} - \overline{x}_{t|t}\right)^T \approx P_{t|t} \tag{2.13}$$

These sample-based approximations reflect the mean ((2.12)) and uncertainty (covariance, (2.13)) of the updated ensemble after assimilating new observations at time step k. As the ensemble size N increases, the approximations converge to the exact Bayesian posterior mean and covariance [40].

Ensemble Kalman Sampling

Ensemble Kalman Filter, which was introduced in Chapter 2.3.2, is primarily designed for sequential state estimation in dynamical systems, it has also been adapted to solve inverse problems. There, the goal is not to follow how a system makes progress over time, but instead

to recover unknown, time-independent quantities—like physical parameters or boundary conditions with using of available data.
In the context of reconstructing past GST from borehole temperature profiles, such inverse approaches are very important. The reconstruction process involves estimating past surface conditions based on the temperature profile of a permafrost borehole, which requires dealing with uncertain and incomplete information, because of sensor uncertainties and more. For this purpose, Ensemble Kalman Inversion (EKI) derived from the EnKF framework [30], has proven particularly useful.
A disadvantage of methods, such as EKI, is the aim to find a single "best" solution, which can be problematic in real-world inverse problems. The reason for the problems are the simplified physical models (e.g., the one-dimensional heat conduction equation) and the observational data (e.g., borehole temperatures) are both approximations of reality. Model simplifications, unknown boundary conditions, and measurement noise can lead to multiple temperature histories that all fit the observed data equally well (ill-posed problem).
Therefore, it is not sufficient to focus only on one "optimal" solution, as this is not possible due to the existing uncertainties, model inaccuracies, and the inherently ill-posed nature of the problem as mentioned before. Instead, it is crucial to use a probabilistic approach that captures the full range of plausible solutions and quantifies the associated uncertainty. Especially in the context of reconstructing historical GST, it is crucial to not only identify a plausible trajectory but also to understand how uncertain it is and whether alternative explanations might exist.
The EKS approach builds on the connection between Langevin dynamics, Bayesian inference and the EKI. In the Bayesian framework, we define the posterior distribution over the parameter space as [36]

$$\pi(u) \propto \exp(-\Phi(u))\pi_0(u), \tag{2.14}$$

where $\pi_0(u)$ is the prior, $\exp(-\Phi(u))$ the likelihood and $\Phi(u)$ denotes the data misfit functional (also called negative log-likelihood), typically defined as

$$\Phi(u) = \frac{1}{2}\|y - G(u)\|_\Gamma^2, \tag{2.15}$$

with $G(u)$ being the forward model (e.g. heat transport in permafrost boreholes) and Γ the covariance matrix of the observational noise.
To sample from this posterior distribution, one can theoretically use the overdamped Langevin dynamics [36] [42]:

$$\dot{u} = -\nabla\Phi_R(u) + \sqrt{2}\,\dot{W}, \tag{2.16}$$

where $\dot{W}$ is the standard Brownian motion and $\Phi_R(u) = \Phi(u) + \frac{1}{2}\|u\|_{\Gamma_0}^2$ includes a Tikhonov-type regularization from the Gaussian prior and is the regularized negative log-posterior [36]. However, computing the gradient $\nabla\Phi(u)$ typically requires evaluating the derivative of the forward model $G(u)$, which is often unavailable or computationally expensive.
To address this, the EKI method—originally introduced as a derivative-free optimization algorithm for inverse problems [30]—uses ensemble statistics to approximate the gradient direction. More precisely, the directional derivative $\nabla\Phi(u)$ is replaced by a regression estimate

using empirical covariances. In continuous time, this leads to the following deterministic update equation for each ensemble member:

$$\dot{u}^{(j)} = \frac{1}{J}\sum_{k=1}^{J}\langle G(u^{(k)}) - \bar{G},\ y - G(u^{(j)})\rangle_\Gamma \cdot (u^{(k)}), \tag{2.17}$$

where $\bar{G}$ and $\bar{u}$ denote the ensemble means of model outputs and parameters, respectively. This formulation avoids the explicit computation of the Jacobian $G'(u)$ and is mathematically equivalent to a preconditioned gradient flow using ensemble-based approximations.
Based on this idea, the Ensemble Kalman Sampler (EKS) introduces an additional stochastic noise term to transform the optimization dynamics into a sampling procedure. The resulting interacting particle system takes the form [36]:

$$\dot{u}^{(j)} = -C(U)\nabla\Phi_R(u^{(j)}) + \sqrt{2C(U)}\,\dot{W}^{(j)}, \tag{2.18}$$

where $C(U)$ is the empirical ensemble covariance matrix and $\Phi_R(u)$ denotes the regularized negative log-posterior. Although a gradient symbol appears, the actual implementation of EKS does not require the derivative of $G(u)$; instead, $\nabla\Phi_R(u)$ is again approximated using ensemble covariances in analogy to the EKI formulation. This makes the method fully derivative-free with respect to the forward model, while preserving the structure of a preconditioned Langevin diffusion. The key idea of EKS is to replace exact gradients with ensemble differences, yielding the derivative-free update [36]:

$$\dot{u}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\langle G(u^{(k)}) - \bar{G}, G(u^{(j)}) - y\rangle_\Gamma\,(u^{(k)} - \bar{u}) - C(U)\Gamma_0^{-1}u^{(j)} + \sqrt{2C(U)}\,\dot{W}^{(j)}, \tag{2.19}$$

where $C(U)$ is the empirical covariance matrix of the ensemble, and $\bar{G}$, $\bar{u}$ denote ensemble means.
This formulation ensures that, in contrast to EKI, the system does not collapse to a single point but instead asymptotically approaches the posterior distribution. The added noise term is crucial for enabling exploration of the posterior with uncertainty quantification.
The EKS approach directly addresses this issue by producing not just one solution, but an ensemble of solutions that collectively approximate the posterior distribution. This allows for a probabilistic interpretation of the results – for example, in the form of confidence intervals, expected values, or variances. These uncertainty estimates provide a clear advantage over deterministic methods, particularly in scientific contexts where a transparent treatment of uncertainty is essential.

2.4 Sources of Uncertainty and Challenges

Reconstructing past ground surface temperatures (GST) from borehole temperature measurements involves several sources of uncertainty. These uncertainties arise from physical limitations of heat transfer, measurement errors, environmental effects, and simplifying model assumptions. Together, they affect the accuracy and timing of reconstructed GST signals. A fundamental limitation comes from heat diffusion in the ground. Heat conduction acts as a low-pass filter, meaning that short-term surface temperature variations are smoothed out with depth. As a result, borehole temperature profiles mainly preserve long-term temperature trends, while short or abrupt surface temperature changes are strongly reduced or completely lost. This makes the inverse problem ill-posed, because different GST histories can lead to very similar temperature profiles in the borehole. Long-term temperature changes can therefore be reconstructed more reliably than short-term events [22, 37]. GST reconstructions also rely on simplified physical models. Many methods assume one-dimensional vertical heat conduction and ignore lateral heat flow or subsurface heterogeneity. In addition, thermal properties such as thermal conductivity and heat capacity are often assumed to be constant with depth and time, although they can change due to variations in soil type, ice content, and moisture. These assumptions introduce structural uncertainties into the reconstruction. Measurement-related uncertainties further affect borehole temperature data. Sensors can drift over time due to moisture, mechanical stress, or aging, leading to artificial warming or cooling trends that are difficult to separate from real ground temperature changes [43]. Phase changes of soil water can also reduce temperature variations because latent heat absorbs or releases energy. Furthermore, drilling the borehole introduces temporary heating. Long recovery times are required before measurements reflect undisturbed ground conditions, especially in deep boreholes [43]. Finally, long-term monitoring is limited by funding constraints. Reliable GST reconstruction requires continuous temperature records over many years, but long-term measurements are difficult to maintain due to limited financial support. As a result, many borehole datasets contain interruptions or inconsistent measurement periods [44]. In summary, GST reconstructions from borehole data are affected by fundamental physical limits and measurement-related uncertainties. These limitations directly influence the stability of the inverse problem and motivate the methodological choices made in this work, particularly the careful treatment of data uncertainty and model assumptions.

3 Datasets and Preprocessing

3.1 Description of Permafrost Borehole Sites in the Lena Delta

Figure 3.1: Location of the Samoylov Borehole, Sardakh Borehole, and Tiksi Borehole in the Lena River Delta, Northeast Siberia (Russia). The Samoylov location represents a single borehole used in this study. Base map: Google Earth satellite imagery [45].Scale bar shown at the bottom of the figure (30 km).

This master's thesis uses data from three different permafrost boreholes in the Russian Arctic: Sardakh, Samoylov, and Tiksi. Figure 3.1 provides an overview of the Lena River Delta, showing the locations of all three borehole sites. This overview helps to contextualize the individual

study sites described below. These three sites were selected because their borehole characteristics and data properties provide a valuable basis for addressing the research questions of this thesis. Differences in borehole depth, temporal coverage, and data quality allow the evaluation of reconstruction methods under varying conditions. In the following sections, each site is described in more detail, along with its specific contribution to the research questions of this thesis.

3.1.1 Sardakh

Sardakh Island is located in the southern Lena River Delta (72.57155°N, 127.24157°E). It features a 100 m deep borehole with temperature data from 2009 to 2021. The site lies in a continuous permafrost zone and is operated by the Alfred Wegener Institute in collaboration with Melnikov Permafrost Institute (PIY) Yakutsk. The subsurface consists of ice-rich loamy sands overlying older sedimentary layers, including pebbles and fossil material.
Climatic conditions are typical of Arctic continental zones with long, cold winters and short summers. The mean annual air temperature is around $-12.5\,^{\circ}\mathrm{C}$, with mean annual precipitation $190\,\mathrm{mm\,yr^{-1}}$[46, 47]. Snowstorms are frequent, and snow cover plays an important role in ground insulation. The thickness of the active layer of the continuous permafrost area in the coastal region of Laptev Sea, to which Sardakh Island belongs, is approximately 30 cm to 40 cm [22].
Sardakh's borehole data are generally well suited for GST reconstructions due to predominantly conductive heat transfer and stable subsurface temperature conditions. However, the site is not completely undisturbed, as thermokarst development occurred following drilling activities. The site shows long-term warming trends that are consistent with regional Arctic climate data and allows for testing inversion methods over a relatively deep profile with 13 years of reliable records.
Additionally, the 100 m borehole at Sardakh Island is a good dataset to check if using multiple, shallower temperature profiles can create similar GST reconstructions as the full-depth data. By cutting the data at different depths and using the multiple-likelihood EKS method, it is possible to study how sensitive the GST reconstructions are to borehole depth and the choice of profiles.

3.1.2 Samoylov

Samoylov is also located in the Lena River Delta. The borehole at Samoylov (72.36945°N, 126.47643°E) provides temperature data from 2006 to 2021 and has a depth of 26 m.
Climatically, Samoylov is characterized by continuous permafrost, low annual precipitation ($\approx$ 125 mm), and strong seasonal contrasts. Average air temperatures range from $-31\,^{\circ}\mathrm{C}$ in winter to $10\,^{\circ}\mathrm{C}$ in summer, with an annual mean of about $-12.5\,^{\circ}\mathrm{C}$. Field measurements conducted in spring 2008 indicate a short snow-free period, typically from June to September, with strong spatial variability in snow depth due to wind redistribution—reaching about 46 cm in polygon centers and approximately 17 cm on [47]. These microtopographic effects are crucial for understanding snow insulation and permafrost dynamics [47].
Geologically, the island lies on the first river terrace, dominated by polygonal tundra, thermokarst lakes, and wetlands. The subsurface consists of ice-rich silty soils, which influence the thermal

properties relevant for GST reconstruction. The active layer is shallow (≈ 50 cm in August), with high spatial variability due to differences in soil moisture and vegetation.
The vegetation cover is typical of Arctic tundra: mosses (e.g., Hylocomium splendens), sedges (Carex chordorrhiza), and dwarf shrubs (Salix spp.) dominate the dry and wet polygons. These vegetation types impact both soil moisture retention and snow accumulation, and thus affect heat transfer into the ground.
Overall, Samoylov provides a well-documented and diverse setting to investigate how snow cover, surface vegetation, and soil properties interact with permafrost and GST inversion techniques [48].
Furthermore, the Samoylov dataset is especially valuable because it has almost continuous, gap-free temperature records between 2010 and 2020. This makes it ideal for testing different combinations of multiple temperature profiles, as the clean data allow a controlled analysis of how the selection and timing of profiles influence the quality of the GST reconstructions.

3.1.3 Tiksi

Tiksi, on the Laptev Sea coast (71.58233°N, 128.91141°E), includes one borehole with a long time interval, from 1992 to 2022, and provides therefore data for 30 years with a depth of 30 m. Unlike the other boreholes, which are located in sedimentary terraces, the Tiksi borehole is situated in bedrock, providing a distinct geological and thermal context for GST reconstruction.
The Tiksi borehole and meteorological data used in this study are provided by the Hydrometeorological Observatory of Tiksi (National Oceanic And Atmospheric Administration (NOAA), Roshydromet, NSF, FMI collaboration) [49]. Located at 71.6°N, 128.9°E, Tiksi offers continuous atmospheric and ground measurements, including air temperature, snow depth, and permafrost conditions, supporting long-term climate monitoring efforts [49].
The borehole is located on a stone ridge in the foothills of the Primorsky Range, approximately 5 km from the town of Tiksi. The surface consists of platy fragments of argillite, and the subsurface is composed of fractured bedrock with limited overlying sediment (1–2 m). This bedrock setting leads to different thermal dynamics compared to soil-dominated sites and provides insight into how geological background influences GST signals [50].
Climatically, Tiksi lies in an excessively wet subarctic zone with cold winters, cool summers. It had an annual precipitation 172 mm to 349 mm in 2010-2021 [50]. The Mean Annual Air Temperature (MAAT) ranged from −3.6 °C in the 1990s to −0.5 °C in recent years, showing a clear warming trend of 0.11 °C/yr from 1991 to 2021. The Mean Annual Ground Temperature (MAGT) has increased by 1.0 °C to 2.4 °C across depths of 3 m to 30 m during the 30-year monitoring period.
Snow cover is highly variable and often thin, especially on the exposed ridge where snow is blown away by strong winds (up to 20 m s^{-1}). Winter snow depths at the borehole site fluctuate between 2 cm to 20 cm, limiting its insulating effect. It an active layer thickness (ALT) of ≈30 cm to 50 cm depending on year and surface type [50].
Vegetation is sparse and consists mainly of clastic mountain tundra, with patches of moss and grasses (e.g., Poa alpigena, Gastrolychnis apetala). The lack of dense vegetation and organic layers, combined with the thin snowpack, makes the Tiksi site particularly responsive to air temperature changes—ideal for studying long-term permafrost warming trends.
In summary, Tiksi provides a unique coastal permafrost environment with long-term data in a

rocky setting, low snow insulation, and clear signs of climate-driven warming. Its contrast to the Lena Delta sites helps evaluate how geology and microclimate shape the GST signal and its reconstruction under different model assumptions.

3.1.4 Comparison of Borehole Characteristics

To summarize the borehole configurations used in this study, Table 3.1 provides an overview of the key parameters for all three locations. These include geographical coordinates, borehole depth, observation period, and geological context.

Site	Coordinates	Depth (m)	Time Span	Notes
Sardakh	72.5716°N, 127.2416°E	100	2009–2021	
Samoylov 1	72.3695°N, 126.4764°E	26	2006–2021	
Tiksi	71.5823°N, 128.9114°E	30	1992–2022	Coastal bedrock site

Table 3.1: Overview of borehole characteristics used for GST reconstruction

3.2 Structure of the Borehole Temperature Datasets

This section outlines the origin, structure, and processing of the borehole temperature datasets used in this study. All datasets originate from long-term permafrost monitoring programs in the Lena River Delta region and include data from Samoylov, Sardakh, and Tiksi. The raw temperature measurements were provided by the Alfred Wegener Institute (AWI) for Samoylov and Sardakh, and by the Global Terrestrial Network for Permafrost (GTN-P) archive for Tiksi.
Each borehole contains a thermistor chain permanently installed into the permafrost, providing vertically resolved temperature measurements across multiple years. A temperature profile $T(t)$ consists of a sequence of N temperature values along a vertical axis perpendicular to the Earth's surface:

$$T(t) = [T(z_1, t), T(z_2, t), \dots, T(z_N, t)] \tag{3.1}$$

where $z_1, z_2, \dots, z_N$ denotes the depths of the individual sensors. These measurements are typically acquired either as point measurements after equilibration or as temporal averages over regular (sub-daily) intervals across multiple years.
The Samoylov borehole (SaHole2006) includes 24 temperature sensors down to 26.75 m, embedded in a Polyvinyl chloride (PVC) tube and backfilled with dry sand.Sensor depths range from 0 m to 26.75 m in irregular intervals (e.g., 0.75, 1.75, 2.75, ..., 26.75). The borehole is still active and part of ongoing AWI/PIY monitoring efforts.
The Sardakh borehole (SdHole2009) has a total depth of 100 m and is equipped with a Geokon thermistor chain comprising 24 sensors installed at predefined depths between 0.0 m and 100 m. The chain is housed within a metal core barrel and connected to a data logger in a protective wooden structure at the surface. Similar to Samoylov, the installation supports continuous logging and is designed for long-term permafrost monitoring.
Compared to Sardakh and Samoylov, only limited metadata are available for the Tiksi borehole.

Table 3.2: Overview of the borehole datasets used in this study. z_{max} denotes the maximum sensor depth of the borehole, Period indicates the range of years with available temperature measurements, N_{obs} is the number of annual mean temperature profiles used in the analysis, Raw res. refers to the original temporal resolution of the measurements in hours, and Provider indicates the data source.

Site	**Borehole** (identifier)	z_{max} [m]	**Period** [yyyy]	N_{obs} [–]	**Raw res.** [h]	**Provider**
Samoylov	SaHole2006	26.8	2006–2024	19	1	AWI / PIY Yakutsk
Sardakh	SdHole2009	100	2009–2024	13	6	AWI / PIY Yakutsk
Tiksi	TIK_BH_30	30	1992–2022	31	24	GTN–P

The site (TIK_BH_30) reaches a depth of 30 m and provides historical temperature records from 1992 to 2022, archived in the GTN–P database. It is part of a Russian national permafrost monitoring initiative.

To reduce seasonal variability and improve signal stability for inversion analysis, the raw data are aggregated into annual mean profiles. These smoothed profiles serve as input for the EKS method [1].

Table 3.2 summarizes the structural dataset parameters relevant for processing and inversion.

The raw datasets for Samoylov and Sardakh are provided as `.dat` files in wide-table format. Each file contains a header row specifying temperature measurements at various depths (e.g., Ts_0, Ts_40, ..., Ts_10000), and subsequent rows with timestamps and corresponding temperature values in °C. Measurements for SaHole2006 are recorded every 6 hours (00:00, 06:00, 12:00, 18:00 UTC) and for SdHole2009 are recorded every hour. Missing values are denoted by NA (missing value indicator) (NA).

The Tiksi dataset is stored in `.csv` format and structured slightly different: each row contains a year, a specific depth (in meters), and the corresponding annual mean temperature.

This long-table format is optimized for compact representation of time–depth profiles.

The borehole temperature datasets show varying degrees of temporal completeness. For the Samoylov and Sardakh sites, the raw measurements include daily and sub-daily records, with occasional missing values and data gaps.

For example, the SaHole2006 dataset lacks most of the winter period for 2008–2009 and the beginning of 2006. The SdHole2009 dataset also contains several extended gaps, including long interruptions in early 2009 and late 2021, which affect the seasonal coverage of some years.

The Tiksi1992 dataset shows even larger gaps in temporal coverage. Around 90 consecutive days are missing in late 2017, the entire years 2018 and 2019 are absent, and approximately 270 days are missing between January and mid-September 2020. These gaps lead to strong differences in data availability and seasonal representation across the datasets.

These characteristics illustrate the heterogeneous temporal structure of the borehole temperature records and motivate the need for explicit data selection criteria, which are described in the methodology section 4.1.

3.3 Description of Additional Meteorological Data

This section describes supplementary meteorological data used in the analysis, such as snow depth, air temperature, and their sources.

3.3.1 ERA5 Reanalysis Data

ERA5 data are used as meteorological input for the construction of a data-driven ground surface temperature prior in combination with an n-factor approach. Both the n-factor concept and the prior formulation are described in detail in Section 4.3.
ERA5 is a global atmospheric reanalysis produced by European Centre for Medium-Range Weather Forecasts (ECMWF) that provides temporally continuous, gridded climate variables for the atmosphere, land surface and ocean waves. It combines observational data with a numerical weather prediction system with coupled land-surface and ocean-wave components through data assimilation to produce physically consistent estimates. The data are provided at an approximate spatial resolution of 31 km and an hourly temporal resolution, representing grid-cell-scale conditions rather than point measurements [51].

3.3.2 Ground Surface Temperature Observations at Samoylov (SaMet and SaSoil)

At the Samoylov site, GST data are available from two long-term monitoring stations: the Samoylov Meteorological Station (SaMet2002) and the Samoylov Soil Station (SaSoil2002). SaMet records meteorological variables and surface temperature, while SaSoil provides soil temperature and related near-surface measurements. Both stations have been operating since 2002 and offer continuous GST measurements [48]. For this study, the GST data from SaMet and SaSoil were processed separately into annual mean values. For each year, the two annual means were averaged to obtain a single GST estimate for the Samoylov site. This combined GST record is used in the evaluation of the reconstruction results.

3.3.3 Limitations and Uncertainties of the Datasets

Despite careful cleaning and filtering, several inherent limitations remain in the borehole temperature datasets, which are used for GST reconstruction. These arise primarily from temporal gaps in the records, potential long-term sensor degradation, heterogeneous sampling intervals, incomplete or inconsistent metadata, and differences in borehole depth and configuration. Each of these factors can introduce uncertainties that influence the interpretation of the reconstructed ground surface temperature histories.

- **Temporal gaps and incomplete records**: Some years contain long gaps in the data due to sensor failures or recording problems. For example, the SdHole2009 dataset is missing large parts of summer 2012, and SaHole2006 has almost no data for winter 2008-2009.

Smaller gaps also appear in several other years across all sites. These gaps mainly limit the choice of likelihood profiles in the multiple-likelihood EKS framework. In this work, years with missing summer or winter months were removed to avoid systematic errors. Missing summer months usually make annual means too low (artificially cold), while missing winter months usually make annual means too high (artificially warm). Both cases would lead to a wrong GST reconstruction.

- **Sensor degradation over time**: Failures of individual sensors, such as the breakdown of the 80 m sensor at Sardakh, reduce the usable depth range of the temperature profile. This is particularly critical for reconstructions targeting earlier periods, since information from deeper layers is required to constrain long-term trends. In practice, this limited the vertical resolution available for the Sardakh dataset in my analysis, and valuable information about past ground surface temperatures was effectively lost, reducing the robustness of long-term reconstructions.
- **Inhomogeneous measurement intervals**: The temporal resolution of the datasets varies substantially. For example, Sardakh records temperatures every 6 hours, while Tiksi provides only one value per day. In the latter case, individual sensor errors or short-term anomalies cannot be averaged out, which increases their influence on derived annual or seasonal means. Consequently, higher-resolution datasets generally yield more reliable input for reconstructions, whereas low-resolution records introduce greater uncertainty.
- **Uncertainty in metadata and stratigraphy**: For several boreholes, particularly Tiksi, the subsurface stratigraphy is not fully documented, which complicates the interpretation of thermal diffusion processes and could introduce modeling bias in the inversion.
- **Uncertainty in surface temperature priors**: The data-driven priors derived from European Centre for Medium-Range Weather Forecasts Reanalysis Atmospheric (ERA) air temperature via n-factors introduce additional uncertainty, since n-factors are empirical and can vary significantly depending on soil, vegetation, and snow conditions.

In summary, these limitations highlight the challenges of working with borehole temperature records. They were therefore explicitly taken into account in the design of the inversion strategy. In sensitivity analyses and additional tests, their effects, such as the impact of data gaps, sensor degradation, and resolution differences, became evident and directly restricted, for instance, the selection of usable temperature profiles for the reconstructions.

3.4 Synthetic CryoGrid Dataset

In addition to the observational borehole datasets, a synthetic dataset was used for controlled validation experiments. The synthetic soil temperature profiles were generated using the CryoGrid Community Model (version 1.0) [52], which provides a physics-based simulation framework for climate-driven permafrost dynamics.
The model configuration was based on the Samoylov Island study setup, including soil stratigraphy and thermal properties derived from field observations [53]. The model was initialized to

obtain thermally consistent subsurface conditions. Subsequently, transient simulations driven by historical ERA forcing were performed starting from 1944, yielding temperature–depth profiles with a fully known surface forcing history. These synthetic temperature profiles were subsequently used in the methodological validation experiments described in Section 4.2.4.

4 Methodology

4.1 Data Preprocessing and Quality Control

The preprocessing focused on ensuring annual temporal completeness, seasonal representativeness, and sensor reliability, which are essential prerequisites for the construction of physically meaningful likelihoods. To ensure a robust foundation for the model-based GST reconstruction, the borehole temperature data from the Samoylov, Sardakh, and Tiksi sites were carefully preprocessed. This step involved several important procedures to improve data quality and compatibility for the inversion model.

1. Outliers and invalid sensor readings were removed. In addition, missing or inconsistent entries were identified and filtered out to avoid biased likelihood estimation. For instance, the sensor at a depth of 80 m in the Sardakh dataset was excluded from the analysis for years in which no measurements were available.

2. The datasets were restructured and reformatted into a form suitable for the inversion process. This included generating daily or annual averages, aligning timestamps across years, and preparing the temperature profiles in the required array formats for likelihood computation

3. All time periods with insufficient data quality were identified and excluded. This applied especially to years with missing months or seasonal gaps (e.g., a missing winter). Only temperature profiles with complete coverage were considered valid inputs for the inversion, ensuring consistency in the multiple-likelihood formulation.
   For instance, A given year was excluded from further analysis if more than one third of the temperature observations were missing. Furthermore, years with incomplete seasonal coverage were rejected: if more than two months within a specific season (e.g., winter or summer) were missing, the corresponding annual profile was excluded to avoid seasonally biased likelihoods.

To study how the available depth affects reconstruction quality, the borehole profiles were also truncated to different depths—such as 30 m, 50 m, and 100 m. This allowed for a comparative analysis of reconstruction stability with increasing depth. Since the temperature signal changes systematically with depth, this aspect was also addressed during preprocessing.
Seasonal temperature fluctuations become weaker with increasing depth, until they almost disappear at a certain level. This level is known as the Zero Annual Amplitude (ZAA) depth and marks the boundary where seasonal surface variations no longer affect the subsurface temperature field. To demonstrate this behavior, Figures 4.1 and 4.2 show, for each borehole, the minimum, maximum, and mean daily average temperatures recorded throughout the year. These profiles reveal how the annual temperature range decreases with depth.

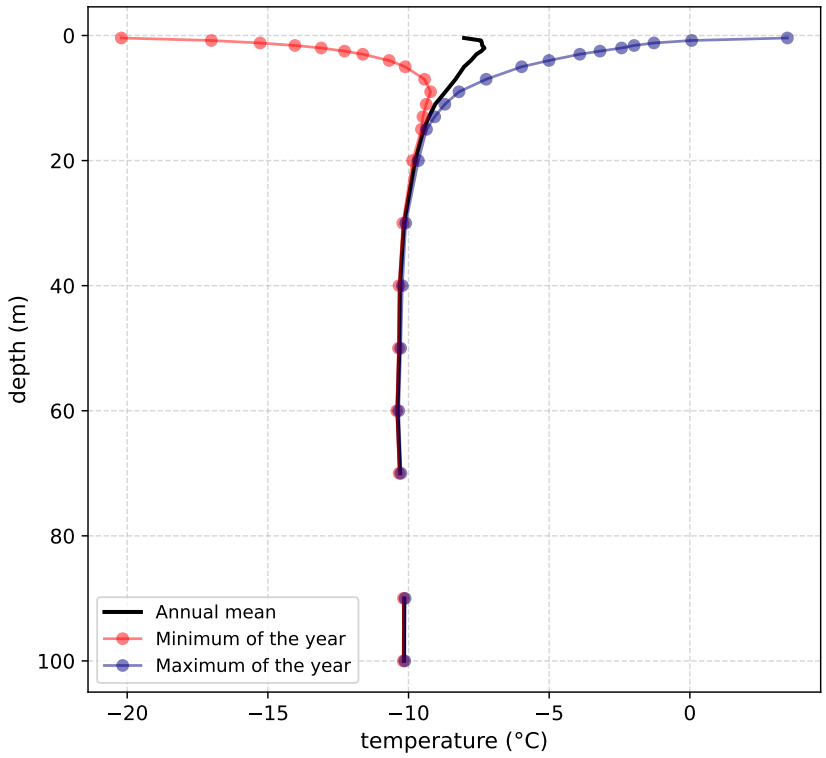


Figure 4.1: Annual ground temperature profile from the Sardakh borehole (SdHole2009) in the Lena River Delta region for the year 2017. The plot shows the annual mean (black), as well as the minimum (blue) and maximum (red) daily averages recorded throughout the year.

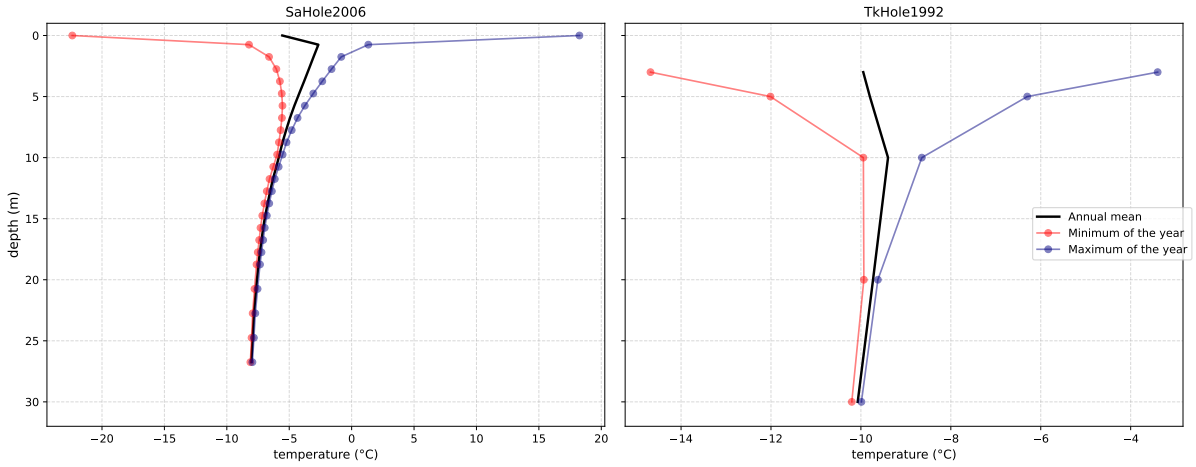


Figure 4.2: Annual ground temperature profiles from the Samoylov and Tiksi boreholes (SaHole2006 and TkHole1992) in the Lena River Delta region for the year 2017. Each plot includes the annual mean (black), as well as the minimum (blue) and maximum (red) daily averages recorded throughout the year.

As shown in these figures, the temperature difference between the maximum (red) and minimum (blue) curves becomes smaller with depth. Both curves converge at about 10-20 m in

Figure 4.1 and Figure 4.2, which indicates the ZAA depth where the annual signal disappears. Based on these observations, I removed the upper 10–15 m of each borehole temperature profile before computing the likelihoods as part of the preprocessing strategy.
This surface-level filtering step reduces the influence of near-surface disturbances such as snow cover, vegetation, and freeze–thaw effects. The goal of this filtering step is to focus the inversion on the more stable subsurface signal and to achieve higher reliability in the reconstructed ground surface temperature history.

4.1.1 Initialization of the Subsurface Temperature Profile

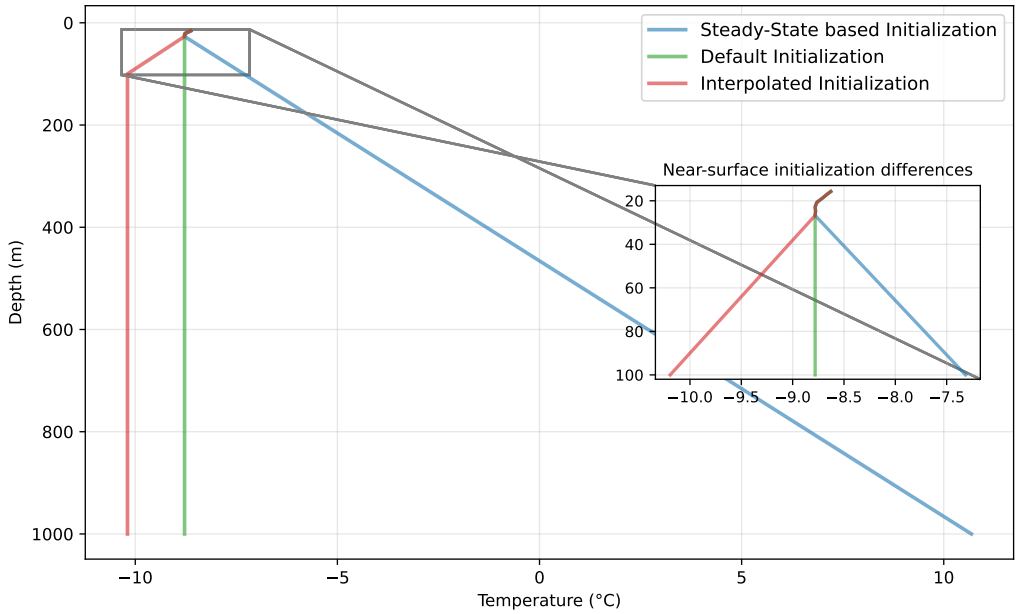


Figure 4.3: Comparison of the three initialization schemes used to extend the borehole temperature profile below the deepest observation depth. The measured temperature segment is continued using (i) a steady-state initialization assuming a constant geothermal gradient (blue), (ii) a default vertically constant continuation of the deepest measured temperature (green), and (iii) an interpolated continuation based on a deeper neighboring borehole (red). The inset highlights the depth interval near the deepest measurement, where the different initialization schemes connect to the observed profile.
Original Profile is based on the profile from the dataset SaHole2006

The initialization of the subsurface temperature profile defines the starting condition of the forward model and therefore directly influences the subsequent reconstruction. Because borehole measurements do not extend to the full modeled depth of approximately 1000 m, the temperature field below the deepest measurement must be specified before running the forward model. In the default configuration, the profile is extended downward by extrapolating the temperature at the deepest sensor, resulting in a vertically constant continuation of the

measured profile. Since the assumed initialization may substantially affect the reconstructed GST history, alternative initialization strategies were tested to evaluate their influence on the reconstruction results. To reduce the potential influence of this simplified assumption, two alternative initialization schemes were evaluated.
First, a steady-state extension of the measured profile was constructed by assuming a constant geothermal heat flux below the deepest sensor. Under this assumption, the temperature gradient at the deepest observation level is continued downward.
Second, a data-informed initialization was tested by incorporating information from a nearby deeper borehole. In this approach, the deepest temperature measurement available for the investigated borehole was linearly interpolated to the next deeper observation level available in the geographically neighboring SdHole2009 borehole dataset, resulting in a more realistic deep-temperature continuation that reflects regional subsurface conditions. This comparison allows assessing the sensitivity of the inversion results, particularly for shallower boreholes where a larger portion of the modeled domain is not constrained by observations.

4.2 Multiple Likelihoods Across Independent Time Slices

4.2.1 Motivation and Assumptions

One of the main objectives of this study is to develop a reconstruction framework that can better capture long-term trends and reduce the impact of short-term fluctuations. To achieve this, I employed a multiple-likelihood approach, in which several independent likelihood terms are constructed from different years of temperature observations from the same borehole. Each likelihood term represents a separate annual mean temperature profile, and all terms are treated as conditionally independent given the model parameters. Similar likelihood factorization strategies are commonly employed in spatial and multi-sensor Bayesian inversion frameworks, where data contributions were modeled as conditionally independent given the model parameters [54, 55].
Although the total time window that can be reconstructed is still limited by the borehole depth, using multiple temperature profiles (from different years) helps the model to make better use of the available information [56]. The changes between these profiles provide the model with additional information about how the surface temperature evolved in the past, which is expected to lead to more stable and accurate reconstructions within the physically possible time range.
The method was tested with varying numbers of profiles and profile arrangements to directly investigate how integration of additional observational data improves GST reconstructions.

4.2.2 Assumption of Conditional Independence

Given the way the multiple likelihoods are combined in this approach, the assumption of conditional independence between the observed temperature, depth profiles $T_{\text{obs}}(t)$ from different years $t \in \mathcal{T}$, given the underlying GST profile θ, becomes a key requirement.
This assumption is justified because:

- the temperature measurements for each year are based on separate time series (for example, annual averages derived from 3-hourly measurements),
- no systematic measurement errors across years have been detected based on available calibration data, and
- the measurement noise is assumed to be independent and identically normally distributed.

Although all likelihood terms depend on the same parameter θ, their conditional independence is mathematically consistent because they do not share any common random influences. This assumption is also widely used in the literature on Bayesian inverse problems involving temporally separated observations [57]. Furthermore, while this assumption simplifies the physical reality of heat conduction, which is inherently cumulative over time, it provides a practical and computationally efficient approximation.

Formulation of the Combined Likelihood

Based on the assumption of conditional independence, the joint likelihood of all temperature–depth observations can be expressed as the product of the individual likelihood terms for each year:

$$\mathcal{L}(\theta) = \prod_{t \in \mathcal{T}} p\left(T_{\text{obs}}^{(t)} \mid \theta\right)$$

and, equivalently, in logarithmic form:

$$\log \mathcal{L}(\theta) = \sum_{t \in \mathcal{T}} \log p\left(T_{\text{obs}}^{(t)} \mid \theta\right)$$

Each likelihood term is based on the following error model:

$$T_i^{\text{obs},(t)} = G_i^{(t)}(\theta) + \epsilon_i^{(t)}, \quad \epsilon_i^{(t)} \sim \mathcal{N}(0, \sigma^2)$$

where $G_i^{(t)}(\theta)$ denotes the simulated temperature profile for year t predicted by the forward model under a given ground surface temperature history θ.
Assuming additive, mutually independent Gaussian noise [56, 58], each likelihood term takes the form

$$p\left(T_{\text{obs}}^{(t)} \mid \theta\right) \propto \exp\left(-\frac{1}{2\sigma^2}\left\|T_{\text{obs}}^{(t)} - G^{(t)}(\theta)\right\|^2\right)$$

We define the observation vector and the corresponding forward model prediction for year t as

$$T_{\text{obs}}^{(t)} = \left(T_1^{\text{obs},(t)}, \ldots, T_n^{\text{obs},(t)}\right)^\top, \qquad G^{(t)}(\theta) = \left(G_1^{(t)}(\theta), \ldots, G_n^{(t)}(\theta)\right)^\top$$

This formulation assumes that annual observations are conditionally independent given θ, and that the measurement errors have constant variance and are uncorrelated across depths. While these assumptions may not perfectly reflect the physical system, they provide a practically useful approximation in Bayesian inversion.
By combining multiple independent likelihoods, the inversion framework incorporates information from different years. Each additional likelihood provides an independent constraint on

the model parameters, effectively increasing the amount of information used in the inversion and potentially allowing the reconstruction to extend further back in time within the depth-constrained limits. Although the total reconstruction window remains limited by borehole depth, the inclusion of multiple annual profiles is intended to better constrain the GST history and to improve the stability of the inversion within the physically plausible time range.

4.2.3 Practical Implementation of the Multiple-Likelihood Approach

The multiple-likelihood approach described in Section 4.2.2 was practically implemented as part of this study. The inversion was carried out using the `SimulationBasedInference.jl` package, where the model was structured as a `SimulatorInferenceProblem`. The observational datasets $T_{\text{obs}}^{(t)}$ for several comparison years $t \in \mathcal{T}$ were treated as separate likelihood components. In this way, the total likelihood across all considered years is represented in product form:

$$\mathcal{L}(\theta) = \prod_{t \in \mathcal{T}} \mathcal{L}_t(\theta)$$

This represents a common approach for Bayesian inverse problems with multiple independent datasets [57]. The employed error distributions are isotropic Gaussian (`IsoNormal`) with an exponential prior for the standard deviation σ.
The implementation developed in this work was further extended and adapted to support the use of data-driven priors and a modular forward-problem framework, allowing flexible integration of different datasets and physical model configurations.

4.2.4 Synthetic Validation Experiments

Before applying the inversion framework with multiple likelihoods to real borehole observations, synthetic validation experiments were performed to assess whether the extended method can recover known temperature signals under controlled conditions.
In these experiments, artificial ground surface temperature (GST) histories were prescribed and used to generate synthetic subsurface temperature profiles with the forward model using the same stratigraphy and model configuration as in the real-data experiments.
This procedure provides a controlled benchmark to assess whether the inversion framework can recover the imposed GST histories under consistent modeling assumptions and to verify the correct implementation of the multiple-likelihood approach. Because the forward and inverse models use the same setup, any reconstruction differences are more likely caused by limitations of the inversion method rather than by inconsistencies in the model configuration.

4.3 Data-driven GST Prior

In Bayesian inversion, the prior represents the initial knowledge or assumptions about the unknown parameters before considering any observational data. In the context of GST reconstruction, one of the priors defines which temperature histories are considered plausible even before information from borehole measurements is incorporated.
Choosing an appropriate prior is crucial, because the inversion problem is ill-posed and highly

sensitive to the assumed long-term variability of GST. A poor or overly restrictive prior can bias the reconstruction, while an uninformative or heuristic prior, such as an educated guess, provides little guidance to the model and fails to constrain the inversion toward realistic temperature histories. In contrast, a well-informed prior helps stabilize the inference and could improve the physical interpretability of the results.
To address this limitation, the present work employs a data-driven prior that is informed by independent observational data [56].
This section first explains why the prior plays such an important role in the inversion and then describes how the data-driven prior is constructed and implemented.
The inversion problem in this study is inherently ill-posed, small differences in data can lead to very different solutions. In particular, shallow boreholes only contain temperature signals from the last few decades and therefore provide limited information about earlier climate variations. In such cases, the prior becomes essential to constrain the inversion toward physically plausible GST histories. By using a data-driven prior, additional information from external data sources can be incorporated, such as local or regional air temperature records (e.g., ERA5). This helps the inversion capture long-term trends and seasonal patterns that would otherwise be lost in the borehole data alone.

4.3.1 n-Factor Approach

The n-factor provides a simplified relationship between air temperature and ground surface temperature by summarizing the effects of surface energy exchanges into a single dimensionless parameter. It reflects the fact that air and ground surface temperatures can differ substantially depending on surface conditions. By condensing complex surface processes into an empirical correction factor, the n-factor offers a practical means of translating atmospheric temperature signals into effective ground surface forcing [59].
Air temperature is not a direct proxy for the ground surface temperature, because the thermal regime for the ground surface is strongly controlled by surface properties and seasonal processes. During summer, ground surface temperatures are mainly influenced by shading, surface albedo and soil moisture, which influence surface energy balance and can result in reduced ground surface temperatures [60, 61]. In winter, the thermal regime is dominated by snow conditions, which strongly control heat exchange between the atmosphere and the ground surface and can effectively decouple ground surface temperatures from atmospheric conditions [61].
There are usually different n-factors for summer and winter [37, 62].
The freezing n_F and thawing n_T factors are defined as ratios of degree-day sums, which quantify the cumulative coldness or warmth over a given period:

$$n_F = \frac{\mathrm{FI}_{\mathrm{soil}}}{\mathrm{FI}_{\mathrm{air}}}, \quad n_T = \frac{\mathrm{TI}_{\mathrm{soil}}}{\mathrm{TI}_{\mathrm{air}}} \tag{4.1}$$

Freezing index (FI) and Thawing index (TI) quantify the cumulative thermal conditions below and above the freezing point, respectively. They are computed as integrals (or discrete sums) over the freezing and thawing seasons θ [37, 62]:

$$\mathrm{TI} = \int_0^{\theta} (T_s - T_f)\, dt \approx \sum_{t=0}^{\theta} \bar{T}_s(t) \quad \text{for } T_s(t) > 0^{\circ}\mathrm{C} \tag{4.2}$$

$$\mathrm{FI} = \int_0^{\theta} (T_s - T_f)\, dt \approx \sum_{t=0}^{\theta} \bar{T}_s(t) \quad \text{for } T_s(t) < 0^{\circ}\mathrm{C} \tag{4.3}$$

where $T_s(t)$ is the daily surface temperature, and θ is the length of the thawing or freezing season. In practice, the thawing index is typically calculated for each calendar year, while the freezing index spans from July of the previous year to June of the current year [37, 62].
Using these n-factors, ground surface temperature can be approximated from air temperature in a simplified form as:

$$T_{\mathrm{GST}}(t) = \begin{cases} n_F \cdot T_{\mathrm{air}}(t), & \text{if } T_{\mathrm{air}}(t) < 0^{\circ}\mathrm{C} \\ n_T \cdot T_{\mathrm{air}}(t), & \text{if } T_{\mathrm{air}}(t) > 0^{\circ}\mathrm{C} \end{cases} \tag{4.4}$$

4.3.2 Construction of the Prior

The data-driven prior is constructed by translating ERA5 air temperature data into ground surface temperature using the n-factor approach. The specific assumptions and parameter choices employed in this study are outlined below.
In this study, I have used a fixed freezing n-factor of $n_F = 0.6$, because has been previously used for the Samoylov area by [37, 61]. A thawing n_T factor was not applied, as the mean annual ERA5 air temperature data did not exceed 0°C in the relevant periods and regions. Therefore, no thawing degree days were present, and only the freezing season contributed to the GST reconstruction. To visually demonstrate the effect of the n-factor transformation, Figure 4.4 shows the ERA5 air temperature series and the resulting ground surface temperature (T_{surf}) estimated using the freezing n-factor. For validation, the ERA5-based GST is compared with observed GST data from the Samoylov site. The results show that the n-factor approach reproduces the overall temperature level and interannual variability of the measured GST, confirming the plausibility of the chosen parameter for representing the air–ground temperature relationship in this region.

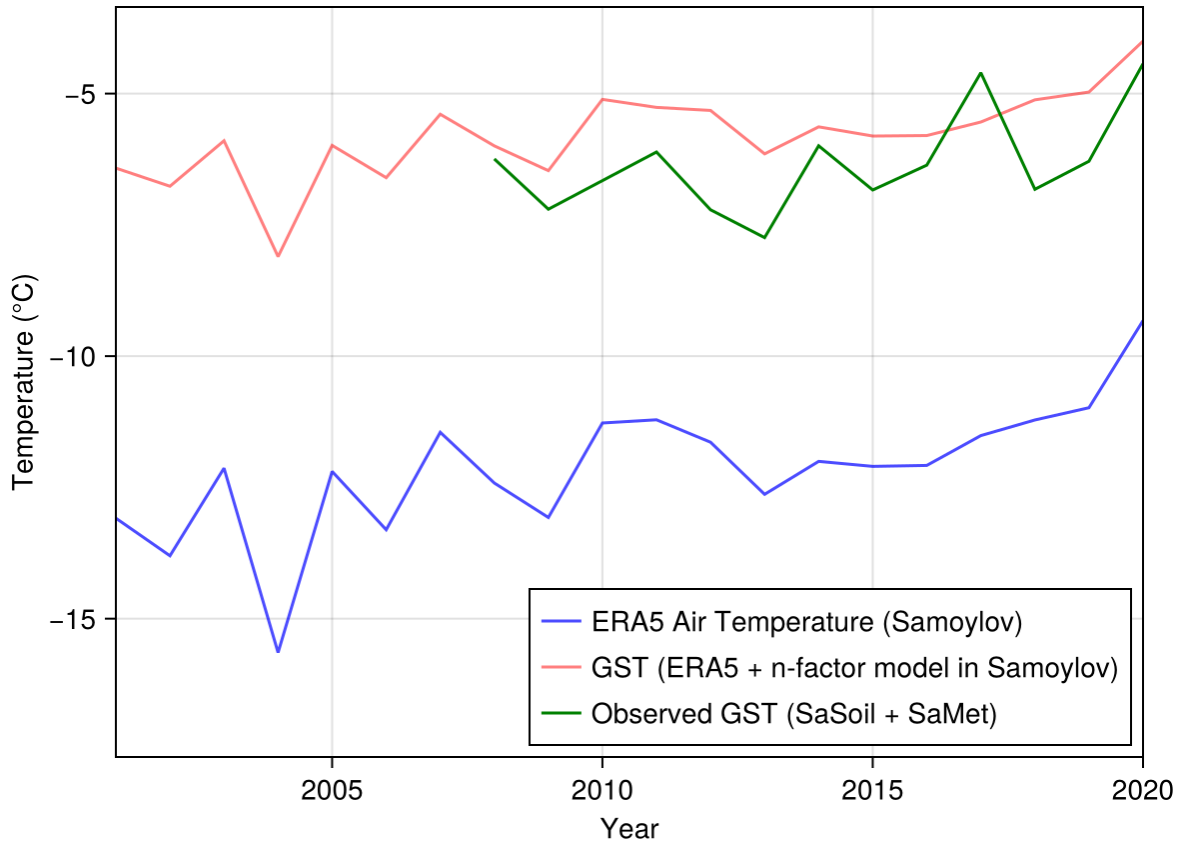


Figure 4.4: Annual mean air temperature (T_{air}) from ERA5 reanalysis for Samoylov (blue), compared with the ground surface temperature (T_{surf}) derived from ERA5 using an n-factor model (red). The green line shows the observed mean GST from the two Samoylov weather stations (SaSoil and SaMet). The applied freezing n-factor was $n_{freezing} = 0.6$.

4.3.3 Combining Data Informed Prior and Educated Guess Prior - (Hybrid Prior)

While ERA5 air temperature data are temporally complete, air temperature information used to construct a data-driven prior can also be obtained from other sources, such as local weather station observations, which may contain temporal gaps. In this case, I used a hybrid approach that mixes two sources of prior information:

- A data-driven prior derived from available air temperature data (e.g., ERA5 or local observations) using n-factors
- A default guess (e.g., a constant GST value or simple trend)

For each reconstruction year, I checked whether the data were available. If air temperature data were available, the data-driven prior value, for the corresponding year, was used; if not, it fell back to the default. I also assigned different uncertainty levels to reflect data reliability: years with available data received a lower uncertainty ($\sigma_{data} = 2.0$°C), while the guessed

years are modeled with a higher uncertainty ($\sigma_{\text{guess}} = 2.5^{\circ}\text{C}$). This created a combined prior distribution, where each year had its own mean and variance.

Formally, the prior mean μ_i for year i is defined as:

$$\mu_i = \begin{cases} T_{\text{GST}}^{\text{data}}(t_i), & \text{if air temperature data are available at } t_i \\ T_0, & \text{otherwise} \end{cases} \qquad \text{for } i = 1, \dots, N\,,$$

where $T_{\text{GST}}^{\text{data}}(t_i)$ denotes the data-informed annual mean surface temperature estimate used as prior information and T_0 represents the default educated-guess temperature.
This leads to a hybrid prior mean vector $\boldsymbol{\mu}_{\text{prior}} \in \mathbb{R}^N$, which contains both empirical and assumed values depending on data availability.
Formally, each year θ_i in the prior is drawn according to the data availability:

$$\theta_i = \begin{cases} \mu_i + \varepsilon_i^{(1)}, & \text{if data is available at } t_i \\ \mu_0 + \varepsilon_i^{(2)}, & \text{otherwise} \end{cases}$$

where $\varepsilon_i^{(1)} \sim \mathcal{N}(0, \sigma_{\text{data}}^2)$ and $\varepsilon_i^{(2)} \sim \mathcal{N}(0, \sigma_{\text{guess}}^2)$.
The final prior vector is constructed year by year based on this rule, resulting in a piecewise Gaussian process with heterogeneous variance.
This reflects that neighboring years are more similar than distant ones, which is typical in long-term GST signals. The Autoregressive (AR(1)) process is used to introduce temporal correlation between consecutive years, enforcing smoothness in the reconstructed temperature histories. This reflects the physical persistence of ground surface temperature variations over time and helps to stabilize the inversion in the presence of noisy or incomplete data [38, 58].

4.3.4 Use of the Prior

In the Ensemble Kalman Sampling framework, the prior defines both the initial ensemble and is used during the ensemble updates. The algorithm begins by generating multiple random GST histories, each sampled from the prior distribution:

$$\phi^{(i)} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{prior}}, \Sigma_{\text{prior}}), \quad i = 1, \dots, N$$

Each GST history was used to calculate a synthetic borehole profile. These were compared with the real profile. Then the EKS updated the ensemble to better match the observations.
In our implementation, the prior samples are not drawn from a single Gaussian distribution but from a piecewise hybrid prior. This prior combines two components: one derived from ERA5-based estimates with smaller uncertainty, and one from a fallback assumption with broader uncertainty. This design allows the prior to flexibly adapt to incomplete data while maintaining physically coherent temporal behavior.

4.4 Sensitivity Analysis

To evaluate how the extended model components, such as the multiple-likelihood formulation and the data-driven prior, affect the quality and stability of the GST reconstruction, this study

performed a systematic sensitivity analysis.
The sensitivity analysis (SA) is used to quantify how variations in model inputs and modeling assumptions influence the reconstruction results [63]. Its purpose is to identify which parameters most affect the quality, robustness, and reliability of the reconstructed GST profiles and to guide the selection of suitable model configurations [64, 65].
In this study, a targeted sensitivity analysis was performed to assess how key modeling choices within the EKS framework influence the GST reconstruction. In particular, the assumed noise level is systematically varied for different likelihood configurations. For a given number of likelihood, only the noise parameter was changed while all other model components were kept fixed. The number of likelihood was increased, and the same procedure was repeated. This approach was chosen to efficiently assess how the number of likelihoods, in combination with different noise levels, influences the robustness and stability of the GST reconstruction within the EKS-based inversion framework.
In all experiments, data from the Samoylov borehole were used. This site was chosen because it provides a continuous and gap-free temperature time series from 2010 to 2020, allowing for a fair comparison across different parameter settings. For each test case, a reconstruction period of ten years was considered. This time window was chosen because, in addition to the Samoylov borehole, the two other boreholes SaMet and SaSoil (see Section 3.3.2) provide reliable temperature records for the same period, which can be used as independent ground-truth references. This allows for a direct validation of the proposed concept and helps to verify whether the methodological extensions indeed lead to measurable improvements. During this period, ERA5-based GST values were available for every year, so a hybrid prior was not required and only the data-driven (ERA5-based) prior was used.
The number of likelihoods was varied between one, two, and three. Each likelihood used borehole data from a different year, typically spaced 2–3 years apart. Additionally, the assumed noise level (i.e., the standard deviation of the measurement error) for each likelihood was tested with the following values: 0.05, 0.1, 0.2, 0.5, 1.0, and 2.0. Although the accuracy of the temperature sensors is typically in the range of 0.05 – 0.2 °C, a broader range of noise levels is considered in this sensitivity analysis. This exploratory choice allows assessing how the inversion behaves under different noise assumptions, particularly when combining multiple likelihoods. All reconstructions were carried out using the EKS algorithm, with 20 iterations and an ensemble size of 128 particles. This setup provided sufficient convergence and variability within the posterior ensemble.
To evaluate the quality of the reconstruction results, several quantitative metrics were computed: the Root Mean Square Error (RMSE), the posterior spread, the coverage rate (i.e., how often the true value lies within the posterior interval), the normalized log-likelihood and the Pearson correlation.
The RMSE is used as a point prediction accuracy measure and is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(T_t^{\mathrm{obs}} - \hat{T}_t\right)^2},$$

where T_t^{obs} denotes the observed annual mean ground surface temperature (GST) for year t, $\hat{T}_t$ denotes the corresponding posterior mean GST reconstruction, and N is the number of reconstruction years. RMSE provides an easily interpretable measure of reconstruction

accuracy [38].
Posterior spread describes the uncertainty of an estimated quantity and is defined as the width of the central posterior interval. In this study, uncertainty is summarized using the 95% central posterior interval, defined by the 2.5% and 97.5% posterior quantiles.
The coverage rate is used to assess the consistency of the posterior uncertainty. It is defined as the fraction of reconstruction years for which the observed ground surface temperature (GST) lies within the corresponding 95% posterior interval. If a 95% posterior interval is used, the coverage rate indicates how often the observed GST values fall inside this interval. Ideally, this fraction should be close to 95%. However, coverage alone is not sufficient to evaluate uncertainty quality and should be therefore interpreted together with the posterior spread [38]
The log predictive density measures how well the model predictions match the observed data, taking uncertainty into account. The log predictive density is also sometimes called the log-likelihood [38]. Finally, the Pearson correlation measures the strength of a linear association between two continuous variables and ranges from -1 to +1, where values close to zero indicate little or no linear relationship [66]. However, correlation quantifies association rather than agreement and should therefore not be interpreted as a direct measure of reconstruction accuracy. All metrics were computed with respect to the ground truth, defined as the mean of the SaMet2002 and SaSoil2002 datasets.
The results were visualized using heatmaps, where the axes represent different parameter combinations (such as likelihood number vs. noise level), and the color scale indicates the value of the chosen evaluation metric (e.g., RMSE). This visualization enabled a clear interpretation of how the reconstruction accuracy and robustness depend on the number of Likelihood terms and the assumed noise level.
Overall, the sensitivity analysis provided insight into the optimal configuration of the inversion model. In particular, it showed under which conditions multiple likelihoods had the highest stability and accuracy of GST reconstructions.

4.5 Comparative Evaluation and Depth Analysis

To analyze how the depth of a borehole affects the potential to reconstruct historical GST, reconstructions were performed at different truncation depths. In this study, the Sardakh borehole (SdHole2009) was used for this purpose due to its original length of 100 m, which allows for flexible testing with shorter profile versions.
Specifically, the original profile was truncated at depths of 30 m, 50 m, and 100 m, and reconstructions were carried out separately for each case. This procedure enables a systematic comparison of how much temporal information can be recovered from profiles of different depths.
Additionally, I explored how using multiple profiles from different years (i.e., multiple likelihoods) compares to using a single profile. This helps evaluate whether the availability of richer time-resolved data can partially compensate for shallower boreholes.
This methodological step is particularly relevant for Arctic fieldwork planning. If reliable reconstructions are possible even with shallower boreholes, future drilling efforts in remote or logistically challenging areas may be optimized accordingly.

5 Results

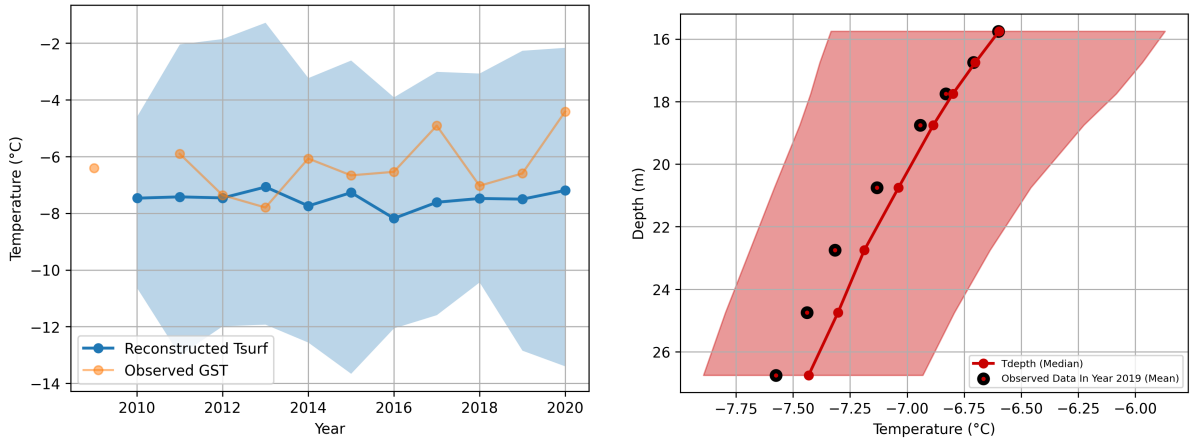
This chapter presents the results of a series of methodological experiments designed to evaluate the proposed inversion framework under realistic data constraints. The focus is on assessing how different modeling choices influence the stability, temporal behavior, and internal consistency of the reconstructed ground surface temperature histories. The results are not intended as a validation of long-term climate signals. Instead, the ERA5+n-factor prior and the Multiple Likelihood approach are tested as proof-of-concept extensions at Samoylov where independent GST observations are available over a limited time window of approximately one decade. This allows a controlled evaluation of methodological behavior close to ground-truth conditions.
The results are organized in a stepwise manner. First, the effect of the data-driven ERA5+n-factor prior is examined in isolation. This prior introduces physically plausible variability without degrading the overall reconstruction quality.
Unless stated otherwise, all reconstruction figures are shown for a fixed noise_scale of 1.0, which was found to yield high log-likelihood values in the majority of test configurations. This choice was made for comparability and visualization purposes. The sensitivity to different noise assumptions is analyzed separately using quantitative metrics.

5.1 Validation of the Data-driven Prior

Traditional priors in Bayesian inversion often rely on theoretical or heuristic assumptions that may not reflect the true variability of the underlying physical system. As discussed by [67], such priors can introduce model-induced biases and limit the interpretability of posterior estimates. To overcome this limitation, this study derives data-driven priors directly from empirical observations, providing a more realistic constraint on the reconstructed ground surface temperature based on empirical observations.
Following this concept, the present study constructs a data-driven prior for ground surface temperature (T_{surf}) using ERA5 reanalysis data and n-factor modeling of the air–ground temperature relationship. The validation presented in this section assesses how well this data-driven prior improves the reconstructed temperature histories compared to the previously used educated-guess prior.
To evaluate how the data-driven prior affects the results, two reconstructions were carried out using the same inversion settings. The first used an educated-guess prior, while the second included the ERA5+n-factor prior. This comparison showed how the use of real data in the prior could affect the stability and accuracy of the reconstructed ground surface temperature (T_{surf}).

5.1.1 Reconstruction without ERA5 Prior

Figure 5.1 shows the reconstruction obtained using the simple educated-guess prior. Removing



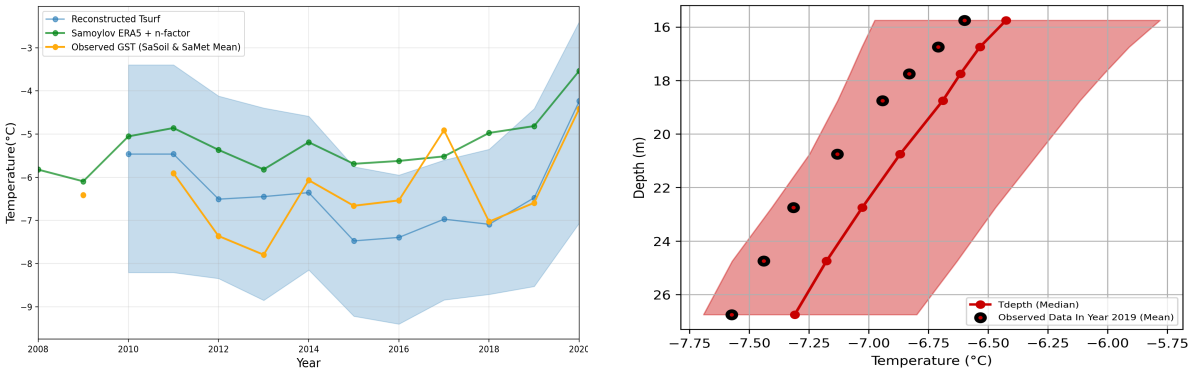
(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.1: Reconstruction with the educated-guess prior. (a) Reconstructed GST history for the available measurement years, with blue shading indicating the uncertainty range and orange points showing observed GST values. (b) Reconstructed borehole temperature profile (red curve and uncertainty band) compared to measured temperatures (black points).

the upper 15 m shifted the inversion to depths where short-period variability is strongly damped. Although the annual cycle may not fully vanish at 15 m, its impact on mean annual temperatures is expected to be small (see Chapter 4.1). Even if the upper part of the borehole were included, only the most recent annual cycle would contain some information about high-frequency behavior, while older years would still appear smooth due to the natural low-pass filtering of the subsurface.
This explained why the reconstructed T_{surf} in Figure 5.1a lacked in the year-to-year fluctuations visible in the observed GST. The general temperature level is captured, but the reconstruction remains smoother than the measurements.
To understand how a real-world, climatology-based prior changes exactly the inversion, I repeated the test using the ERA5+n-factor prior.

5.1.2 Reconstruction with ERA5 Prior

Figure 5.2 shows the reconstruction obtained using the data-driven ERA5+n-factor prior. Compared to the simple educated-guess prior, the reconstructed T_{surf} followed the observed GST more closely in terms of temporal structure. The curve remained smoother than the measurements, as expected from the strong low-pass filtering of the ground, but the overall temporal pattern matched more to the ground truth. In panel 5.2a, the reconstruction exhibits interannual to multi-annual variability that is constrained by the borehole data. In contrast, strictly seasonal (sub-annual) temperature fluctuations are strongly attenuated by the low-pass filtering effect of heat conduction in the subsurface. The use of annually resolved ERA5-based priors can introduce additional year-to-year structure in the reconstruction. The posterior



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.2: Reconstruction using the data-driven prior. (a) Reconstructed ground surface temperature history (T_{surf}) with observed GST data and the data-driven ERA5 + n-factor prior for comparison. (b) Corresponding predicted and measured borehole temperature profile. The data-driven prior is based on ERA5 air temperatures scaled with local n-factors, ensuring that the inversion remains consistent with site-specific climatic conditions.

therefore represented a combination of the smoothed information carried by the borehole data and the annual variability supplied by the prior.
Although both inversions use the same noise scale with a value of 1, the predicted temperature profile with the ERA5+n-factor prior shows a larger deviation from the measured borehole profile (see panel 5.2b). This behavior indicates that the ERA5+n-factor prior provides additional information that partially counteracts the influence of likelihood, reducing the overly strong conditioning on a single borehole profile and promoting a more balanced trade-off between data and prior information.

5.1.3 Effect of Prior Choice under Different Noise Assumptions

To further evaluate the effect of the data-driven prior, the normalized log-likelihood was computed for different assumed noise scales. The test was performed with the following noise scale values: 0.05, 0.1, 0.2, 0.5, 1, 2, and 5. These noise scale values from 0.5 °C to 5 °C do not represent measured sensor uncertainty, but were chosen to explore how different noise assumptions influence the inversion behavior under the assumed forward model. As shown in Figure 5.3, both priors achieved their highest likelihoods at moderate noise scales between 0.5 and 1, after which the fit gradually decreased. The best overall score is achieved by the ERA5+n-factor prior at a noise scale of 1.0 with a normalized log-likelihood of -1.81, slightly better than the educated-guess prior (-1.84). Overall, the results for both priors were very similar, indicating that the educated-guess and ERA5-based approaches lead to comparable levels of internal model–data consistency and similar temporal patterns in the reconstructed GST. It should be noted again that the sensor-related measurement uncertainty is substantially smaller (0.05 °C to 0.2 °C). The interpretation of the higher log-likelihood values obtained at larger noise scales (e.g. noise_scale = 1.0) is discussed in Section 6.2.6.

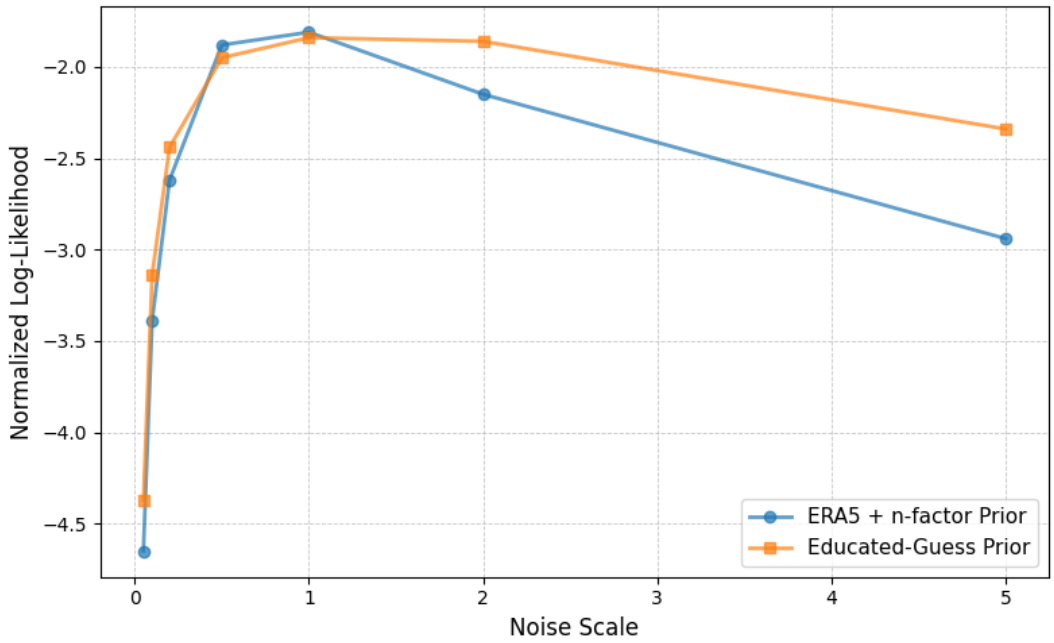


Figure 5.3: Comparison of normalized log-likelihood values for different noise scales at the Samoylov site. Higher (less negative) values indicate a better fit between simulated and observed borehole temperatures. Across most noise levels, the log-likelihood values of both priors are very similar, with only small differences and a slight advantage for the ERA5+n-factor prior at moderate noise scales.

Table 5.1: Reconstruction metrics for different noise scales at the Samoylov site using the educated-guess prior. Reported metrics include RMSE, coverage, posterior spread, and Pearson correlation between reconstructed and observed GST over the available validation period.

Noise Scale	RMSE [°C]	Coverage [%]	Spread [°C]	Pearson Correlation
0.05	1.90	82	1.91	0.17
0.1	1.81	91	2.07	0.17
0.2	1.67	91	2.27	0.31
0.5	1.70	91	2.55	-0.19
1.0	1.62	91	2.86	-0.37
2.0	1.22	91	3.54	-0.06
5.0	1.16	91	7.88	0.01

Table 5.2: Reconstruction metrics for different noise scales at the Samoylov site using the data-driven ERA5+n-factor prior. The table reports RMSE, coverage, posterior spread, and Pearson correlation with observed GST.

Noise Scale	RMSE [°C]	Coverage [%]	Spread [°C]	Pearson Correlation
0.05	1.83	82	1.68	0.44
0.1	1.76	82	2.17	0.43
0.2	1.49	82	2.39	0.47
0.5	1.23	82	2.78	0.48
1.0	0.99	91	2.98	0.46
2.0	1.48	91	4.64	0.17
5.0	1.30	91	12.30	0.42

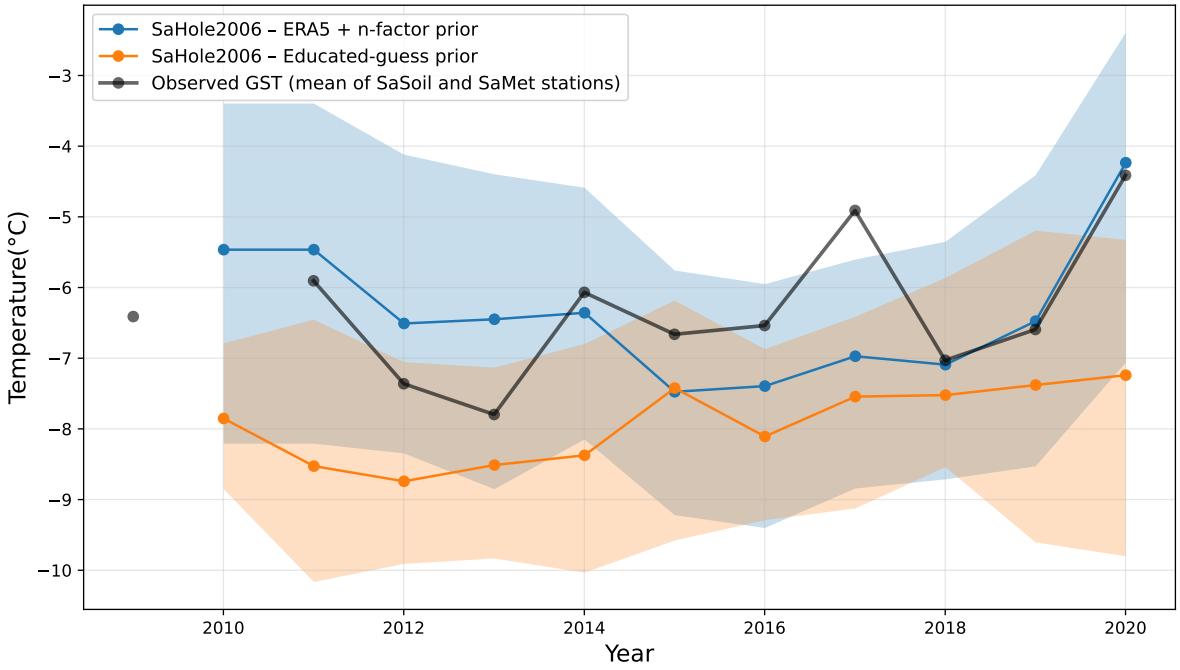


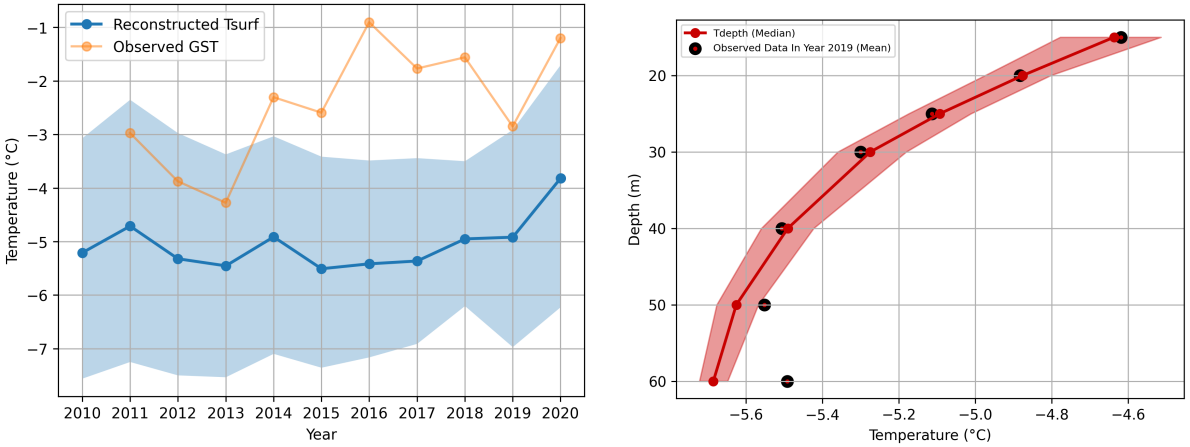
Figure 5.4: Direct comparison of SaHole2006 GST reconstructions using an educated-guess prior and a data-driven ERA5 + n-factor prior. Solid lines show posterior mean temperatures, and shaded regions represent 95 % credible intervals. The data-driven prior aligns more closely with observed GST and provides more realistic variability across the reconstruction period.

The quantitative comparison of Table 5.1 and Table 5.2 shows that both priors achieve similar overall performance across the tested noise scales. At low noise levels (0.05–0.2), the RMSE with respect to the available GST reference is around 1.5–1.9 °C for both approaches, indicating that the borehole likelihood dominates the solution. However, at moderate noise scales (0.5–1.0), the ERA5 + n-factor prior consistently achieves lower RMSE values, reaching a minimum of 0.99 °C at a noise scale of 1.0, compared to 1.62 °C for the educated-guess

prior. This is consistent with the log-likelihood analysis, which shows similar behavior at comparable noise scales.
A clear difference between the two priors can be seen in the Pearson correlation values. The educated-guess prior shows weak or even negative correlations for most noise scales, while the ERA5+n-factor prior consistently shows higher positive correlations (up to 0.48). Pearson correlation measures how well the temporal pattern of the reconstructed GST follows the observed GST. In borehole-based reconstructions, the exact temperature level is often shifted due to model simplifications and regularization. Therefore, it is more important to reproduce the correct warming and cooling phases than to match the exact absolute values. For this reason, the higher correlation values obtained with the ERA5-based prior indicate an improved reconstruction of the temporal evolution, even when RMSE differences are moderate.
The coverage rates remain nearly constant (around 82–91%) across all noise levels, showing that both priors yield well-calibrated uncertainty estimates. In contrast, the spread values increased with higher noise scales, as expected from the growing posterior variance. The ERA5-based prior tended to produce slightly larger spreads at higher noise levels, suggesting that the climatological prior allows more flexibility in regions where the data constraint is weak, while maintaining realistic temperature variability.
The figure 5.4 highlights the impact of prior choice on GST reconstruction quality. This visual impression is supported by the Pearson correlation values, which are consistently higher for the ERA5-based prior, indicating a stronger linear association with the observed GST trends, while not implying improved point-wise agreement.

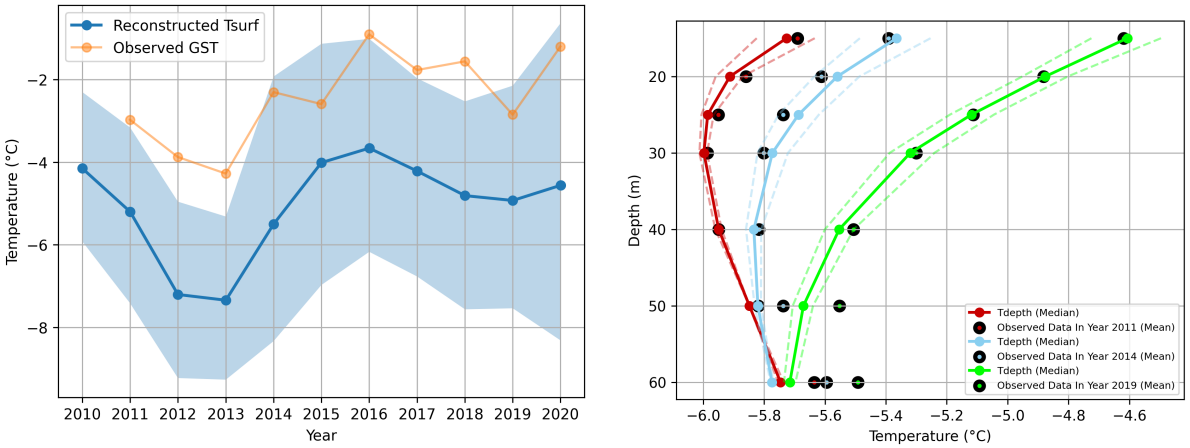
5.2 Synthetic Validation of the Multiple-Likelihood Framework

To examine the behavior of the proposed inversion framework under controlled conditions, synthetic temperature–depth datasets were generated using the CryoGrid community model [52]. The synthetic data was produced using the same stratigraphic configuration assumed in the inversion. However, the forward model used for data generation differs slightly from the forward model used within the inversion procedure. Since the true ground surface temperature (GST) history is known in the synthetic setup, the experiment allows a direct comparison between reconstructed and prescribed GST histories. The focus of this validation is on the behavior of the inversion algorithm and the multiple likelihood formulation, in particular its stability and sensitivity to the number of likelihoods and how much the multiple likelihood framework improves the reconstruction. Reconstructions were performed using one and three likelihood while keeping all other inversion settings identical. Figure 5.5 shows the GST reconstruction obtained with a single likelihood term, while Figure 5.6 presents the result using three likelihood terms.
The single-likelihood reconstruction already captures the general temporal evolution of the imposed GST signal but exhibited noticeable deviations in amplitude and temporal alignment. When additional likelihood terms are included, the reconstructed GST histories showed a progressively improved agreement with the reference signal, particularly in terms of temporal



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.5: Synthetic validation using one likelihood term. (a) Reconstructed ground surface temperature (T_{surf}) compared to the known synthetic reference signal. (b) Corresponding predicted and synthetic borehole temperature profile.



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.6: Synthetic validation using three likelihood terms. (a) Reconstructed ground surface temperature (T_{surf}) compared to the known synthetic reference signal. (b) Corresponding predicted and synthetic borehole temperature profiles for the selected likelihood years.

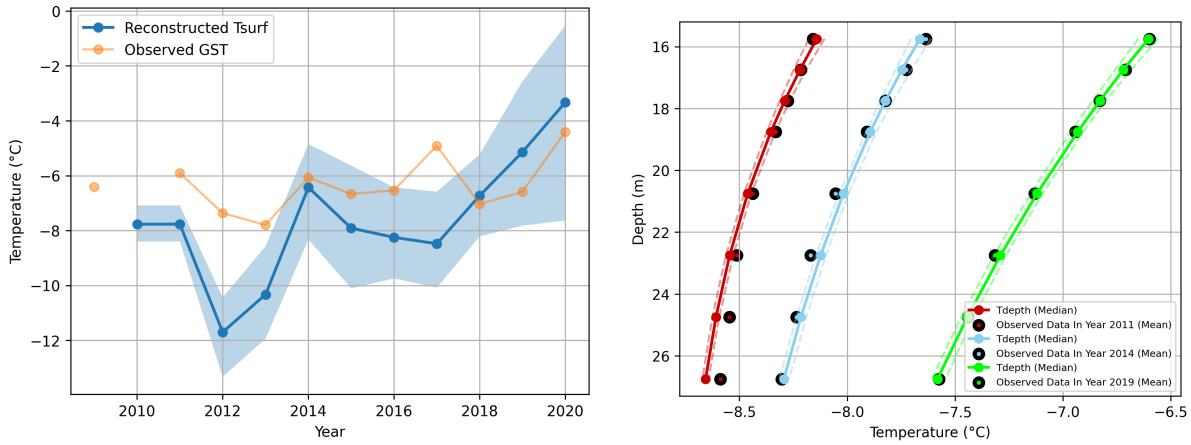
correlation and overall stability.
Although a systematic temperature offset of approximately 1–3 °C remains visible in all configurations, the consistent improvement in correlation and temporal structure demonstrated that the multiple-likelihood framework successfully extracts additional information from independent observation years. These results confirm that the extended inversion setup is functioning as intended and that the inclusion of multiple likelihood terms enhances the recoverable temporal signal even under idealized synthetic conditions.
A closer look at the predicted temperature profiles helps to better understand the behavior of the inversion. In the three-likelihood case, the fit of the predicted profiles (Figure 5.6b) appears slightly worse compared to the single-likelihood reconstruction (Figure 5.6a. Since the noise level is kept constant (noise_scale=0.1), adding more likelihood terms places a stronger overall constraint on the inversion. As a result, the inversion may find a compromise solution that explains all likelihood terms together, but does not perfectly match each individual profile. In addition, the predicted profiles show larger deviations below approximately 50–60 m depth. The initial temperature field was constructed using the measured sensor values and then linearly extrapolated below the deepest sensor. This simplified treatment of the deeper subsurface could contribute to the observed differences.

5.3 Effect of the Multiple-Likelihood Framework with Real Borehole Data

After evaluating the data-driven prior and the multiple-likelihood framework separately, both methodological extensions are combined and applied to the real borehole datasets. Each likelihood term represents a separate annual temperature depth profile from the same borehole, allowing the inversion to include information from multiple years simultaneously. I expect that the year-specific noise, which appears independently in each measurement and motivates the assumption of conditional independence, may be partly averaged out when several profiles are considered separately within the inversion. This could help stabilize the reconstruction, even though the physical limits of the borehole signal remain unchanged. Moreover, by including information from several observation years, the approach effectively increases the amount of data constraining the inversion and may help to better exploit the limited temporal signal preserved in shallow boreholes when reconstructing older temperature changes.
The analysis investigates reconstructions performed with different numbers of likelihood, while keeping all other settings identical. In each configuration, the same prior and inversion parameters were used to ensure that any observed differences are due solely to the number of likelihoods included in the inference process.
This setup allows the method to investigate how the integration of multiple time-dependent likelihoods influences the reconstruction of long-term temperature variations, overall uncertainty, and temporal consistency within a single inversion framework.

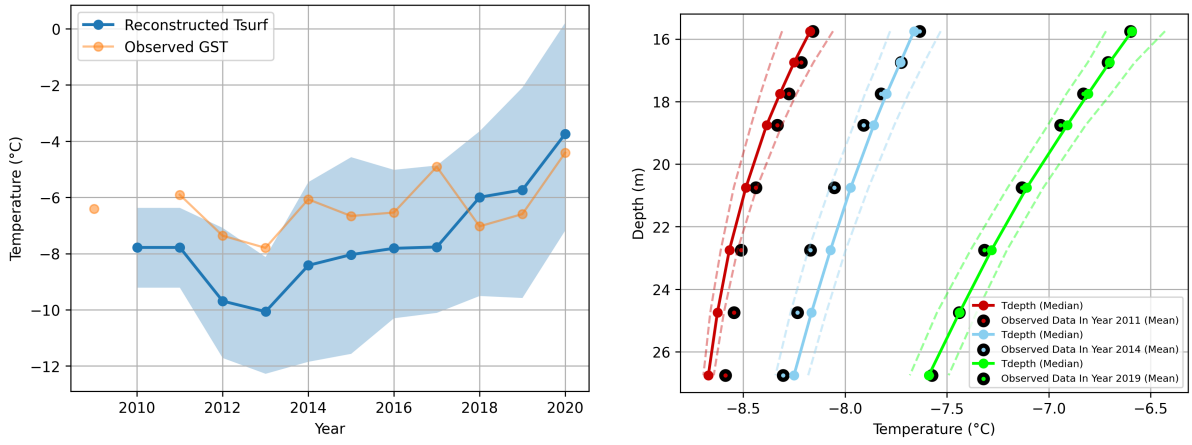
5.3.1 Sensitivity to noise scale in the multiple likelihood framework

Before analyzing the performance of the multiple-likelihood framework in detail, the sensitivity of the reconstruction to the assumed noise scale is first examined. The assumed noise scale controls how strongly the likelihood constrains the inversion. The following figures illustrate the effect of the assumed observation noise scale on the inversion results. For small noise scales (noise_scale = 0.05), the posterior uncertainty is narrow and the inversion is forced to closely match all likelihood profiles (see Figure 5.7). In this regime, even small mismatches between propagated and observed profiles lead to strong adjustments of the reconstructed surface temperature history. This results in pronounced year-to-year variability and reduced smoothness. At intermediate noise scales (noise_scale = 0.2), this effect is partially reduced, as we can see in the Figure 5.8. The inversion becomes less restrictive and allows for moderate discrepancies between model predictions and observations. As a result, the reconstructed surface temperature history appears smoother, while still being influenced by the individual likelihoods. At noise_scale = 1.0, the reconstructed GST appears smoother and the internal consistency between predicted and measured profiles improves across years (Figure 5.9). Based on these observations, noise_scale = 1.0 is used as the reference setting for the following methodological comparisons. The implications of employing noise values that exceed realistic sensor uncertainty are discussed in Section 6.2.6.



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.7: Reconstruction using the data-driven prior and multiple likelihoods with noise_scale = 0.05. (a) Ground surface temperature history and (b) corresponding borehole profiles for the measurement years 2011, 2014, and 2019.

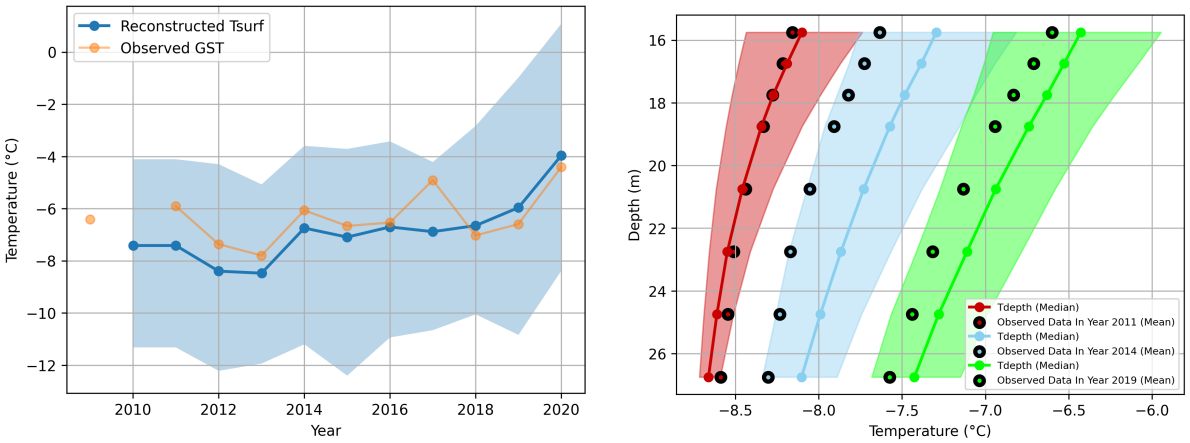


(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.8: Reconstruction using the data-driven prior and multiple likelihoods with noise_scale = 0.2. (a) Ground surface temperature history and (b) corresponding borehole profiles for the measurement years 2011, 2014, and 2019.

5.3.2 Non-consecutive Likelihood

To test how the inclusion of additional and independent observation periods affects the inversion, reconstructions were first performed using non-consecutive likelihoods.
Figure 5.2 shows the reconstruction using a single likelihood, while Figure 5.9 presents the



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.9: Reconstruction using the data-driven prior and multiple likelihoods. (a) Ground surface temperature history and (b) corresponding borehole profiles for the measurement years 2011, 2014, and 2019.

result with three likelihoods included in the inversion. Although both reconstructions use the same ERA5 + *n*-factor prior and identical inversion settings, the version with multiple likelihoods produced a visibly smoother T_{surf} history. For the short period where ground truth is available for comparison, the reconstruction with three likelihoods follows the observed GST more closely and the uncertainty band appears slightly narrower. This suggests that including several yearly profiles helps the inversion integrate information from different observation years, which may stabilize the temporal behavior of the reconstruction. Figure 5.10 illustrates how the normalized log-likelihood varies with both the noise scale and the number of likelihood terms. A clear trend can be observed: while all reconstructions perform similarly at low noise scales, the advantage of using multiple likelihoods became increasingly evident as the noise scale rose. Starting from noise_scale ≥ 1, reconstructions that incorporate two or more likelihoods consistently achieved higher (less negative) log-likelihood values compared to those using a single likelihood. This effect was most evident in the moderate noise range (1.0-2.0), where the inversion reaches a higher log-likelihood score. At very low noise scales, using several likelihoods can slightly degrade the results, because the inversion becomes overly constrained by the forward model and is forced to reconcile small inconsistencies between multiple profiles that cannot be fully explained by the simplified model (see Chapter 6.2.6).

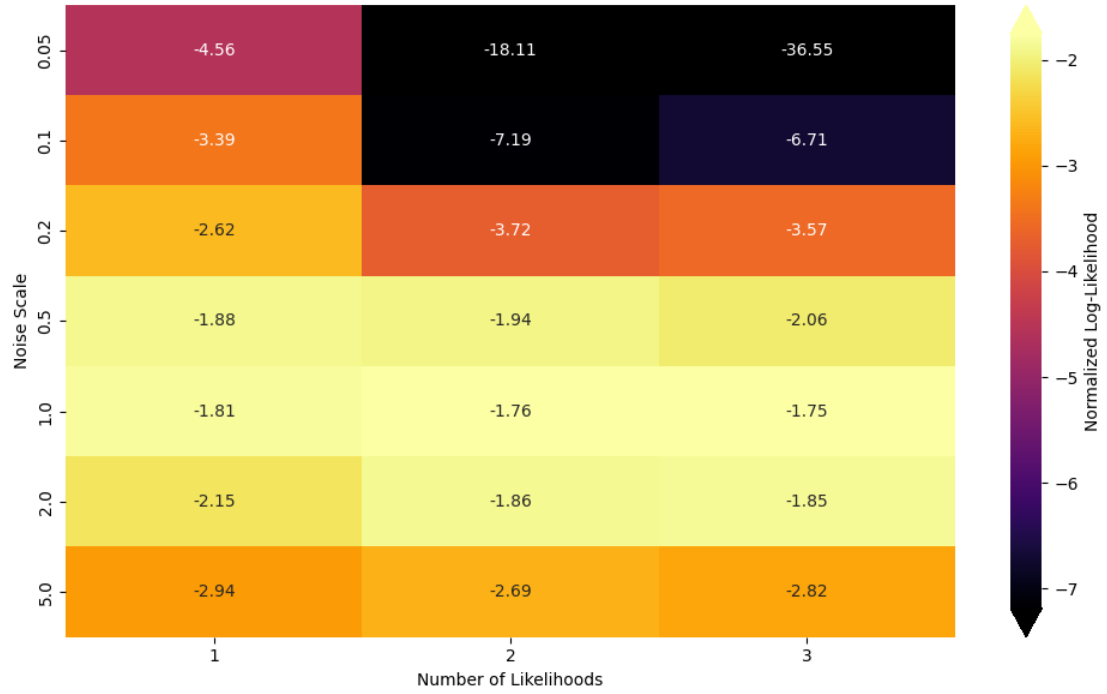


Figure 5.10: Heatmap showing the normalized log-likelihood values as a function of the assumed noise scale and the number of likelihoods. Higher (less negative) values indicate a improved internal consistency between the predicted and measured borehole temperatures. The results demonstrate that increasing the number of likelihoods generally stabilizes the inversion and leads to slightly higher likelihood values in the optimal noise range of 1.0-2.0 , indicating a more robust fit to the data.

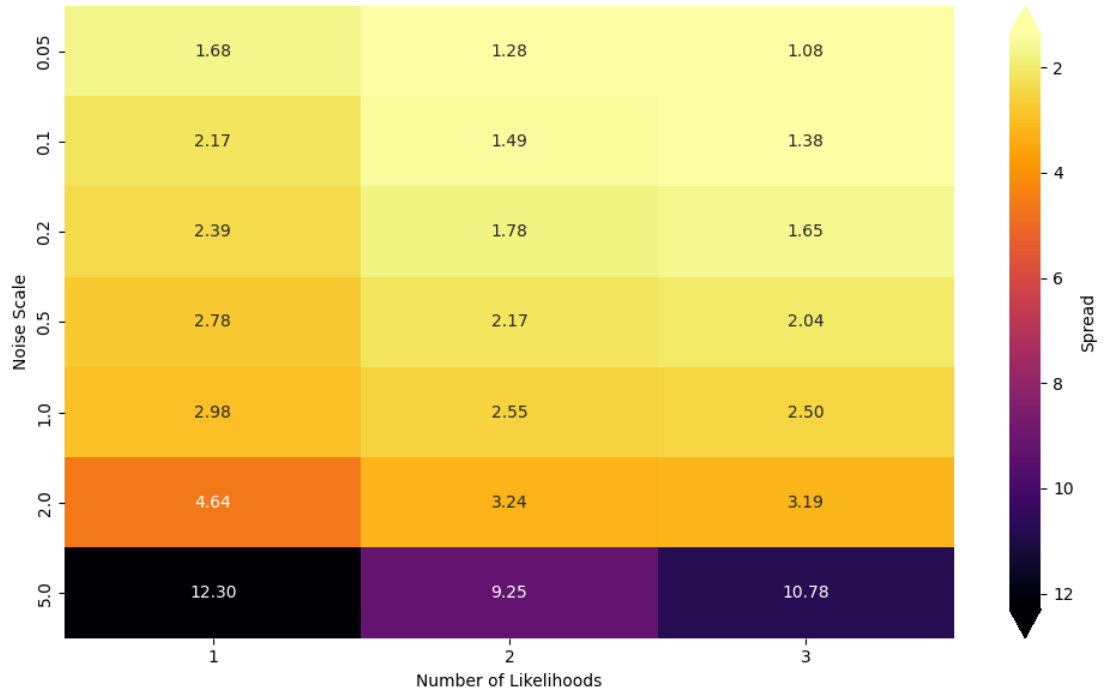


Figure 5.11: Heatmap showing the posterior spread depending on the noise scale and the number of likelihoods. The results show that, in general, using a higher number of likelihoods leads to a smaller spread, indicating a more confident reconstruction.

The posterior spread provides insight into the uncertainty of the reconstructed surface temperature (T_{surf}). A smaller spread means that the ensemble members of the inversion agree more closely, which indicates higher posterior certainty under the assumed model and a more stable reconstruction.

As shown in Figure 5.11, the spread decreased when more likelihoods were used, particularly for moderate noise scales between 0.5 and 2.0. This behavior suggests that adding independent likelihood terms introduces additional information from different observation periods, which helps the inversion to better constrain the solution space.

Although the spread provides useful information about the uncertainty of the inversion, it cannot be interpreted in isolation. A low spread does not necessarily indicate a good reconstruction. For example, at very small noise scales (e.g., noise_scale = 0.1), the spread values are minimal, but as seen in the log-likelihood heatmap, the corresponding reconstructions deviate strongly from the ground truth. This combination of low spread and poor data fit suggests that the posterior distribution is overconfident and does not correctly capture the true variability of the system. In contrast, at noise_scale = 1.0, the spread remains comparably low, while the log-likelihood is high. In general, the results show that the use of multiple likelihoods can reduce posterior spread and may improve posterior stability in the moderate noise range (noise_scale ≈1-2), while the benefit is limited at very low noise.

Figure 5.12 shows the RMSE distribution for different noise scales and numbers of likelihoods. At a noise scale of 1.0, the single-likelihood configuration achieves the lowest RMSE ($\approx 0.99\,^{\circ}\text{C}$), slightly lower than the multi-likelihood setups. This behavior likely reflects the fact that a

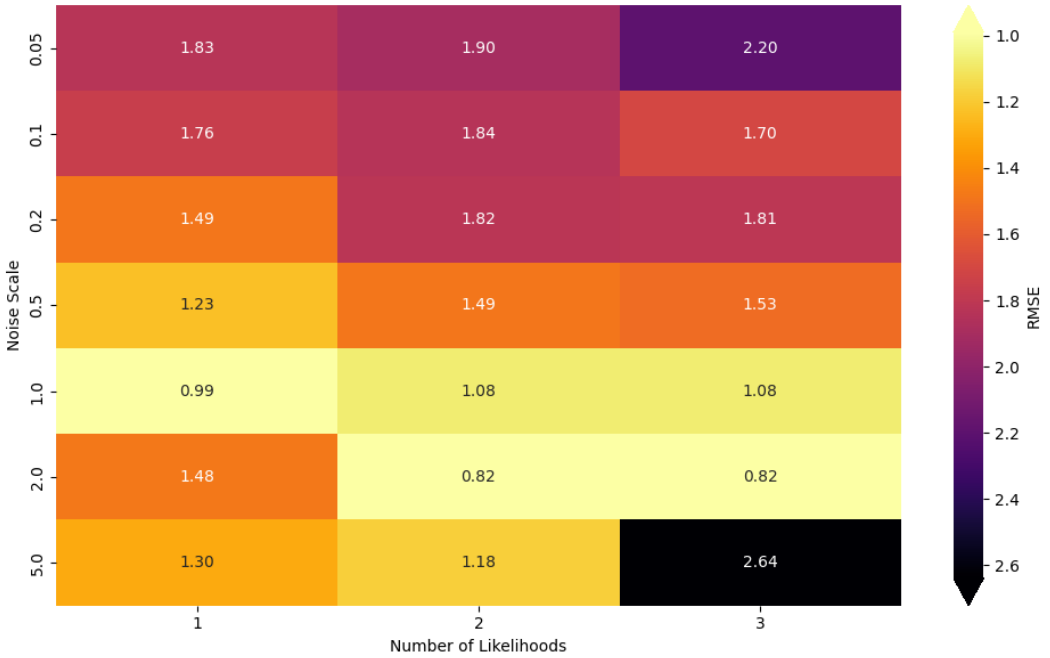


Figure 5.12: Root Mean Square Error of reconstructed ground surface temperatures for different combinations of noise scales and numbers of likelihoods. Lower RMSE values indicate a better agreement between reconstructed and observed GST. The heat map highlights how noise assumptions and the number of likelihoods influence reconstruction accuracy.

single likelihood can more directly fit the short validation interval. For larger noise scales (e.g. noise_scale = 2.0), reconstructions using two or three likelihoods yield the lowest RMSE values ($\approx 0.82\,^{\circ}\mathrm{C}$). In this regime, the influence of individual profiles is reduced, and combining information from multiple observation years leads to a more balanced reconstruction with improved agreement over the validation period. Overall, RMSE results indicated that multiple likelihoods can improve short-term reconstruction accuracy under certain noise assumptions. However, RMSE should be interpreted with caution, as low RMSE values do not necessarily imply a physically more realistic reconstruction, particularly when they occur at noise levels that exceed the actual measurement uncertainty.

The Pearson correlation analysis in Figure 5.13 shows that configurations with multiple likelihoods tended to achieve higher temporal agreement between the reconstructed and observed GST than the single-likelihood case. For a configuration with two likelihoods, the Pearson correlation values lie between 0.52 and 0.69 across the noise scales. For 3 likelihoods Pearson correlation values lie between 0.64 and 0.69 across the most noise scales. Although three likelihoods in general achieve the highest correlations, this behavior is not consistently robust. In particular, at very large noise scales (e.g. noise_scale = 5), the correlation for three likelihoods drops sharply to negative values (−0.55). However, this reflects a regime in which the likelihood contribution becomes weak and the reconstruction is increasingly dominated by the prior and forward model propagation rather than observational constraints.
In general, the Pearson correlation indicated that multiple likelihoods improved the reconstruction of temporal trends. In contrast, RMSE and log-likelihood reach their optimal values

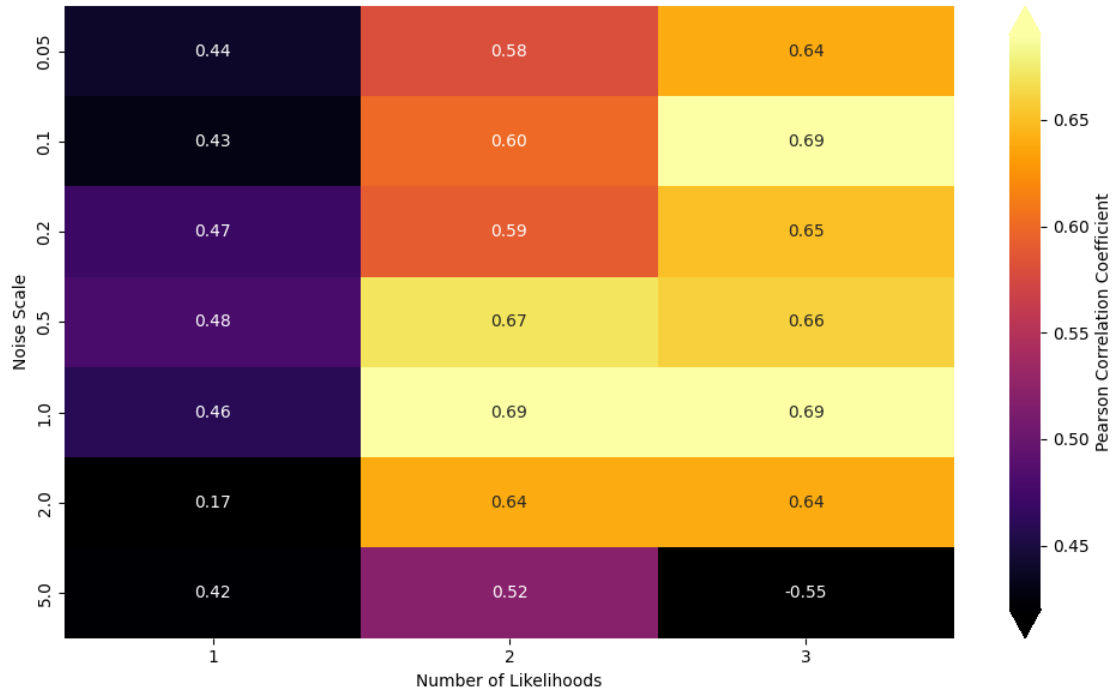


Figure 5.13: Pearson correlation of reconstructed ground surface temperatures for different combinations of noise scales and numbers of likelihoods. Values close to 1 indicate a strong linear agreement in both phase and amplitude, while lower values indicate weaker agreement.

primarily at unrealistically large noise scales, which exceed the actual sensor uncertainty. This discrepancy suggests that improved numerical scores at high noise levels reflect compensation for structural model limitations rather than increased physical accuracy. The implications of this model–data mismatch are discussed in Section 6.2.6.

Figure 5.14 illustrates the effect of incorporating multiple independent likelihoods into the inversion and provides a direct comparison between using three non-consecutive likelihoods and using only a single likelihood, both combined with the ERA5 + n-factor prior. Both approaches capture the overall trend of the observed GST. However, qualitative comparison to ground truth shows, that the 3 likelihood results yields a clearly improved reconstruction, showing a slightly smaller absolute deviations and a clear higher temporal correlation with the observations, especially for the years between 2012-2016. This indicates that incorporating multiple likelihoods effectively extends the time interval over which the GST reconstruction remains data-constrained rather than prior-dominated. These results suggest that including several non-consecutive likelihoods can produce a closer qualitative alignment and improve the internal consistency of the inversion.

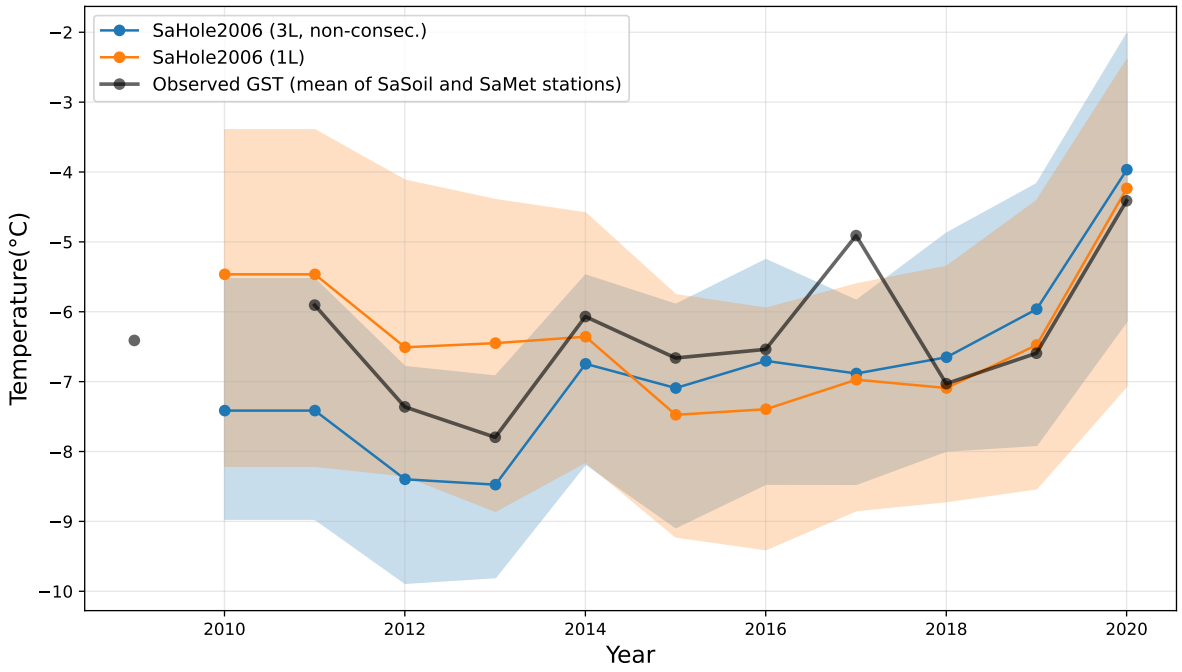


Figure 5.14: Direct comparison of SaHole2006 GST reconstructions using one likelihood (1L) and three non-consecutive likelihoods (3L). Solid lines represent posterior means, and shaded areas denote 95 % credible intervals. The 3L configuration produces smoother and visually more consistent reconstructions, while maintaining uncertainty ranges comparable to the 1L case.

5.3.3 Consecutive Likelihoods

After analysing the general effect of multiple likelihoods across independent years, the next section focuses on reconstructions that use consecutive years as likelihoods and therefore share a stronger thermal memory in the ground temperature signal. This allows examining whether the temporal continuity between observation years provides additional constraints that improve the inversion, or if the strong correlation between consecutive profiles leads to redundant information and a reduced benefit from adding more likelihoods.

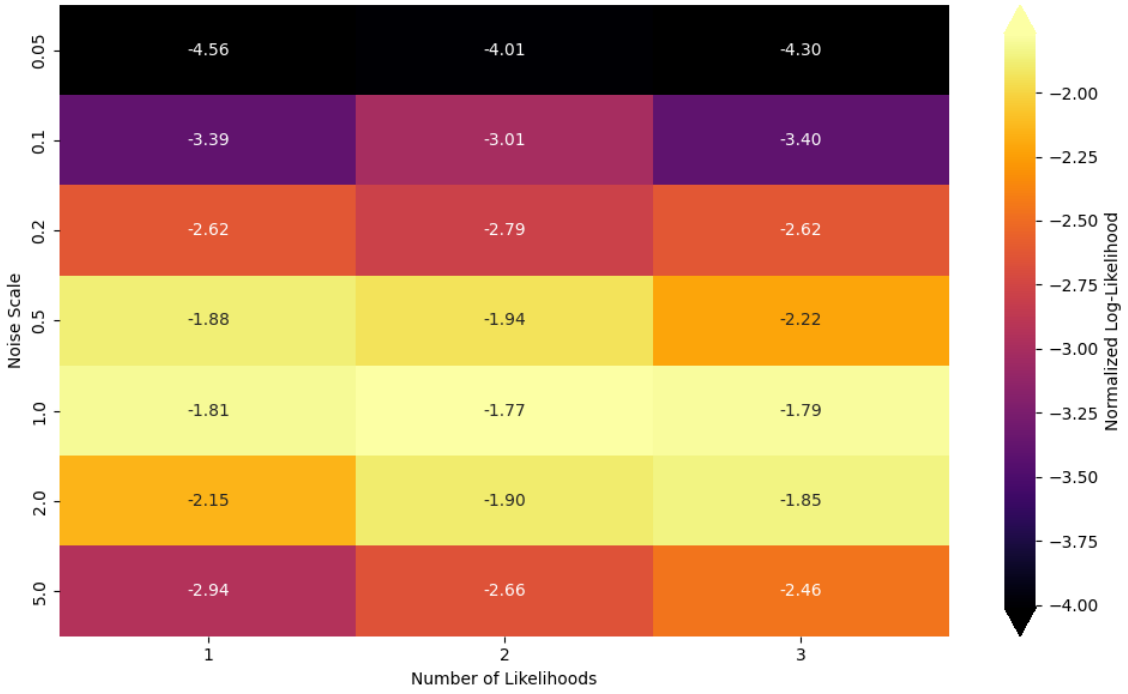


Figure 5.15: Heatmap showing the log-likelihood values as a function of noise scale and the number of consecutive likelihoods. Higher (less negative) values indicate a improved internal consistency between reconstructed and observed borehole temperatures. The results reveal that the best performance is obtained with two likelihoods in the moderate noise range between 1.0 and 2.0, whereas adding a third likelihood provides little or no further improvement.

Figure 5.15 illustrates how the log-likelihood changes with both noise scale and the number of consecutive likelihoods. Similar to the non-consecutive setup, the inversion achieved its best log-likelihood score at moderate noise levels between 1.0 and 2.0. However, unlike the previous case, the configuration with two likelihoods provides the highest log-likelihood value with -1.77, while the three-likelihood setup did not lead to a higher log-likelihood. This result suggests that consecutive years carry highly correlated temperature information, so the benefit of adding further likelihoods saturates much earlier. In this case, two likelihoods already capture most of the usable information, and additional profiles provide only marginal improvement or even slightly reduce the internal model fit due to redundancy, over-constraining effects. At very small noise scales, this leads to a strong decrease in log-likelihood values,

which is consistent with the model–data mismatch effects discussed in the previous section.

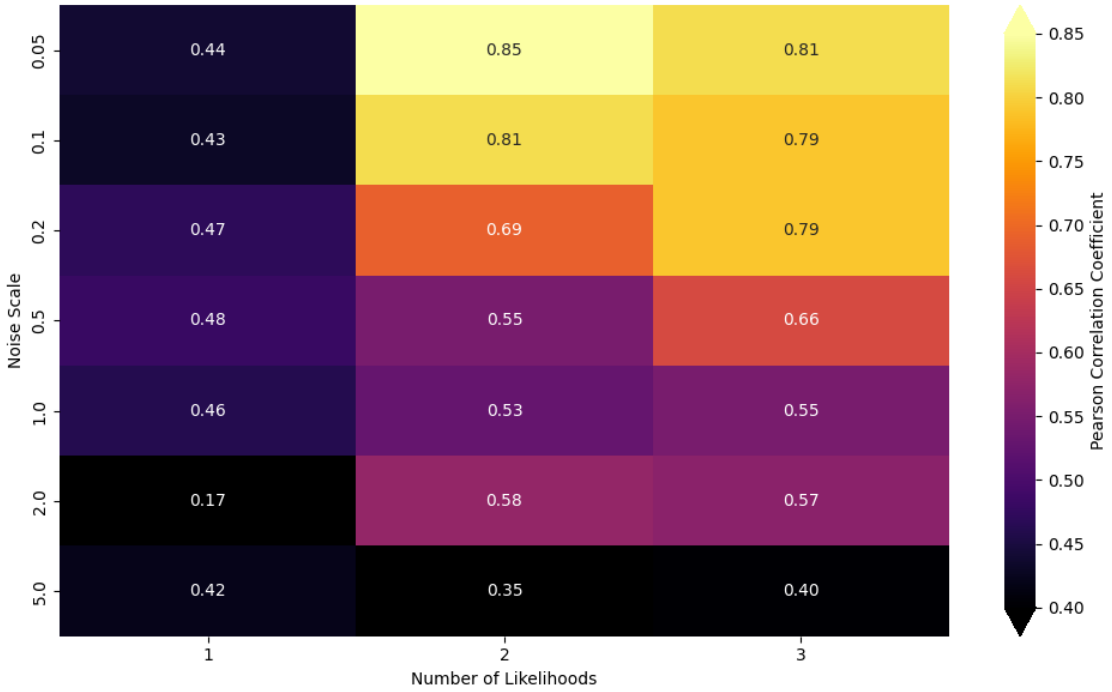
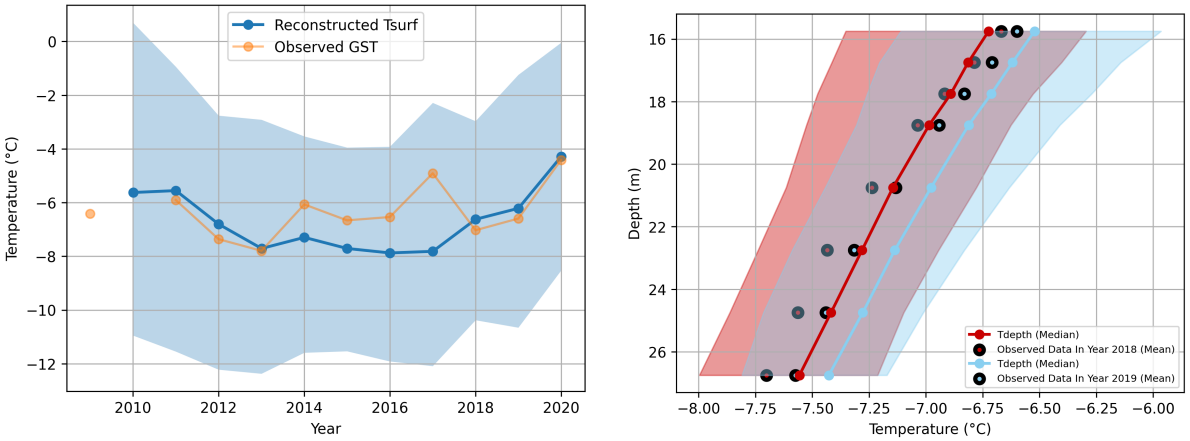


Figure 5.16: Root Mean Square Error of reconstructed ground surface temperatures for different combinations of noise scales and numbers of likelihoods. Lower RMSE values indicate a better agreement between reconstructed and observed GST. The heat map highlights how noise assumptions and the number of likelihoods influence reconstruction accuracy.

The Pearson correlation analysis in Figure 5.16 shows how well the reconstructed GST matches the observed GST in time, focusing on the temporal pattern. For the single-likelihood case, the correlation values remain moderate across all noise scales (around 0.43–0.48). Using two likelihoods leads to the highest correlation values. In particular, for small to moderate noise scales (0.05–0.2), correlation values reach up to 0.85 and remain relatively high (about 0.55–0.69) for larger noise scales. This indicates that two likelihoods already improve the reconstruction in terms of the temporal pattern. For three likelihoods, the correlation values are comparable but not consistently higher. While high correlations are also observed at small noise scales, the improvement compared to two likelihoods is limited. One possible explanation is that the third likelihood year contains anomalous features that are not fully consistent with the other profiles. When propagated through the simplified forward model, these anomalies can reduce the overall temporal agreement. However, on average, the three-likelihood configuration remains relatively stable in terms of temporal pattern at higher noise scales.
Figure 5.17 therefore presents the corresponding reconstruction result. Compared to the single-likelihood case, the reconstruction captures the temporal variation of the surface temperature visually more clearly and provides an improved match to the short observed GST data. The reconstructed borehole temperature profiles in Figure 5.17b also align closely

with the measured values for both time periods, indicating that the inversion successfully reproduces the observed subsurface thermal gradients.



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.17: Reconstruction using two consecutive likelihoods with the data-driven prior. (a) Reconstructed ground surface temperature history (T_{surf}) and (b) corresponding reconstructed and measured borehole temperature profile for Samoylov.

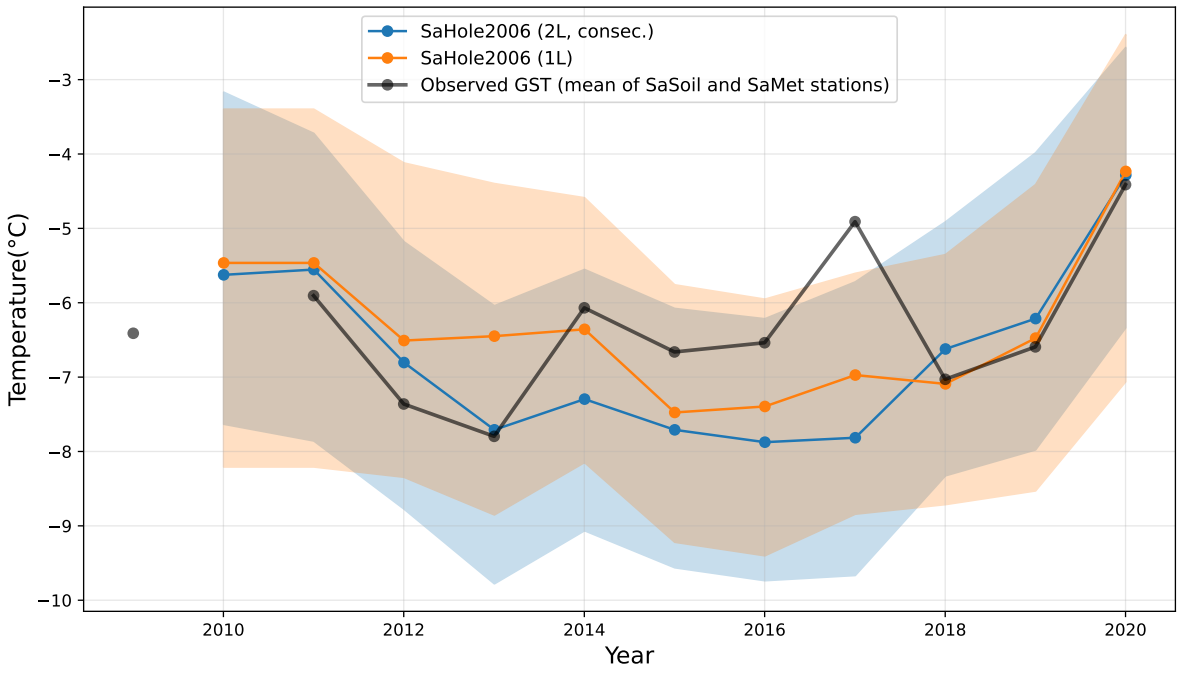


Figure 5.18: Comparison of SaHole2006 GST reconstructions using one likelihood (1L) and two consecutive likelihoods (2L). Solid lines show the mean reconstruction, and shaded areas show the 95% uncertainty range.The two-likelihood case looks very similar to the one-likelihood case.

The uncertainty range (blue shading) in Figure 5.17b remains moderate, showing that the inversion is well constrained without being overly confident.
The figure 5.18 shows that adding a second consecutive likelihood does not change the reconstruction much. The curves almost overlap, and there is no clear visual improvement or worsening compared to the inversion with 1 likelihood and ERA5 + n-factor. However, the heatmap in Figure 5.15 shows that the reconstruction error becomes slightly smaller when using two consecutive likelihoods. This means the second likelihood still adds a little information, even if the improvement is very small.

5.4 Depth Analysis and Recoverable Past

The ability to reconstruct past ground surface temperature histories from borehole data is fundamentally limited by the physical properties of heat diffusion in the subsurface. Temperature signals at the surface gradually attenuate and smooth out with depth, so that deeper boreholes contain longer but more diffused records of past climate variations, while shallow boreholes retain only the most recent temperature changes. In addition, the number of available likelihoods, representing conditionally independent observation periods, determines how much information can be extracted from the data.

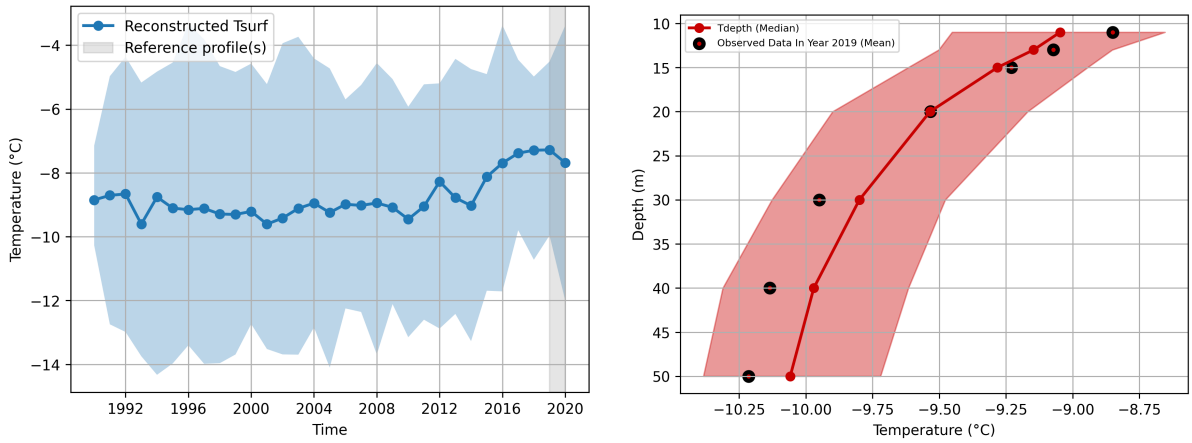
To explore these relationships, two complementary experiments were conducted. First, the Sardakh borehole was systematically truncated from its original 100 m depth to 50 m, 30 m, and 20 m, in order to examine how borehole depth influences the recoverable past under otherwise identical inversion settings. This analysis provides a direct view of how thermal memory diminishes with decreasing depth and whether the use of multiple likelihoods can partially compensate for the information loss.

In the second part, the Tiksi borehole was used to test the temporal reach of the method itself. Here, the inversion was performed using 1, 4, and 6 consecutive likelihoods to investigate how far back in time reliable reconstructions can be achieved before the GST prior starts to dominate the solution.

In both depth experiments, only consecutive likelihoods were used, since the available profiles at Sardakh come from neighboring years. This allows testing whether additional observational information can partly compensate for the loss of depth, while still remaining within the physical limits of the borehole data. By using consecutive years, the added information remains strongly correlated and realistic, ensuring that any improvement reflects the true recoverable signal.

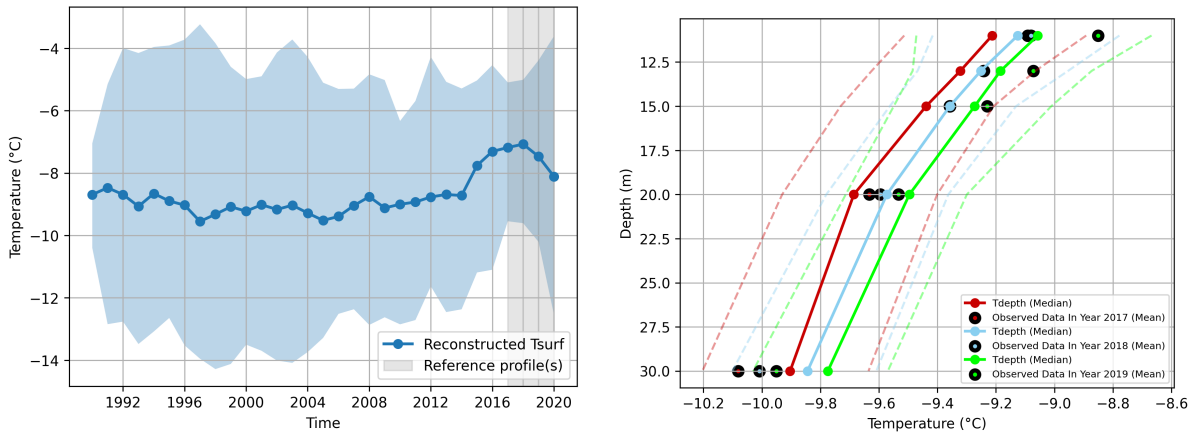
Together, these two experiments quantify both the physical and methodological limits of the proposed approach and illustrate how borehole depth and data redundancy jointly control the amount of recoverable climatic information. The following results illustrate how varying borehole depth affects the reconstruction of past surface temperatures. For each depth level, the inversion was performed using both single and multiple likelihood configurations to reveal how additional observational constraints can influence the recoverable past and overall reconstruction stability. Figure 5.19 shows the reconstruction result for the 50 m borehole using a single likelihood. This configuration serves as the reference for the following depth tests, as it provides the most complete thermal record and a well-constrained inversion result. When the borehole depth is reduced to 30 m but three likelihoods are used (Figure 5.20), the reconstruction remains smooth and shows a stable long-term evolution. The uncertainty range remains relatively the same as Figure 5.19, and the reconstructed profile fits the observed temperatures well across the full depth in Figure 5.20b. The 30 m reconstruction using a single likelihood (Figure 5.21) appears broadly similar to the deeper and multi-likelihood cases in terms of its mean temperature structure. However, the posterior uncertainty is noticeably larger, both in the surface temperature reconstruction and in the borehole profile. This reflects the reduced thermal memory of the shallower borehole and the limited observational constraint when only a single likelihood year is used.

A more detailed comparison between the three configurations (50 m with one likelihood, 30 m with three likelihoods, and 30 m with one likelihood) is provided in Figure 5.22.



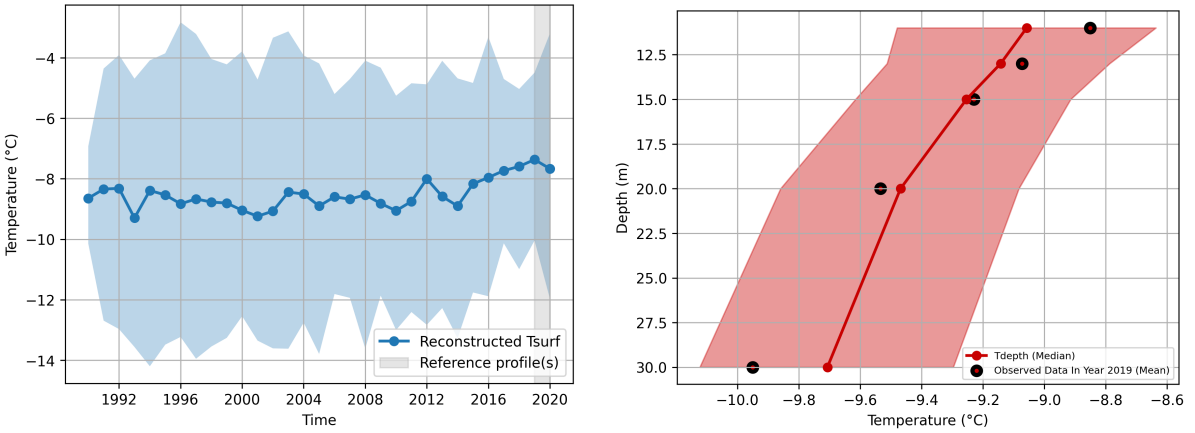
(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.19: Reconstruction results for the 50 m Sardakh borehole using a single likelihood. (a) Reconstructed ground surface temperature (T_{surf}) history and (b) corresponding modeled and observed borehole temperature profile. This configuration serves as the reference case, as the greater depth provides a more complete record of past thermal variations and thus represents the physically most constrained reconstruction.



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.20: Reconstruction results for the 30 m Sardakh borehole using three likelihoods. (a) Reconstructed ground surface temperature (T_{surf}) history and (b) corresponding modeled and observed borehole temperature profile.



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.21: Reconstruction results for the 30 m Sardakh borehole using a single likelihood. (a) Reconstructed ground surface temperature (T_{surf}) history and (b) corresponding modeled and observed borehole temperature profile.

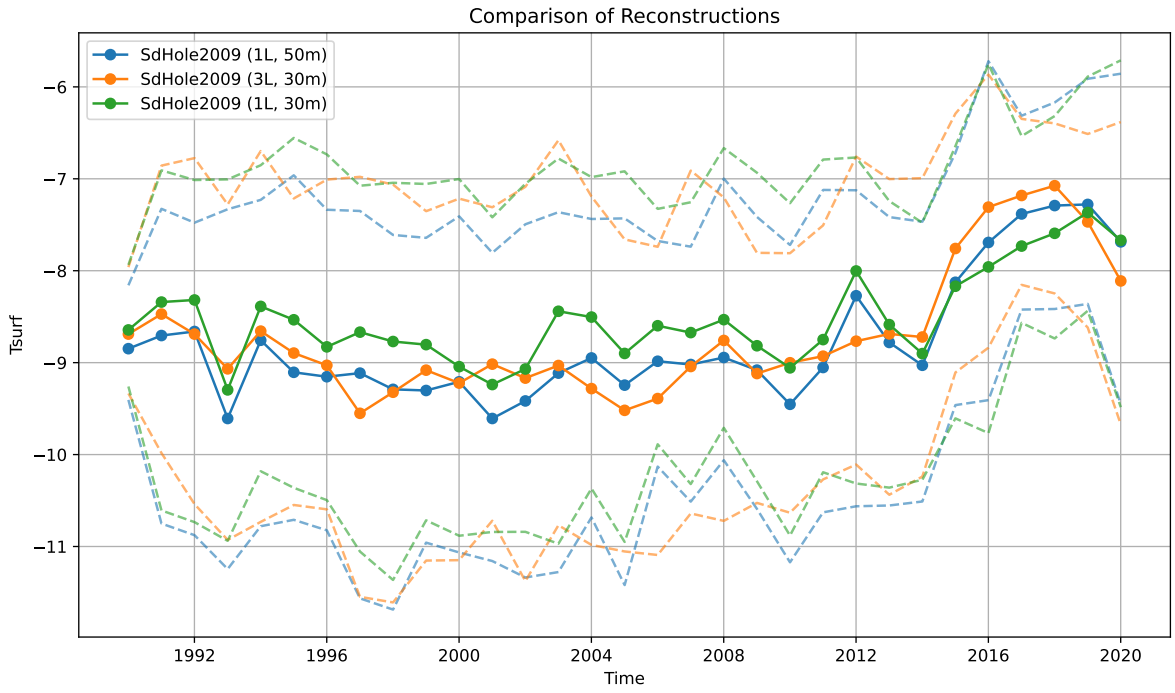


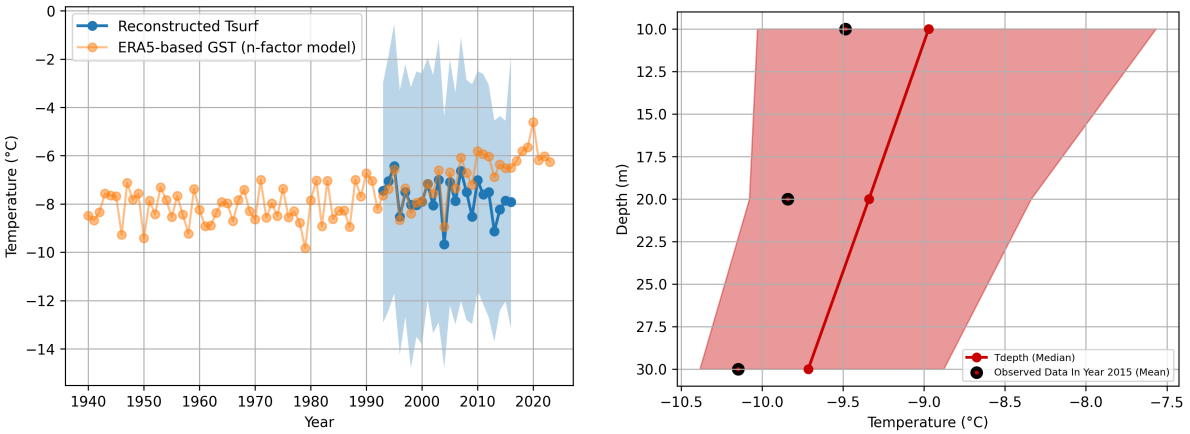
Figure 5.22: Direct comparison of SdHole2009 GST reconstructions using one likelihood at 50 m depth, three likelihoods at 30 m depth, and one likelihood at 30 m depth.Solid lines show mean reconstructions, dashed lines show the 95% uncertainty limits

Figure 5.22 provides a direct comparison of the three Sardakh reconstructions and shows how depth and the number of likelihoods influence the recovered GST history. The 30 m reconstruction with three likelihoods (orange) aligns most closely with the 50 m reference case (blue) in terms of its mean structure. In contrast, the 30 m single-likelihood reconstruction

(green) deviates more strongly in the mean and shows a generally higher temperature level. Interestingly, the year-to-year variability behaves differently: the 30 m single-likelihood curve (Figure 5.21) resembles the 50 m reconstruction (Figure 5.19) more closely than the 30 m three-likelihood case. One possible explanation is that both 1L configurations are dominated by the same prior and a single borehole profile, so their interannual variability is shaped by the same combination of prior structure and data. The uncertainty ranges follow the expected pattern: the 50 m reconstruction shows the narrowest band, the 30 m three-likelihood case remains moderately constrained, and the 30 m single-likelihood reconstruction exhibits the widest uncertainty envelope, as discussed before.

5.4.1 Validation using the Tiksi borehole

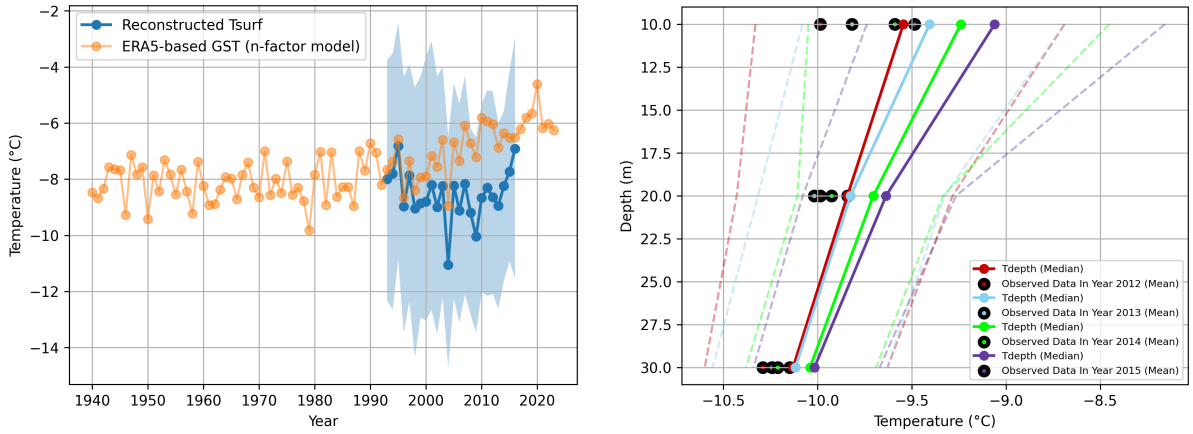
To further examine the temporal behavior of the Multiple Likelihood approach under long-term observational conditions, a complementary experiment was conducted using the Tiksi borehole. In contrast to the deeper Sardakh site, Tiksi is relatively shallow (approximately 30 m) but provides a much longer temperature record, covering the period from 1993 to 2016. This makes it particularly well-suited for evaluating how far back in time the inversion remains sensitive to observational information before the prior begins to dominate.
For this analysis, reconstructions were performed using between one and six likelihoods, each representing independent time windows within the available data period. In the main text, the cases with one, four, and six likelihoods are presented as representative examples. All inversions were performed using the same data-driven GST prior derived from local ERA5 air temperatures for Tiksi, combined with n-factors. This setup allows investigating how strongly the likelihood influences the reconstruction. To illustrate this effect, the prior was plotted together with each reconstructed surface temperature curve in the figures.



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.23: Reconstruction results for the Tiksi borehole using a single likelihood. (a) Reconstructed ground surface temperature (T_{surf}) history and (b) corresponding modeled and observed borehole temperature profile. In this configuration, the reconstruction begins to follow the GST prior after roughly eight years, indicating an early onset of prior dominance. The inversion captured only the most recent surface variations, reflecting the limited temporal sensitivity imposed by both the shallow depth and the single data constraint.

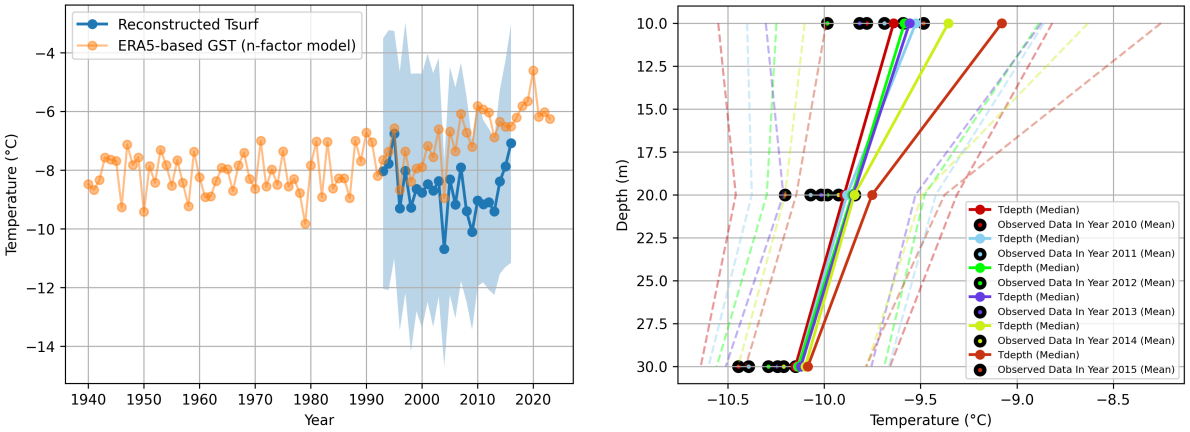
Figure 5.23 shows the reconstruction using a single likelihood. The inversion captured the most recent surface variability but quickly aligns with the GST prior. It can be seen that the likelihood contributed to the reconstruction for only about eight years, until the year 2008, indicating that the information from a single observation period is not sufficient to constrain earlier parts of the temperature history. The recoverable signal length is therefore short, reflecting the limited thermal memory contained in one borehole profile, especially for shallow boreholes. By increasing the number of likelihoods to four (Figure 5.24), the recoverable past extended noticeably further. The reconstructed T_{surf} remains distinct from the GST prior to the year 1998, covering a period of approximately 19 years. As we can see, the



(a) Reconstructed surface temperature history (T_{surf}). **(b)** Predicted and measured borehole profile.

Figure 5.24: Reconstruction results for the Tiksi borehole using four consecutive likelihoods. (a) Reconstructed ground surface temperature (T_{surf}) history and (b) corresponding modeled and observed borehole temperature profile. Increasing the number of likelihoods extends the recoverable past and strengthens the data influence without overfitting. The reconstruction diverges from the GST prior for around 19 years, after which it gradually converges back, demonstrating that the available information in the borehole data has been effectively exploited.

new multiple-likelihood approach helps to extract more information from the borehole data. With only three additional years of likelihood, it was possible to reconstruct approximately 11 more years of temperature history. This clearly showed that including multiple likelihoods enhances the temporal reach of the inversion while still maintaining physical plausibility and avoiding overfitting. Using six likelihoods (Figure 5.25), the inversion no longer yielded significant improvements. The reconstructed GST curve becomes almost identical to the four-likelihood result, demonstrating that the inversion has reached a saturation point.
Overall, the Tiksi experiment shows that while multiple likelihoods can extend the recoverable past and improve the stability of the inversion, this effect is inherently limited by the physics of heat diffusion. After a certain point, additional likelihoods no longer contribute meaningful independent information, marking the effective boundary of the method's temporal sensitivity. The figure 5.26 shows a direct comparison between the 1-likelihood and the 4-likelihood reconstruction for Tiksi. We can clearly see that the 1-likelihood case starts to follow the prior much earlier. In fact, the prior dominance begins about 11 years earlier in the 1-likelihood reconstruction than in the 4-likelihood case. With four likelihoods, the reconstruction keeps useful information from the borehole for a much longer period and stays independent from the prior for almost two decades. This highlights how adding multiple likelihoods delays prior dominance and increases the amount of recoverable past information.



(a) Reconstructed surface temperature history (T_{surf}).

(b) Predicted and measured borehole profile.

Figure 5.25: Reconstruction results for the Tiksi borehole using six consecutive likelihoods. (a) Reconstructed ground surface temperature (T_{surf}) history and (b) corresponding modeled and observed borehole temperature profile. The higher number of likelihoods does not further improve the fit, indicating that the reconstruction has reached a saturation point. Beyond approximately 10–12 years before the most recent profile, the reconstructed temperatures converge again towards the GST prior, confirming that the physical information content of the borehole has been fully utilized.

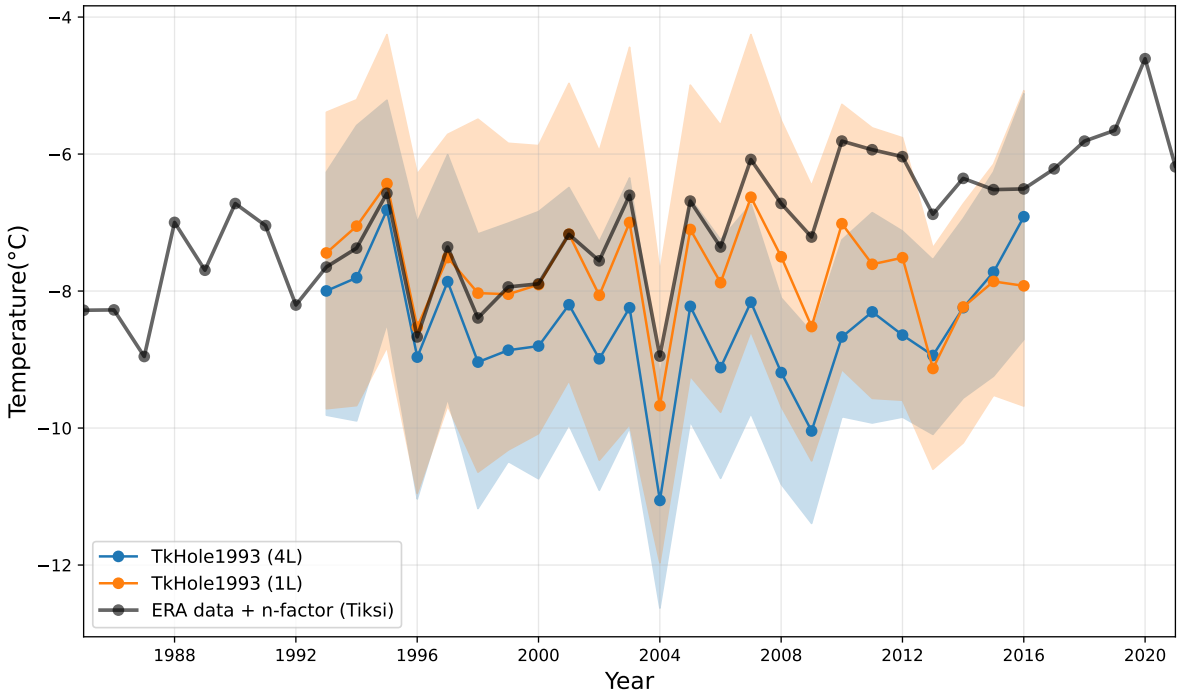


Figure 5.26: Direct comparison of TkHole1993 reconstructions using one likelihood and four likelihoods. The shaded regions show the 95% uncertainty range for each reconstruction.The 1-likelihood reconstruction aligns with the GST prior much earlier, while the 4-likelihood reconstruction remains data-driven for a longer period before gradually converging toward the prior.

5.4.2 Sensitivity to Initial Profile Initialization

To assess the influence of subsurface initialization on the reconstructed ground surface temperature (GST), a sensitivity analysis was performed using three alternative initialization schemes:

1. a steady-state extension assuming a constant geothermal gradient
2. a vertically constant continuation of the deepest measured temperature (default initialization)
3. an interpolated continuation based on a deeper neighboring borehole

All reconstructions were carried out using three likelihood terms, the ERA5 + n-factor prior, and the exclusion of the upper 15 m of the borehole profile, consistent with the main reconstruction setup. In contrast to the previous likelihood-comparison experiments, a fixed noise_scale of 0.1 was applied to reflect realistic sensor uncertainty and to ensure strong data constraint. The purpose of this analysis is to evaluate the robustness of the inversion with respect to different initialization assumptions.

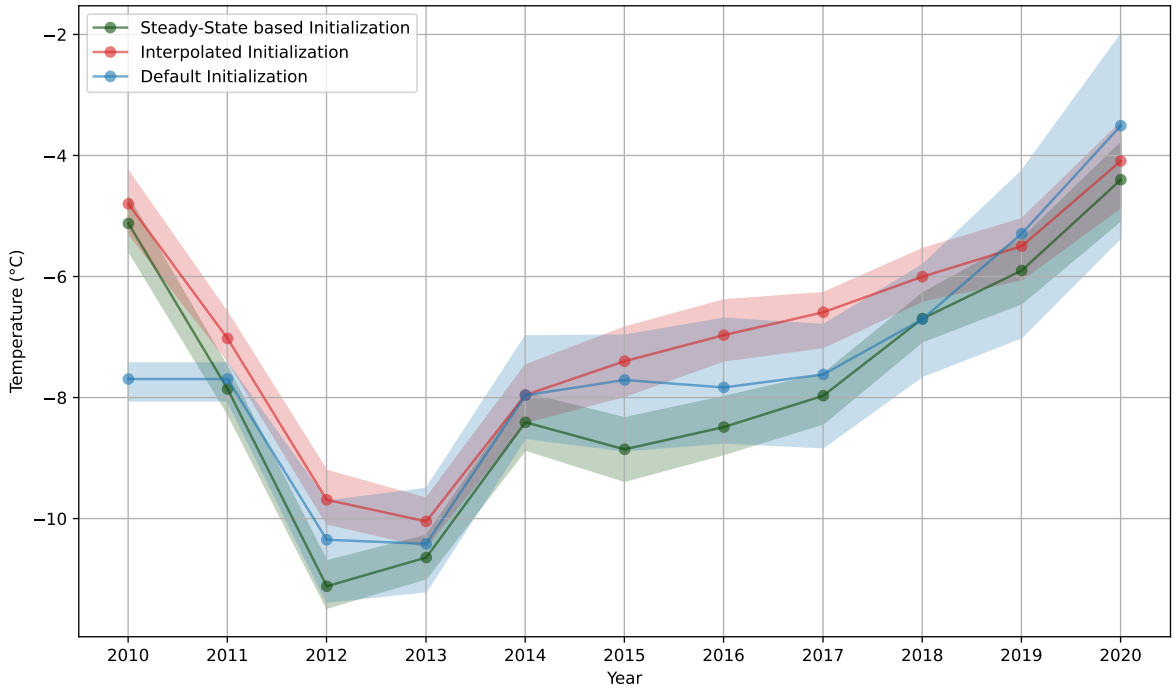


Figure 5.27: Comparison of reconstructed ground surface temperature (GST) histories obtained with three different subsurface initialization schemes: steady-state extension (green), interpolated continuation (red), and default vertically constant continuation (blue). Shaded regions indicate the 95% posterior uncertainty range for each reconstruction. All reconstructions were performed using three likelihood terms, the ERA5+n-factor prior, and a fixed noise_scale of 0.1. The temporal evolution and uncertainty ranges show strong overlap across configurations, indicating limited sensitivity of the reconstructed GST to the chosen initialization scheme.

Figure 5.27 compares the reconstructed GST histories obtained with the three initialization schemes. Despite the differences in the assumed initial subsurface temperature profiles, the reconstructed surface temperature histories remain highly consistent across all configurations. The temporal structure and interannual variability are nearly identical across setups. However, the posterior uncertainty is noticeably higher for the default initialization compared to the steady-state configuration.
Most reconstruction experiments in this study were conducted using the default initialization scheme to ensure consistency across sensitivity analyses. Small differences in absolute temperature level are visible in some years; however, these deviations remain minor compared to the posterior uncertainty range. Overall, the results indicate that the reconstruction is relatively insensitive to the specific choice of subsurface initialization within the tested configurations, at least for short-time reconstructions. Nevertheless, given the slightly reduced posterior uncertainty observed for the steady-state initialization, this configuration may be preferable in future applications where minimizing uncertainty is of particular importance.

5.5 Long Term Reconstructions with Methodological Extensions

This final results section provides a visual and qualitative comparison of the long-term GST reconstructions before and after applying the methodological extensions. The goal of this section is not to validate long-term climate trends, but to demonstrate how the methodological extensions behave under realistic data constraints. The two plots shown in this chapter summarize the combined effect of the extended likelihood structure, the data-driven prior, and the hybrid time grid.

5.5.1 Baseline Reconstructions

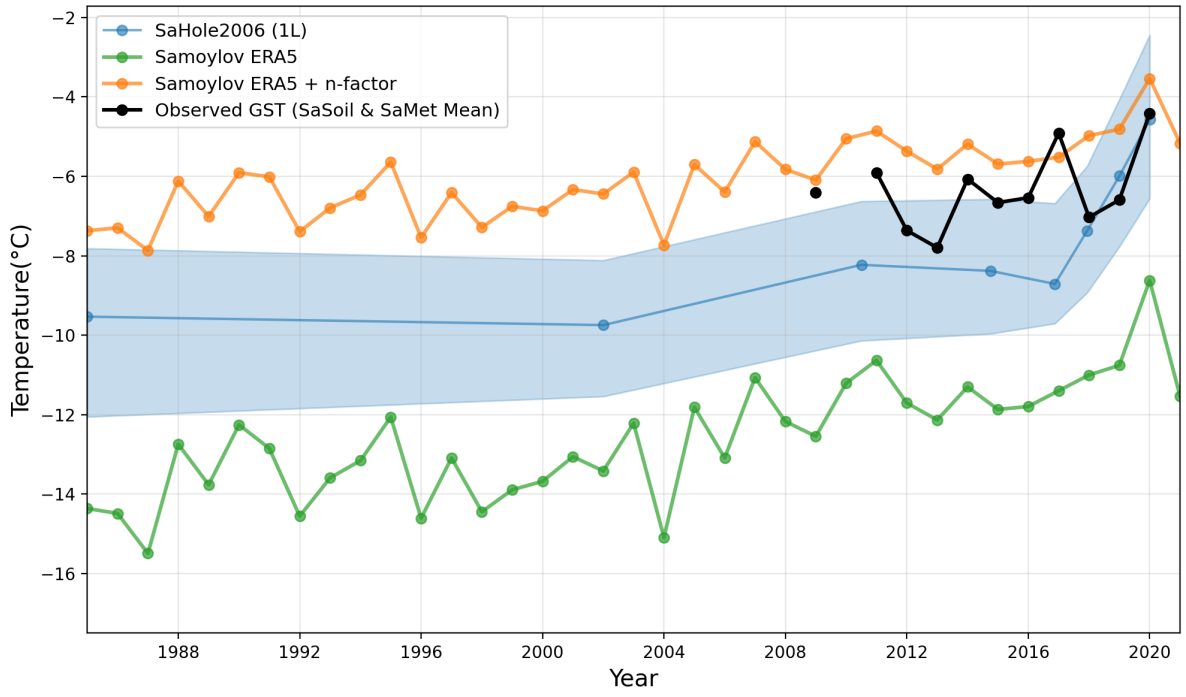


Figure 5.28: Baseline GST reconstruction for Samoylov (SaHole2006) using one likelihood year (1L). The figure shows the reconstructed GST (posterior mean and 95% credible interval) together with observed GST (2010–2020), ERA5 air temperature (1985–2020), and the ERA5+n-factor GST estimate. The baseline setup reproduces the large-scale cooling–warming pattern but remains strongly smoothed and exhibits wide uncertainties, especially during the recent warming period.

The baseline reconstruction (Figure 5.28) serves as a reference for evaluating the effect of the methodological extensions. Using only a single likelihood year and a simple prior, the reconstruction reproduced the general cooling and warming trend but lacks sensitivity to shorter-term variations. In the overlap period with observed GST (2010-2020), the reconstruction follows the broad trend but remains noticeably smoother and more uncertain. This

behavior reflects the limited information content of a single likelihood, as well as the absence of climatological structure in the prior.

5.5.2 Extended Reconstructions

The extended reconstructions combine two key methodological components:

1. a data-driven prior based on ERA5 air temperatures transformed with local *n*-factors
2. three conditionally independent likelihood years

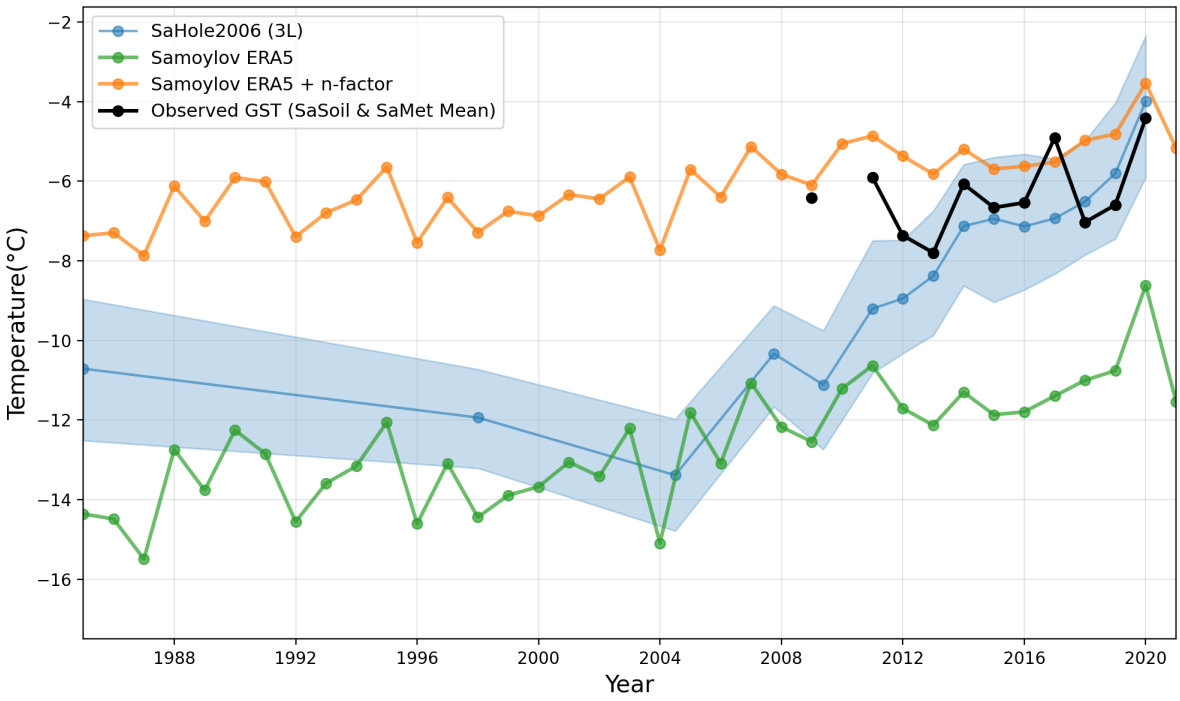


Figure 5.29: Extended GST reconstruction for Samoylov (SaHole2006) using three non-consecutive likelihood years (3L), a data-driven ERA5+*n*-factor prior, and a hybrid time grid. Compared to the baseline case, the reconstruction follows the observed GST more closely in 2010–2020, resolves the recent warming more clearly but shows only slightly reduced uncertainties. A pronounced cooling feature is visible in the late 1990s to early 2000s.

The extended reconstructions are shown in Figure 5.29. The effect of the methodological extensions is most pronounced in the period during which ground-truth GST measurements exist. Compared to the baseline, the reconstructed GST more closely followed the observed interannual variations, and the warming after approximately the early 2000s is captured more distinctly. However, the reduction in uncertainty remains modest. Additionally, between approximately 1996 and 2005, the reconstruction exhibits a cooling feature. This period is not directly constrained by observational data and coincides with a phase dominated by forward model propagation. Although no ground-truth GST observations are available for this interval, such a strong and isolated cooling is not supported by climate indicators, such as the ERA5 air

temperature, which show no comparable anomaly. Therefore, this cooling period between 1996 and 2005 likely reflects a model–data mismatch arising from the interaction between the simplified forward model and the multi-profile inversion setup, rather than a physically meaningful climate signal.
The improvement arose from the complementary effects of the methodological extensions:

- The **ERA5+*n*-factor prior** provides climatologically plausible year-to-year variability in the annual grid, supplying climatological structure where the borehole lacks information.
- **Multiple likelihoods** increase the effective information content by incorporating several independent borehole profiles, which delays the onset of prior dominance, while the reconstruction remains limited by model assumptions.

Hybrid Temporal Discretization

The reconstruction in Chapter 5.5.2 also uses a hybrid time grid, in which the temporal resolution is adapted over time. The most important time period is the section that contains the likelihood years, because this part of the reconstruction is strongly supported by the data. For this reason, annual time nodes are used until the last likelihood year. After that point, the grid becomes coarser and uses geometrically spaced time nodes.
This hybrid structure avoids unnecessary degrees of freedom in the older parts of the reconstruction. The deeper past is strongly smoothed by heat diffusion, and the inverse problem becomes more ill-posed as it goes further back in time. A fine grid would not be supported by the available information and could lead to unstable results.

6 Discussion

6.1 Summary of Key Findings

The results demonstrate that reconstruction quality is controlled by the interaction between prior design, the number of likelihoods, and borehole depth.
First, the data-driven ERA5 + n-factor prior improves the physical plausibility and temporal structure of the reconstructed GST histories. By anchoring the prior in site-specific climatology, it introduces realistic interannual variability and prevents implausible long-term drift. The quantitative gains in RMSE and log-likelihood remain modest, indicating that climatologically informed priors primarily enhance qualitative realism rather than systematically reducing numerical error metrics. However, the main improvement lies in enhancing the stability and temporal consistency of the reconstructed GST histories.
Second, incorporating multiple likelihoods improves temporal consistency and stabilizes the reconstruction and temporal pattern, particularly when the included profiles provide sufficiently independent information. At Tiksi, this extended the recoverable past by approximately one decade. However, the benefit saturates once additional likelihoods become strongly correlated, as observed for consecutive years. Under very low noise assumptions, redundant likelihoods can even degrade reconstruction quality, reflecting sensitivity to model–data mismatch.
Third, borehole depth remains the dominant physical constraint. Deeper profiles retain longer thermal memory, while shallow boreholes quickly lose climatic information and become prior-dominated. Multiple likelihoods can partly compensate for reduced depth by extending the data-constrained interval, but they cannot overcome the fundamental limits imposed by heat diffusion.
Overall, the findings show that stable and physically consistent GST reconstructions require a balanced combination of informative but non-dominant priors, realistically chosen noise assumptions, and likelihood terms that contribute complementary information. Within these constraints, the method enhances temporal stability while making the physical limits of the available information transparent.

6.2 Interpretation and Limitations

The results of this study show how the interaction between prior information, the number of likelihoods, and the borehole depth jointly determines the stability and quality of ground surface temperature reconstructions. Each factor affects the inversion in a different, yet interconnected, way.

6.2.1 Effect of the data-driven prior

The ERA5 + *n*-factor prior generally improves the physical consistency of the reconstructions by providing boundary conditions that approximate local processes such as snow insulation and vegetation. it defines a realistic climatological framework within which the posterior can evolve, preventing the reconstruction from drifting toward implausible temperature trends, especially when observational data are sparse. In addition, this type of prior is particularly beneficial when higher temporal resolution is desired, such as yearly nodes, because ERA data also provide yearly climate information. In such cases, the data-driven prior supports a smoother and more realistic year-to-year structure, which would be harder to achieve with a purely heuristic prior.
However, the influence of this prior is not always beneficial, if the prior uncertainty is too small or if ERA5 data contain local biases, the posterior can become overly constrained and too similar to the climatological mean. The results therefore highlight that the prior should be informative but not dominant—broad enough to allow variability, yet constrained enough to ensure physical plausibility.
A further limitation arises from the use of the *n*-factor approach itself. The *n*-factors, which are used to approximate ground surface temperatures from air temperatures, are often derived from literature values or empirical estimates based on local vegetation and snow conditions. Obtaining accurate and site-specific *n*-factors is challenging, as they vary across terrain types and environmental settings. Moreover, *n*-factors are not constant over time, and changes in vegetation cover, snow depth, or surface moisture due to climate change can alter the ground–air temperature relationship.
If such temporal variations are ignored, or if incorrect assumptions about local vegetation and surface properties are made, the resulting *n*-factors may be inaccurate. This can introduce systematic biases into the GST prior, leading to reconstructions that are either too warm or too cold relative to reality.
For completeness, it should be noted that a sensitivity analysis of the ERA5 + n-factor prior itself, particularly with respect to changes in prior uncertainty would be scientifically relevant. Such an analysis could reveal how strongly the reconstruction depends on the assumed prior variance. However, this would require additional experiments beyond the main methodological focus of this thesis and is therefore not included. Future work should address this aspect to better quantify the interplay between prior strength and data support.

6.2.2 Role of multiple likelihoods

Using multiple likelihoods allows the inversion to combine information from different observation periods. This generally slightly strengthens the reconstruction because the model can better distinguish persistent long-term signals from year-specific noise (Figure 5.10). This effect could especially be confirmed qualitatively. In the non-consecutive setup, the improvements were more clearly visible compared to the consecutive method: for the short period where ground truth exists, the reconstructions were closer to the observed GST, especially at moderate noise levels (1–2).
However, the results also show important limitations. At very low noise scales, adding more likelihoods reduced reconstruction quality due to systematic model bias. In this regime, the

inversion tries to match small annual fluctuations too closely, even though these fluctuations partly reflect measurement noise, instrument drift, or short-term surface processes that are not meaningful indicators of long-term climatic change. This leads to reconstructed GST histories that appear confident but drift away from physically plausible behaviour, as confirmed by the likelihood maps in the Figure 5.10.

The benefit further decreases when consecutive likelihoods are used. Profiles from adjacent years share similar seasonal patterns and long-term trends, which means their information content is strongly correlated. Adding such years increases redundancy instead of expanding the effective data support. In the experiments, two consecutive likelihoods still improved slightly the reconstruction (Figure 5.15), but additional years contributed little new information and in some cases inflated posterior confidence without improving the actual reconstruction quality as we have seen.

Taken together, the results show that the multiple likelihood method offers meaningful qualitative improvements, especially when the included likelihoods contribute information that is as independent as possible. While true independence cannot be achieved in a physically cumulative system like heat conduction, the assumption of conditional independence, as introduced earlier 4.2.2, remains valid. Noise assumptions must remain realistic, and the simplifications of the physical model must always be kept in mind (see Section 6.2.6).

The synthetic validation experiments provided additional information on the behaviour of the inversion framework. Under controlled conditions with identical stratigraphy and model configuration, the reconstructed GST signal was reproduced with good temporal agreement, and correlation increased consistently when additional likelihood terms were included.

However, even in the synthetic case a systematic temperature offset remained visible. Since the synthetic experiments were conducted with an identical forward-model structure and stratigraphy, a substantial structural model mismatch is unlikely. The remaining bias therefore most plausibly reflects regularization effects inherent to the inversion, limitations in the GST parameterization, or the ill-posed nature of the inverse heat conduction problem.

In addition, the functional parameterization of the GST history may contribute to this behavior. If the assumed GST formulation cannot fully represent the true temporal structure of the imposed signal, the inversion may compensate through a systematic temperature shift.

With respect to RQ1, the results provide conditional evidence that incorporating multiple likelihoods can improve GST reconstructions under specific noise assumptions, primarily by reducing posterior spread and stabilising temporal behaviour. In contrast, quantitative error metrics such as RMSE and normalized log-likelihood do not improve consistently for realistic noise scales. This indicates that, for the forward model used in this study, the main benefit of multiple likelihoods lies in improved temporal consistency and reduced posterior variability, rather than in a systematic reduction of reconstruction error.

Addressing RQ2, the experiments demonstrate that both the number and temporal distribution of likelihoods play a critical role. For consecutive likelihoods, the best internal consistency scores—such as Pearson correlation or normalized log-likelihood—were most often achieved with two likelihoods. Adding a third consecutive profile provided little additional benefit and in some cases reduced the quality of the reconstruction. In contrast, for non-consecutive likelihoods, the best scores were generally obtained with three likelihoods. In this case, the larger temporal separation between profiles increases the effective independence of the likelihood

terms, allowing the inversion to integrate more complementary information from different years.

6.2.3 Influence of borehole depth

Borehole depth remains the most fundamental physical constraint on reconstructing past surface temperatures. Because thermal diffusion dampens and delays older signals exponentially with depth, deeper boreholes preserve a longer memory of the surface temperature history. The depth experiments at Sardakh confirm this relationship: the 50 m case reproduced long-term variability more reliably, whereas reconstructions from 20–30 m depths lost the signal earlier and followed the prior more closely. Nonetheless, using multiple likelihoods partly compensated for this information loss in shallower boreholes, extending the recoverable past by about ten years without introducing instability. Still, the results underline that this compensation remains limited by the physics of heat conduction—depth cannot be fully replaced by additional data.
With respect to RQ3, the results show that borehole depth remains the dominant physical constraint on GST reconstructions. Deeper profiles retain long-term thermal memory, while shallow boreholes lose climatic information due to thermal diffusion. The multiple-likelihood formulation can partly compensate for reduced depth by allowing the likelihoods to constrain the reconstruction over a longer time window before the prior becomes dominant. However, it cannot replace the information content provided by deeper boreholes.

6.2.4 Validation and data limitations

Despite these consistent findings, the quantitative evaluation of the reconstruction performance remains constrained by the available data. Reliable ground surface temperature observations are limited, and where such datasets exist, they usually cover only short time spans or contain temporal gaps. As a result, it is not possible to rigorously validate the reconstructed long-term trends against continuous ground truth data.
Therefore, the improvements observed in this study—such as reduced spread and enhanced temporal consistency—should be interpreted primarily as qualitative indicators of methodological progress rather than as fully verified quantitative results.
Future research would benefit from longer and better-resolved GST time series, ideally co-located with borehole measurements, to enable a more comprehensive validation of the inversion framework under real-world conditions.

6.2.5 Validation of Sensitivity to Initial Profile Initialization

The sensitivity experiment shows that the short-time reconstructed GST histories are largely unaffected by the choice of subsurface initialization. Although the initial deep-temperature profiles differ substantially, the resulting surface temperature reconstructions exhibit nearly identical temporal behaviour and strongly overlapping uncertainty ranges. Therefore, the systematic temperature offset observed in the results remains present across all initialization schemes. This indicates that the discrepancy does not originate from the assumed initial temperature but from other aspects of the inversion framework.

Overall, the reconstruction appears robust with respect to reasonable variations in subsurface initialization.

6.2.6 Model limitations and noise assumptions

A fundamental limitation of the proposed extended inversion framework arises from the interaction between model simplifications, and the assumed observation noise scale. The forward model used in this study is a simplified representation of the thermal dynamics of the ground. While it captures the dominant conductive processes, it neglects or idealizes several important factors, such as snow dynamics, spatial variability in soil properties, radiative and turbulent heat fluxes, and vegetation dynamics [68]. As a result, the model output does not perfectly match reality.

The sensitivity experiments presented in Section 5.3.1 illustrate how these structural simplifications interact with the assumed observation noise scale. In particular, the pronounced year-to-year variability observed at very small noise scales does not necessarily reflect true climatic variability, but rather the inversion's attempt to satisfy mutually inconsistent constraints under strict likelihood assumptions. The limitations of the forward model become particularly apparent with multiple temperature profiles. When several temperature profiles from different years are assimilated, the reconstructed surface temperature history is sequentially propagated through the forward model from one likelihood year to the next. Due to the simplified representation by the model, the propagated temperature profile generally differs from the observed profile at the likelihood time. These differences arise consistently from structural simplifications in the forward model and reflect limitations of the forward model.

In summary, increasing the noise scale relaxes this constraint and allows the inversion to tolerate mismatches between propagated and observed profiles. In this case, the reconstructed temperature history becomes smoother and avoids extreme adjustments driven by individual likelihoods. However, this increased smoothness comes at the cost of larger posterior uncertainty. The smoother behaviour observed at noise_scale = 1.0 therefore does not necessarily imply improved physical realism, but rather reflects a relaxation of the likelihood constraint.

It is important to note that, in the current implementation, the noise scale mainly controls how strongly the likelihood constrains the inversion. However, using a larger noise scale is reducing the impact of model-data mismatches on the reconstruction, which partially masks the model limitations. Furthermore, it should be emphasized that noise scales in this range (0.5 °C to 5 °C) exceed the realistic measurement uncertainty of the sensors and therefore should not be interpreted as physically meaningful error estimates.

Addressing RQ4, the results show that the multiple-likelihood EKS reconstruction is highly sensitive to the assumed observation noise scale. Small noise values enforce strict consistency between the forward model and the likelihoods, which amplifies model–data mismatches and can lead to non-physical adjustments in the reconstructed GST history. Larger noise scales relax this constraint, resulting in smoother reconstructions but substantially increased posterior uncertainty. These findings indicate that reconstruction behaviour depends strongly on noise assumptions and that the noise scale influences how model–data mismatches are handled by the inversion. This highlights the need for future developments that explicitly separate measurement noise from model error.

6.3 Comparison with Previous Studies

The depth-dependent behavior observed in this study closely aligns with earlier findings on borehole-based GST reconstructions. Classical work by Lachenbruch and Marshall [69] showed that thermal diffusion acts as a strong low-pass filter, progressively damping high-frequency surface temperature variations with increasing depth. This means that shallow zones of the subsurface predominantly contain seasonal and short-term signals, whereas deeper layers preserve only long-term climatic trends. Beltrami [70] quantified this frequency–depth relationship and demonstrated that daily variations are detectable only within the upper meters (≈ 1 m), annual variations down to approximately 15 m to 20 m, and century-scale changes to depths of roughly 100 m. These physical limits explain why shallow boreholes lose climatic information rapidly, whereas deeper profiles retain a clearer imprint of long-term variations. Following this rationale, the upper 10 m to 15 m of all boreholes were excluded in this study, as these depths are strongly influenced by seasonal forcing and short-term fluctuations that are not relevant for recovering multi-decadal GST histories. This filtering step is consistent with previous inversion studies, which typically avoid near-surface layers to prevent the annual cycle from dominating the inversion and to enhance sensitivity to long-term climatic trends.

6.3.1 Data-driven GST priors versus heuristic priors

A key difference to earlier Bayesian-based borehole inversion studies [1, 8, 18, 34] is the explicit construction and validation of a data-driven GST prior. Previous work typically relied on mathematically convenient priors or educated guesses rather than climate-informed ones. For example, Woodbury and Ferguson (2006) used an empirical Bayesian approach in which the strength of the prior is estimated from the data itself, but the prior structure remains purely statistical and contains no physical climate information. Likewise, Hopcroft et al. (2007) used a trans-dimensional Bayesian inversion with a Gaussian prior centred on the present-day GST, and Hopcroft et al. (2023) expanded this framework with hierarchical hyperpriors, which remain statistical in nature and do not include explicit site-specific climate information. Although these methods are more flexible, these priors remain statistically defined and generally do not incorporate explicit local climate information.

Classical GST reconstructions therefore often rely on priors such as smoothness constraints or piecewise-constant temperature segments, sometimes combined with heuristic assumptions about long-term warming. These choices stabilize the inversion but do not necessarily represent realistic GST behavior at specific sites.

In contrast, the present study derives the prior for T_{surf} directly from ERA5 air temperatures using an n-factor model for air–ground coupling. This anchors the prior in observationally constrained local climatology and ensures that both its mean and variability reflect site-specific conditions. The validation experiment at Samoylov shows that this data-driven prior improves the physical consistency of the inversion especially qualitatively, because reconstructions become smoother while maintaining or slightly improving the fit to the borehole temperatures, as seen in marginally higher normalized log-likelihood values and lower RMSE at moderate noise scales. At the same time, the overall reconstruction accuracy remains comparable to the educated-guess prior, indicating that the ERA5+n-factor prior provides physically meaningful

regularization without introducing systematic bias.
However, this also means that reconstruction quality now depends partly on the accuracy of the reanalysis data and the chosen n-factor parameterization, which should be considered when comparing results across different sites and methods.

6.3.2 Multiple likelihoods and temporal information content

Most borehole inversion studies use only one temperature–depth profile per site. This is usually either a yearly mean or a snapshot from one specific year. When only one profile is used, the amount of temporal information is limited. Repeated measurements are then used mainly for quality control or validation, not as part of the inversion itself.
A study that is methodologically close to the present work is Guglielmin et al. (2018) [20]. They also used more than one temperature profile, but in a simpler way, because they combined the mean annual profiles from three consecutive years (2013, 2014, 2015) into a single averaged profile before running the inversion. This improves data quality but does not increase the temporal information content, because all three years are merged and treated as one dataset. In contrast, the multiple-likelihood method introduced here treats several annual profiles from the same borehole as separate and conditionally independent likelihood terms within a single Bayesian inversion. Instead of averaging the profiles first, each year contributes its own likelihood, which increases the amount of information available to the inversion.
The experiments show that multiple likelihoods provide modest but meaningful improvements in reconstruction stability, but only when the individual years contribute sufficiently different information. At Samoylov, non-consecutive years perform best because they differ slightly in their measurement noise and in their year-specific upper boundary conditions, which results in partially independent likelihood contributions. The small differences between the annual profiles allow the inversion to identify which parts of the signal represent long-term climatic trends and which are only short-term fluctuations or measurement noise. In contrast, consecutive years show only small improvements, as their temperature–depth profiles are highly correlated and therefore introduce little additional information beyond the first year.
The depth experiments and the Tiksi results further illustrate that adding more likelihoods does not overcome the physical information limit imposed by thermal diffusion. At Sardakh, additional likelihoods at 30 m depth help the reconstruction move closer to the 50 m reference but cannot restore long-term signals that are only preserved at greater depths. At Tiksi, more likelihoods extend the period during which the reconstruction remains independent from the prior, showing that the information limit is not reached with one likelihood alone. However, this benefit saturates after several likelihoods, once the borehole's thermal memory is fully used.
Compared to earlier work, including Guglielmin et al. (2018), which relied on an averaged multi-year profile—this study shows that treating repeated measurements as separate likelihoods provides additional, usable temporal information. Instead of simply improving data quality through averaging, multiple likelihoods allow the inversion to extract more of the recoverable past and delay the point where the prior starts to dominate.
Overall, this demonstrates that treating repeated measurements as separate likelihoods allows the inversion to exploit temporal information that is typically ignored in borehole studies. This situates the method between traditional single-profile inversions and more data-intensive

observational approaches, enabling a larger fraction of the physically available information of the borehole to be extracted before the prior begins to dominate.

6.4 Implications for Permafrost Research

The findings of this study have several implications for how borehole data can be used to understand permafrost dynamics and past climate variability more effectively.
A central insight is that repeated borehole temperature measurements contain additional information that is not used in traditional single-profile inversions. When several annual profiles are treated as separate likelihood terms, the inversion can make use of small year-to-year differences caused by variations in surface conditions or measurement noise. These subtle differences are usually lost when profiles are averaged into a single climatological curve. Within the physical limits set by thermal diffusion, this allows shallow and medium-depth boreholes to support more stable and internally consistent GST reconstructions than previously assumed.
A second implication concerns the role of climatologically informed priors. The ERA5+n-factor prior acts as a local climatic reference: it links the prior mean and variability to reanalysis-based air temperatures, ensuring that reconstructed GST histories reflect site-specific climate characteristics. While this does not increase the physical information in the borehole itself, it provides a realistic and climatologically consistent regularization. This helps prevent unrealistic behaviour when the observational signal is weak, shallow, or noisy, and ensures that the reconstructed temperatures remain anchored to the local climate regime.
Another implication concerns the temporal scale of reconstructions. While traditional borehole inversions mainly aim at multi-decadal climate trends, the multiple-likelihood formulation presented here is also well suited for higher-resolution reconstructions over shorter periods, within the shallow-time window that remains resolvable by thermal diffusion. If several years of borehole profiles exist—especially if they are spaced apart—the inversion can recover finer temporal variations than classical single-profile methods allow. This makes the approach useful not only for long-term climate analysis, but also for studying medium-term processes such as:

- changes in snow insulation
- shifts in vegetation or surface moisture
- impacts of exceptional warming events
- short-term transitions in the active layer and upper permafrost

Finally, the results emphazise the physical limits imposed by thermal diffusion, reminding us that even advanced statistical methods cannot compensate for missing depth or insufficient thermal memory. Nonetheless, the combination of multiple likelihoods and data-driven priors makes it possible to exploit the available information more efficiently.

7 Conclusion & Outlook

This study investigated how GST histories can be reconstructed from permafrost borehole measurements based on the Bayesian inversion framework of Groenke et al. (2024) [1], and extended their approach with two methodological additions, an ERA5-based data-driven prior and a multiple-likelihood formulation, to improve the GST reconstructions.

7.1 Conclusion

A key result is that the multiple-likelihood provides demonstrable practical benefits in a controlled Proof-of-Concept setting. In the Proof-of-Concept test, where reconstructed GST histories were compared against synthetic ground truth, configurations with two to three likelihoods showed conditional improvements in the moderate noise range. In particular, they tended to reduce posterior spread and produced more temporally consistent reconstructions, while changes in RMSE and normalized log-likelihood were not uniform across all configurations. Non-consecutive years often performed better than consecutive years, likely because they provide less redundant information and reduce the impact of correlated profile structure. Within the physically available information budget of each borehole, the multiple-likelihood therefore enables a quantitatively closer reconstruction in the Proof of Concept (PoC) and a qualitatively more consistent behavior in the heatmap analyzes. Depth experiments further confirm that additional likelihoods cannot compensate for the loss of deep thermal memory: deeper profiles retain older climate signals, while shallow boreholes lose them rapidly. Multiple likelihoods can extend the recoverable past by several years but remain bounded by the fundamental limits of heat conduction.
Secondly, the study demonstrates the value of using a climatologically informed prior. The ERA5+n-factor prior anchors the GST reconstruction to local climate conditions without introducing strong/obvious systematic bias. While borehole temperature profiles does not contain explicit seasonal variability, the data-driven prior provides a physically meaningful regularisation. This is particularly important for shallow or noisy profiles, where the inversion would otherwise drift toward unrealistic behaviour.
Overall, the study shows that multiple likelihoods and a data-driven prior mainly improve the temporal consistency of reconstructed GST histories, as reflected in consistently higher Pearson correlation values (see Figure 5.13). However, the extended framework does not lead to a systematic reduction in reconstruction error.

7.2 Outlook: Future Work & Applications

The presented framework opens several avenues for future research aimed at improving borehole-based GST reconstructions and expanding their interpretative value for permafrost science.

First, the method can be extended by incorporating more realistic boundary-layer processes. Explicit modeling of snow insulation, seasonally varying n-factors, vegetation effects, or soil moisture dynamics could refine the link between air temperature and GST, and reduce systematic biases in the prior. Introducing depth-dependent thermal properties or including latent heat processes may further improve physical consistency, especially in areas with active layers or heterogeneous substrates.

Second, the multiple-likelihood formulation provides a promising path for utilising repeated observations more effectively. Future studies could combine profiles from several boreholes into a joint Bayesian framework, analyse how much independent information different combinations of years provide, or explore hierarchical setups that allow site-specific priors to be informed by regional climate patterns. Such extensions could be particularly valuable in observatories with irregular measurement intervals or incomplete data records.

Third, the method is well suited for reconstructing medium-term climatic variations on decadal timescales. Because repeated profiles help to stabilise the inversion against short-term noise, the approach enables finer temporal interpretation of recent climatic changes, such as shifts in snow regimes, extreme warming events, or rapid changes in vegetation cover. This makes the method applicable not only for long-term palaeoclimatic studies but also for monitoring contemporary permafrost evolution.

A key limitation of the present framework is that structural model error is not treated explicitly. The forward model relies on simplified boundary conditions, constant thermal properties, and neglects processes such as snow dynamics or moisture transport. As a result, systematic model–data mismatches arise and become more visible when the assumed observation noise is small.

In the current implementation, the observation noise scale partly compensates for these unresolved model errors. This effect becomes particularly visible in the multiple-likelihood setting, where small systematic biases accumulate during sequential forward propagation between likelihood years. When the noise scale is chosen too small, the inversion is forced to correct these mismatches through strong adjustments of the reconstructed surface temperature history, which can lead to non-physical variability. Therefore, future extensions of the method with the forward model used in this thesis, should introduce an explicit model-error component to account for unresolved processes and systematic bias.

A further important direction for future work is the validation of the multiple-likelihood method using long and continuous GST observations. Because borehole inversions lack multi-decadal ground truth, the qualitative improvements found in this study cannot yet be confirmed in a quantitative way. At least 15–20 years of co-located GST measurements would be needed to test whether the method truly makes better use of the physical information available in the borehole signal, within the limits set by thermal diffusion. Such validation would provide clearer evidence for the method's effectiveness and help separate real physical improvement

Finally, several open research questions remain. These include the year-to-year variability of

n-factors, the impact of ERA5 uncertainties on data-driven priors, the degree to which the assumption of conditional independence holds across different sites, and the identification of optimal likelihood combinations that maximize independent information. Addressing these questions will help to refine the method further and to better quantify the physical and statistical limits of GST reconstructions across different Arctic landscapes.

Taken together, this study provides a practical and extensible framework for improving GST reconstructions from shallow and intermediate-depth boreholes. By combining climatological knowledge with repeated thermal observations, the method brings borehole inversions closer to their physical information limit and offers a solid foundation for future developments in permafrost research.

Bibliography

1. GROENKE, B. et al.: Robust reconstruction of historical climate change from permafrost boreholes. *Journal of Geophysical Research: Earth Surface*. 2024. Available from DOI: https://doi.org/10.1029/2024JF007734.
2. FRENCH, H. M.: *The Periglacial Environment*. The University of Ottawa, 2007.
3. OBU, J.: How much of the Earth's Surface is Underlain by Permafrost ? *Journal of Geophysical Research: Earth Surface*. 2021. Available from DOI: https://doi.org/10.1029/2021JF006123.
4. BOIKE, J. et al.: Arctic Permafrost. 2023. Available from DOI: https://doi.org/10.1016/B978-0-12-822974-3.00141-5.
5. RISEBOROUGH, D. et al.: Recent Advances in Permafrost Modelling. *Permafrost and Periglacial Processes*. 2008. Available from DOI: https://doi.org/10.1002/ppp.615.
6. BISKABORN, B. K.: Permafrost is warming at a global scale. *nature communications*. 2019. Available from DOI: https://doi.org/10.1038/s41467-018-08240-4.
7. HUANG, S. et al.: Temperature trends over the past five centuries reconstructed from borehole temperatures. *nature*. 2000. Available from DOI: https://doi.org/10.1038/35001556.
8. HOPCROFT, P. O.; GALLAGHER, K.: Global Variability in Multi-Century Ground Warming Inferred From Geothermal Data. *Journal of Geophysical Research: Earth Surface*. 2023. Available from DOI: https://doi.org/10.1029/2023GL104631.
9. LACHENBRUCH, A. H.; MARSHALL, B. V.: Changing climate: geothermal evidence from permafrost in the Alaskan Arctic. *Science*. 1986. Available from DOI: https://doi.org/10.1126/science.234.4777.689.
10. CLOW, G. D.: The extent of temporal smearing in surface-temperature histories derived from borehole temperature measurements. *Global and Planetary Change*. 1992. Available from DOI: https://doi.org/10.1016/0921-8181(92)90027-8.
11. HELD, L.; BOVE, D. S.: *Likelihood and Bayes Inference*. Springer, 2020.
12. CHOUINARD, C.; MARESCHAL, J.-C.: Ground surface temperature history in southern Canada: Temperature at the base of the Laurentide ice sheet and during the Holocene. *Earth and Planetary Science Letters*. 2008. Available from DOI: https://doi.org/10.1016/j.epsl.2008.10.026.
13. PICKLER, C. et al.: Recent climate variations in Chile: constraints from borehole temperature profiles. *Climate of the Past*. 2018. Available from DOI: https://doi.org/10.5194/cp-14-559-2018.
14. BELTRAMI, H.: Ground Temperature Histories for Central and Eastern Canada from Geothermal Measurements: Little Ice Age Signature. *Geophysical Research Letters*. 1992. Available from DOI: https://doi.org/10.1029/92GL00671.

15. LIU, J. et al.: Application of Tikhinov regularization to reconstruct past climate record from borehole tmeperature. *Inverse Problems in Science and Engineering*. 2021. Available from DOI: https://doi.org/10.1080/17415977.2021.1975700.

16. JAUME-SANTERO, F. et al.: North American regional climate reconstruction from ground surface temperature histories. *Climate of the Past*. 2016. Available from DOI: https://doi.org/10.5194/cp-12-2181-2016.

17. CUESTA-VALERO, F. J. et al.: A new bootstrap technique to quantify uncertainty in estimates of ground surface temperature and ground heat flux histories from geothermal data. *Geoscientific Model Development*. 2022. Available from DOI: https://doi.org/10.5194/gmd-15-7913-2022.

18. HOPCROFT, P. O. et al.: Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo. *Geophysical Journal International*. 2007. Available from DOI: https://doi.org/10.1111/j.1365-246X.2007.03596.x.

19. RANTANEN, M. et al.: The Arctic has warmed nearly four times faster than the globe since 1979. *communications earth and communications*. 2022. Available from DOI: https://doi.org/10.1038/s43247-022-00498-3.

20. GUGLIELMIN, M. et al.: Ground surface temperature reconstruction for the last 500 years obtained from permafrost temperatures observed in the SHARESTELVIO Borehole, Italian Alps. *copernicus*. 2018. Available from DOI: https://doi.org/10.5194/cp-14-709-2018.

21. SMITH, S. L. et al.: The changing thermal state of permafrost. *nature communications*. 2022. Available from DOI: https://doi.org/10.1038/s43017-021-00240-1.

22. KNEIER, F. et al.: Borehole temperature reconstructions reveal differences in past surface temperature trends for the permafrost in the Laptev Sea region, Russian Arctic. *Polar Science*. 2018. Available from DOI: https://10.1007/s41063-018-0041-3.

23. ANDERSLAND, O. B.; LADANYI, B.: *Frozen Ground Engineering*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.

24. GRIGULL, U.; SANDNER, H.: *Wärmeleitung*. Springer Verlag, 1990.

25. CARSLAW, H. S.; JAEGER, J. C.: *Conduction of Heat in Solids*. Oxford, Clarendon Press, 1959.

26. NARASIMHAN, T. N.: Fourier's Heat Conduction Equation: History, Influence, And Connections. *Journal of Geophysical Research: Earth Surface*. 1999. Available from DOI: https://doi.org/10.1029/1998RG900006.

27. JURY, W.; HORTON, R.: *Soil physics*. John Wiley & Sons., 2004.

28. RISEBOROUGH, D. W.: Soil latent heat as a filter of the climate signal in permafrost. *In Proceedings of the fifth Canadian permafrost conference, collection nordicana*. 1990.

29. IGLESIAS, M. A. et al.: Ensemble Kalman Methods for Inverse Problems. *Inverse Problems*. 2013. Available from DOI: https://10.1088/0266-5611/29/4/045001.

30. SCHILLINS, C.; STUART, A. M.: Analysis of the Ensemble Kalman Filter for Inverse Problems. *SIAM Journal on Numerical Analysis*. 2016. Available from DOI: https://doi.org/10.48550/arXiv.1602.02020.

31. ALIFANOV: *Inverse Heat Transfer Problems*. Springer Science and Business Media, 2012.

32. SRINIVASAMURTHY, S. B.: *Methods Of Solving Ill-Posed Problems*. 2004. MA thesis. Kansas State University.

33. CALVETTI, D.; SOMERSALO, E.: Inverse problems: From regularization to Bayesian inference. *WIREs Computational Statistics*. 2017. Available from DOI: https://doi.org/10.1002/wics.1427.

34. WOODBURY, A. D.; FERGUSON, G.: Ground surface paleotemperature reconstruction using information measures and empirical Bayes. *Journal of Geophysical Research: Earth Surface*. 2006. Available from DOI: https://doi.org/10.1029/2005GL025243.

35. OVERSTALL, A. M.; WOODS, D. C.: A Strategy for Bayesian Inference for Computationally Expensive Models with Application to the Estimation of Stem Cell Properties. *Biometrics*. 2013. Available from DOI: https://doi.org/10.1111/biom.12017.

36. GARBUNO-INIGO, A. et al.: Interacting Langevin Diffusions:Gradient Structure And Ensemble Kalman Sampler. *SIAM Journal on Applied Mathematics*. 2019. Available from DOI: https://doi.org/10.48550/arXiv.1903.08866.

37. GROENKE, B. et al.: Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer. *The Cryosphere*. 2023. Available from DOI: https://doi.org/10.5194/tc-17-3505-2023.

38. GELMAN, A. et al.: *Bayesian Data Analysis*. Columbia University, 2025. with error fixed as of 20 February 2025.

39. WELCH, G.; BISHOP, G.: An Introduction to the Kalman Filter. *Computer Science Papers - University of North Carolina at Chapel Hill*. 2006.

40. EVENSEN, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*. 2002. Available from DOI: https://doi.org/10.1007/s10236-003-0036-9.

41. ROTH, M. et al.: The Ensemble Kalman filter: a signal processing perspective. *EURASIP Journal of Advances in Signal Processing*. 2017. Available from DOI: https://doi.org/10.48550/arXiv.1702.08061.

42. [N.d.]. Last access: 26.03.2025.

43. LUETHI, R.; PHILLIPS, M.: Challenges and solutions for long-term permafrost borehole temperature monitoring and data interpretation. *Geographica Helvetica*. 2016. Available from DOI: https://doi.org/10.5194/gh-71-121-2016.

44. NOETZLI, J. et al.: Best Practice for Measuring Permafrost Temperature in Boreholes Based on the Experience in the Swiss Alps. *frontiers in Earth Science*. 2021. Available from DOI: https://doi.org/10.3389/feart.2021.607875.

45. GOOGLE EARTH: *Siberia, Russia* [https://earth.google.com/web/@72.17977351,127.75199874,164.98966799a,296687.8206623d,35y,360h,0t,0r/]. 2025. Accessed on 25 April 2025.

46. WAGNER, D. et al.: Microbial Controls on Methane Fluxes from a Polygonal Tundra of the Lena Delta, Siberia. *John Wiley & Sons, Ltd.* 2003. Available from DOI: https://doi.org/10.1002/ppp.443.

47. BOIKE, J. et al.: Baseline characteristics of clime, permafrost and land cover from a new permafrost observatory in the Lena River Delta, Siberia (1998 - 2011). *Biogesciences.* 2013. Available from DOI: https://doi.org/10.5194/bg-10-2105-2013.

48. BOIKE, J. et al.: Continuous measurements in soil and air at the permafrost long-term observatory at Samoylov Station. 2022. Available from DOI: https://doi.org/10.1594/PANGAEA.947032.

49. LABORATORY, N. P. S.: *Tiksi Observatory.* 2024. Available also from: https://psl.noaa.gov/arctic/observatories/tiksi/. Accessed: April 2025.

50. KONSTANTINOV, P. et al.: Impact of Climate Change on the Ground Thermal Regime in the Lower Lena Region, Arctic Central Siberia. *Arctic Central Siberia.* 2022. Available from DOI: https://doi.org/10.3390/land12010019.

51. HERSBACH, H. et al.: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society.* 2020. Available from DOI: https://10.1002/qj.3803.

52. WESTERMANN, S. et al.: The CryoGrid community model (version 1.0): A multi-physics toolbox for climate-driven simulations in the terrestrial cryosphere. *Geoscientific Model Development.* 2023. Available from DOI: 10.5194/gmd-16-2607-2023.

53. MARTIN, J.: *Influence of the inter-annual variability of snow physical properties on the ground thermal regime - through observations and modelling (Samoylov Island, Siberia).* 2022. MA thesis. University of Bremen and Alfred-Wegener Institut.

54. MCCALMAN, L. et al.: Distributed Bayesian Geophysical Inversions. In: *Proceedings of the Thirty-Ninth Workshop on Geothermal Reservoir Engineering.* Stanford, California, 2014.

55. WALKER, M.; CURTIS, A.: Spatial Bayesian inversion with localized likelihoods: An exact sampling alternative to MCMC. *Journal of Geophysical Research: Solid Earth.* 2014. Available from DOI: 10.1002/2014JB011010.

56. TARANTOLA, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics (SIAM), 2005.

57. WAQAR, F. G. et al.: A tutorial on the Bayesian statistical approach to inverse problems. 2023. Available from DOI: https://doi.org/10.48550/arXiv.2304.07610.

58. KAIPIO, J.; SOMERSALO, E.: *Statisticaland Computational Inverse Problems.* Springer the language of science, 2004.

59. KLENE, A. E. et al.: The N-factor in Natural Landscapes: Variability of Air and Soil-Surface Temperatures, Kuparuk River Basin, Alaska, U.S.A. *Arctic, Antarctic, and Alpine Research.* 2001. Available from DOI: https://doi.org/10.1080/15230430.2001.12003416.

60. ZHANG, T.: Influence of the seasonal snow cover on the ground thermal regime: an overview. *Journal of Geophysical Research: Earth Surface.* 2005. Available from DOI: https://doi.org/10.1029/2004RG000157.

61. GISNÅS, K. et al.: CryoGRID 1.0: Permafrost Distribution in Norway estimated by a Spatial Numerical Model. *Permafrost and Periglacial Processes*. 2013. Available from DOI: https://doi.org/10.1002/ppp.1765.

62. LUO, D. et al.: Difference between near-surface air, land surface and ground surface temperatures and their influences on the frozen ground on the Qinghai-Tibet Plateau. *Geoderma*. 2017. Available from DOI: https://doi.org/10.1016/j.geoderma.2017.09.037.

63. IOOSS, B.; SALTELLI, A.: Introduction: Sensitivity Analysis. 2008.

64. KLEIJNEN, J. P.: Sensitivity analysis and related analysis: A review of some statistical techniques. *Journal of Statistical Computation and Simulation*. 1997. Available from DOI: https://doi.org/10.1080/00949659708811805.

65. BORGONOVO, E.; PLISCHKE, E.: Sensitivity Analysis: A review of recent advances. *European Journal of Operational Research*. 2015. Available from DOI: https://doi.org/10.1016/j.ejor.2015.06.032.

66. SCHOBER, P. et al.: Correlation Coefficients: Appropriate Use and Interpretation. *ANESTHESIA & ANALGESIA*. 2018. Available from DOI: https://10.1213/ANE.0000000000002864.

67. MARTIN, R.; WALKER, S. G.: Data-driven priors and their posterior concentration rates. *Electronic Journals of Statistics*. 2019. Available from DOI: https://doi.org/10.48550/arXiv.1604.05734.

68. GROENKE, B. et al.: Investigating the thermal state of permafrost with Bayesian inverse modeling of heat transfer. *The Cryosphere*. 2023. Available from DOI: https://doi.org/10.5194/tc-17-3505-2023.

69. LACHENBRUCH, A. H.; MARSHALL, B. V.: Heat Flow in the Arctic. 1969.

70. BELTRAMI, H.: Surface heat flux histories from inversion of geothermal data: Energy balance at the Earth's surface. *Journal of Geophysical Research: Earth Surface*. 2001. Available from DOI: https://doi.org/10.1029/2000JB000065.

Acronyms

ALT active layer thickness

AR(1) Autoregressive

AWI Alfred Wegener Institute

ECMWF European Centre for Medium-Range Weather Forecasts

EKI Ensemble Kalman Inversion

EKS Ensemble Kalman Sampling

EnKF Ensemble Kalman Filter

ERA European Centre for Medium-Range Weather Forecasts Reanalysis Atmospheric

FI Freezing index

GST Ground Surface Temperatures

GTN-P Global Terrestrial Network for Permafrost

KF Kalman Filter

MAAT Mean Annual Air Temperature

MAGT Mean Annual Ground Temperature

MCMC Markov Chain Monte Carlo

NA Missing values are denoted by NA (missing value indicator)

NOAA National Oceanic And Atmospheric Administration

PIY Melnikov Permafrost Institute

PoC Proof of Concept

PVC Polyvinyl chloride

rjMCMC reversible-jump Markov Chain Monte Carlo

RMSE Root Mean Square Error

SA sensitivity analysis

TI Thawing index

ZAA Zero Annual Amplitude

List of Figures

List of Tables