Global variability in multi-century ground warming inferred from geothermal data

Hopcroft, Peter; Gallagher, Kerry

Global Variability in Multi-Century Ground Warming Inferred From Geothermal Data

Peter O. Hopcroft[1] and **Kerry Gallagher[2]**

[1]School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, UK, [2]Géosciences Rennes, Université de Rennes, Rennes, France

Abstract Geothermal temperature-depth measurements can be used to reconstruct past climate without the need for calibration against meteorological data. The global geothermal database is one of opportunity and is therefore subject to variations in measurement protocol, quality control and potential non-climatic influences. As the climatic history recovery is sensitive to these factors, we developed a Bayesian hierarchical model that allows us to treat errors and uncertainty formally. This way we may better isolate the climate signal. For the Northern Hemisphere extra-tropics our reconstruction shows a warming beginning in CE 1750 and it captures the observed two-phase warming over the past century. We clearly identify the northern tropics as a region of greatest ground warming and hypothesize that this reflects land-use change. For the Southern Hemisphere, the inclusion of newer data leads to a modest cooling until CE 1750. Outside the tropics, agreement with multi-proxy reconstructions has improved relative to earlier studies.

Plain Language Summary Understanding past climate is invaluable for evaluating the natural context of man-made warming. Long term surface-air temperature records only exist at a few locations. To reconstruct global trends further back in time proxies must then be used. Measurements from such systems are then calibrated against observed climate variations. Temperatures measured in the ground can provide more direct information on past variations because sustained trends at the surface drive thermal perturbations that penetrate into the subsurface and which can be measured today. These geothermal data have the advantage that they do not require calibration and so are independent of meteorological observations. However, recovering the climate signal is not trivial. For this reason we developed a new statistical approach to infer past temperature variations from a database of 1012 temperature-depth profiles distributed near-globally. The results show excellent agreement with observed temperatures and also demonstrate improved agreement with proxy-based records. One exception is noted over equatorial regions in the northern hemisphere where a potential influence of historical land-use may be significant.

1. Introduction

Measurements of Earth's surface air temperatures covering the last 150 years show a pronounced warming (I. Harris et al., 2014) that is largely the result of human influence (Bindoff et al., 2013; Hegerl et al., 2019). Longer records of surface temperature are invaluable for placing the recent anthropogenically perturbed period in the broader context of Earth's pre-industrial climate variability (Neukom et al., 2019). The instrumental surface air temperature record is limited in the years before CE 1850 (Benestad et al., 2019; I. Harris et al., 2014; Morice et al., 2012) and only a few longer instrumental records are available from historical records (Brönnimann et al., 2019). For this reason temperature sensitive proxies such as tree and coral growth rings, ice-cores and lake sediments are used to reconstruct earlier climate variations (e.g., Jacoby & D'Arrigo, 1989; Mann et al., 1998, 1999, 2008; PAGES 2k Consortium, 2013, 2019) showing the unprecedented nature of globally synchronous warming since CE 1850 (e.g., Mann et al., 1998, 1999; Neukom et al., 2019).

An alternative method for reconstructing climate makes use of geothermal temperature measurements (e.g., Beltrami & Mareschal, 1992; Beltrami et al., 2005; Chouinard & Mareschal, 2009; R. Harris & Chapman, 2005; Jaume-Santero et al., 2016; Lachenbruch & Marshall, 1986; Mareschal & Beltrami, 1992; Pickler et al., 2018; Pollack & Huang, 2000; Pollack et al., 1998). This relies on changes at the surface being conducted into the ground, causing measurable perturbations in the temperature-depth profile (Beltrami et al., 2015; Pollack et al., 1998). Globally averaged results from geothermal data consistently show a continuous warming since around CE 1500 of around 1°C up to the late 20th Century (Beltrami & Bourlon, 2004; Cuesta-Valero et al., 2021; Huang et al., 2000;
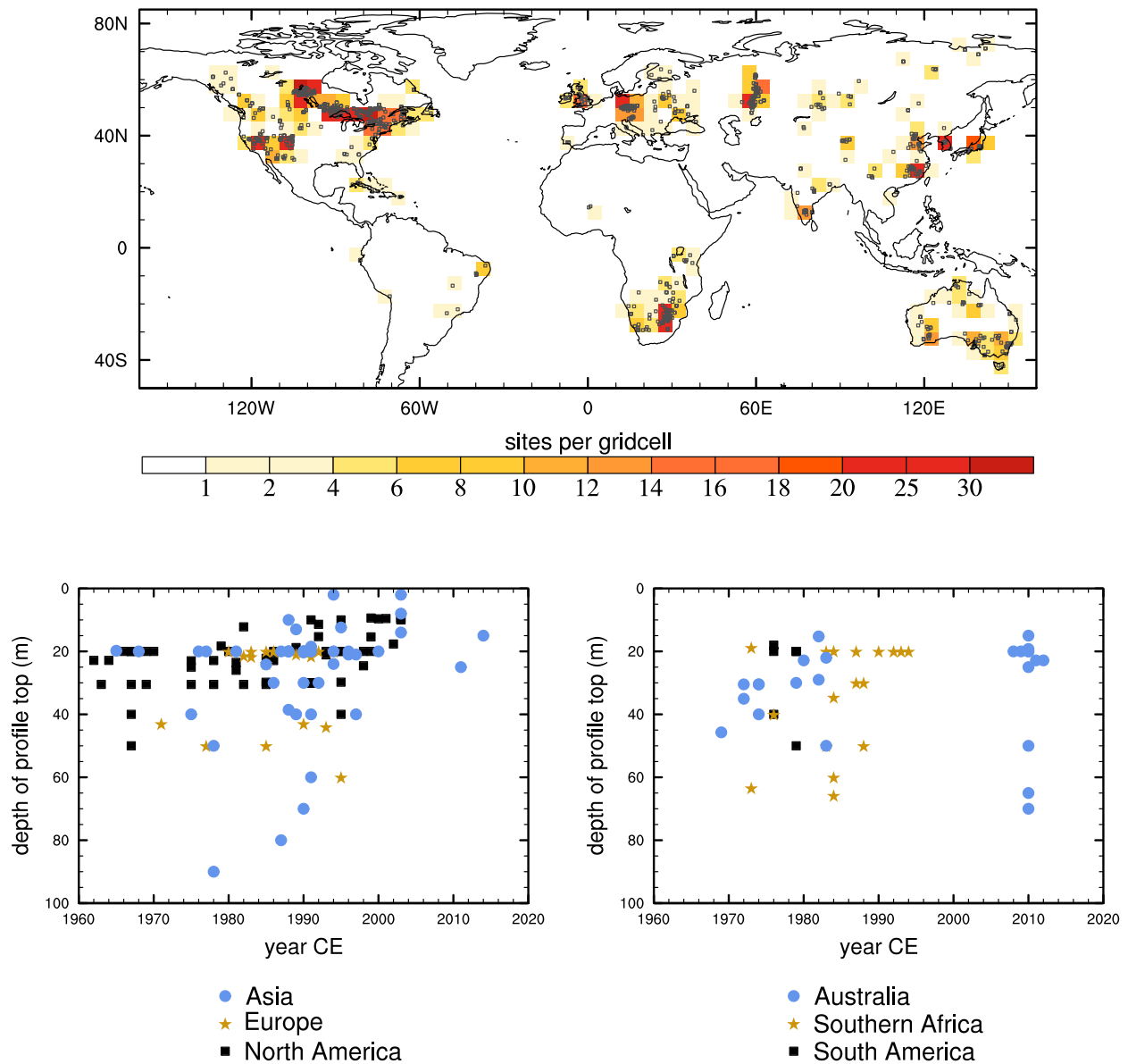
Figure 1. Sampling density of the 1012 borehole temperature profiles on the 5° × 5° grid and a comparison of the profile top depth against measurement date across six sub-regions.

Pollack & Smerdon, 2004). In contrast, proxy-based methods mostly indicate a long-term cooling from CE 1000 to around CE 1700–1800 (Klippel et al., 2020; Mann, 2021; PAGES 2k Consortium, 2013, 2019).

The global geothermal climate database was originally compiled with the International Heatflow Commission (IHFC) (Huang & Pollack, 1998; Huang et al., 2000). It is a database of opportunity that has been compiled from measurements taken over several decades with different instrumentation, in varied settings and with no commonly adopted protocol or quality control. It is also relatively clustered spatially. Around half of sites (471 out of 1012) were logged before CE 1985, with profile tops mostly 20 m or deeper (see Figure 1), meaning the observed post CE 1970 surface air temperature rise may not be recorded. There are also potentially non-climatic influences which have been investigated with modeling, direct observations and comparisons with instrumental data (Bartlett et al., 2006; Beltrami et al., 2005; Cermak et al., 2017; Gonzàlez-Rouco et al., 2006, 2009; Majorowicz & Skinner, 1997; Melo-Aguilar et al., 2018, 2020; Skinner & Majorowicz, 1999). This means that the global database is potentially subject to variable and unquantified levels of measurement error and non-climatic influence, termed as noise hereafter (see also Text S1 in Supporting Information S1). Partly for this reason but also

because the ill-posed nature of the inverse problem (Menke, 1989), a smoothing operator is required to prevent over-fitting. Without this, a reconstruction with unrealistically large swings in temperature could result. Although there are empirical approaches to choose this smoothing, there is no single value that can apply equally well in the range of situations encountered.

We identify four main reasons for studying this data set here: (a) spatial variability in the geothermal reconstruction has not been evaluated or attributed in detail (e.g., Beltrami & Bourlon, 2004; Huang et al., 2000); (b) it remains unclear why proxy-based reconstruction show cooling until Industrialization (e.g., Esper et al., 2012; Klippel et al., 2020; Mann, 2021; Neukom et al., 2019) whereas geothermal reconstruction show the opposite (e.g., Cuesta-Valero et al., 2021; Masson-Delmotte et al., 2013; Pollack & Smerdon, 2004); (c) it is unclear whether the heterogeneity in the geothermal database has been fully accounted for and for example, whether noisier data have caused reconstructions to be smoother than necessary (Clow, 1992; Hopcroft et al., 2007); and (d) 91 profiles have been added to the IHFC database since CE 2000, with recent additions in Australia which was otherwise dominated by much older measurements (Figure 1).

For these reasons we have developed a Bayesian hierarchical model to infer past surface temperature histories from the global geothermal database (Huang & Pollack, 1998). This approach allows uncertainties to be robustly treated in a probabilistic framework (Denison et al., 2002). Both the observational noise level at each site and the required smoothing of the climatic history are inferred directly from the data. A key advantage of this is that noisier data are given less weight in the final result and data which can support more detailed climate histories are not subject to excessive smoothing. We applied this method to the available global IHFC database and evaluate spatial variability and uncertainty in the resultant reconstructions.

2. Methods: Bayesian Inference of Past Climate From Geothermal Data

Extracting climate from geothermal data is an ill-posed and non-linear inverse problem (Menke, 1989). As such, a Bayesian approach is a natural choice (Tarantola, 2005) and the hierarchical aspect also allows us to treat some necessary, but not directly relevant, parameters, such as data errors, as unknowns. We expanded on the model described by Hopcroft et al. (2007, 2009) to infer the surface temperature history at each location from down-hole geothermal data. The posterior distribution from Bayes' law is reconstructed using a Markov chain Monte Carlo (MCMC) algorithm (Denison et al., 2002; Geyer, 2011; Gilks et al., 1995). We use a trans-dimensional or reversible jump MCMC (Denison et al., 2002; Fan & Sisson, 2011; Green, 1995) which is widely used in Earth Science problems (e.g., Charvin et al., 2009; Gallagher et al., 2009; Jeong et al., 2012; Malinverno, 2002; Sambridge et al., 2006, 2013).

2.1. Model Setup

We use 1012 profiles in the University of Michigan/International Heatflow commission database (Huang & Pollack, 1998), see http://heatflow.world/project (IHFC henceforth). Our global analysis is performed on a $5° \times 5°$ latitude/longitude grid shown in Figure 1. Given the heterogeneous distribution of sites, this results in 172 occupied grid cells. A one-dimensional finite element heat conduction model (Lewis et al., 1996) is used to calculate the temperature-depth profile at all 1012 sites over time (Hopcroft et al., 2007). The model starts at 1400 CE and runs forward in time with a 5-year timestep up to the timestep that is closest to the latest date of logging among the profiles that are contained within each grid cell. The spatial discretization uses the depths of the site measurements, with five padding layers above the shallowest measurement and a total of 480 elements vertically. We use IHFC estimated site thermal conductivity data and follow previous work (Hopcroft et al., 2007) by prescribing fixed values for density and heat capacity at each site.

The upper and lower boundary conditions are the surface temperature and basal heat flux respectively. For the inverse problem, the past surface temperature history is parameterized separately in each of the $5° \times 5°$ gridcells as a series of k time-temperature (t-T) points interpolated onto the 5-year timestep (Hopcroft et al., 2007). The temperature is assumed to vary linearly between each adjacent pair of t-T points (Hopcroft et al., 2007). Together with the sampled values of the parameters, basal heat flux q_0 and pre-reconstruction mean surface temperature T_{pre} defining the initial condition for the model, the sampled temperature history is used to predict the temperature depth profile at the date of measurement for a given borehole data set. The surface temperature history model is allowed between 2 and 30 points over 600 years, which is consistent with the temporal resolution of geothermal climate reconstruction method (Pollack & Huang, 2000). The model parameters are listed in Table S1 of Support-
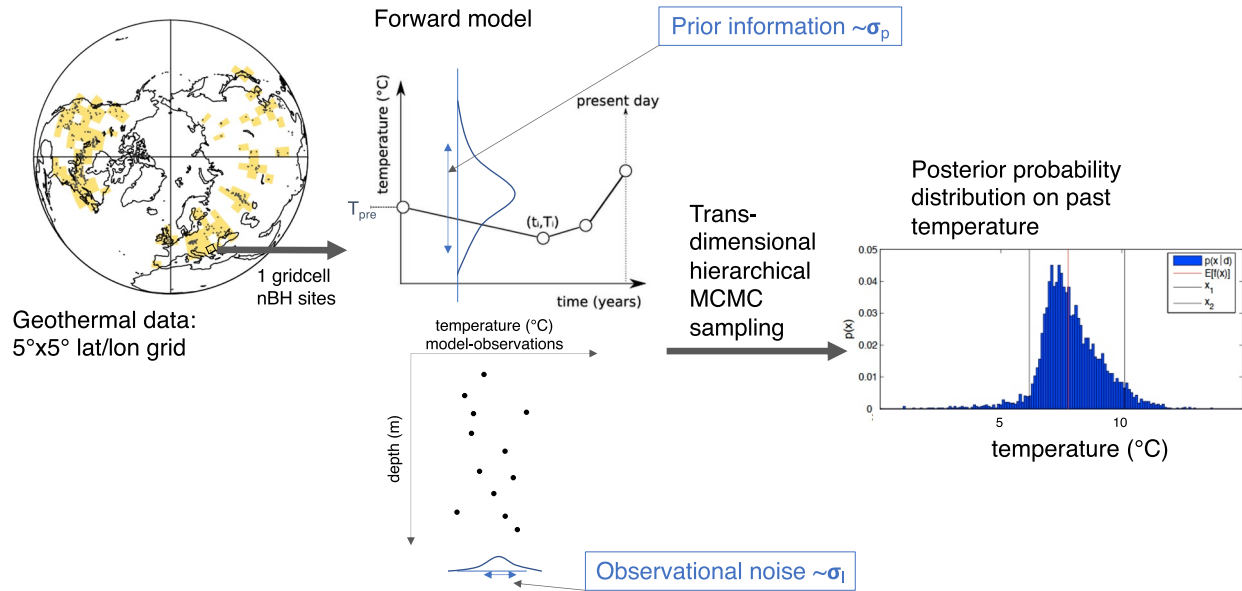
Figure 2. Schematic of the hierarchical model. The key model parameters in each gridcell are the basal heat flux and pre-reconstruction surface temperature (T_{pre}), the surface temperature history (T), the prior information width (σ_p) and the observational noise (σ_l), all of which are jointly estimated from the data.

ing Information S1. All borehole temperature depth profiles that lie in a particular grid cell (see Figure 1) are used jointly when sampling the values of the time and surface temperature, q_0 and T_{pre}. The accepted samples are then collated to produce a posterior probability density function for each parameter. This model setup is shown schematically in Figure 2.

The trans-dimensional (rjMCMC Green, 1995; Hopcroft et al., 2007) formulation means that the number of timepoints in the reconstruction (k, the resolution or model complexity) adapts to the information in the data at each site and is sampled as an additional model parameter (see Figure 2). Although the model can increase the number of parameters to better fit the data, such over-fitting is avoided because models with more parameters or with wider a priori probability distributions will integrate over a larger region of model state-space and hence give a lower marginal likelihood (Denison et al., 2002; MacKay, 2003). In practice, this allows the complexity of the model (surface temperature history) to self-adapt to the information contained in the borehole temperature data. The hierarchical approach means that several important parameters are assigned prior probability distributions rather than using fixed values for all locations—this is key difference with our earlier work (Hopcroft et al., 2007). A potential advantage of this is that we do not need to specify one value for either the measurement noise or the expected variance in the past climate at a given site. This is potentially important because some profiles are likely to be less noisy than others. This allows the algorithm to treat different data sets accordingly.

We treat the following key parameters in this way:

1. σ_l—the uncertainty on the borehole temperature observations at each site.
2. σ_p—the width of the prior distribution for past temperature variations in each gridcell.
3. λ_S—the a priori favored value of k, the number of time-temperature nodes in the reconstruction in a given gridcell.

The variance of these prior distributions is inferred together with the other model parameters, using hyper-priors (e.g., Denison et al., 2002), see Figure 2. We use hyper-prior parameters as listed in Table S1 of Supporting Information S1 that encompass a realistic range of values in each case. Full details of the prior, hyperprior and likelihood distributions and the methods for sampling with the hierarchical rjMCMC algorithm are given in the Supporting Information S1 together with an evaluation using two global test cases that mimic the real database with 1012 synthetic temperature-depth profiles.

2.2. Initial Conditions and Parallel Sampling

Under suitable conditions a MCMC algorithm will give an excellent approximation of the underlying posterior probability distribution. However, inefficient sampling can lead to biases. This is analogous to being trapped

in a local minima in a multi-dimensional optimization problem. A widely adopted method to deal with this involves evaluating convergence through comparison of several parallel MCMC algorithms (Gelman, 1995; Gelman & Rubin, 1992; Gelman & Shirley, 2011). After an initial period of the sampling is discarded, the ratio of the variance within and between parallel changes should be less than around 1.2 (Gelman, 1995; Gelman & Rubin, 1992; Gelman & Shirley, 2011) indicating that different initialized MCMC chains are covering a similar range in state-space and have therefore converged on a similar solution.

We follow this approach and initialized parallel rjMCMC algorithms with randomly chosen values of the heat flow at each site (q_0), the pre-reconstruction temperature (T_{pre}) and the temperature values of the reconstruction (T). The model is initialized with 5 equally spaced time-temperature nodes, that is, $k = 5$ (see Table S1 in Supporting Information S1) and with randomly selected values for the temperatures at these times. Ten parallel chains are configured and each is run for half a million rj-MCMC iterations with the first half of these discarded. The model parameters are saved every 50 iterations to reduce the computational storage requirements. The convergence diagnostic is calculated for the reconstructed surface temperature anomaly between years CE 1705–1850 and CE 1855–2000 as such, evaluating the pre- to post-industrial change. Only gridcells that have converged are considered in the following discussion.

2.3. Summarizing the Posterior

For each grid cell, the inferred CE 1955–1990 mean of the temperature reconstruction is subtracted and the resultant temperature anomaly timeseries is area-weighted to produce regional averages. The reconstruction runs to the year CE 2015, but the inferred temperatures are only included in hemispheric averages up to the latest date of measurement of any profile within that grid cell minus the effective time of the shallowest measurement (z_0). The characteristic time for a 5% perturbation is given by Turcotte and Schubert (2002) (their equations 4–115):

$$t_{d0} = z_0^2 / \left(2.32^2 \kappa \right). \tag{1}$$

For z_0 of 50 m and $\kappa = 1 \times 10^{-6}$, this is around 15 years. Thus, the end date (t_*) is given by:

$$t_* = t_{meas} - t_{d0} \tag{2}$$

where t_{meas} is the date of measurement of the profile, and t_{d0} is the effective response time of the shallowest measurement given above. Thus a profile measured in CE 1985 with a shallowest measurement at 50 m is averaged up to the year CE 1970.

3. Results

3.1. Sampling and Convergence

For the real data case the rj-MCMC algorithm converges in 94% of the gridcells (162 of the 172). An example of the convergence across parallel chains is shown for one gridcell in Figure S3 of Supporting Information S1. When the temperature history proposals were adapted slightly (as described in Text S2 of Supporting Information S1), this increased to 98% (6 more cells) convergence. The remaining four gridcells all contain multiple profiles and the lack of convergence is most likely caused by a combination of noisier/inconsistent data with possibly relatively limited constraints on climate and/or inefficiency of the rj-MCMC algorithm when faced with many profiles in one gridcell. Using adaptive proposal distributions (e.g., Rosenthal, 2011) or a more advanced Bayesian sampling approaches such as partition modeling (Denison et al., 2002) may help to resolve this, but this is beyond the scope of the present study. Thus in the following only results from the 168 converged gridcells are analyzed.

3.2. Hemispheric Mean Temperature Trends

The results of the global inversion are shown for the two hemispheres in Figure 3, plotted using only gridcells in which the rjMCMC sampling converged (as described above). Two results are shown, one in which the prior on the temperature history is constant through time and a second in which 1°C of warming is assumed for all locations over the 20th Century since this is known to have occurred and it thereby better approximates our knowledge of more recent climate change. The posterior results from both are compared with the meteorological record from
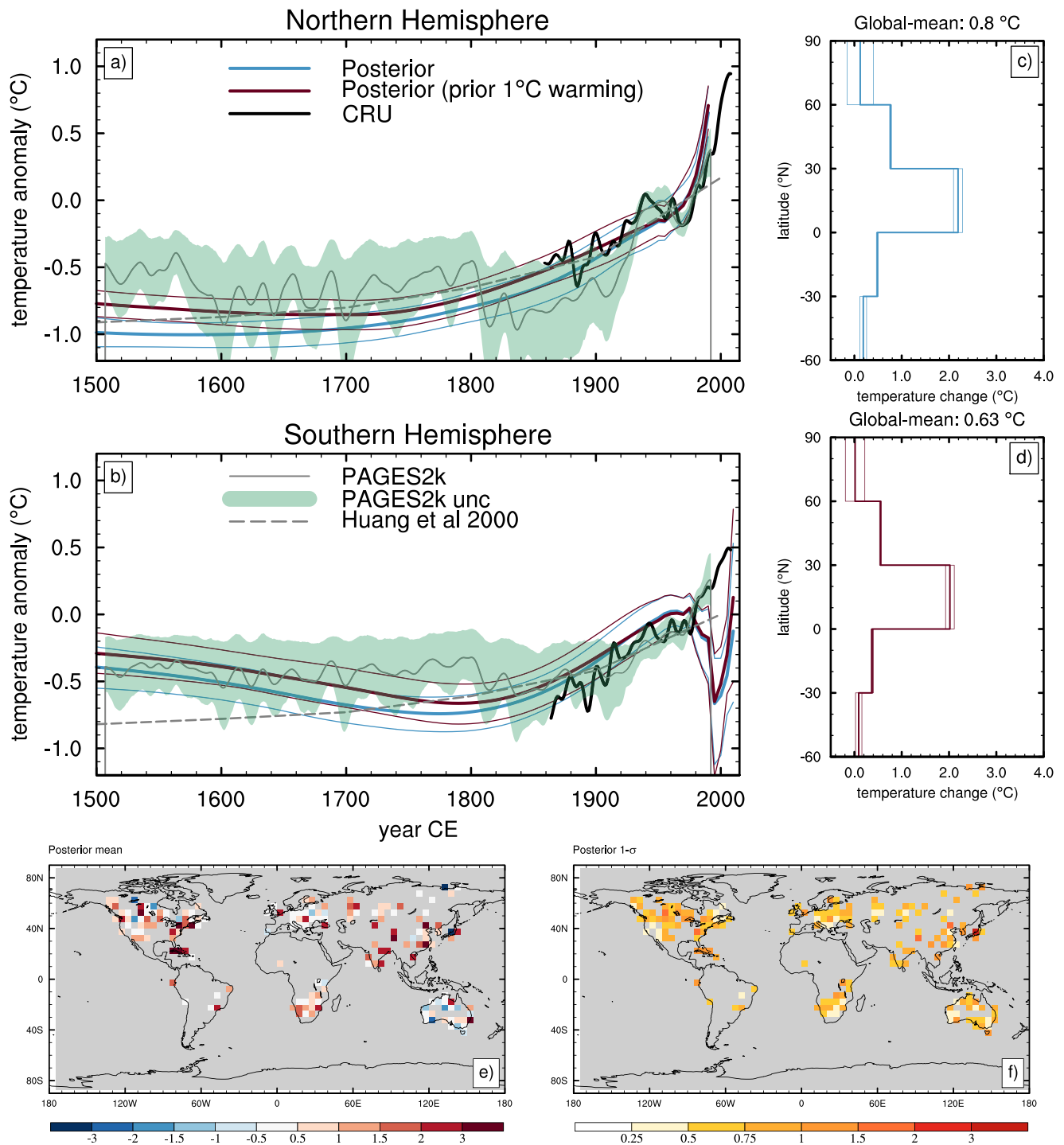
Figure 3. Observed and reconstructed temperature variations for (a) the Northern hemisphere (0°–70°N) and (b) the Southern hemisphere (0°–45°S) relative to the late twentieth century mean. The geothermal posterior mean (thick blue) and 95% confidence intervals (thin blue) and a second case where the a priori model includes a global historical era warming of 1°C. Observed land station temperature (I. Harris et al., 2014) filtered with a 20-year moving window (CRU: black). Multi-proxy terrestrial temperature reconstructions (Neukom et al., 2019; PAGES2k Consortium, 2017), mean: thin gray lines, green shading: 10%–90% uncertainty limits. Hemispheric-mean geothermal reconstruction digitized from Huang et al. (2000) (dashed gray). (c and d) The latitudinally averaged warming signal (1955–1980 minus 1500–1550) inferred from the geothermal data. (e) Posterior mean temperature change CE 1950–1980 minus 1500–1550. (f) Posterior 1σ uncertainty on the temperature change CE 1950–1980 minus 1500–1550.
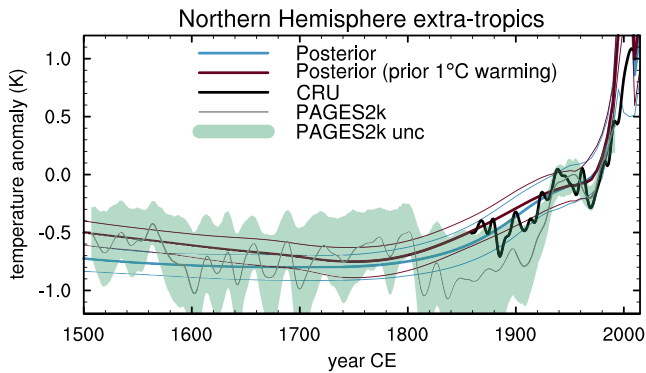
Figure 4. As in Figure 3 but for areas between 30° and 70°N only. The peak temperature anomaly in the posterior mean is 1.9 K.

land stations (against which it has not been calibrated), the terrestrial signal derived from PAGES2k reconstructions (v2.0.0) (Neukom et al., 2019) and previous geothermal reconstructions for the two hemispheres (Huang et al., 2000). We calculated the regional temperature reconstruction from the PAGES2k version 2.0.0 multi-method global temperature field reconstructions (Neukom et al., 2019; PAGES2k Consortium, 2017) masked here to the same gridcell coverage as the geothermal data (as shown in Figure 1).

The northern hemisphere reconstruction shows a continuous warming since around CE 1500 which has accelerated in the last 150 years in good agreement with the station-based observed temperature history between CE 1860 and 2000. The new results confirm that the very rapid warming in the past two decades is evident in geothermal data (Cuesta-Valero et al., 2021), although this part of the reconstruction is more uncertain because it is conditioned on fewer gridcells as most of the geothermal data predate CE 2000 (see Figure 1). Differences caused by the prior specification are relatively modest. The overall amplitude of warming for the Northern hemisphere is 1.0°C up to the late 20th Century mean. This is within the uncertainties of the work by Huang et al. (2000), Beltrami and Bourlon (2004), and the more recent work by Cuesta-Valero et al. (2021) who use a similar data set. Spatially, Figure 3c shows that the warming signal is much greater in the northern hemisphere tropics than elsewhere, a feature that we discuss further below.

Excluding the signal over 0°–30°N (shown in Figure 4) alters the geothermal result significantly. In this region we find that the continuous warming started later, at around CE 1750 with a cooling preceding this. This transition at around CE 1750 from cooling to warming is different from previous results from the northern extra-tropics (e.g., Beltrami & Bourlon, 2004) and highlights where use of the Bayesian hierarchical model has diverged from earlier studies. Over this region (>30°N) the result is in very good agreement with the observed meteorological temperature record, and in this case resolves the observed slowdown in warming during the mid-20th Century. The differences between the two cases in terms of the prior information used indicates that the data may not translate to a unique posterior reconstruction, namely that the influence of noise versus the prior information is important (e.g., Li et al., 2010).

In the Southern hemisphere the reconstruction is also different from previous studies. Figure 3 shows a different trend to the Northern Hemisphere with a cooling until CE 1750–1800. The amplitude of warming is smaller than in earlier geothermal studies (Huang et al., 2000). This contrasts with the results for the Northern Hemisphere and is most likely because of the inclusion of a large number of new profiles in the IHFC database in recent years as shown in Figure 1. The last 35 years of the SH reconstruction shows a trend that does not agree with the preceding decades or observations. This is because the number of sites included in the average is very low and is therefore biased toward the results in few gridcells. The alternative prior makes very little difference to the results for the Southern Hemisphere. In contrast with the North, over the Southern Hemisphere there is no two-phase warming over the 20th Century, in agreement with the observed temperature record.

We performed three additional tests to isolate the impact of the developments made in this study. We tested the impacts of (a) of fixing the parameter σ_l, the estimated noise on the temperature observations to a globally constant value of 0.05°C, (b) keeping σ_p, the prior width on the temperature histories constant at 1.0°C and (c) fixing both of these parameters (σ_l and σ_p) and keeping the dimensionality of the temperature history (k) constant at one temperature-point per 100 years. The noise value is chosen as an approximately intermediate value (Beardsmore & Cull, 2001; Wang, 1992) that is similar to previous studies (see also Text S1 in Supporting Information S1). For example, Huang et al. (2000) used 0.03°C (Shen et al., 1995) and Cuesta-Valero et al. (2021) used a fixed number of eigenvalues in their inversion, which is equivalent to a fixed noise level. By keeping σ_l fixed globally, we are assigning an equal weighting to all observational data sets in the posterior distributions. Fixing σ_p globally means that a priori we expect similar scale climate variations in all locations. In the third test, we only allow the heatflow, pre-reconstruction mean temperature and the temperature history at fixed 100-year intervals to vary. The results of the sensitivity tests are compared with the standard result in Figures S4–S6 of Supporting Information S1. The first two cases show relatively limited impacts on the hemispheric means although differences do emerge. Holding all three parameters constant (k, σ_l, and σ_p) leads to a very similar result at the global scale, though the fixed noise case does result in larger warming polewards of 60°N. Thus we conclude that the noise parameter estimation and adapting the time-resolution are not critical for this problem, especially for earlier centuries where the data's resolving power in time is very limited.

4. Discussion

Focusing on spatial trends reveals a significant difference in the northern hemisphere between the tropics (0°–30°N) and extra-tropics (>30°N) in all cases we analyzed. A review of past work (Beltrami & Bourlon, 2004; Beltrami et al., 2015; Cuesta-Valero et al., 2021; Huang et al., 2000; Pollack & Smerdon, 2004) reveals that, to our knowledge, this has not been clearly discussed before. The significantly higher amplitude of warming in the tropics may be caused by localized effects of land use which have been shown to be important in both Cuba (Cermak & Bodri, 2001; Cermak et al., 1992) and to some extent in India (Roy & Chapman, 2012), two regions of the most intense warming (Figure 3e). Geothermal data are likely more sensitive to land-cover change than other proxy-based methods. Unlike tree-rings they are not implicitly biased toward areas of time-continuous forest presence. Also changes in the sensible versus latent heating at the ground-air interface can be very large following clearance and these tend to dominate the increase in albedo (MacDougall & Beltrami, 2017). Anthropogenic land-use may also have impacted other regions. For example, Majorowicz and Skinner (1997) and Skinner and Majorowicz (1999) found that geothermal data overestimated recent warming in the North America Great Plains in comparison with meteorological data which they attributed to historical land clearance. Therefore, future work should address this potential source of bias and aim to clarify why the geothermal results show more warming in this region and at the same time display good agreement with the instrumental records which have protocols around setting and are screened for changes in surroundings (e.g., Trewin, 2010).

Our geothermal results show a similar level of cooling in the late 19th Century (CE 1850–1900) relative to the CE 2000 as the instrumental record (Figures 3 and 4). This is potentially significant because the multiproxy records diverge with the instrumental record at this time and it is unclear which is correct (Frank et al., 2007). It is possible that the instrumental records are biased because of limited coverage outside of Europe, North America and Asia or because of changes in meteorological protocols and equipment. It has also been suggested that both the geothermal (Mann & Schmidt, 2003) and multiproxy records (R. Harris & Chapman, 2005) are seasonally biased. Thus, a focus on records from similar areas across this early instrumental time-interval may help to improve understanding of the relative merits of sparse instrumental observations (Hegerl et al., 2018) in comparison with geothermal and multiproxy records.

The agreement with the PAGES2k composite is better for the northern extra-tropics particularly when the prior includes the known warming (labeled as *prior* 1°C *warming* in Figure 4). Both the geothermal and the PAGES2k multi-proxy reconstruction show similar levels of warming from the pre-industrial to the present-day and approximately similar timing of minimum temperatures around CE 1750. In the full northern hemisphere the geothermal result shows more warming overall, and similar to previous geothermal studies shows continuous warming, whereas the multi-proxy results do not (Mann, 2021). In the Southern Hemisphere the overall trend is in good agreement with proxy-based PAGES2k result and, up to around CE 1980 is in good agreement with observational temperature record. However, the PAGES2k southern hemisphere result is based on a much smaller collection of sites than the northern hemisphere and so should be interpreted with caution.

The improved agreement with the multi-proxy reconstruction can be attributed to the addition of newer geothermal data in the Southern Hemisphere which has altered the reconstruction in comparison with older results (e.g., Huang et al., 2000; Pollack et al., 2006) and because our method gives different results over the northern extra-tropics (Beltrami & Bourlon, 2004), most likely because of the adaptive nature of the method, since the fixed noise case shown in Figure S4 of Supporting Information S1 shows a larger amplitude of warming north of 60°N.

The differing trends in the two hemispheres is evident in the instrumental record and in multi-proxy reconstructions (Abram et al., 2016; Neukom et al., 2019). It is therefore plausible that changes in ocean circulation or other modes of atmospheric variability contributed to differences between the hemispheres (Goosse, 2017). However, this is not clearly seen in forced climate model simulations and data assimilation methods have not been able to unambiguously identify missing factors that could explain these trends (e.g., Goosse, 2017). Geothermal data do not allow us to pinpoint the underlying mechanism, but the fact that this feature arises in this independent data set adds to the evidence for distinct long-term variations in the two hemispheres and perhaps can motivate further data acquisition in the Southern Hemisphere (e.g., Pickler et al., 2018).

A critical consideration for Bayesian methods is the specification of prior information. We have considered two choices. The first assumes no temperature variation through time and the second accounting for known warming (I. Harris et al., 2014). The 'flat' prior is not a good reflection of our a priori knowledge but the uncertainty

on this prior information is not fixed and is jointly inferred alongside other model parameters. This means that prior does not rule out larger climate variations in particular locations where they can be supported by the data. In future work the prior could be adapted so that its spatial variability is based on meteorological observations and the prior width on the temperature history could be scaled using the observed spatial footprint of long-term variability from meteorological records.

The uncertainties on the hemispheric reconstructions are derived from the rj-MCMC statistical sampling which captures the full uncertainty involved in the inverse problem as posed. A notable feature here is the agreement between different geothermal approaches (at least for the Northern Hemisphere) but also the relatively narrow uncertainties when the results are aggregated to the hemispheric scale. These uncertainties are substantially smaller than those associated with the multi-proxy reconstructions (e.g., Neukom et al., 2019). It may be the case that the true uncertainty in the geothermal method is larger than indicated because of factors not directly considered directly. This question of relative uncertainties could be studied in future work.

Future work could aim for a multi-method approach from geothermal data following that recommended for calibrated multi-proxy reconstructions (Neukom et al., 2019; PAGES 2k Consortium, 2019). The geothermal data could also be merged with the non-geothermal reconstructions which would improve spatial coverage and potentially reduce uncertainties. Alternatively, the Bayesian method here could be extended to exploit the observed spatial coherence in the meteorological surface temperature record in order to generate a complete global field from the geothermal data through time.

5. Conclusions

We applied a new statistical method to extract the climatic signal from a global data set of 1012 temperature-depth profiles. This approach is more flexible than typical inverse methods because it is able to tailor the model to the data in each location, potentially improving the information retrieved through natural down-weighting of noisier data sources. Our results show good agreement with much shorter station-based observational records in resolving both the warming slow-down in the mid-20th century and subsequent acceleration. The hemispheric means do show potentially major divergences with the reconstructions from multi-proxies in the Northern hemisphere which is driven by greater inferred ground warming in the northern tropics. This difference may be related to the effects of land use. Outside of the tropics the geothermal and multi-proxy data show good agreement. Overall the geothermal data are consistent with the finding (e.g., Mann et al., 1998; Neukom et al., 2019; PAGES 2k consortium, 2019) that long-term temperature variations in the 500 years before Industrialization were relatively small and that 20th century warming was unprecedented in this context.

Data Availability Statement

Code developed for this study is available from https://dx.doi.org/10.5281/zenodo.7108225. This comprises shell scripts for downloading the IHFC database and processing the files and the Bayesian hierarchical model written in C++. NetCDF files containing the posterior mean and credible intervals of the reconstructed ground surface temperatures on the $5° \times 5°$ grid are available from https://doi.org/10.5281/zenodo.7761459 for all cases discussed in this work. The IHFC borehole database is available from https://geothermal.earth.lsa.umich.edu or from the US National Oceanographic and Atmospheric Administration https://www.ncei.noaa.gov/products/paleoclimatology/borehole. The PAGES2k v2.0.0 global temperature field reconstructions are available from https://doi.org/10.25921/fbzb-1n14. CRU observational data is available from https://crudata.uea.ac.uk/cru/data/temperature.

References

Abram, N., McGregor, H., Tierney, J., Evans, M., McKay, N., Kaufman, D., et al. (2016). Early onset of industrial-era warming across the oceans and continents. *Nature*, *536*(7617), 411–418. https://doi.org/10.1038/nature19082

Bartlett, M., Chapman, D., & Harris, R. (2006). A decade of ground-air temperature tracking at Emigrant Pass Observatory, Utah. *Journal of Climate*, *19*(15), 3722–3731. https://doi.org/10.1175/jcli3808.1

Beardsmore, G., & Cull, J. (2001). *Crustal heat flow: A guide to measurement and modelling*. Cambridge University Press.

Beltrami, H., & Bourlon, E. (2004). Ground warming patterns in the Northern Hemisphere during the last five centuries. *Earth and Planetary Science Letters*, *27*(3–4), 169–177. https://doi.org/10.1016/j.epsl.2004.09.014

Beltrami, H., Ferguson, G., & Harris, R. N. (2005). Long-term tracking of climate change by underground temperatures. *Geophysical Research Letters*, *32*(19), L19707. https://doi.org/10.1029/2005GL023714

Beltrami, H., & Mareschal, J. (1992). Ground temperature histories for central and eastern Canada from geothermal measurements: Little Ice Age signature. *Geophysical Research Letters*, *19*(7), 689–692. https://doi.org/10.1029/92gl00671

Beltrami, H., Matharoo, G., & Smerdon, J. (2015). Ground surface temperature and continental heat gain: Uncertainties from underground. *Environmental Research Letters*, *10*(1), 014009. https://doi.org/10.1088/1748-9326/10/1/014009

Benestad, R., Erlandsen, H., Mezghani, A., & Parding, K. (2019). Geographical distribution of thermometers gives the appearance of lower historical global warming. *Geophysical Research Letters*, *46*(13), 7654–7662. https://doi.org/10.1029/2019GL083474

Bindoff, N., Stott, P., AchutRao, K., Allen, M., Giillett, N., Gutzler, D., et al. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker, D. Qin, G. K. Plattner, M. M. H. L. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 867–952). C.U.P.

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., et al. (2019). Unlocking pre-1850 instrumental meterological records: A global inventory. *Bulletin of the American Meteorological Society*, *100*(12), ES389–ES413. https://doi.org/10.1175/BAMS-D-19-0040.1

Cermak, V., & Bodri, L. (2001). Climate reconstruction from subsurface temperatures demonstrated on example of Cuba. *Physics of the Earth and Planetary Interiors*, *126*(3–4), 295–310. https://doi.org/10.1016/s0031-9201(01)00262-x

Cermak, V., Bodri, L., Kresl, M., Dedecek, P., & Safanda, J. (2017). Eleven years of ground-air temperature tracking over different land cover types. *International Journal of Climatology*, *37*(2), 1084–1099. https://doi.org/10.1002/joc.4764

Cermak, V., Bodri, L., & Safanda, J. (1992). Underground temperature fields and changing climate: Evidence from Cuba. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *97*(4), 325–337. https://doi.org/10.1016/0031-0182(92)90215-q

Charvin, K., Gallagher, K., Hampson, G., & Labourdette, R. (2009). A Bayesian approach to inverse modelling of stratigraphy, part 1: Method. *Basin Research*, *21*(1), 5–25. https://doi.org/10.1111/j.1365-2117.2008.00369.x

Chouinard, C., & Mareschal, J.-C. (2009). Ground surface temperature history in southern Canada: Temperatures at the base of the Laurentide ice sheet and during the Holocene. *Earth and Planetary Science Letters*, *277*(1–2), 280–289. https://doi.org/10.1016/j.epsl.2008.10.026

Clow, G. (1992). The extent of temporal smearing in surface-temperature histories derived from borehole temperature measurements. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *98*(2–4), 81–86. https://doi.org/10.1016/0031-0182(92)90189-c

Cuesta-Valero, F., Garcia-Garcia, A., Beltrami, H., Gonzalez-Ruoco, F., & Garcia-Bustamante, E. (2021). Long-term global ground heat flux and continental heat storage from geothermal data. *Climate of the Past*, *17*(1), 451–468. https://doi.org/10.5194/cp-17-451-2021

Denison, D. T., Homes, C., Mallick, B., & Smith, F. (2002). *Bayesian methods for non-linear classification and regression*. John Wiley and Sons.

Esper, J., Frank, D., Timonen, M., Zorita, E., Wilson, R. J. S., Luterbacher, J., et al. (2012). Orbital forcing of tree-ring data. *Nature Climate Change*, *2*(12), 862–866. https://doi.org/10.1038/NCLIMATE1589

Fan, Y., & Sisson, S. (2011). Reversible jump MCMC. In S. Brook, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

Frank, D., Buntgen, U., Bohm, R., Maugeri, M., & Esper, J. (2007). Warmer early instrumental measurements versus colder reconstructed temperatures: Shooting at a moving target. *Quaternary Science Reviews*, *26*(25–28), 3298–3310. https://doi.org/10.1016/j.quascirev.2007.08.002

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M., & Stephenson, J. (2009). Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems. *Marine & Petroleum Geology*, *26*(4), 525–535. https://doi.org/10.1016/j.marpetgeo.2009.01.003

Gelman, A. (1995). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.

Gelman, A., & Shirley, K. (2011). Inference from simulations and monitoring convergence. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

Geyer, C. (2011). Introduction to Markov chain Monte Carlo. In S. Brook, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. Routledge.

Gilks, W., Richardson, S., & Spiegelhalter, D. (1995). Introducing Markov chain Monte Carlo methods. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

Gonzàlez-Rouco, J., Beltrami, H., Zorita, E., & Stevens, M. (2009). Borehole climatology: A discussion based on contributions from climate modeling. *Climate of the Past*, *5*(1), 97–127. https://doi.org/10.5194/cp-5-97-2009

Gonzàlez-Rouco, J., Beltrami, H., Zorita, E., & von Storch, H. (2006). Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling. *Geophysical Research Letters*, *33*(1), L01703. https://doi.org/10.1029/2005gl024693

Goosse, H. (2017). Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries. *Climate Dynamics*, *48*(5–6), 1483–1501. https://doi.org/10.1007/s00382-016-3154-z

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732. https://doi.org/10.1093/biomet/82.4.711

Harris, I., Jones, P., Osborn, T., & Lister, D. (2014). Updated high-resolution grids of monthly climatic observations - The CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Harris, R., & Chapman, D. (2005). Borehole temperatures and tree rings: Seasonality and estimates of extratropical Northern Hemispheric warming. *Journal of Geophysical Research*, *110*(F4), F04003. https://doi.org/10.1029/2005JF000303

Hegerl, G., Bronnimann, S., Cowan, T., Friedman, A. R., Hawkins, E., Iles, C., et al. (2019). Causes of climate change over the historical record. *Environmental Research Letters*, *14*(12), 123006. https://doi.org/10.1088/1748-9326/ab4557

Hegerl, G., Bronnimann, S., Schurer, A., & Cowan, T. (2018). The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(4), e522. https://doi.org/10.1002/wcc.522

Hopcroft, P., Gallagher, K., & Pain, C. (2007). Inference of past climate from borehole temperature data using Bayesian Reversible Jump Markov chain Monte Carlo. *Geophysical Journal International*, *171*(3), 1430–1439. https://doi.org/10.1111/j.1365-246X.2007.03596.x

Hopcroft, P., Gallagher, K., & Pain, C. (2009). A Bayesian Partition Modelling approach to resolve spatial variability in climate records from borehole temperature inversion. *Geophysical Journal International*, *178*(2), 651–666. https://doi.org/10.1111/j.1365-246X.2009.04192.x

Huang, S., & Pollack, H. (1998). *Global Borehole Temperature Database for Climate Reconstruction. IGBP PAGES-World Data Center-A for Paleoclimatology. 1998-044*. NOAA-NGDC Paleoclimatology Program.

Huang, S., Pollack, H., & Shen, P.-Y. (2000). Temperature trends over the last five centuries reconstructed from borehole temperatures. *Nature*, *403*(6771), 756–758. https://doi.org/10.1038/35001556

Jacoby, G., & D'Arrigo, R. (1989). Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America. *Climatic Change*, *14*(1), 39–59. https://doi.org/10.1007/bf00140174

Jaume-Santero, C., PicklerBeltrami, H., & Mareschal, J.-C. (2016). North American regional climate reconstruction from ground surface temperature histories. *Climate of the Past*, *12*, 2181–2194. https://doi.org/10.5194/cp-12-2181-2016

Jeong, S.-J., Medvigy, D., Shevliakova, E., & Malyshev, S. (2012). Uncertainties in terrestrial carbon budgets related to spring phenology. *Journal of Geophysical Research*, *117*(G1), G01030. https://doi.org/10.1029/2011JG001868

Klippel, L., St. George, S., Buntgen, U., Krusic, P. J., & Esper, J. (2020). Differing pre-industrial cooling trends between tree rings and lower-resolution temperature proxies. *Climate of the Past*, *16*(2), 729–742. https://doi.org/10.5194/cp-16-729-2020

Lachenbruch, A., & Marshall, B. (1986). Changing climate: Geothermal evidence. *Science*, *234*(4777), 689–696. https://doi.org/10.1126/science.234.4777.689

Lewis, R., Morgan, K., Thomas, H., & Seetharamu, K. (1996). *The finite element method in heat transfer analysis*. John Wiley & Sons.

Li, B., Nychka, D., & Ammann, C. (2010). The value of multiproxy reconstruction of past climate. *Journal of the American Statistical Association*, *105*(491), 883–895. https://doi.org/10.1198/jasa.2010.ap09379

MacDougall, A., & Beltrami, H. (2017). Impact of deforestation on subsurface temperature profiles: Implications for the borehole paleoclimate record. *Environmental Research Letters*, *12*(7), 074014. https://doi.org/10.1088/1748-9326/aa7394

MacKay, D. (2003). *Information theory, inference, and learning algorithms*. Cambridge University Press.

Majorowicz, J., & Skinner, W. (1997). Potential causes of differences between ground and surface air temperature warming across different ecozones in Alberta, Canada. *Global and Planetary Change*, *15*(3–4), 79–91. https://doi.org/10.1016/s0921-8181(97)00005-2

Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, *151*(3), 675–688. https://doi.org/10.1046/j.1365-246X.2002.01847.x

Mann, M. (2021). Beyond the hockey stick: Climate lessons from the Common Era. *Proceedings of the National Academy of Sciences*, *118*(39), e2112797118. https://doi.org/10.1073/pnas.2112797118

Mann, M., Bradley, R., & Hughes, M. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, *392*(6678), 779–787. https://doi.org/10.1038/33859

Mann, M., Bradley, R., & Hughes, M. (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters*, *26*(6), 759–762. https://doi.org/10.1029/1999gl900070

Mann, M., & Schmidt, G. (2003). Ground versus surface air temperature trends: Implications for borehole surface temperature reconstructions. *Geophysical Research Letters*, *30*(12), 1607. https://doi.org/10.1029/2003gl017170

Mann, M., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., & Ni, F. (2008). Proxy-based reconstructions of hemispheric and global surface temperature variations. *Proceedings of the National Academy of Sciences*, *105*(36), 13252–13257. https://doi.org/10.1073/pnas.0805721105

Mareschal, J., & Beltrami, H. (1992). Evidence for recent warming from perturbed thermal gradients: Examples from eastern Canada. *Climate Dynamics*, *6*(3–4), 135–143. https://doi.org/10.1007/bf00193525

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gonzalez-Rouco, J., et al. (2013). Information from paleoclimate archives. In T. F. Stocker, D. Qin, G. K. Plattner, M. M. H. L. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 383–464).

Melo-Aguilar, C., González-Rouco, J. F., García-Bustamante, E., Navarro-Montesinos, J., & Steinert, N. (2018). Influence of radiative forcing factors on ground–air temperature coupling during the last millennium: Implications for borehole climatology. *Climate of the Past*, *14*(11), 1583–1606. https://doi.org/10.5194/cp-14-1583-2018

Melo-Aguilar, C., González-Rouco, J. F., García-Bustamante, E., Steinert, N., Jungclaus, J. H., Navarro, J., & Roldán-Gómez, P. J. (2020). Methodological and physical biases in global to sub-continental borehole temperature reconstructions: An assessment from a pseudo-proxy perspective. *Climate of the Past*, *16*(2), 453–474. https://doi.org/10.5194/cp-16-453-2020

Menke, W. (1989). *Geophysical data analysis: Discrete inverse theory*. Academic Press.

Morice, C., Kennedy, J., Rayner, N., & Jones, P. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*(D8), D08101. https://doi.org/10.1029/2011JD017187

Neukom, R., Steiger, N., Gomez-Navarro, J., Wang, J., & Werner, J. P. (2019). No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, *571*(7766), 550–554. https://doi.org/10.1038/s41586-019-1401-2

PAGES 2k Consortium. (2013). Continental-scale temperature variability during the past two millennia. *Nature Geoscience*, *6*(5), 339–345. https://doi.org/10.1038/NGEO1797

PAGES2k Consortium. (2017). A globalmultiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*(1), 170088. https://doi.org/10.1038/sdata.2017.88

PAGES 2k Consortium. (2019). Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nature Geoscience*, *12*(8), 643–649. https://doi.org/10.1038/s41561-019-0400-0

Pickler, C., Fausto, E., Beltrami, H., Mareschal, J. C., Suárez, F., Chacon-Oecklers, A., et al. (2018). Recent climate variations in Chile: Constraints from borehole temperature profiles. *Climate of the Past*, *15*(4), 559–575. https://doi.org/10.5194/cp-14-559-2018

Pollack, H., & Huang, S. (2000). Climate reconstruction from subsurface temperatures. *Annual Review of Earth and Planetary Sciences*, *28*(1), 339–365. https://doi.org/10.1146/annurev.earth.28.1.339

Pollack, H., Huang, S., & Shen, P.-Y. (1998). Climate change record in subsurface temperatures: A global perspective. *Science*, *282*(5387), 279–281. https://doi.org/10.1126/science.282.5387.279

Pollack, H., Huang, S., & Smerdon, J. (2006). Five centuries of climate change in Australia: The view from underground. *Journal of Quaternary Science*, *21*(7), 701–706. https://doi.org/10.1002/jqs.1060

Pollack, H., & Smerdon, J. (2004). Borehole climate reconstructions: Spatial structure and hemispheric averages. *Journal of Geophysical Research*, *109*(D11), D11106. https://doi.org/10.1029/2003JG004163

Rosenthal, J. (2011). Optimal proposal distributions and adaptive MCMC. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

Roy, S., & Chapman, D. (2012). Borehole temperatures and climate change: Ground temperature change in south India over the past two centuries. *Journal of Geophysical Research*, *117*(D11), D11105. https://doi.org/10.1029/2011jd017224

Sambridge, M., Bodin, T., Gallagher, K., & Tkalcic, H. (2013). Transdimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 20110547. https://doi.org/10.1098/rsta.2011.0547

Sambridge, M., Gallagher, K., Jackson, A., & Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, *167*(2), 528–542. https://doi.org/10.1111/j.1365-246x.2006.03155.x

Shen, P., Pollack, H., Huang, S., & Wang, K. (1995). Effects of subsurface heterogeneity on the inference of climate change from borehole temperature data: Model studies and field examples from Canada. *Journal of Geophysical Research*, *100*(B4), 6383–6396. https://doi.org/10.1029/94jb03136

Skinner, W., & Majorowicz, J. (1999). Regional climatic warming and associated twentieth century land-cover changes in north-western North America. *Climate Research*, *12*, 39–52. https://doi.org/10.3354/cr012039

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Trewin, B. (2010). Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIREs Climate Change*, *1*(4), 490–506. https://doi.org/10.1002/wcc.46

Turcotte, D., & Schubert, G. (2002). *Geodynamics* (2nd ed.). CUP.

Wang, K. (1992). Estimation of ground surface temperatures from borehole temperature data. *Journal of Geophysical Research*, *97*(B2), 2095–2106. https://doi.org/10.1029/91jb02716

References From the Supporting Information

Biskaborn, B., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., et al. (2019). Permafrost is warming at a global scale. *Nature Communications*, *10*(1), 254. https://doi.org/10.1038/s41467-018-08240-4

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A Practical Introduction to Bayesian Analysis*. CRC Press.