Author's Response Document for "Long-term Surface Temperature (LoST) Database as a complement for GCM preindustrial simulations" by Francisco José Cuesta-Valero, Almudena García-García, Hugo Beltrami, Eduardo Zorita and Fernando Jaume-Santero.

We thank Dmitry Demezhko, the anonymous reviewer and Irina Rogozhina for their thoughtful and constructive feedback.

This Author's Response file provides a complete documentation of the changes that have been made in response to each individual Reviewer comment. Reviewer comments are shown in plain text. Author responses are shown in bold text. Corrections within the revised manuscript are shown in blue text. All page and line numbers in the author responses refer to locations in the revised manuscript.

The response to the Editor's comments is also included in this document.

Anonymous Referee #1

Received and published: 4 January 2019

Review of "Long-term Surface Temperature (LoST) Database as a complement for GCM preindustrial simulations" By: Cuesta-Valero, et al. Corresponding author: Beltrami

This manuscript addresses an important problem – large variation, and therefore uncertainty, in climate sensitivity estimates, that is the change in surface temperature accompanying a doubling of the atmospheric concentration of CO_2. Since the 1970s the variation in estimates of ECS has not improved much; still from 1.5 to 4.5 ^oC.

The paper introduction provides a useful summary of why the existing GCMs have not provided convergence to a narrow band of ECS: differences in model parameterization, in particular radiative transfer and model tuning; feedback mechanisms such as ice-albedo and water vapor effects; and permafrost stability and permafrost carbon feedback. The authors suggest "a constrained preindustrial control simulation may improve the representation of those feedbacks in transient climate experiments, reducing the uncertainty of ECS estimates from model simulations, as well as reducing the spread in projections of future climate change."

The proposed constraint for such a preindustrial simulation is a database they have assembled called LoST (LoST = Long Term Surface Temperature). The paleo surface temperature at a given site is estimated by extrapolating a subsurface temperature profile from a depth range of 200 to 300 m, most sensitive to surface temperatures between about 1300 and 1700 CE, to the surface. The database is based on 514 temperature-depth profiles in North America. Creation of a LoST database offers the possibility of a better estimate of a preindustrial reference temperature field and thus an improvement in the ECS estimate.

The database is compared with five past millennium and five preindustrial control simulations from the PMIP3/CMIP5 archive to assess the realism of the simulated preindustrial equilibrium state by the current generation of global climate models. The paper is consistent with many previous papers in advocating that borehole temperatures are a robust complement to observational (met data) and model studies of past climate.

I recommend publishing the paper after some minor to modest revisions.

Details.

1. Appropriate databases are used. With respect to borehole temperatures, it is safe to neglect heat production (does not introduce appreciable curvature into the temperature-depth profile at the depths considered) but rock heterogeneity can be more of a problem. Extrapolation of a temperature gradient from 200 m to surface to arrive at To has an error that should be discussed.

Rock heterogeneity affects the T0 estimates via variations in thermal conductivity. In the case of inhomogeneous thermal properties, the ground temperature at a depth z (T(z)) is described as

$$T(z)=T_0+q_0\cdot R(z)+T_t(t), \qquad \text{Eq. (R1)}$$

where T_t is the surface transient perturbation, T0 is the long-term surface temperature, q_0 is the quasi-equilibrium heat flux and $R(z)$ the thermal resistance (Bullard, 1939). The thermal resistance requires measurements of thermal conductivity through the profile to be estimated. Unfortunately, the majority of borehole temperature profiles (BTPs) do not have conductivity measurements, which hampers the quantification of errors in T0 estimates arising from variations in thermal conductivity. Additionally, large variations of thermal conductivities can be inferred from the lithological log for each site. Generally, when large variations of thermal properties are indicated from the lithological log descriptions, rock samples are obtained, and if this is not possible, then the temperature log is not used in climate analysis. The BTP database employed in this work has been screened by Jaume-Santero et al. (2016), and only profiles suitable for climate studies were retained for the analysis.

As an example of a borehole site with both temperature and conductivity measurements, we provide here the T0 estimates using data from the Neil well (Canadian Arctic, see Beltrami and Taylor, 1995 for a full description of the data and the site). We find a T0 estimate of $-11.6\pm0.4\,^oC$ assuming a constant thermal conductivity for the linear regression analysis for the depth range from 200 m to 300 m. If we introduce corrections by computing the thermal resistance from the thermal conductivity as a function of depth for the same depth range, we obtain a T0 estimate of $-11.4\pm0.2\,^oC$. The error due to rock heterogeneity, therefore, is not large for the Neil well site. This result cannot be directly extrapolated to the rest of BTPs, since thermal properties vary from site to site, but it contextualizes the magnitude of the errors in T0 estimates in a typical borehole site due to variations of thermal conductivity with depth.

The new version of the manuscript describes the role of thermal conductivity in the determination of T0 values and the scarcity of thermal conductivity measurements (page 4, lines 18-25; page 8, lines 13-19).

2. Temperatures in the depth range 200 to 300 m are largely affected by surface temperatures from 300 to 700 years prior to the temperature logging as the manuscript points out, corresponding to surface temperatures from about 1300 to 1700 CE. I would like to see a comment on how much of the signal in that depth range comes from surface temperatures outside of that time window.

As the reviewer points out, the depth range of BTPs is fundamental to provide temporal context for the reconstructed surface temperature histories and T0 temperatures. We identify the depth range of 200-300 m with the temporal period 1300-1700 of the Common Era, but the effect of the Little Ice Age and the Medieval Warm Period may also affect the T0 estimates. However, the spatial extent of both events is not homogeneous over North America, which implies that not all BTPs employed here are affected by these events. Additionally, their influence should be part of any transient millennial-scale climate simulation and thus, these climate events should be represented within both the transient simulations and the BTPs.

We have added a few lines commenting on the impact of the Little Ice Age and the Medieval Warm Period on the T0 estimates (page 8, lines 19-25).

3. Figure 1 is a good illustration of the extrapolation of the borehole temperature profile to the surface. The term in the caption "linear fit of the last 100 m" is ambiguous. Say, "linear fit of bottom 100 m" or better still, "linear fit of temperatures between 200 and 300 m." The paper should also say that any thermal conductivity heterogeneity in the depth range 0 to 200 m would affect the zero depth (i.e. Surface) extrapolated temperature and ideally give a bound for how big an error that would introduce.

We have changed the caption of Figure 1 in the new version of the manuscript. Regarding the effect of the thermal conductivity heterogeneity on T0 estimates, please see our answer to the first comment.

4. Figure 2. Is the temperature scale on Fig 2(b) mislabeled or is it some kind of a non-linear scale? The colored temperature scale for Fig 2(c) and (d) needs a label and units.

Indeed, Fig. 2(b) was mislabeled in the previous version of the manuscript, and the temperature scale of Fig. 2(c, d) should read "Surface Temperature (°C)". We have corrected all these issues in the new version of the figure.

5. The paper would be improved by a discussion of various kinds of uncertainties in LoST and whether the magnitude of those uncertainties detract significantly from the goal of providing a robust preindustrial surface temperature field. Include: (a) extrapolation uncertainties for a typical borehole site. (b) whether the 514 sites are generally representative of the topography (elevation and site azimuth) of the region being modeled (scatter in extrapolated borehole temperatures in a region can vary by ~ 4 oC). (c) Are the BTT's in Fig 2(b) corrected for elevation or are elevation differences at particular latitude (considerable in North America) the cause of about 10 oC scatter at constant elevation?

We have included a paragraph in the Discussion section addressing the reviewer's points (from line 12 in page 8 to line 2 in page 9).

Overall this is a refreshing new approach of showing how borehole temperature profiles can be used to complement the more conventional meteorological and GCM modeling studies to reveal the long-term evolution of surface temperature on the planet.

Dmitry Demezhko (Referee)

Review of "Long-term Surface Temperature (LoST) Database as a complement for GCM preindustrial simulations" By: Cuesta-Valero, et al.

General comments: The development and improvement of climate models (GSM) is a leading scientific method for understanding the Earth's climate system and its forecast- ing. Despite considerable efforts in the development of these models, there remains a large uncertainty in GCM scenarios. Authors suggest a new Long-term Surface Temperature (LoST) Database as "a reference to narrow down the spread of surface temperature climatologies on GCM preindustrial control and past millennium simulations". Preindustrial (1300-1700) ground surface temperatures cover North America and obtained from borehole temperature profiles (BTP) analysis. A robust paleotemperature evaluation technique based on extrapolating the temperature profile from the interval 200-300 m to the Earth surface has been used. Unlike BTP inversion methods, this simple technique is not burdened by the uncertainties associated with the choice of the algorithm for the inverse problem solving, and provides comparable estimates of paleotemperatures. I suppose the paper describes a new and important result for the development of climatology and can be published in the CP.

Specific comments

1. The simplicity of the technique used does not obviate the need to justify it. Strictly speaking, the extrapolation of temperature profiles from the interval of 200-300 m provides a very approximate estimate of the mean ground surface temperature in 1300-1700. It is necessary to provide a justification or refer to the paper where it was done (for example, “First-order estimate of the GST history” technique by Pickler et al., 2016).

Although the original manuscript already cited the work of Pickler et al., we have added another reference to this paper in the description of the T0 estimates as suggested by the reviewer (page 5, line 2).

First-order estimate technique is based on the use of formula (2), but its description “. . .the recorded temperature at a depth z can be related to an estimate of time (t)” is incorrect. Correctly: t is the time after which the temperature anomaly dT appeared at the surface reach 0.16dT at a depth of z and 0.005 at a depth of 2z. Therefore, if we assume that 0.16 dT is a negligible part of the anomaly, we should replace the description on Figure 1

from: z=300m->t≈1300,z=200m->t≈1700

to: z=300m→t>700yearsago,z=200m→t>300yearsago

We have made the suggested changes in Figure 1 and we have modified the conflicting definition of t in the new version of the manuscript (page 5, lines 4-5).

2. A large sample of BTP data was used. Obviously, many temperature profiles revealed evidences of non-climatic influences within the studied interval (hydrogeology, heterogeneity of thermal properties). Did the authors select (or correct) the initial data and by what criteria?

Yes, the BTP database was previously filtered to remove profiles containing non-climatic signals as described in Jaume-Santero et al. (2016).

We have acknowledged this in the description of the borehole data (page 4, lines 9-12).

3. P2,L32-34: “. . .BTP measurements have been employed to estimate . . . surface flux histories over the last centuries (e.g., Beltrami, 2002; Beltrami et al., 2002, 2006)”. Here the authors refer only to themselves. Meanwhile, the possibility of estimating the surface heat flux changes from ground surface temperature changes was formulated by Wang, and Bras (1999).

With regard to borehole temperature data, this technique (besides the mentioned papers) was developed in (Huang, 2006; Demezhko and Gornostaeva.2015a,b).

We have expanded our references as indicated by the reviewer, excluding the Huang (2004) work which does not employ BTP measurements for his flux estimates (page 2, lines 33-35).

In the last two papers an alternative measure of the Earth's climatic sensitivity has been proposed as the ratio between the ground surface flux changes and external fluxes changes. I believe that estimates of preindustrial surface heat flux changes can also be useful for GCM simulations, as well as estimates of paleotemperatures. I would like the authors to raise this question in the "Discussion" section.

The reviewer rises an interesting idea here; the possible estimation of the climate sensitivity using reconstructions of past changes in surface heat fluxes from BTP measurements. Such method for estimating the climate sensitivity should be further investigated, but we think that such investigation is beyond of the scope of this work.

References

Pickler, C., Beltrami, H., and Mareschal, J.-C.: Laurentide Ice Sheet basal temperatures during the last glacial cycle as inferred from borehole data, Climate of the Past, 12, 115–127, 2016.

Wang, J. and Bras, R. L.: Ground heat flux estimated from surface soil temperature, J. Hydrol., 216, 214–226, 1999.

Huang, S.: 1851–2004 annual heat budget of the continental landmasses, Geophys. Res. Lett., 33, L04707, doi:10.1029/2005GL025300, 2006

Demezhko, D. Y., & Gornostaeva, A. A. Late Pleistocene–Holocene ground surface heat flux changes reconstructed from borehole temperature data (the Urals, Russia). Climate of the Past, 11(4), 647-652, 2015a.

Demezhko, D. Y., & Gornostaeva, A. A. Reconstructions of ground surface heat flux variations in the urals from geothermal and meteorological data. Izvestiya, Atmospheric and Oceanic Physics, 51(7), 723-736, 2015b.

Editor Decision: Publish subject to minor revisions (review by editor) (13 Apr 2019) by Irina Rogozhina.

Comments to the Author:

Dear Francisco and co-authors

I have now evaluated your responses to reviews and suggest that in addition to modifying your initial version of the manuscript by addressing the reviewers' comments, you should consider including a discussion of other paleoclimate proxy data available in North America (e.g., PAGES 2k consortium, 2017) and how your new dataset can complement them when it comes to the validation of climate model reconstructions.

Good luck with the revisions. I look forward to seeing the new version of your manuscript.

Kind regards,

Irina

References:

PAGES 2k consortium (2017). A global multiproxy database for temperature reconstructions of the Common Era. Scientific Data, 4, 170088.

As suggested by the Editor, we have included a paragraph in the Discussion section commenting on the differences between proxy and borehole techniques and the role of the LoST database as a complementary tool for evaluating climate simulations of preindustrial times (from line 33 in page 9 to line 10 in page 10).

References

Beltrami, H. and Taylor, A. E.: Records of climatic change in the Canadian Arctic: towards calibrating oxygen isotope data with geothermal data. *Global and Planetary Change*, 11, 127-138, 1995.

Bullard, E. C. and Schonland, B. F. J.: Heat flow in South Africa. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 173, 474–502, 1939.

Cook, B.I., Ault, T.R. and Smerdon, J.E.: Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Science Advances*, 1, e1400082, 2015.

Jaume-Santero, F., Pickler, C., Beltrami, H., and Mareschal, J.-C.: North American regional climate reconstruction from ground surface temperature histories. *Climate of the Past*, 12, 2181-2194, 2016.

PAGES 2K-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Climate of the Past,* 11, 1673-1699, 2015.

Long-term Surface Temperature (LoST) Database as a complement for GCM preindustrial simulations

Francisco José Cuesta-Valero[1,2], Almudena García-García[1,2], Hugo Beltrami[2], Eduardo Zorita[3], and Fernando Jaume-Santero[2,4]

[1]Environmental Sciences Program, Memorial University of Newfoundland, St. John's, NL, Canada.
[2]Climate & Atmospheric Sciences Institute, St. Francis Xavier University, Antigonish, NS, Canada.
[3]Institute of Coastal Research, Hemlholtz-Zentrum Geesthacht, Germany.
[4]Departamento de Física de la Tierra y Astrofísica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040, Madrid, Spain.

Correspondence: H. Beltrami, Climate & Atmospheric Sciences Institute, Department of Earth Sciences, St. Francis Xavier University, 5009 Chapel Square, Antigonish, NS, B2G 2W5, Canada (hugo@stfx.ca).

Abstract. Estimates of climate sensitivity from General Circulation Model (GCM) simulations still present a large spread despite the continued improvements in climate modeling since the 1970s. This variability is partially caused by the dependence of several long-term feedback mechanisms on the reference climate state. Indeed, state-of-the-art GCMs present a large spread of control climate states probably due to the lack of a suitable reference for constraining the climatology of preindustrial simulations. We assemble a new gridded database of long-term ground surface temperatures (LoST database) obtained from geothermal data over North America, and we explore its use as a potential reference for the evaluation of GCM preindustrial simulations. We compare the LoST database with observations from the CRU database, as well as with five past millennium transient climate simulations and five preindustrial control simulations from the third phase of the Paleoclimate Modelling Intercomparison Project (PMIP3) and the fifth phase of the Coupled Model Intercomparison Project (CMIP5). The database is consistent with meteorological observations as well as with both types of preindustrial simulations, which suggests that LoST temperatures can be employed as a reference to narrow down the spread of surface temperature climatologies on GCM preindustrial control and past millennium simulations.

1 Introduction

General Circulation Model (GCM) simulations of the Earth's climate are sophisticated tools that reproduce many physical processes of the climate system, helping to understand and characterize the dynamics of the climate system both at global and regional scales, as well as from decadal to millennial timescales (Flato et al., 2013). Despite the large number of different GCMs developed and maintained by modeling groups around the world, future projections of climate change still present a large degree of uncertainty (Knutti and Sedláček, 2012), mainly due to the different climate sensitivity achieved by each model. The Equilibrium Climate Sensitivity (ECS) is typically defined as the change in equilibrium temperature given a doubling of atmospheric CO_2 concentration (Gregory et al., 2002), and it is considered one of the most important metrics to understand the long-term evolution of the climate system. Numerous studies, nonetheless, have unsuccessfully tried to narrow the uncertainty range of ECS using observations, simulations and paleoreconstructions, the best estimates of ECS remaining between $1.5-4.5\ ^\circ\mathrm{C}$ since the 1970s (Knutti et al., 2017).

The large uncertainty in ECS estimates is also present in state-of-the-art GCMs (Andrews et al., 2012; Flato et al., 2013; Forster et al., 2013; Knutti et al., 2017), mainly as result of approximating the description of several climate phenomena, tuning practices, and the spread in control climate states. Each GCM approximates and resolves the differential equations governing the evolution of the climate system using different numerical methods and algorithms, leading to a diverse representation of the climate evolution within the array of models (Dommenget, 2016). Additionally, each GCM employs different parameterizations for approximating processes that cannot be resolved due to the lack of spatial resolution or computational resources, such as radiative transfer, convection or clouds (McFarlane, 2011; Sen Gupta et al., 2013). All these necessary approximations add inconsistencies to simulations, affecting the simulated climate state and trajectory. Parameterizations of radiative forcing by CO_2 in climate models are of special importance, being responsible for nearly 50% of the uncertainties in the estimated values of ECS (Soden et al., 2018). Another practice related to parameterizations that affects the simulated ECS is model tuning (Mauritsen et al., 2012; Hourdin et al., 2017; Schmidt et al., 2017). Tuning practices consist in varying model parameters, whose values are poorly constrained by theory or observations or not constrained at all, to obtain a simulated climate evolution compatible with observations. Thereby, this parameter adjustment affects the representation of feedback mechanisms and other physical processes within the model, altering the response to external forcings (Mauritsen et al., 2012; Schmidt et al., 2017).

Furthermore, the magnitude of some important long-term feedback mechanisms depends on the mean climate state - i.e., the model response to external forcings is itself mean state dependent (Dommenget, 2016; Hu et al., 2017, and references therein). Ice-albedo and water vapor feedbacks are two important processes affected by the control climate state (Hu et al., 2017). The strength of both feedbacks is associated with simulated absolute values of surface temperature, since absolute temperature is the main factor governing water phase changes on the Earth. Permafrost stability, and thus permafrost carbon feedback, also depends on the reference climatology and the simulated climate trajectory (Slater and Lawrence, 2013). Although many GCMs are still in the process of implementing permafrost modules in their code, several studies have suggested that the impact of the permafrost carbon feedback on climate evolution would be important (e.g., Koven et al., 2011; MacDougall et al., 2012). Therefore, a constrained preindustrial control simulation may improve the representation of those feedbacks in transient climate experiments, reducing the uncertainty of ECS estimates from model simulations, as well as reducing the spread in projections of future climate change (Dommenget, 2016; Hu et al., 2017). At this point, estimates of preindustrial long-term surface temperatures from geothermal data may be an useful reference for assessing wether the simulated surface temperature climatology is realistic enough in preindustrial climate simulations. Additionally, such preindustrial long-term absolute temperatures may be employed to define a preindustrial baseline from which to evaluate the temperature change due to the anthropogenic influence on climate (Hawkins et al., 2017).

Borehole Temperature Profile (BTP) measurements have been employed to estimate both global and regional past trends of surface temperature (e.g., **Vasseur et al. 1983**; Huang et al. 2000; Harris and Chapman 2001; Beltrami 2002; Beltrami and Bourlon 2004) and surface flux histories over the last centuries (e.g., **Wang and Bras 1999**; Beltrami 2002; Beltrami et al. 2002, 2006; **Demezhko and Gornostaeva 2015a, b**). Several studies have validated the borehole methodology using past millennium simulations from the ECHO-G GCM (González-Rouco et al., 2006; González-Rouco et al., 2009) and the PMIP3/CMIP5 GCMs (García-García et al., 2016), reinforcing results retrieved from subsurface temperature. Reconstructions of surface temperature and surface flux from borehole measurements have been compared with ECHO-G millennial simulations (Stevens et al., 2008; MacDougall et al., 2010), as well as with estimates of continental heat storage from CMIP5 GCM simulations (Cuesta-Valero et al., 2016). All these direct comparisons between BTP estimates and GCM simulations have revealed several strengths and weaknesses of GCM simulations, and have contributed to the improvement of the represented physical processes relevant for the climate evolution within land surface model components (e.g., Alexeev et al., 2007; MacDougall and Beltrami, 2017).

Here, we propose the use of long-term surface temperatures estimated from BTP measurements as an additional tool to better evaluate the realism of surface temperature climatology within GCM preindustrial simulations, and thereby to help to improve the representation of mean state dependent feedbacks. These long-term surface temperatures are retrieved from the quasi-equilibrium state of the subsurface thermal regime at the location of each BTP measurement. This is estimated from the deepest section of the temperature profile, which is the part least affected by the recent changes in the surface energy balance. The subsurface temperature at the bottom part of each temperature profile depends linearly on depth, and the extrapolation of this linear behavior to the surface is interpreted as the long-term mean surface temperature at each borehole site (e.g. Huang et al., 2000; Harris and Chapman, 2001; Beltrami, 2002). We present here a gridded Long-term Surface Temperature (LoST) database for most of continental North America and three Caribbean islands (Cuba, Hispaniola and Puerto Rico) using 514 BTP measurements. This database is freely available for the scientific community at (Add link to download LoST database). The database is compared with five past millennium and five preindustrial control simulations from the PMIP3/CMIP5 archive to assess the realism of the simulated preindustrial equilibrium state by the current generation of global climate models.

2 Data

2.1 Meteorological data: Climate Research Unit (CRU) data

We employ surface air temperatures from the University of East Anglia Climatic Research Unit's (CRU) TS4.01 gridded dataset (Harris et al., 2014) for evaluation purposes. This dataset consists in a gridded set of climate variables derived from meteorological observations worldwide. Sources of meteorological data include several national meteorological services, CRU archives, the World Meteorological Organization (WMO) and the National Oceanic and Atmospheric Administration (NOAA). Surface air temperature are supplied on a monthly resolution for continental areas except for Antarctica from 1901 to 2016 of the Common Era (CE).

2.2 GCM data

We use five Past Millennium (PM) and five preindustrial control (piControl) GCM simulations (see Table 1 for references) from the third phase of the Paleoclimate Modelling Intercomparison Project and the fifth phase of the Coupled Model Intercomparison Project (PMIP3/CMIP5) (Braconnot et al., 2012; Taylor et al., 2011) to test the LoST database. PM simulations (Past1000 experiment in the PMIP3/CMIP5 database) simulate the climate response to prescribed external forcings from Schmidt et al. (2011) for the period 850-1850 CE, including orbital variations, changes in solar activity, greenhouse gas concentrations, volcanic eruptions and changes in land use and land cover. Each PMIP3/CMIP5 GCM also performs a piControl simulation forced with agreed preindustrial forcings to provide a baseline from which to start transient climate experiments. For more details about the PMIP3/CMIP5 control simulations and initialization procedures see Sen Gupta et al. (2013) and Séférian et al. (2016).

2.3 Borehole data

Here, we use estimates of long-term surface temperatures ~~from 514 BTP measurements over North America~~ **from the database described in** Jaume-Santero et al. (2016). **The BTP measurements of this database have been previously filtered excluding profiles with non-climatic signals and artifacts, thus providing 514 BTPs suitable for climate studies over North America.** The standard methodology to retrieve past temperature and flux histories from geothermal data assumes that each borehole temperature profile results from the superposition of a transient perturbation due to the changes in the surface energy balance with time and the quasi steady-state of the subsurface thermal regime (e.g. Huang et al., 2000; Harris and Chapman, 2001; Beltrami, 2002). Therefore, considering the subsurface as a ~~homogeneous~~ half space without heat production from radioactive decay or advection, the solution of the heat diffusion equation for a temperature profile can be approximated as (e.g., Jaume-Santero et al., 2016) ~~(Carslaw and Jaeger, 1959)~~ :

$$\cancel{T(z) = T_0 + \Gamma \cdot z + T_t(t),} T(z) = T_0 + q_0 \cdot R(z) + T_t(t), \tag{1}$$

where $\mathrm{T_t}$ is the surface transient perturbation, $\mathrm{T_0}$ is the long-term surface temperature ($\mathrm{T_0}$ temperature hereafter)**, q_0 is the subsurface flux at equilibrium** and ~~Γ is the subsurface thermal gradient at equilibrium~~ **$R(z)$ is the thermal resistance (Bullard and Schonland, 1939)**. **Estimates of thermal resistance require measurements of thermal conductivity through the subsurface profile, but the majority of available BTPs does not present such conductivity data. Thus, the thermal conductivity is assumed to be constant and Eq. (1) is rewritten as**

$$T(z) = T_0 + \Gamma \cdot z + T_t(t), \tag{2}$$

where Γ is the subsurface thermal gradient at equilibrium. The recorded surface transient perturbation ($\mathrm{T_t}$) can be retrieved from each temperature profile, once the subsurface thermal equilibrium is estimated (for more details about the borehole methodology, see Mareschal and Beltrami 1992; Bodri and Cermak 2007; Jaume-Santero et al. 2016). As the heat flux from the Earth's interior remains stable at time scales of millions of years and the deepest part of a BTP is the least affected by the recent changes in the surface energy balance, the quasi-equilibrium state of the subsurface thermal regime can be estimated from the deepest temperatures of each borehole profile (see scheme in Fig. 1). Once vertical variations in thermal properties of the subsurface rocks are taken into account, temperature depends linearly on depth at the bottom part of the temperature profile, allowing to approximate the subsurface thermal equilibrium by a linear least-squares regression. The extrapolation of this linear behavior to the surface can be interpreted as the long-term mean surface temperature at each borehole location (T_0 temperature in Eq. 2 and Fig. 1, see Pickler et al. 2016 for further details). Depending on the profile's depth, the T_0 temperatures represent the long-term ground surface temperature for a determined period of time. Due to the nature of heat diffusion through the ground, ~~the recorded temperature at a depth z can be related to an estimate of time (t) following the equation~~ **the required time (t) for a change in the surface energy balance to reach a certain depth (z) is given by** (Carslaw and Jaeger, 1959; Pickler et al., 2016):

$$t \approx \frac{z^2}{4\kappa}, \tag{3}$$

where κ is the thermal diffusivity of the subsurface. We consider $\kappa = 1 \times 10^{-6}\ \mathrm{m^2 s^{-1}}$ for all BTP measurements (Cermak and Rybach, 1982). In this study, all BTPs are truncated at the same depth ($300\ \mathrm{m}$) to ensure that all T_0 temperatures are estimated for the same temporal period. We use the last hundred meters of each BTP to estimate the subsurface thermal equilibrium, obtaining an estimated temporal period that approximately ranges from ~ 1300 CE ($z = 300\ \mathrm{m}$) to ~ 1700 CE ($z = 200\ \mathrm{m}$). Thereby, this period of time provides a baseline to compare with long-term temperatures from the PMIP3/CMIP5 PM simulations. However, the estimated temporal period is not homogeneous as result of the non-linear relationship between time and depth (Beltrami and Mareschal, 1995), and thus estimates of recent years (i.e., 1700 CE) are better determined than estimates of past years (1300 CE). **Influences of long-term perturbations of the past surface energy budget outside of that temporal window can also be detected in the employed depth range, see Section 5 for more details.**

3 The LoST database

In order to provide with a gridded dataset over continental North America, T_0 temperatures from BTP measurements are spatially interpolated to a $0.5° \times 0.5°$ grid using the Gradient plus Inverse Distance Squared (GIDS) technique. The GIDS method (Nalder and Wein, 1998) relies on the multiple linear regression of observed climate variables to retrieve longitudinal, latitudinal and altitudinal gradients that are employed to estimate values for gridded nodes. The contribution of each measurement is inverse-weighted by their squared distance to the target node, while the coefficients from the regression analysis allow to correct for the location of each measurement:

$$V_0 = \frac{\sum_{i=1}^{N} \left[V_i + (lat_0 - lat_i)C_{lat} + (lon_0 - lon_i)C_{lon} + (z_0 - z_i)C_z\right] d_i^{-2}}{\sum_{i=1}^{N} d_i^{-2}}, \tag{4}$$

where $\mathrm{V_0}$ is the predicted variable at the target node, $\mathrm{V_i}$, $\mathrm{lat_i}$, $\mathrm{lon_i}$ and $\mathrm{z_i}$ represent the variable, latitude, longitude and altitude of the $\mathrm{i^{th}}$ measurement respectively, $\mathrm{lat_0}$, $\mathrm{lon_0}$ and $\mathrm{z_0}$ represent the latitude, longitude and altitude of the target node respectively, $\mathrm{C_{lat}}$, $\mathrm{C_{lon}}$ and $\mathrm{C_z}$ are the coefficients from the regression analysis, and $\mathrm{d_i}$ is the distance from the $\mathrm{i^{th}}$ measurement to the target node. The propagation of known errors in the GIDS algorithm is described in Section S1. The GIDS technique has been used to interpolate surface temperature, precipitation, evapotranspiration and other climate variables in several zones of the world including North America (e.g., Price et al., 2000; Mardikis et al., 2005). Furthermore, the GIDS method performs well in comparison with other broadly used interpolation techniques like co-kriging or smoothing splines (ANUSPLIN suite)

(Nalder and Wein, 1998; Price et al., 2000; Li and Heap, 2011), and it has been previously employed to downscale CMIP5 simulations (McCullough et al., 2016).

Since the $\mathrm{T_0}$ dataset employed here provides latitudes and longitudes for each temperature profile, we expand the database estimating the altitude above sea level for each BTP measurement from the second version of the 2-minute Gridded Global Relief Data (ETOPO2) of the National Oceanic and Atmospheric Administration (National Geophysical Data Center, 2006. Two-minute Gridded Global Relief Data (ETOPO2) v2. National Geophysical Data Center, NOAA. doi:10.7289/V5J1012Q [last accessed on July 7th, 2017]). For this study, the regression analysis of $\mathrm{T_0}$ temperatures considering latitude, longitude and altitude yields robust results, with a R^2 value of 0.865 and a p-Value $\ll 0.05$. The distance from the measurements to the nodes is computed using the Vincenty's formula for an ellipsoid with different major and minor axes (Vincenty, 1975), and therefore the altitude of both measurements and grid nodes are not considered in our distance calculations.

We performed a pseudo-proxy experiment (e.g., Smerdon, 2012) to determine which is the maximum appropriate distance from a grid node to a BTP measurement to interpolate the $\mathrm{T_0}$ temperatures. That is, we use the long-term mean ground surface temperatures for the period 1300-1700 CE from the five PMIP3/CMIP5 PM simulations as surrogate realities, and apply the interpolation methodology employed to create the LoST database. Thereby, these GCM simulations were regridded to a $0.5° \times 0.5°$ grid, considering grid cells containing BTP measurements as reference for applying Eq. 4 to the rest of grid cells. Then, Root-Mean Squared Errors (RMSEs) between the interpolated data and the remapped simulations were computed (Fig. S1). We set $650\ \mathrm{km}$ as maximum distance criterion since this is the maximum distance at which the RMSE is lower than $1.0\ °\mathrm{C}$ for the five simulations. Such distance criterion, nevertheless, produces results for three grid cells in the Yucatan peninsula (Mexico), which we consider unjustifiable as there are no BTP measurements in or near that part of Mexico. Those grid cells are therefore masked out from our analysis.

4 Results

The distribution of LoST temperatures at grid cells containing BTP measurements reproduces the shape of the distribution of raw $\mathrm{T_0}$ temperatures (Fig. 2a), indicating that the GIDS interpolation does not substantially modify the shape of the original distribution of temperatures retrieved from BTP measurements. However, the distribution of the entire LoST database resembles the distribution of CRU temperatures, differing from the distribution of the raw $\mathrm{T_0}$ temperatures. This change in the temperature distribution after the spatial interpolation may be related to the inclusion of interpolated temperatures at higher and lower latitudes than the raw $\mathrm{T_0}$ temperatures, as the majority of BTP measurements cover from $35\ °\mathrm{N}$ to $60\ °\mathrm{N}$. Nonetheless, the latitudinal mean temperatures from the LoST database are consistent with $\mathrm{T_0}$ temperatures from BTP measurements, either considering only grid cells with BTP measurements or the entire LoST database (Fig. 2b). The latitudinal mean temperatures from the LoST database reach higher values than the CRU database at latitudes higher than $\sim 50\ °\mathrm{N}$, while both datasets achieve similar mean temperatures at lower latitudes (Fig. 2b). Previous studies have found warmer ground temperatures than air temperatures in meteorological observations over North America, probably due to the insulating effect of snow cover during winter (e.g., Beltrami and Kellman, 2003; Smerdon et al., 2003). That is, warmer temperatures should be expected for the

LoST database than for the CRU database, as our results show (Fig. 2a and 2b). It should be noted, nevertheless, that the CRU database covers a period with a marked global temperature increase (Hartmann et al., 2013). Therefore, estimates of long-term surface temperatures from CRU data reflect such temperature increase, hindering the direct comparison between both datasets. Despite this difference in the climatology of both databases, the long-term surface temperature from the LoST dataset reproduces the expected spatial pattern of temperatures for North America (Figs. 2c and 2d), in agreement with long-term surface temperatures estimated from BTP measurements and with long-term surface temperatures from CRU data.

The LoST temperatures were also compared with long-term surface temperature estimates from five Past Millennium (PM) and five piControl simulations (Table 1) included in the PMIP3/CMIP5 archive to test the realism of forced and control GCM simulations in reproducing estimates of long-term surface temperatures. Long-term surface temperatures from the PM simulations are estimated as the mean surface air temperature for the period 1300-1700 CE ($\mathrm{SAT_0}$) and the mean ground surface temperature linearly interpolated at $1.0\,\mathrm{m}$ depth for the same period ($\mathrm{GST_0}$), in order to be consistent with the estimated temporal range for $\mathrm{T_0}$ temperatures in Section 2.3. The PMIP3/CMIP5 simulations are interpolated onto the grid of the LoST database; $\mathrm{SAT_0}$ and $\mathrm{GST_0}$ values are estimated only at grid cells containing LoST temperatures. $\mathrm{SAT_0}$ and $\mathrm{GST_0}$ values are also estimated for piControl simulations following the same method, but averaging over each entire control simulation.

Surface temperatures from PMIP3/CMIP5 PM and piControl simulations show similar latitudinal patterns to that from the LoST database, with lower temperatures at northern latitudes and higher temperatures at southern latitudes (Figs. S2 and S3). $\mathrm{SAT_0}$ estimates from the CCSM4, the MRI-CGCM3 and the BCC-CSM1.1 models show generally lower values than LoST temperatures for both piControl and PM simulations, while $\mathrm{GST_0}$ estimates show higher values than LoST temperatures at high latitudes for the same GCM simulations (Figs. S4 and S5). Such result is in agreement with previous analyses of air and ground temperature relationship within GCM simulations (González-Rouco et al., 2003, 2006; Stieglitz and Smerdon, 2007; Koven et al., 2013; García-García et al., 2016) and meteorological observations over North America (e.g., Smerdon et al., 2003; Beltrami and Kellman, 2003). In contrast, MPI-ESM-P and GISS-E2-R simulations present lower $\mathrm{SAT_0}$ and $\mathrm{GST_0}$ values than LoST temperatures, indicating lower long-term ground surface temperatures than the rest of the models (Table 1 and Figs. S4 and S5). The comparison of the mean LoST temperature over North America with the simulated temperature evolution by each GCM shows three different behaviors within the PMIP3/CMIP5 ensemble. The CCSM4 and the BCC-CSM1.1 simulations present lower mean air temperatures and higher mean ground temperatures than the mean LoST temperature (Fig. 3 and Table 1). The similar $\mathrm{GST_0}$ and mean ground surface temperatures for the CCSM4 and the BCC-CSM1.1 GCMs in both PM and piControl simulations were expected since these models use a similar land surface model component (Wu et al., 2014) and the simulated ground temperatures by CMIP5 models are highly dependent on the employed land surface model component (Slater and Lawrence, 2013, García-García et al., submitted to *Journal of Geophysical Research - Atmospheres*). In contrast, the GISS-E2-R and the MPI-ESM-P models produce lower mean $\mathrm{GST_0}$ values than the mean LoST temperature and the rest of models, while simulating similar $\mathrm{SAT_0}$ values to those from the rest of the PMIP3/CMIP5 GCMs. Previous results have shown that the MPI-ESM-P PM simulation yields a high air-ground temperature coupling (García-García et al., 2016), probably due to the omission of latent heat of fusion in soil water (Koven et al., 2013). This could cause the low ground surface temperature simulated by the MPI-ESM-P model in both PM and piControl simulations in comparison with the mean LoST temperature

(Fig 3). A strong air-ground coupling may also cause the low ground surface temperature in the GISS-E2-R simulations, since the magnitude of the difference between GST_0 and SAT_0 is similar to that from the MPI-ESM-P simulations (Table 1). Finally, the MRI-CGCM3 PM simulation yields GST_0 values below the LoST climatology, but only by $0.3\ ^\circ\mathrm{C}$ ($0.1\ ^\circ\mathrm{C}$ if considering the 2σ range of the LoST climatology, Fig. S6), which are relatively small in comparison with the differences between the LoST climatology and the GST_0 values from MPI-ESM-P and GISS-E2-R simulations ($> 2.0\ ^\circ\mathrm{C}$, Table 1). Thus, we can consider that three of the five PMIP3/CMIP5 GCMs (the CCSM4, the MRI-CGCM3 and the BCC-CSM1.1) simulate a surface temperature climatology, in the PM (1300-1700 CE) and piControl simulations, comparable to that from the LoST dataset, which is an unexpected result as none of the PMIP3/CMIP5 GCM simulations studied here were specifically tuned to match this climatology.

5 Discussion

Our results demonstrate that LoST temperatures can be used as reference for assessing the represented climatology in both PM and piControl simulations. **The determination of T_0 temperatures, nevertheless, presents some uncertainties that should be discussed. The extrapolation of each quasi-equilibrium temperature profile to the surface introduces a small error in the LoST estimates, averaging less than $0.15\ ^\circ\mathrm{C}$ from the 514 BTPs evaluated here (see Section S1 for details about the error treatment in the LoST database). Rock heterogeneity should also be considered for estimating T_0 temperatures. We assume, nevertheless, homogenous thermal properties for all borehole profiles, which is another source of uncertainty for LoST temperatures. The ideal approach consists in estimating the thermal resistance with depth (Eq. 1), but the absence of thermal conductivity measurements for the vast majority of the employed BTPs makes that approach impractical. Long-term alterations of the surface energy balance out of the 1300-1700 CE period may also affect the LoST estimates. Particularly, possible transient temperatures in BTPs due to the Little Ice Age and the Medieval Warm Period add a certain degree of uncertainty in the determination of T_0 values. However, the spatial extend of both climate events was not homogeneous over North America, meaning that not all BTPs were affected by the events. Additionally, the influence of both the Little Ice Age and the Medieval Warm Period should be part of any millennial-scale transient climate simulation, and therefore the effect of such climate events is taken into account in the comparison between LoST results and transient climate simulations. Another factor that may impact the retrieved quasi-equilibrium temperature profile is the heterogeneity of North American topography (e.g., Kohl, 1999). To our knowledge, all analyzed BTPs are located in plain terrain, and were not corrected for elevation since the employed BTP database does not provide elevation data. Therefore, we use the ETOPO2 database to assess if the altitude distribution of BTPs is enough for representing the topography of the LoST domain. The altitude distribution over the LoST domain and at grid cells containing boreholes sites are displayed in Fig. S8. Both histograms present a similar shape for altitudes up to ~ 430 m, showing a lack of borehole locations at altitudes between ~ 430 m and ~ 1013 m. The uneven latitudinal distribution of borehole sites is probably causing this gap in the distribution of altitudes, as well as a small excess of**

BTP locations at high altitudes. Despite these differences, both distributions are generally in agreement, indicating a sufficient altitude distribution from the borehole database to represent the North American broad-scale topography.

There are, however, two main limitations **for the application of the LoST database** at this stage of the study: the supplied variable and the regional character of the database. The LoST database is constituted by estimates of long-term ground surface temperatures, while GCM simulations are typically evaluated against observations of surface air temperature (SAT) (e.g., Mauritsen et al., 2012; Flato et al., 2013; Séférian et al., 2016; Schmidt et al., 2017). We can provide a reference for simulated long-term SAT by accounting for the offset between simulated air and ground temperatures and using the LoST temperatures. As an example, SAT references are estimated for the five PM and five piControl simulations employed in this study (dashed blue line in Fig. 3). SAT references for PM simulations are estimated from the offset between air and ground temperatures in piControl simulations, while SAT references for piControl simulations are estimated from the offset between air and ground temperatures in PM simulations. Such offsets show a constant behavior in both simulations (Fig. S7). GCM simulations in disagreement with the estimated SAT reference (the MPI-ESM-P and the GISS-E2-R simulations) may be representing a strong air-ground coupling, as discussed in Section 4. Therefore, although the LoST database contains estimates of ground surface temperatures, it may be also used to assess simulated long-term surface air temperatures on a first order approach.

The regional character of the presented LoST database poses some caveats for analyzing the global climatology of preindustrial simulations. Indeed, results of the simulated regional climatology cannot be globally extrapolated since the magnitude of the potential spurious drifts in control simulations varies markedly at regional scales and these regional drifts could be larger than the global-averaged drift (Sen Gupta et al., 2012, 2013). Further work would consist in generating a global LoST database from the existing global network of BTP measurements, helping to minimize the effect of possible regional drifts on the simulated climatology. However, BTP measurements are scarce in the Southern Hemisphere, a potential burden that needs to be considered for assembling such global version of this database. Additionally, the temperature profiles employed in this study to estimate T_0 temperatures were truncated to $300\ \mathrm{m}$ of depth, which determines the temporal period of reference for the comparison with PM simulations. Deeper BTP measurements can retrieve the climatology of previous time periods, although the global BTP network contains fewer temperature profiles deeper than $300\ \mathrm{m}$ (see Fig. 1 in Beltrami et al., 2015).

Despite the regional character of the LoST temperatures, the northern BTPs contained in this database allow to evaluate the long-term stability of permafrost over North America. That is, the northern temperatures in this database can be compared with regional and global simulations as a reference to the preindustrial permafrost stability (*Jaume-Santero et al.*, in preparation). Furthermore, previous studies have found that the CMIP5 GCM simulations have difficulties to properly represent permafrost evolution (Koven et al., 2013; Slater and Lawrence, 2013), partially due to the broad range of simulated climate trajectories by each GCM and the differences between the employed land surface model components (Slater and Lawrence, 2013). Using LoST temperatures to improve the surface temperature climatology of global and regional simulations may enhance the simulated long-term preindustrial $0\ ^{\circ}\mathrm{C}$ isotherm, which is important to correctly represent permafrost evolution.

Numerous proxy-data based reconstructions of temperature, precipitation and other climate related variables exist for North America, providing a reference for the evaluation of important aspects of past and future climate model simulations (e.g., PAGES 2k-PMIP3 Group 2015; Cook et al. 2015). Proxy-data temperature reconstructions have already

been compared against borehole temperature records of past variations in surface temperature over North America (e.g., Jaume-Santero et al. 2016). It is worthy to note that proxy systems are indirect sources of climate information requiring a calibration procedure with modern meteorological data, while borehole temperature data consist of direct measurements of the thermal regime of the subsurface in the recent past. That is, the LoST database contains information derived from direct measurements of subsurface temperatures, constituting the first estimates of long-term absolute surface temperatures in North America. Another important difference between proxy and borehole reconstructions is that most proxy systems generally captures high-frequency variations of climate conditions (Moberg et al. 2005), while borehole temperature profiles record long-term changes in the surface conditions, filtering out short-period signals. In this context, LoST temperatures provide a complementary reference to the multiproxy database over North America for evaluating the performance of climate model simulations.

6 Conclusions

A gridded database of past long-term surface temperatures over most part of continental North America has been assembled from geothermal measurements. Our results show that this database can be used as reference to evaluate the realism of GCM preindustrial control and past millennium simulations and possibly to improve the reference climate state by adjusting key parameters or preindustrial forcings in control experiments. Thereby, spread in ECS estimates by GCM simulations may be reduced given the relationship between control temperature climatology and three long-term powerful feedbacks as the ice-albedo feedback, the water vapor feedback and the permafrost carbon feedback. Future work would consist in generating a global version of the LoST database using the rest of the global network of borehole temperature profile measurements and following the described methodology, as well as generating new versions of this global database including future temperature profile measurements.

Data availability. The data will be available once the manuscript is accepted.

Author contributions. FJCV, AGG and HB designed the study. FJCV generated the results and pictures. FJS provided the T0 data. All authors analyzed the results, discussed the outcome, and drafted the paper.

Competing interests. The authors declare no competing interests.

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups responsibles for the model simulations used herein (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. **We are grateful to Dmitry Demezhko, an anonymous reviewer and Irina Rogozhina for their constructive comments.** We **also** acknowledge fruitful discussions with J. Fidel González-Rouco on a previous version of the manuscript and we thank Timothy Osborn for his guidance on using the CRU database. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada Discovery Grant (NSERC DG 140576948), the Canada Research Chairs Program (CRC 230687), and the Canada Foundation for Innovation (CFI) to H. Beltrami. H. Beltrami holds Canada Research Chair in Climate Dynamics. A.G.G. and F.J.C.V. are funded by H. Beltrami's Canada Research Chair program, the School of Graduate Students at Memorial University of Newfoundland and the Research Office at St. Francis Xavier University.

References

Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E., and Lawrence, D. M.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, Geophysical Research Letters, 34, n/a–n/a, https://doi.org/10.1029/2007GL029536, http://dx.doi.org/10.1029/2007GL029536, l09502, 2007.

Andrews, T., Gregory, J. M., Webb, M. J., and Taylor, K. E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophysical Research Letters, 39, n/a–n/a, https://doi.org/10.1029/2012GL051607, http://dx.doi.org/10.1029/2012GL051607, l09712, 2012.

Beltrami, H.: Climate from borehole data: Energy fluxes and temperatures since 1500, Geophysical Research Letters, 29, 26–1–26–4, https://doi.org/10.1029/2002GL015702, http://dx.doi.org/10.1029/2002GL015702, 2111, 2002.

Beltrami, H. and Bourlon, E.: Ground warming patterns in the Northern Hemisphere during the last five centuries, Earth and Planetary Science Letters, 227, 169 – 177, https://doi.org/http://dx.doi.org/10.1016/j.epsl.2004.09.014, http://www.sciencedirect.com/science/article/pii/S0012821X04005576, 2004.

Beltrami, H. and Kellman, L.: An examination of short- and long-term air–ground temperature coupling, Global and Planetary Change, 38, 291 – 303, https://doi.org/http://dx.doi.org/10.1016/S0921-8181(03)00112-7, //www.sciencedirect.com/science/article/pii/S0921818103001127, 2003.

Beltrami, H. and Mareschal, J.-C.: Resolution of ground temperature histories inverted from borehole temperature data, Global and Planetary Change, 11, 57–70, https://doi.org/https://doi.org/10.1016/0921-8181(95)00002-9, 1995.

Beltrami, H., Smerdon, J. E., Pollack, H. N., and Huang, S.: Continental heat gain in the global climate system, Geophysical Research Letters, 29, 8–1–8–3, https://doi.org/10.1029/2001GL014310, http://dx.doi.org/10.1029/2001GL014310, 2002.

Beltrami, H., Bourlon, E., Kellman, L., and González-Rouco, J. F.: Spatial patterns of ground heat gain in the Northern Hemisphere, Geophysical Research Letters, 33, n/a–n/a, https://doi.org/10.1029/2006GL025676, http://dx.doi.org/10.1029/2006GL025676, l06717, 2006.

Beltrami, H., Matharoo, G. S., and Smerdon, J. E.: Impact of borehole depths on reconstructed estimates of ground surface temperature histories and energy storage, Journal of Geophysical Research: Earth Surface, 120, 763–778, https://doi.org/10.1002/2014JF003382, http://dx.doi.org/10.1002/2014JF003382, 2014JF003382, 2015.

Bodri, L. and Cermak, V.: Borehole Climatology, Elsevier Science Ltd, Oxford, https://doi.org/https://doi.org/10.1016/B978-008045320-0/50001-6, http://www.sciencedirect.com/science/article/pii/B9780080453200500016, 2007.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nature Clim. Change, 2, 417–424, http://dx.doi.org/10.1038/nclimate1456, 2012.

Bullard, E. C. and Schonland, B. F. J.: Heat flow in South Africa, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 173, 474–502, https://doi.org/10.1098/rspa.1939.0159, https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1939.0159, 1939.

Carslaw, H. and Jaeger, J.: Conduction of Heat in Solids, Clarendon Press, Oxford, 1959.

Cermak, V. and Rybach, L.: Thermal conductivity and specific heat of minerals and rocks, in: Landolt Börnstein: Physical Properties of Rocks, Group V, Geophysics, Volume 1a, edited by Angenheister, G., Springer-Verlag Berlin Heidelberg, https://doi.org/10.1007/10201894_62, http://materials.springer.com/lb/docs/sm_lbs_978-3-540-31364-9_62, accessed 2017-06-06, 1982.

Cook, B. I., Ault, T. R., and Smerdon, J. E.: Unprecedented 21st century drought risk in the American Southwest and Central Plains, Science Advances, 1, https://doi.org/10.1126/sciadv.1400082, http://advances.sciencemag.org/content/1/1/e1400082, 2015.

Cuesta-Valero, F. J., García-García, A., Beltrami, H., and Smerdon, J. E.: First assessment of continental energy storage in CMIP5 simulations, Geophysical Research Letters, pp. n/a–n/a, https://doi.org/10.1002/2016GL068496, http://dx.doi.org/10.1002/2016GL068496, 2016GL068496, 2016.

Demezhko, D. Y. and Gornostaeva, A. A.: Late Pleistocene–Holocene ground surface heat flux changes reconstructed from borehole temperature data (the Urals, Russia), Climate of the Past, 11, 647–652, https://doi.org/10.5194/cp-11-647-2015, https://www.clim-past.net/11/647/2015/, 2015a.

Demezhko, D. Y. and Gornostaeva, A. A.: Reconstructions of ground surface heat flux variations in the urals from geothermal and meteorological data, Izvestiya, Atmospheric and Oceanic Physics, 51, 723–736, https://doi.org/10.1134/S0001433815070026, https://doi.org/10.1134/S0001433815070026, 2015b.

Dommenget, D.: A simple model perturbed physics study of the simulated climate sensitivity uncertainty and its relation to control climate biases, Climate Dynamics, 46, 427–447, https://doi.org/10.1007/s00382-015-2591-4, https://doi.org/10.1007/s00382-015-2591-4, 2016.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., book section 9, pp. 741–866, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.020, www.climatechange2013.org, 2013.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., and Zelinka, M.: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, Journal of Geophysical Research: Atmospheres, 118, 1139–1150, https://doi.org/10.1002/jgrd.50174, http://dx.doi.org/10.1002/jgrd.50174, 2013.

García-García, A., Cuesta-Valero, F. J., Beltrami, H., and Smerdon, J. E.: Simulation of air and ground temperatures in PMIP3/CMIP5 last millennium simulations: implications for climate reconstructions from borehole temperature profiles, Environmental Research Letters, 11, 044 022, http://stacks.iop.org/1748-9326/11/i=4/a=044022, 2016.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model Version 4, Journal of Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, http://dx.doi.org/10.1175/2011JCLI4083.1, 2011.

González-Rouco, F., von Storch, H., and Zorita, E.: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, Geophysical Research Letters, 30, n/a–n/a, https://doi.org/10.1029/2003GL018264, http://dx.doi.org/10.1029/2003GL018264, 2116, 2003.

González-Rouco, J. F., Beltrami, H., Zorita, E., and von Storch, H.: Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling, Geophysical Research Letters, 33, n/a–n/a, https://doi.org/10.1029/2005GL024693, http://dx.doi.org/10.1029/2005GL024693, l01703, 2006.

González-Rouco, J. F., Beltrami, H., Zorita, E., and Stevens, M. B.: Borehole climatology: a discussion based on contributions from climate modeling, Climate of the Past, 5, 97–127, https://doi.org/10.5194/cp-5-97-2009, http://www.clim-past.net/5/97/2009/, 2009.

Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A., and Rayner, N. A.: An Observationally Based Estimate of the Climate Sensitivity, Journal of Climate, 15, 3117–3121, https://doi.org/10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2, https://doi.org/10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2, 2002.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, https://doi.org/10.1002/joc.3711, http://dx.doi.org/10.1002/joc.3711, 2014.

Harris, R. N. and Chapman, D. S.: Mid-latitude (30°–60° N) climatic warming inferred by combining borehole temperatures with surface air temperatures, Geophysical Research Letters, 28, 747–750, https://doi.org/10.1029/2000GL012348, http://dx.doi.org/10.1029/2000GL012348, 2001.

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P.: Observations: Atmosphere and Surface, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., book section 2, pp. 159–254, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.008, www.climatechange2013.org, 2013.

Hawkins, E., Ortega, P., Suckling, E., Schurer, A., Hegerl, G., Jones, P., Joshi, M., Osborn, T. J., Masson-Delmotte, V., Mignot, J., Thorne, P., and van Oldenborgh, G. J.: Estimating changes in global temperature since the pre-industrial period, https://doi.org/10.1175/BAMS-D-16-0007.1, https://doi.org/10.1175/BAMS-D-16-0007.1, in press, 2017.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning, Bulletin of the American Meteorological Society, 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1, http://dx.doi.org/10.1175/BAMS-D-15-00135.1, 2017.

Hu, X., Taylor, P. C., Cai, M., Yang, S., Deng, Y., and Sejas, S.: Inter-Model Warming Projection Spread: Inherited Traits from Control Climate Diversity, Scientific Reports, 7, 4300, https://doi.org/10.1038/s41598-017-04623-7, https://doi.org/10.1038/s41598-017-04623-7, 2017.

Huang, S., Pollack, H. N., and Shen, P.-Y.: Temperature trends over the past five centuries reconstructed from borehole temperatures, Nature, 403, 756–758, http://dx.doi.org/10.1038/35001556, 2000.

Jaume-Santero, F., Pickler, C., Beltrami, H., and Mareschal, J.-C.: 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, http://www.clim-past.net/12/2181/2016/, 2016.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, Journal of Advances in Modeling Earth Systems, 5, 422–446, https://doi.org/10.1002/jame.20023, http://dx.doi.org/10.1002/jame.20023, 2013.

Jungclaus, J. H., Lohmann, K., and Zanchettin, D.: Enhanced 20th-century heat transfer to the Arctic simulated in the context of climate variations over the last millennium, Climate of the Past, 10, 2201–2213, https://doi.org/10.5194/cp-10-2201-2014, http://www.clim-past.net/10/2201/2014/, 2014.

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nature Climate Change, 3, 369 EP –, https://doi.org/10.1038/nclimate1716, http://dx.doi.org/10.1038/nclimate1716, 2012.

Knutti, R., Rugenstein, M. A. A., and Hegerl, G. C.: Beyond equilibrium climate sensitivity, Nature Geoscience, 10, 727 EP –, https://doi.org/10.1038/ngeo3017, http://dx.doi.org/10.1038/ngeo3017, 2017.

Kohl, T.: Transient thermal effects below complex topographies, Tectonophysics, 306, 311 – 324, https://doi.org/https://doi.org/10.1016/S0040-1951(99)00063-3, http://www.sciencedirect.com/science/article/pii/S0040195199000633, 1999.

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: Permafrost carbon-climate feedbacks accelerate global warming, Proceedings of the National Academy of Sciences, 108, 14 769–14 774, https://doi.org/10.1073/pnas.1103910108, http://www.pnas.org/content/108/36/14769, 2011.

Koven, C. D., Riley, W. J., and Stern, A.: Analysis of Permafrost Thermal Dynamics and Response to Climate Change in the CMIP5 Earth System Models, Journal of Climate, 26, 1877–1900, https://doi.org/10.1175/JCLI-D-12-00228.1, http://dx.doi.org/10.1175/JCLI-D-12-00228.1, 2013.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., and Teng, H.: Last Millennium Climate and Its Variability in CCSM4, Journal of Climate, 26, 1085–1111, https://doi.org/10.1175/JCLI-D-11-00326.1, http://dx.doi.org/10.1175/JCLI-D-11-00326.1, 2013.

Li, J. and Heap, A. D.: A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors, Ecological Informatics, 6, 228 – 241, https://doi.org/http://dx.doi.org/10.1016/j.ecoinf.2010.12.003, http://www.sciencedirect.com/science/article/pii/S1574954110001147, 2011.

MacDougall, A. H. and Beltrami, H.: Impact of deforestation on subsurface temperature profiles: implications for the borehole paleoclimate record, Environmental Research Letters, 12, 074 014, http://stacks.iop.org/1748-9326/12/i=7/a=074014, 2017.

MacDougall, A. H., Beltrami, H., González-Rouco, J. F., Stevens, M. B., and Bourlon, E.: Comparison of observed and general circulation model derived continental subsurface heat flux in the Northern Hemisphere, Journal of Geophysical Research: Atmospheres (1984–2012), 115, 2010.

MacDougall, A. H., Avis, C. A., and Weaver, A. J.: Significant contribution to climate warming from the permafrost carbon feedback, Nature Geosci, 5, 719–721, https://doi.org/10.1038/ngeo1573, http://dx.doi.org/10.1038/ngeo1573, 2012.

Mardikis, M. G., Kalivas, D. P., and Kollias, V. J.: Comparison of Interpolation Methods for the Prediction of Reference Evapotranspiration—An Application in Greece, Water Resources Management, 19, 251–278, https://doi.org/10.1007/s11269-005-3179-2, http://dx.doi.org/10.1007/s11269-005-3179-2, 2005.

Mareschal, J.-C. and Beltrami, H.: Evidence for recent warming from perturbed geothermal gradients: examples from eastern Canada, Climate Dynamics, 6, 135–143, https://doi.org/10.1007/BF00193525, http://dx.doi.org/10.1007/BF00193525, 1992.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, Journal of Advances in Modeling Earth Systems, 4, n/a–n/a, https://doi.org/10.1029/2012MS000154, http://dx.doi.org/10.1029/2012MS000154, m00A01, 2012.

McCullough, I. M., Davis, F. W., Dingman, J. R., Flint, L. E., Flint, A. L., Serra-Diaz, J. M., Syphard, A. D., Moritz, M. A., Hannah, L., and Franklin, J.: High and dry: high elevations disproportionately exposed to regional climate change in Mediterranean-climate landscapes, Landscape Ecology, 31, 1063–1075, https://doi.org/10.1007/s10980-015-0318-x, http://dx.doi.org/10.1007/s10980-015-0318-x, 2016.

McFarlane, N.: Parameterizations: representing key processes in climate models without resolving them, Wiley Interdisciplinary Reviews: Climate Change, 2, 482–497, https://doi.org/10.1002/wcc.122, http://dx.doi.org/10.1002/wcc.122, 2011.

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., DelGenio, A. D., Kelley, M., Lo, K. K., Ruedy, R., Shindell, D. T., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Menon, S., Oinas, V., Pérez García-Pando, C., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Sun, S., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: CMIP5 historical simulations (1850–2012) with GISS ModelE2, Journal of Advances in Modeling Earth Systems, 6, 441–477, https://doi.org/10.1002/2013MS000266, http://dx.doi.org/10.1002/2013MS000266, 2014.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, Nature, 433, 613 EP –, https://doi.org/10.1038/nature03265, 2005.

Nalder, I. A. and Wein, R. W.: Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest, Agricultural and Forest Meteorology, 92, 211 – 225, https://doi.org/http://dx.doi.org/10.1016/S0168-1923(98)00102-6, http://www.sciencedirect.com/science/article/pii/S0168192398001026, 1998.

PAGES 2k-PMIP3 Group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Climate of the Past, 11, 1673–1699, https://doi.org/10.5194/cp-11-1673-2015, https://www.clim-past.net/11/1673/2015/, 2015.

Pickler, C., Beltrami, H., and Mareschal, J.-C.: Laurentide Ice Sheet basal temperatures during the last glacial cycle as inferred from borehole data, Climate of the Past, 12, 115–127, https://doi.org/10.5194/cp-12-115-2016, http://www.clim-past.net/12/115/2016/, 2016.

Price, D. T., McKenney, D. W., Nalder, I. A., Hutchinson, M. F., and Kesteven, J. L.: A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data, Agricultural and Forest Meteorology, 101, 81 – 94, https://doi.org/http://dx.doi.org/10.1016/S0168-1923(99)00169-0, http://www.sciencedirect.com/science/article/pii/S0168192399001690, 2000.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geoscientific Model Development, 4, 33–45, https://doi.org/10.5194/gmd-4-33-2011, http://www.geosci-model-dev.net/4/33/2011/, 2011.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y.-H., Cheng, Y., Clune, T. L., Del Genio, A., de Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, Journal of Advances in Modeling Earth Systems, 6, 141–184, https://doi.org/10.1002/2013MS000265, http://dx.doi.org/10.1002/2013MS000265, 2014.

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., Molod, A., Neale, R. B., and Saha, S.: Practice and philosophy of climate model tuning across six US modeling centers, Geoscientific Model Development, 10, 3207–3223, https://doi.org/10.5194/gmd-10-3207-2017, https://www.geosci-model-dev.net/10/3207/2017/, 2017.

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., Dunne, J. P., Christian, J. R., Doney, S. C., Ilyina, T., Lindsay, K., Halloran, P. R., Heinze, C., Segschneider, J., Tjiputra, J., Aumont, O., and Romanou, A.: Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment, Geoscientific Model Development, 9, 1827–1851, https://doi.org/10.5194/gmd-9-1827-2016, http://www.geosci-model-dev.net/9/1827/2016/, 2016.

Sen Gupta, A., Muir, L. C., Brown, J. N., Phipps, S. J., Durack, P. J., Monselesan, D., and Wijffels, S. E.: Climate Drift in the CMIP3 Models, Journal of Climate, 25, 4621–4640, https://doi.org/10.1175/JCLI-D-11-00312.1, http://dx.doi.org/10.1175/JCLI-D-11-00312.1, 2012.

Sen Gupta, A., Jourdain, N. C., Brown, J. N., and Monselesan, D.: Climate Drift in the CMIP5 Models, Journal of Climate, 26, 8597–8615, https://doi.org/10.1175/JCLI-D-12-00521.1, http://dx.doi.org/10.1175/JCLI-D-12-00521.1, 2013.

Slater, A. G. and Lawrence, D. M.: Diagnosing Present and Future Permafrost from Climate Models, Journal of Climate, 26, 5608–5623, https://doi.org/10.1175/JCLI-D-12-00341.1, http://dx.doi.org/10.1175/JCLI-D-12-00341.1, 2013.

Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, Wiley Interdisciplinary Reviews: Climate Change, 3, 63–77, https://doi.org/10.1002/wcc.149, http://dx.doi.org/10.1002/wcc.149, 2012.

Smerdon, J. E., Pollack, H. N., Enz, J. W., and Lewis, M. J.: Conduction-dominated heat transport of the annual temperature signal in soil, Journal of Geophysical Research: Solid Earth, 108, n/a–n/a, https://doi.org/10.1029/2002JB002351, http://dx.doi.org/10.1029/2002JB002351, 2431, 2003.

Soden, B. J., Collins, W. D., and Feldman, D. R.: Reducing uncertainties in climate models, Science, 361, 326–327, https://doi.org/10.1126/science.aau1864, http://science.sciencemag.org/content/361/6400/326, 2018.

Stevens, M. B., González-Rouco, J. F., and Beltrami, H.: North American climate of the last millennium: Underground temperatures and model comparison, Journal of Geophysical Research: Earth Surface, 113, n/a–n/a, https://doi.org/10.1029/2006JF000705, http://dx.doi.org/10.1029/2006JF000705, f01008, 2008.

Stieglitz, M. and Smerdon, J. E.: Characterizing Land–Atmosphere Coupling and the Implications for Subsurface Thermodynamics, Journal of Climate, 20, 21–37, https://doi.org/10.1175/JCLI3982.1, http://dx.doi.org/10.1175/JCLI3982.1, 2007.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, http://dx.doi.org/10.1175/BAMS-D-11-00094.1, 2011.

Vasseur, G., Bernard, P., de Meulebrouck, J. V., Kast, Y., and Jolivet, J.: Holocene paleotemperatures deduced from geothermal measurements, Palaeogeography, Palaeoclimatology, Palaeoecology, 43, 237 – 259, https://doi.org/https://doi.org/10.1016/0031-0182(83)90013-5, http://www.sciencedirect.com/science/article/pii/0031018283900135, 1983.

Vincenty, T.: Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations, Survey Review, 23, 88–93, https://doi.org/10.1179/sre.1975.23.176.88, http://dx.doi.org/10.1179/sre.1975.23.176.88, 1975.

Wang, J. and Bras, R.: Ground heat flux estimated from surface soil temperature, Journal of Hydrology, 216, 214 – 226, https://doi.org/https://doi.org/10.1016/S0022-1694(99)00008-6, http://www.sciencedirect.com/science/article/pii/S0022169499000086, 1999.

Wu, T., Li, W., Ji, J., Xin, X., Li, L., Wang, Z., Zhang, Y., Li, J., Zhang, F., Wei, M., Shi, X., Wu, F., Zhang, L., Chu, M., Jie, W., Liu, Y., Wang, F., Liu, X., Li, Q., Dong, M., Liang, X., Gao, Y., and Zhang, J.: Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century, Journal of Geophysical Research: Atmospheres, 118, 4326–4347, https://doi.org/10.1002/jgrd.50320, http://dx.doi.org/10.1002/jgrd.50320, 2013.

Wu, T., Song, L., Li, W., Wang, Z., Zhang, H., Xin, X., Zhang, Y., Zhang, L., Li, J., Wu, F., Liu, Y., Zhang, F., Shi, X., Chu, M., Zhang, J., Fang, Y., Wang, F., Lu, Y., Liu, X., Wei, M., Liu, Q., Zhou, W., Dong, M., Zhao, Q., Ji, J., Li, L., and Zhou, M.: An overview of BCC climate system model development and application for climate change studies, Journal of Meteorological Research, 28, 34–56, https://doi.org/10.1007/s13351-014-3041-7, http://dx.doi.org/10.1007/s13351-014-3041-7, 2014.

Xiao-Ge, X., Tong-Wen, W., and Jie, Z.: Introduction of {CMIP5} Experiments Carried out with the Climate System Models of Beijing Climate Center, Advances in Climate Change Research, 4, 41 – 49, https://doi.org/http://dx.doi.org/10.3724/SP.J.1248.2013.041, http://www.sciencedirect.com/science/article/pii/S1674927813500050, 2013.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., Mizuta, R., Yabu, S., Obata, A., Nakano, H., Koshiro, T., Ose, T., and Kitoh, A.: A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM; -Model Description and Basic Performance-, Journal of the Meteorological Society of Japan. Ser. II, 90A, 23–64, https://doi.org/10.2151/jmsj.2012-A02, 2012.

Table 1. Model name, SAT_0 estimates, GST_0 estimates, SAT_0 and GST_0 differences with the mean LoST temperatures and references for each PMIP3/CMIP5 GCM simulation. All results in $^\circ\mathrm{C}$. Ground temperatures for MRI-CGCM3 piControl simulation could not be retrieved from the PMIP3/CMIP5 data servers. Temperature average of the LoST database is $5.2\,^\circ\mathrm{C}$, with a 95% confidence interval between 5.0 and $5.4\,^\circ\mathrm{C}$.

Model	Past Millennium					Preindustrial Control				
	SAT_0	GST_0	SAT_0-LoST	GST_0-LoST	Reference	SAT_0	GST_0	SAT_0-LoST	GST_0-LoST	Reference
CCSM4	1.53	5.60	-3.65	0.37	Landrum et al. 2013	2.12	6.03	-3.07	0.80	Gent et al. 2011
MRI-CGCM3	1.38	4.84	-3.81	-0.31	Yukimoto et al. 2012	1.39	-	-3.80	-	Yukimoto et al. 2012
MPI-ESM-P	1.63	2.75	-3.56	-2.91	Jungclaus et al. 2014	2.00	3.10	-3.19	-2.56	Jungclaus et al. 2013
GISS-E2-R	1.96	3.10	-3.23	-2.42	Schmidt et al. 2014	2.02	3.14	-3.16	-2.35	Miller et al. 2014
BCC-CSM1.1	0.75	5.39	-4.44	0.22	Xiao-Ge et al. 2013	1.03	5.58	-4.15	0.42	Wu et al. 2013

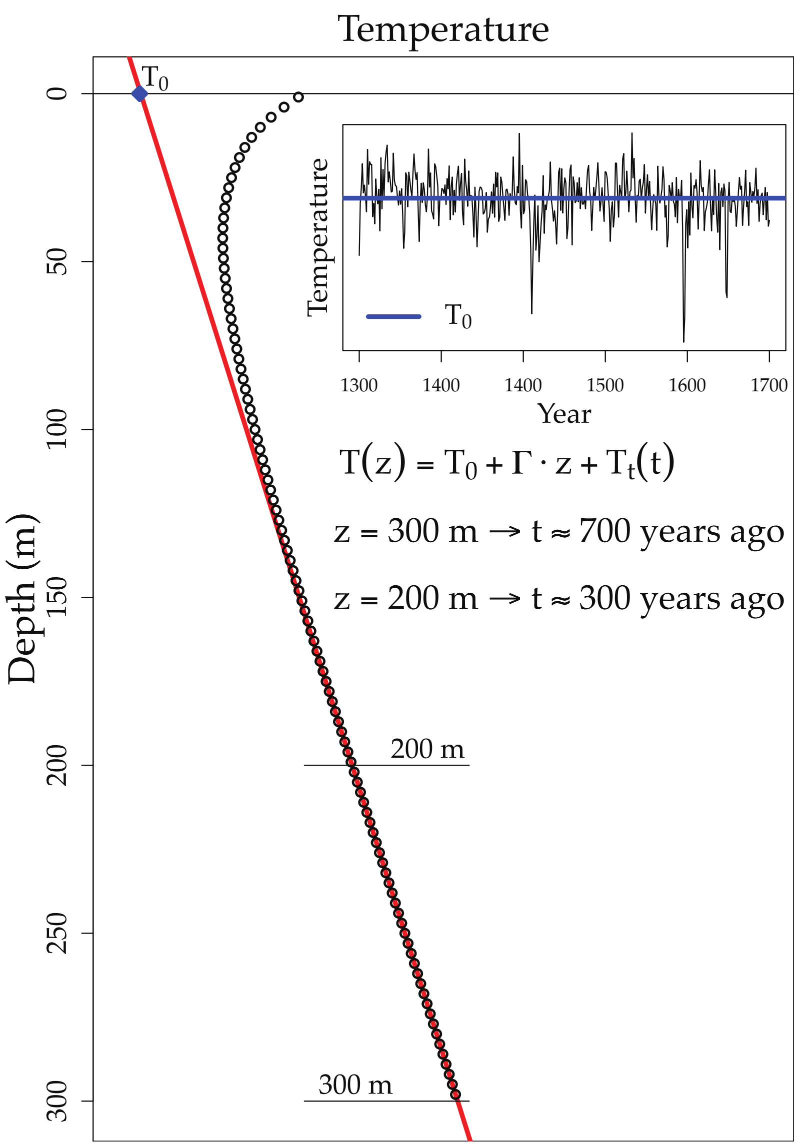

Figure 1. Synthetic borehole temperature profile (black dots) using data from the CCSM4 PM simulation (inset) and linear fit of ~~the last 100 m of the profile~~ **temperatures between** 200 m **and** 300 m (red line). The synthetic temperature profile is generated using the simulated global ground temperature anomaly at 1.0 m depth for the period 1300-1700 CE as transient perturbation ($\mathrm{T_t}$), mean ground temperature as long-term surface temperature ($\mathrm{T_0}$) and a typical thermal gradient (Γ) of $0.01\ \mathrm{K\,m^{-1}}$ (Jaume-Santero et al., 2016). The equivalence between depth (z) and time (t) is given by Eq. 3. Thermal diffusivity is considered as $\kappa = 1 \cdot 10^{-6}\ \mathrm{m^2\,s^{-1}}$ (Cermak and Rybach, 1982).

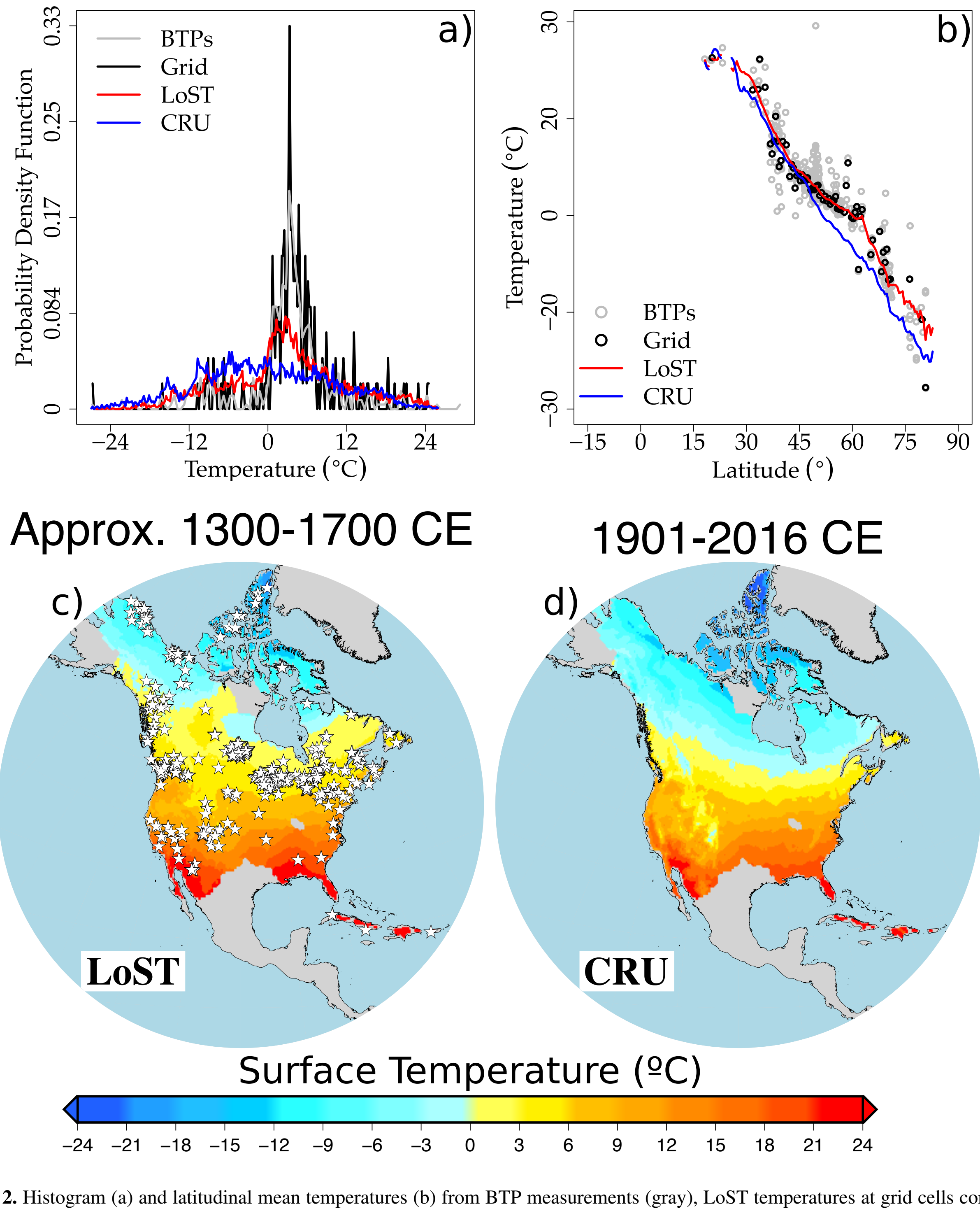

Figure 2. Histogram (a) and latitudinal mean temperatures (b) from BTP measurements (gray), LoST temperatures at grid cells containing BTP measurements (black), LoST temperatures (red) and mean surface air temperature from the CRU database (blue). LoST temperatures (∼1300-1700 CE) (c) in comparison with mean surface air temperature from CRU data (1901-2015 CE) (d). White stars in (c) indicate the location of the 514 BTP measurements.

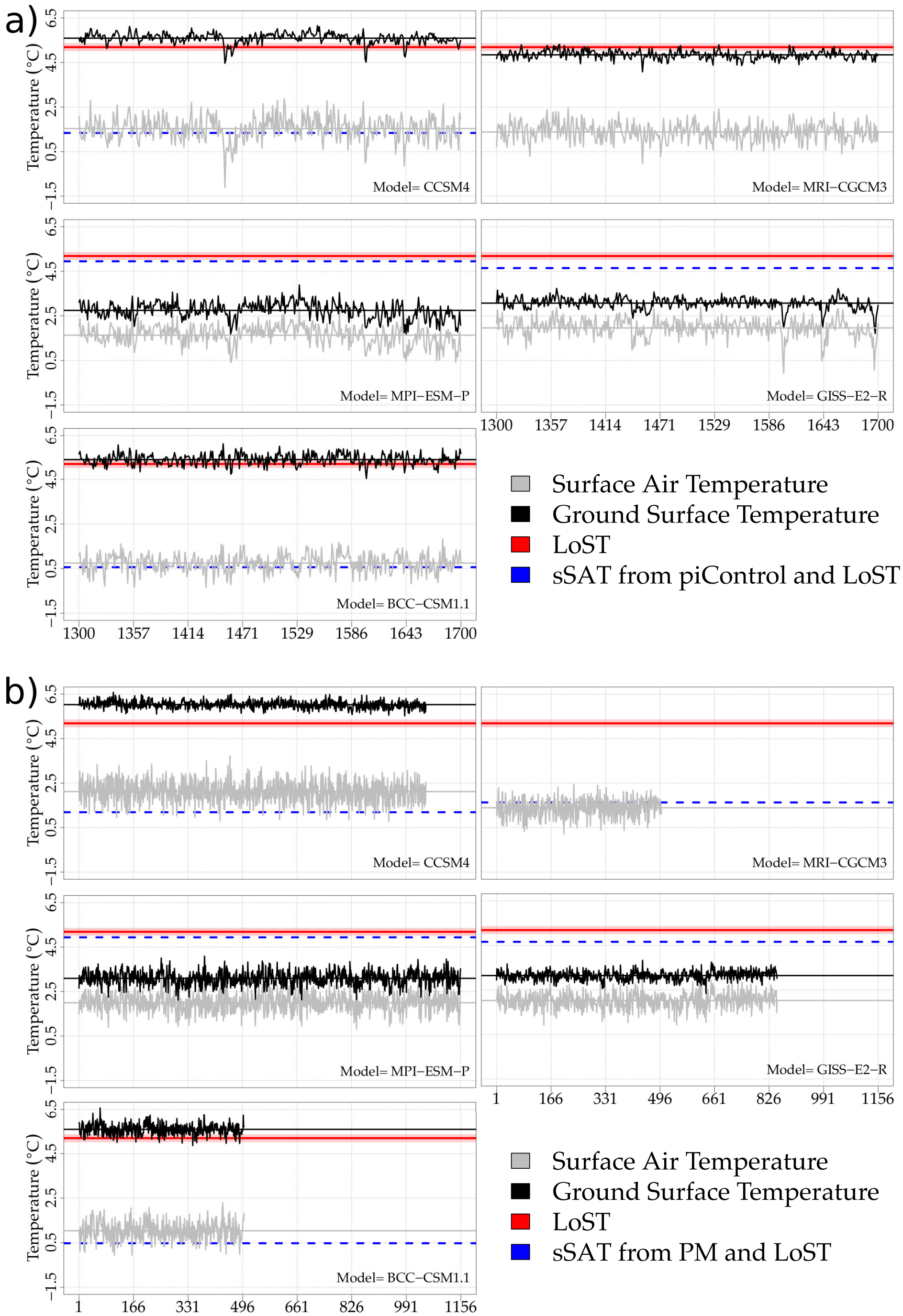

Figure 3. Surface air temperature evolution (gray solid line), ground surface temperature evolution (black solid line), SAT_0 (gray horizontal line) and GST_0 (black horizontal line) for (a) PMIP3/CMIP5 PM and (b) PMIP3/CMIP5 piControl simulations. Solid red lines represent the mean LoST temperature and the red shadow represents the 95% confidence interval (Section S1, Fig. S6). Dashed blue lines represent estimated references for long-term surface air temperatures from the LoST climatology and the simulated air-ground temperature offset in (a) piControl and (b) PM simulations. Ground temperatures for the MRI-CGCM3 piControl simulation could not be retrieved from the PMIP3/CMIP5 data servers.

Supporting Information for "Long-term Surface Temperature (LoST) Database as a complement for GCM preindustrial simulations"

Francisco José Cuesta-Valero[1,2], Almudena García-García[1,2], Hugo Beltrami[2], Eduardo Zorita[3], and Fernando Jaume-Santero[2,4]

[1]Environmental Sciences Program, Memorial University of Newfoundland, St. John's, NL, Canada.
[2]Climate & Atmospheric Sciences Institute, St. Francis Xavier University, Antigonish, NS, Canada.
[3]Institute of Coastal Research, Hemlholtz-Zentrum Geesthacht, Germany.
[4]Departamento de Física de la Tierra y Astrofísica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040, Madrid, Spain.

S1 Error Propagation for LoST Database

The GIDS algorithm (Eq. 3) incorporates errors from the determination of the latitudinal, longitudinal and altitudinal gradients as well as errors from the $\mathrm{T_0}$ estimates. Errors in $\mathrm{T_0}$ temperatures are specified by the linear regression analysis employed to determine the $\mathrm{T_0}$ values from each BTP measurement, while the linear regression analysis of the geographical distribution of $\mathrm{T_0}$ temperatures provides the latitudinal, longitudinal and altitudinal gradients and their errors (see Section 3). Therefore, an estimate of the error in LoST temperatures at each grid cell of the database can by computed just by applying basic error propagation theory to Eq. 3, which results in:

$$\Delta V_0 = \frac{\sqrt{\sum_{i=1}^{N}\left\{\left(\Delta V_i^2 + \left(|lat_0 - lat_i|\,\Delta C_{lat}\right)^2 + \left(|lon_0 - lon_i|\,\Delta C_{lon}\right)^2 + \left(|z_0 - z_i|\,\Delta C_z\right)^2\right)^{\frac{1}{2}}\left|d_i^{-2}\right|\right\}^2}}{\left|\sum_{i=1}^{N} d_i^{-1}\right|} \tag{S1}$$

where $\Delta \mathrm{V_0}$ is the error of the predicted temperature at the target node, $\Delta \mathrm{V_i}$ represents the $\mathrm{T_0}$ error from the $\mathrm{i^{th}}$ BTP measurement, $\Delta \mathrm{C_{lat}}$, $\Delta \mathrm{C_{lon}}$ and $\Delta \mathrm{C_z}$ are the errors in the gradients from the regression analysis of the geographical distribution of $\mathrm{T_0}$ data, $\mathrm{lat_i}$, $\mathrm{lon_i}$ and $\mathrm{z_i}$ represent latitude, longitude and altitude of the $\mathrm{i^{th}}$ measurement respectively, $\mathrm{lat_0}$, $\mathrm{lon_0}$ and $\mathrm{z_0}$ represent the latitude, longitude and altitude of the target node respectively, $\mathrm{d_i}$ is the distance from the $\mathrm{i^{th}}$ measurement to the target node, and N are the number of BTP measurements within a distance of $650\ \mathrm{km}$ to the target node. Errors in latitude, longitude and altitude are considered negligible, as well as the error in the distance between measurements and target nodes. Fig. S6 shows the errors as 2σ values (i.e., $2 \times \Delta \mathrm{V_0}$) for each grid cell, with a spatial average of $0.2\ ^{\circ}\mathrm{C}$.

Beyond the propagation of known errors, other sources of ~~error~~ **uncertainty** are possible but difficult to characterize given the limited temporal resolution of the LoST database. The most probable additional source of error is the distance criterion for the interpolation. This criterion was determined using a pseudo-proxy experiment and five PMIP3/CMIP5 PM simulations, obtaining different results for each model (Fig. S1). However, we did not find any adequate method to characterize such error in the LoST database, and further sources of error are possible. **Section 5 in the main text also discusses additional sources of uncertainty in LoST temperatures, but data limitations prevent us to characterize the error caused by those factors.**

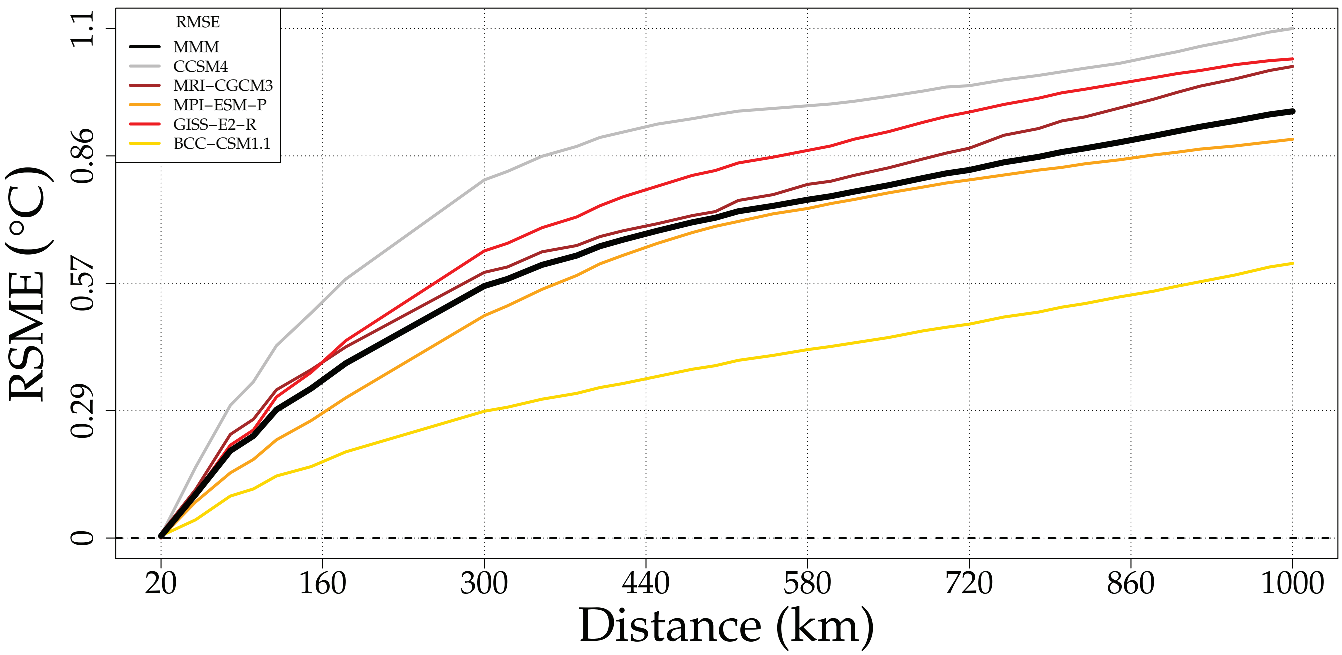

Figure S1. Root-Mean Squared Errors of the GIDS interpolation using ground surface temperatures at 1.0 m depth for the period 1300-1700 CE from the PMIP3/CMIP5 PM simulations to obtain a maximum distance criterion to interpolate each BTP measurement. The black line represents the multimodel mean (MMM).

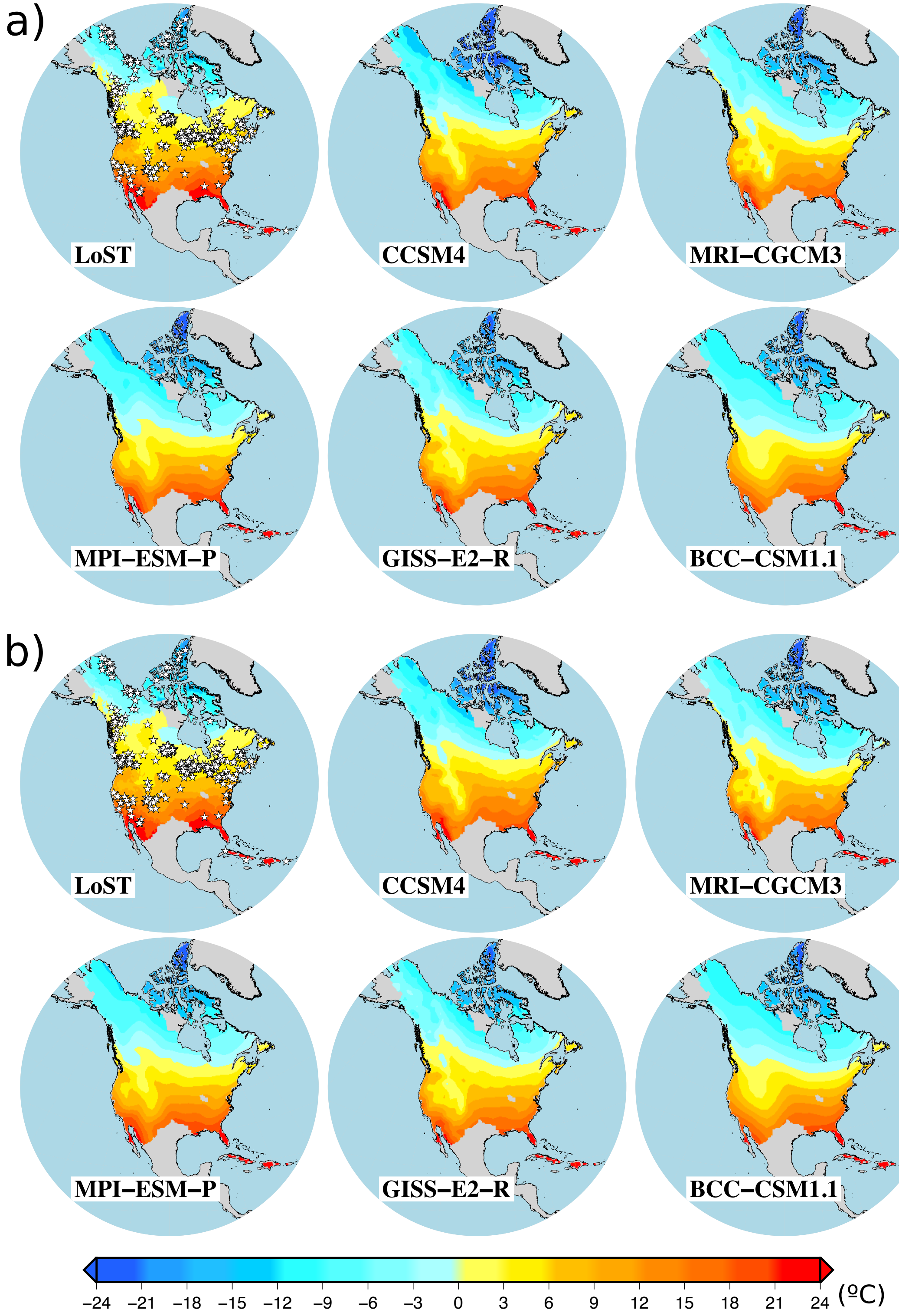

Figure S2. SAT_0 estimates from (a) PMIP3/CMIP5 PM simulations (1300-1700 CE) and (b) PMIP3/CMIP5 piControl simulations together with LoST temperatures. White stars show the location of the employed BTP measurements for the GIDS interpolation.

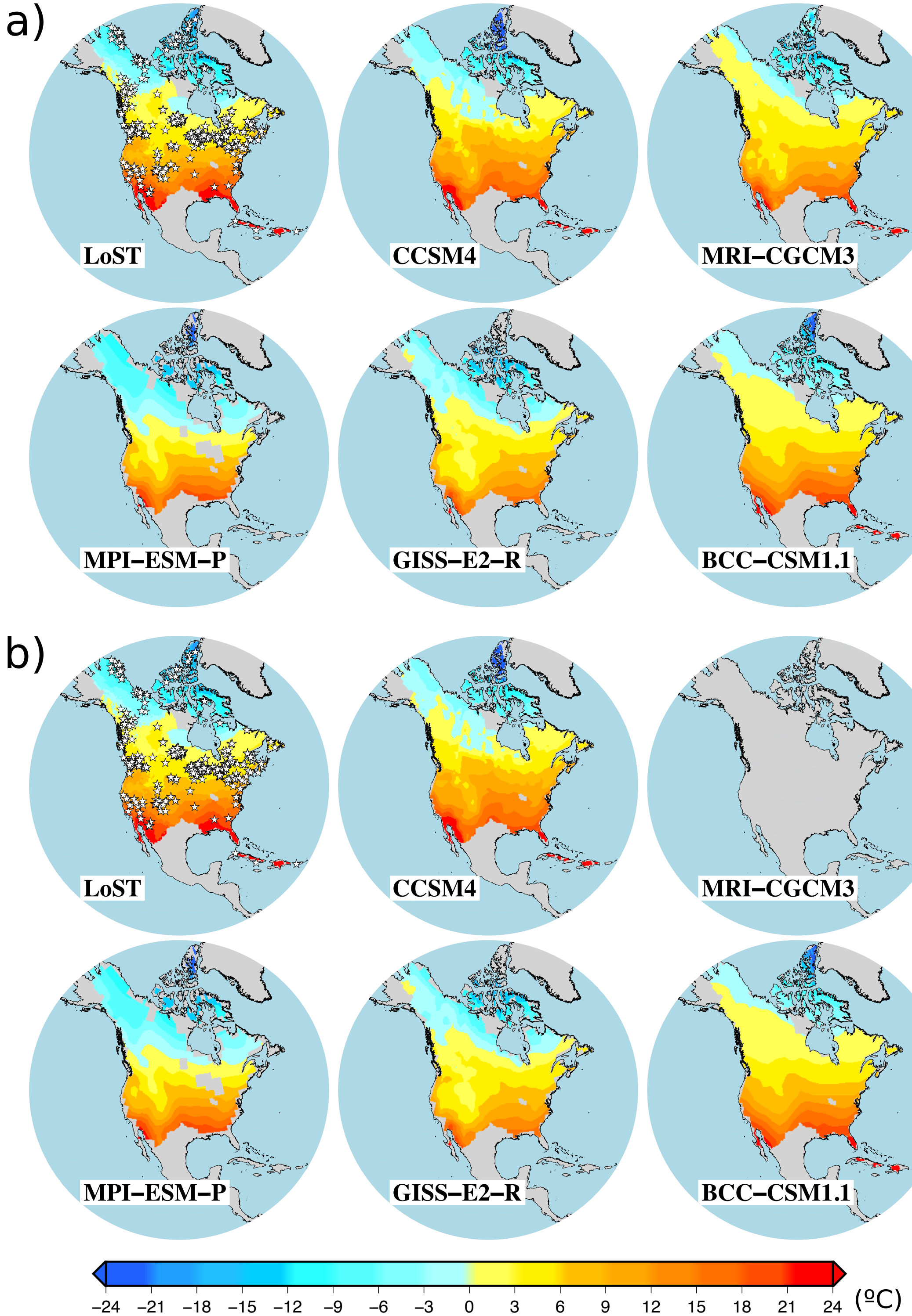

Figure S3. GST_0 estimates from (a) PMIP3/CMIP5 PM simulations (1300-1700 CE) and (b) PMIP3/CMIP5 piControl simulations together with T_0 temperatures. White stars shown the location of the employed BTP measurements for the GIDS interpolation.

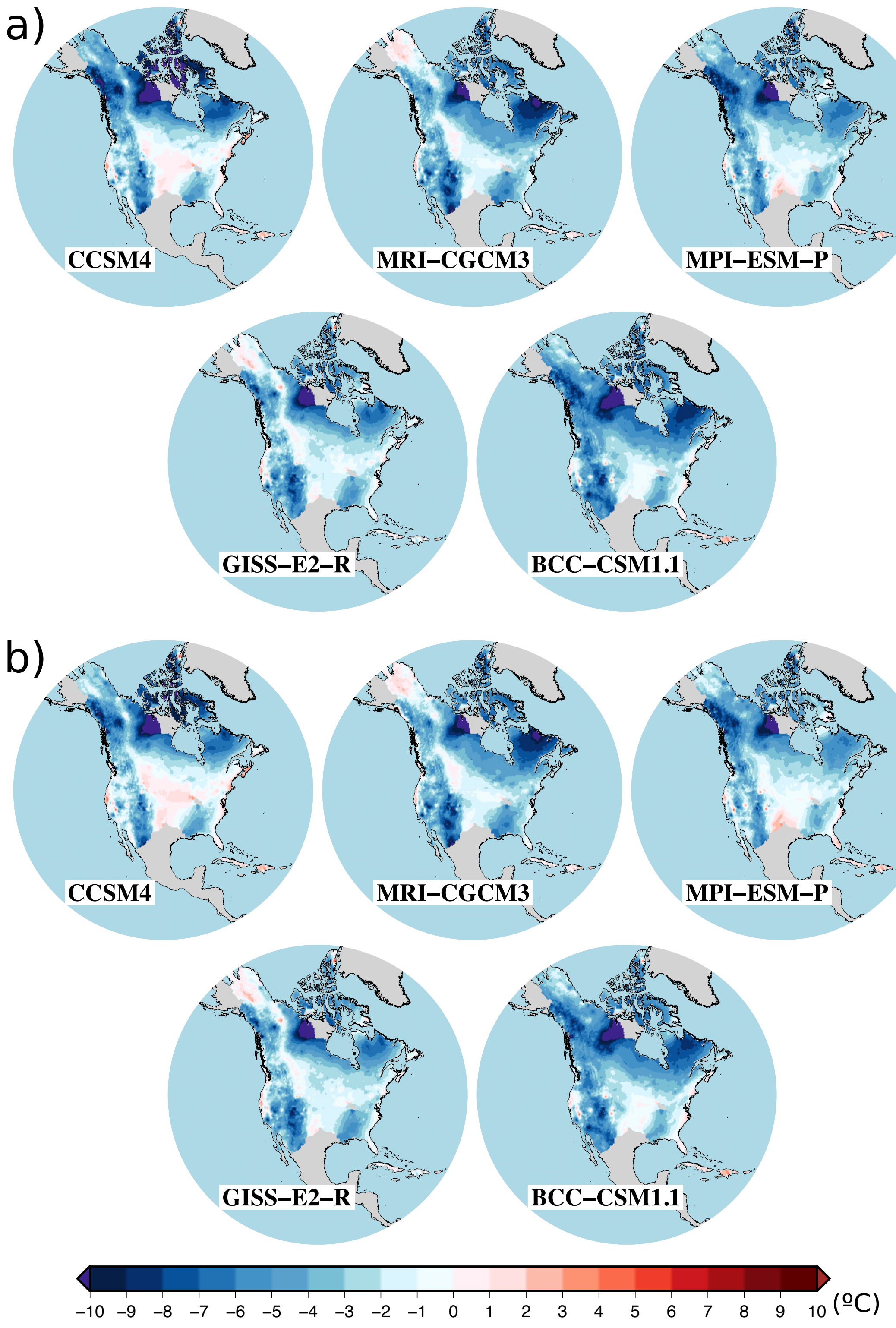

Figure S4. Difference between SAT_0 values from PMIP3/CMIP5 simulations and LoST temperatures. (a) Results for PM simulations (1300-1700 CE). (b) Results piControl simulations.

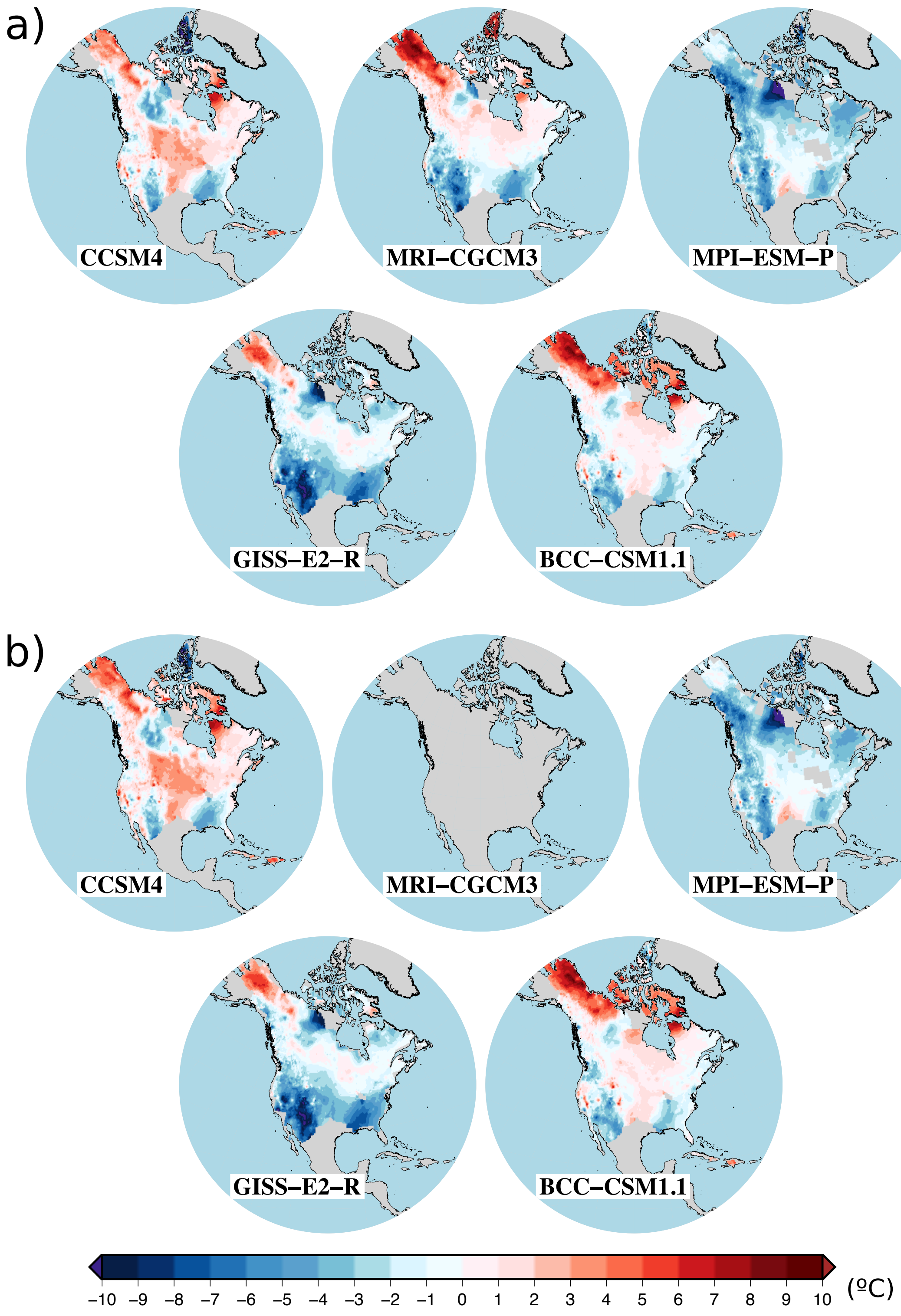

Figure S5. Difference between GST_0 values from PMIP3/CMIP5 simulations and LoST temperatures. (a) Results for PM simulations (1300-1700 CE). (b) Results for piControl simulations.

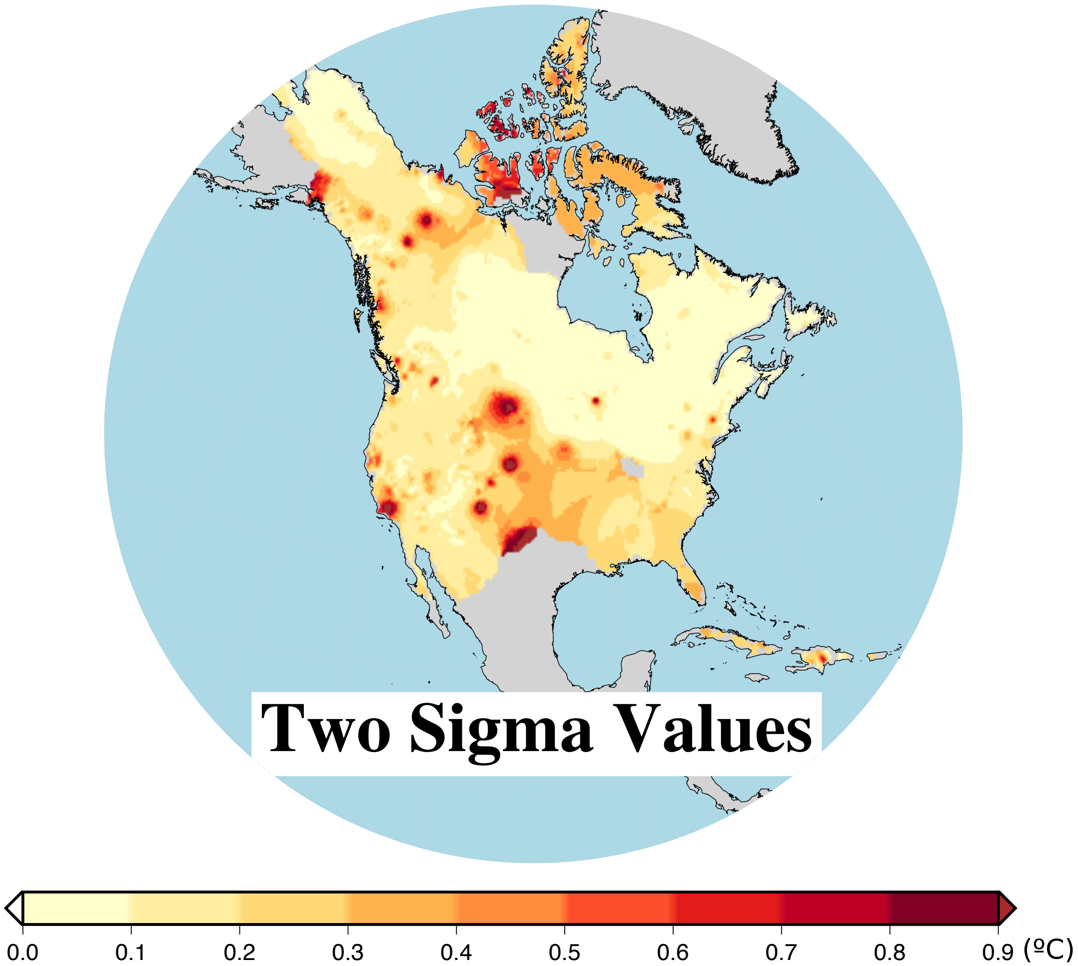

Figure S6. Errors (2σ values) of LoST temperatures estimated as described in Section S1. The spatial average is $0.2\ ^{\circ}\mathrm{C}$.

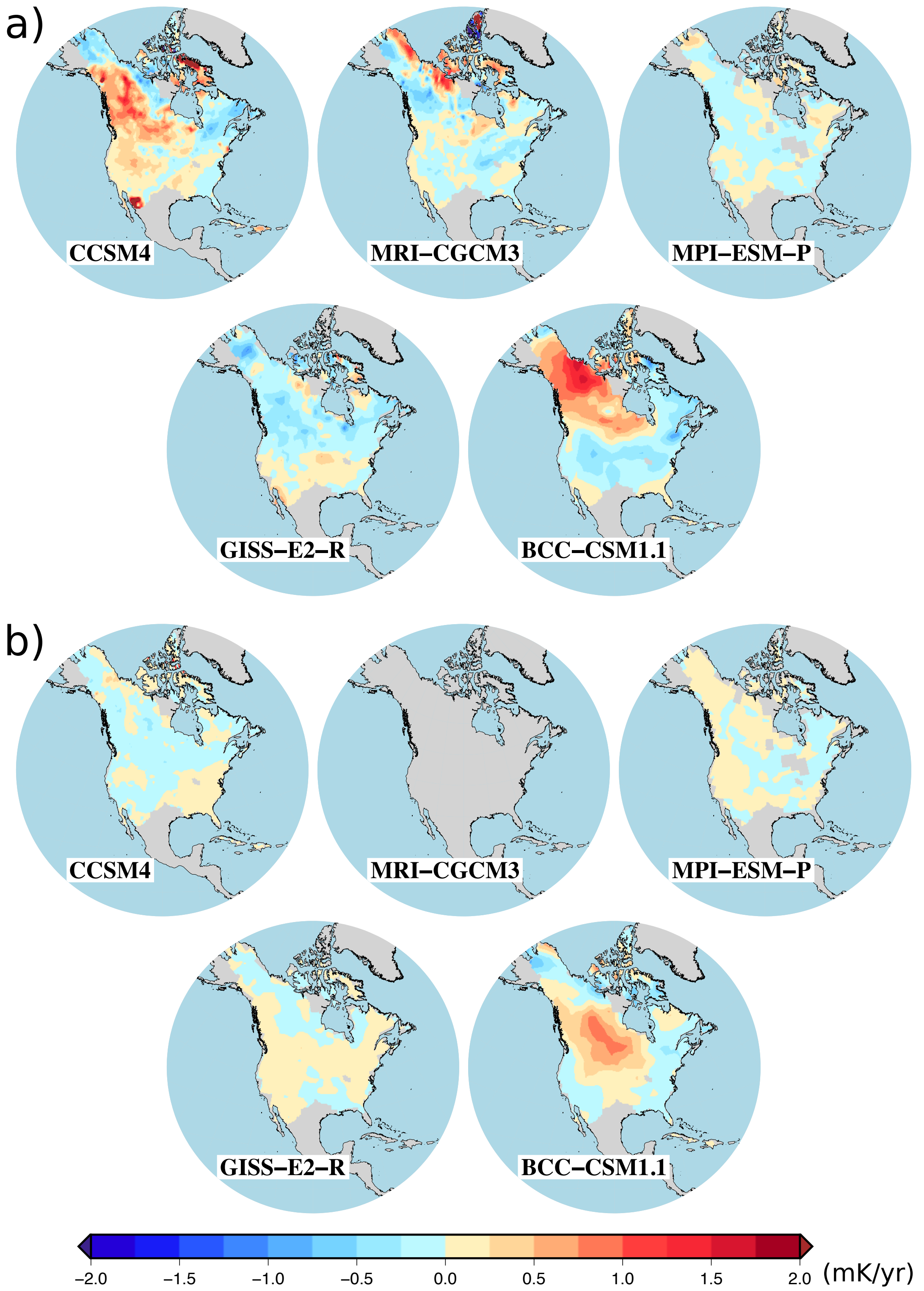

Figure S7. Trends of the difference between air and ground (1.0 m depth) temperatures from PMIP3/CMIP5 simulations. (a) Results for PM simulations (1300-1700 CE). (b) Results for piControl simulations.

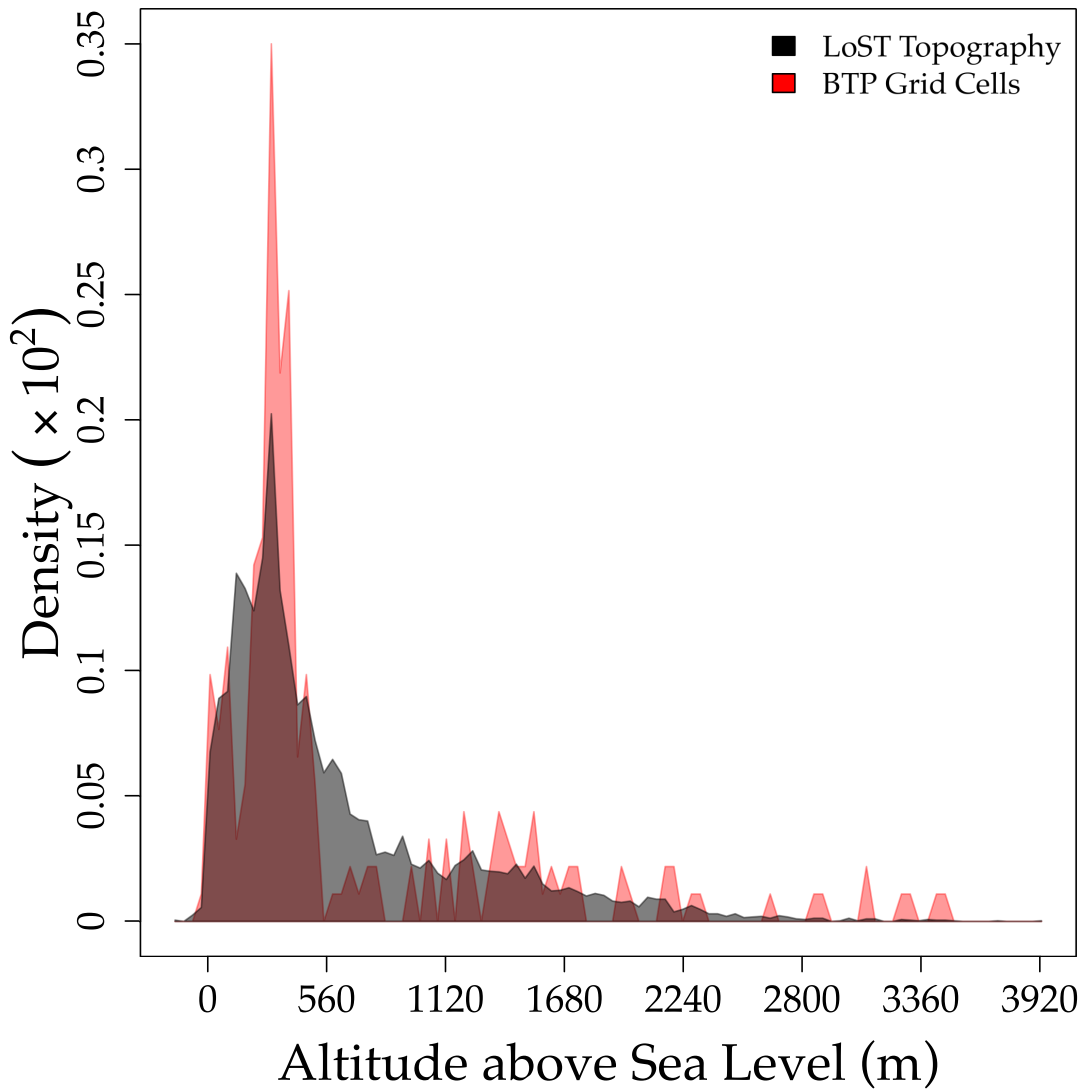

Figure S8. Altitude distribution over the LoST domain (black histogram) and at grid cells containing BTPs (red histogram) from the ETOPO2 product.