



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

Francisco José Cuesta-Valero<sup>1,2</sup>, Almudena García-García<sup>1,2</sup>, Hugo Beltrami<sup>2</sup>, Eduardo Zorita<sup>3</sup>, and Fernando Jaume-Santero<sup>2,4</sup>

<sup>1</sup>Environmental Sciences Program, Memorial University of Newfoundland, St. John's, NL, Canada

<sup>2</sup>Climate & Atmospheric Sciences Institute, St. Francis Xavier University, Antigonish, NS, Canada

<sup>3</sup>Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

<sup>4</sup>Departamento de Física de la Tierra y Astrofísica, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040, Madrid, Spain

**Correspondence:** Hugo Beltrami (hugo@stfx.ca)

Received: 5 October 2018 – Discussion started: 22 October 2018

Revised: 18 May 2019 – Accepted: 2 June 2019 – Published: 19 June 2019

**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 Climate Research Unit (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 the atmospheric CO<sub>2</sub> concentration (Gregory et al., 2002), and it is considered to be one of the most important metrics to understand the long-term evolution of the climate system. Numerous studies have unsuccessfully tried to narrow the uncertainty range of the ECS using observations, simulations and paleoreconstructions; however, the best estimates of the ECS have remained between 1.5 and 4.5 °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 a 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 (Dommenges, 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 of these necessary approximations add inconsistencies to simulations, affecting the simulated climate state and trajectory. Parameterizations of radiative forcing by CO<sub>2</sub> in climate models are of special importance, as they are responsible for nearly 50 % of the uncertainties in the estimated values of the 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 of varying model parameters, for which the values are poorly constrained by theory or observations or not constrained at all, to obtain a simulated climate evolution compatible with observations. Hence, the adjustment of this parameter 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 (Dommenges, 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, as 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 these 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 (Dommenges, 2016; Hu et al., 2017). At this point, estimates of preindustrial long-term surface temperatures from geothermal data may be a useful reference for assessing whether 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 in 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; Demezhko and Gornostaeva, 2015b). Several studies have validated the borehole methodology using past millennium simulations from the ECHO-G GCM (González-Rouco et al., 2006, 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 of 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, in turn, to help 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). Here, we present 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 <https://figshare.com/s/f20d6002a57cf3279a1c> (last access: 12 June 2019). 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 of a gridded set of climate variables derived from worldwide meteorological observations. 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 temperatures are supplied at 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 from 850 to 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 S  f  rian et al. (2016).

### 2.3 Borehole data

Here, we use estimates of long-term surface temperatures 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 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 follows (e.g., Jaume-Santero

et al., 2016):

$$T(z) = T_0 + q_0 \cdot R(z) + T_t(t), \quad (1)$$

where  $T_t$  is the surface transient perturbation,  $T_0$  is the long-term surface temperature ( $T_0$  temperature hereafter),  $q_0$  is the subsurface flux at equilibrium and  $R(z)$  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 BTPs available does not present such conductivity data. Thus, the thermal conductivity is assumed to be constant and Eq. (1) is rewritten as

$$T(z) = T_0 + \Gamma \cdot z + T_t(t), \quad (2)$$

where  $\Gamma$  is the subsurface thermal gradient at equilibrium. The recorded surface transient perturbation ( $T_t$ ) 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 timescales 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 the thermal properties of the subsurface rocks are taken into account, temperature depends linearly on depth at the bottom part of the temperature profile, allowing for the approximation of the subsurface thermal equilibrium by a 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 required time ( $t$ ) for a change in the surface energy balance to reach a certain depth ( $z$ ) is given by the following equation (Carslaw and Jaeger, 1959; Pickler et al., 2016):

$$t \approx \frac{z^2}{4\kappa}, \quad (3)$$

where  $\kappa$  is the thermal diffusivity of the subsurface. We consider  $\kappa = 1 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$  for all BTP measurements (Cermak and Rybach, 1982). In this study, all BTPs are truncated at the same depth (300 m) to ensure that all  $T_0$  temperatures are estimated for the same temporal period. We use the last 100 m of each BTP to estimate the subsurface thermal equilibrium, obtaining an estimated temporal period that approximately ranges from  $\sim 1300$  CE ( $z = 300$  m) to  $\sim 1700$  CE ( $z = 200$  m). Hence, this period of time provides a baseline to compare with long-term temperatures

**Table 1.** Model name, SAT<sub>0</sub> estimates, GST<sub>0</sub> estimates, SAT<sub>0</sub> and GST<sub>0</sub> differences with the mean LoST temperatures and references for each PMIP3/CMIP5 GCM simulation. All results are in degrees Celsius. Ground temperatures for the MRI-CGCM3 piControl simulation could not be retrieved from the PMIP3/CMIP5 data servers. The temperature average of the LoST database is 5.2°C, with a 95 % confidence interval between 5.0 and 5.4 °C.

Model	Past millennium					Preindustrial control				
	SAT <sub>0</sub>	GST <sub>0</sub>	SAT <sub>0</sub> –LoST	GST <sub>0</sub> –LoST	Reference	SAT <sub>0</sub>	GST <sub>0</sub>	SAT <sub>0</sub> –LoST	GST <sub>0</sub> –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)

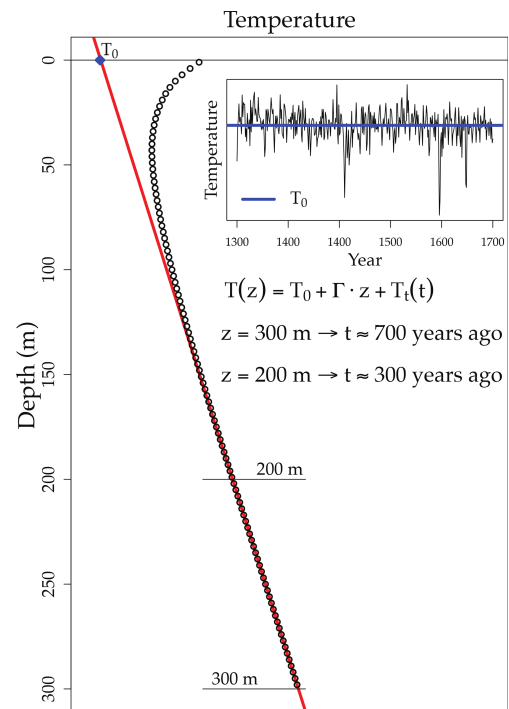
from the PMIP3/CMIP5 PM simulations. However, the estimated temporal period is not homogeneous as result of the nonlinear relationship between time and depth (Beltrami and Mareschal, 1995); thus, the estimates of recent years (i.e., 1700 CE) are better determined than the estimates of past years (1300 CE). Influences of long-term perturbations of the past surface energy budget outside of this temporal window may also affect the temperature within the depth range used here (see Sect. 5 for more details).

### 3 The LoST database

In order to provide a gridded dataset over continental North America,  $T_0$  temperatures from BTP measurements are spatially interpolated to a  $0.5^\circ \times 0.5^\circ$  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:

$$V_0 = \frac{1}{\sum_{i=1}^N d_i^{-2}} \left[ \sum_i^N \{V_i + (\text{lat}_0 - \text{lat}_i)C_{\text{lat}} + (\text{long}_0 - \text{long}_i)C_{\text{long}} + (z_0 - z_i)C_z\} d_i^{-2} \right], \quad (4)$$

where  $V_0$  is the predicted variable at the target node,  $V_i$ ;  $\text{lat}_i$ ,  $\text{long}_i$  and  $z_i$  represent the variable, latitude, longitude and altitude of the  $i$ th measurement respectively;  $\text{lat}_0$ ,  $\text{long}_0$  and  $z_0$  represent the latitude, longitude and altitude of the target node respectively;  $C_{\text{lat}}$ ,  $C_{\text{long}}$  and  $C_z$  are the coefficients from the regression analysis; and  $d_i$  is the distance from the  $i$ th measurement to the target node. The propagation of known errors in the GIDS algorithm is described in Sect. S1 in the Supplement. 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;



**Figure 1.** Synthetic borehole temperature profile (black dots) using data from the CCSM4 PM simulation (inset) and the linear fit of temperatures between 200 and 300 m (red line). The synthetic temperature profile is generated using the simulated global ground temperature anomaly at a depth of 1.0 m for the period from 1300 to 1700 CE as the transient perturbation ( $T_1$ ), mean ground temperature as the long-term surface temperature ( $T_0$ ) and a typical thermal gradient ( $\Gamma$ ) of  $0.01 \text{ K m}^{-1}$  (Jaume-Santero et al., 2016). The equivalence between depth ( $z$ ) and time ( $t$ ) is given by Eq. (3). Thermal diffusivity is considered as  $\kappa = 1 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$  (Cermak and Rybach, 1982).

Mardikis et al., 2005). Furthermore, the GIDS method performs well in comparison with other broadly used interpolation techniques like co-kriging or smoothing splines (ANUS-PLIN 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).

As the  $T_0$  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 min Gridded Global Relief Data (ETOPO2 v2) of the National Oceanic and Atmospheric Administration (National Geophysical Data Center, 2006; Two-minute Gridded Global Relief Data, NOAA, <https://doi.org/10.7289/V5J1012Q>, last access: on 7 July 2017). For this study, the regression analysis of  $T_0$  temperatures considering latitude, longitude and altitude yields robust results, with an  $R^2$  value of 0.865 and a  $p$  value of  $\ll 0.05$ . The distance from the measurements to the nodes is computed using the Vincenty formula for an ellipsoid with different major and minor axes (Vincenty, 1975); 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 the maximum appropriate distance from a grid node to a BTP measurement to interpolate the  $T_0$  temperatures. That is, we used the long-term mean ground surface temperatures for the period from 1300 to 1700 CE from the five PMIP3/CMIP5 PM simulations as surrogate realities, and applied the interpolation methodology employed to create the LoST database. Therefore, these GCM simulations were regridded to a  $0.5^\circ \times 0.5^\circ$  grid, considering grid cells containing BTP measurements as a 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 in the Supplement). We set 650 km as the maximum distance criterion as this is the maximum distance at which the RMSE is lower than  $1.0^\circ\text{C}$  for the five simulations. Nevertheless, such a distance criterion produces results for three grid cells in the Yucatán Peninsula (Mexico), which we consider unjustifiable as there are no BTP measurements in or near this part of Mexico. Therefore, these grid cells are masked out from our analysis.

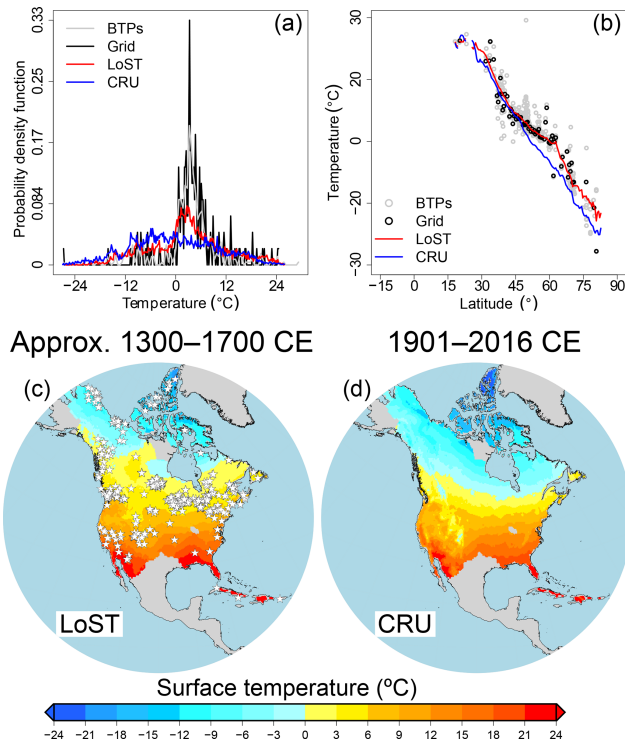
## 4 Results

The distribution of LoST temperatures at grid cells containing BTP measurements reproduces the shape of the distribution of raw  $T_0$  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  $T_0$  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 for the raw  $T_0$  temperatures, as the majority of BTP measurements cover from  $35$  to  $60^\circ\text{N}$ . Nonetheless, the latitudinal mean temperatures from the LoST database are consistent

with  $T_0$  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  $\sim 50^\circ\text{N}$ , whereas 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, which is 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 b). Nevertheless, it should be noted 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 a 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 (Fig. 2c, d), which is 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 from 1300 to 1700 CE ( $\text{SAT}_0$ ) and the mean ground surface temperature linearly interpolated at 1.0 m depth for the same period ( $\text{GST}_0$ ), in order to be consistent with the estimated temporal range for  $T_0$  temperatures in Sect. 2.3. The PMIP3/CMIP5 simulations are interpolated onto the grid of the LoST database;  $\text{SAT}_0$  and  $\text{GST}_0$  values are only estimated at grid cells containing LoST temperatures.  $\text{SAT}_0$  and  $\text{GST}_0$  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, S3).  $\text{SAT}_0$  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, whereas  $\text{GST}_0$  estimates show higher values than LoST temperatures at high latitudes for the same GCM simulations (Figs. S4, S5). Such results are in agreement with previous analyses of the 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



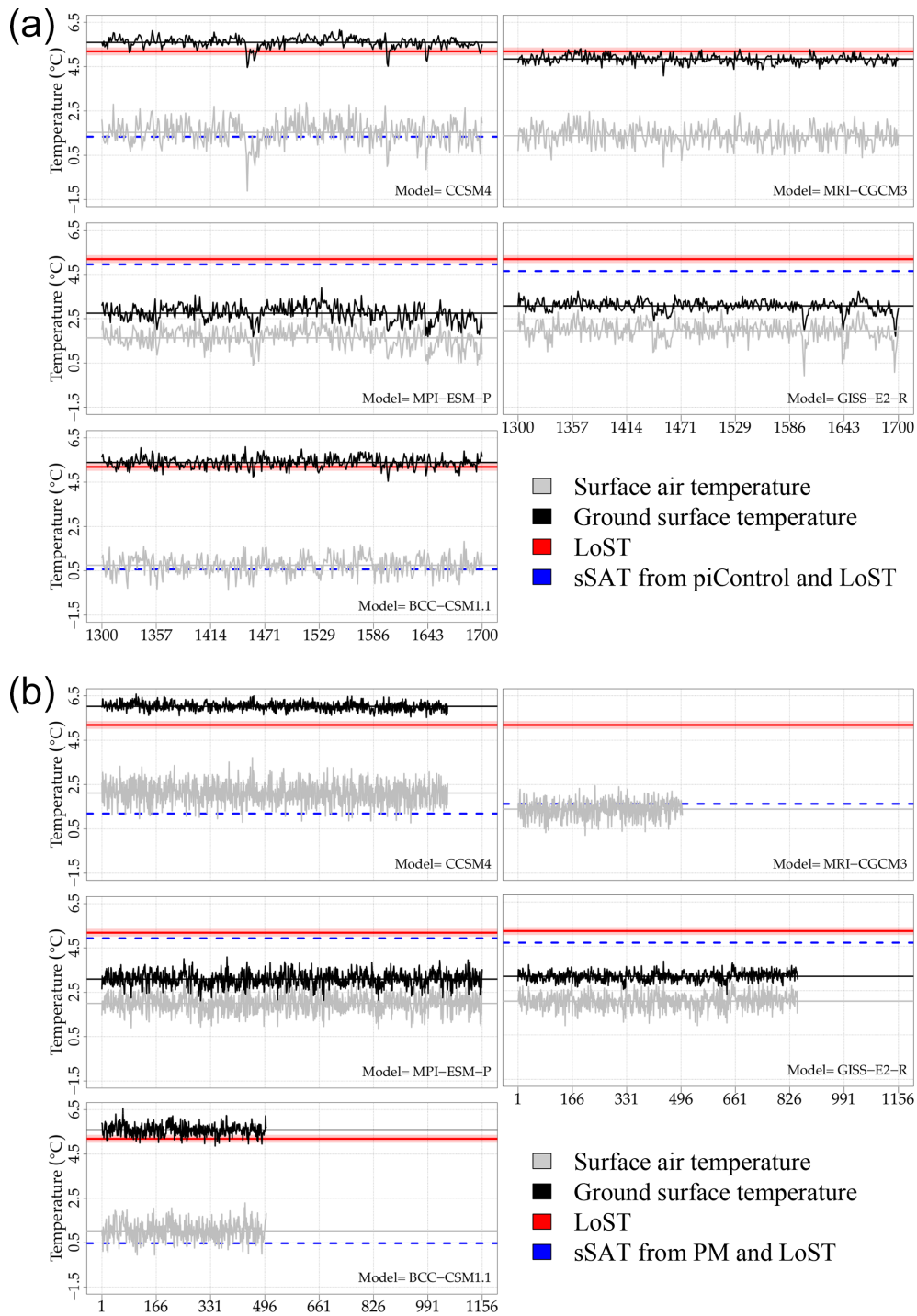
**Figure 2.** Histogram (a) and latitudinal mean temperatures (b) from BTP measurements (gray), LoST temperatures at grid cells containing BTP measurements (black), LoST temperatures (red) and mean surface air temperature from the CRU database (blue). LoST temperatures ( $\sim 1300$ – $1700$  CE) (c)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

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  $SAT_0$  and  $GST_0$  values than LoST temperatures, indicating lower long-term ground surface temperatures than the rest of the models (Table 1; Figs. S4, 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, Table 1). The similar  $GST_0$  and mean ground surface temperatures for the CCSM4 and the BCC-CSM1.1 GCMs in both PM and piControl simulations were expected as 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 land surface model component employed (García-García et al., 2019; Slater and Lawrence, 2013). In contrast, the GISS-E2-R and the MPI-ESM-P models produce lower mean  $GST_0$  values than the mean LoST temperature and the rest of models, while simulating similar  $SAT_0$  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, as 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  $0.3$  °C ( $0.1$  °C if considering the  $2\sigma$  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 the MPI-ESM-P and GISS-E2-R simulations ( $> 2.0$  °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 a reference for assessing the represented climatology in both PM and piControl simulations. Nevertheless, the determination of  $T_0$  temperatures 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  $0.15$  °C from the 514 BTPs evaluated here (see Sect. S1 for details about the error treatment in the LoST database). Rock heterogeneity should also be considered for estimating  $T_0$  temperatures. We assume homogenous thermal properties for all borehole profiles, which is another source of uncertainty for LoST temperatures. The ideal approach consists of estimating the thermal resistance with depth (Eq. 1), but the absence of thermal conductivity measurements for the BTPs employed (Jaume-Santero et al., 2016) makes this approach impractical. Additionally, measurements of thermal conductivity tend to be distributed around a central value (e.g., the measurements at the Neil Well, Beltrami and Taylor, 1995). If the thermal conductivity varies systematically with depth at a certain location, such variation will be reflected in the temperature profile as an unphysical signal. Such logs were removed from the database employed in this analysis, as explained in Jaume-Santero et al. (2016). Therefore, it is reasonable to assume a homogenous conductivity with depth. Long-term alterations of the surface energy balance out of the



**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 (Sect. 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.

1300–1700 CE period may also affect the LoST estimates. Particularly, possible transient temperatures in BTPs due to the Little Ice Age (LIA) and the Medieval Warm Period (MWP) add a certain degree of uncertainty in the determination of  $T_0$  values. However, the spatial extent of both the LIA and the MWP was not homogeneous over North America (e.g., Masson-Delmotte et al., 2013, and references therein), meaning that not all BTPs were affected by the events (Beltrami and Mareschal, 1992; Chouinard et al., 2007; Jaume-Santero et al., 2016). Additionally, the influence of the LIA and the MWP should be part of any millennial-scale transient climate simulation; therefore, the effect of such climate events is taken into account in the comparison between LoST results and transient climate simulations. The absence of these two periods in piControl simulations probably contributes to the slightly poorer agreement between LoST temperatures and piControl temperatures in comparison with results for the PM simulation (Fig. 3). 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 as the BTP database employed does not provide elevation data. Therefore, we use the ETOPO2 database to assess if the altitude distribution of the BTPs is sufficient to represent the topography of the LoST domain. The altitude distribution over the LoST domain and at grid cells containing borehole sites are displayed in Fig. S8. Both histograms present a similar shape for altitudes up to  $\sim 430$  m, showing a lack of borehole locations at altitudes between  $\sim 430$  and  $\sim 1013$  m. The uneven latitudinal distribution of borehole sites probably causes 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, whereas 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 (synthetic SAT, sSAT). 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, whereas 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 represent a strong air–ground coupling, as discussed in Sect. 4. Therefore, although the LoST database contains estimates of ground surface temperatures, it may also be used to assess simulated long-term surface air temperatures as a first-order approach.

The regional character of the LoST database presented poses some caveats for analyzing the global climatology of preindustrial simulations. Indeed, results of the simulated regional climatology cannot be globally extrapolated as 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 of generating a global LoST database from the existing global network of BTP measurements, which would help to minimize the effect of possible regional drifts on the simulated climatology. However, BTP measurements are scarce in the Southern Hemisphere, which is a potential burden that needs to be considered when assembling a global version of this database. Additionally, the temperature profiles employed in this study to estimate  $T_0$  temperatures were truncated to a depth of 300 m, 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 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 for the evaluation of 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. Furthermore, previous studies have found that the CMIP5 GCM simulations have difficulties properly representing 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 land surface model components employed (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  $0^\circ\text{C}$  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, whereas 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 capture high-frequency variations of climate conditions (Moberg et al., 2005), whereas 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 parts of continental North America has been assembled from geothermal measurements. Our results show that this database can be used as a 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. Hence, the spread in ECS estimates by GCM simulations may be reduced given the relationship between control temperature climatology and three long-term powerful feedbacks: the ice–albedo feedback, the water vapor feedback and the permafrost–carbon feedback. Future work would consist of 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 LoST database can be downloaded from <https://figshare.com/s/f20d6002a57cf3279a1c> (last access: 12 June 2019), with <https://doi.org/10.6084/m9.figshare.8124887> (Cuesta-Valero et al., 2019).

**Supplement.** The supplement related to this article is available online at: <https://doi.org/10.5194/cp-15-1099-2019-supplement>.

**Author contributions.** FJCV, AGG and HB designed the study. FJCV generated the results and figures. FJS provided the  $T_0$  data. All authors analyzed the results, discussed the outcome and drafted the paper.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP. We thank the climate modeling groups responsible for the model simulations used herein (listed in Table 1 of this paper) for producing and making their model output available. 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 paper, 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 Grants Program (grant no. NSERC DG 140576948), the Canada Research Chairs Program (grant no. CRC 230687), and the Canada Foundation for Innovation (CFI) to Hugo Beltrami. Hugo Beltrami holds the position of Canada Research Chair in Climate Dynamics. Almudena García-García and Francisco José Cuesta-Valero are funded by Hugo 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.

**Financial support.** This research has been supported by the Canada Research Chairs Program (grant no. CRC 230687) and the Natural Sciences and Engineering Research Council (grant no. NSERC DG 140576948).

**Review statement.** This paper was edited by Irina Rogozhina and reviewed by Dmitry Demezhko and one anonymous referee.

## 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, *Geophys. Res. Lett.*, 34, 109502, <https://doi.org/10.1029/2007GL029536>, 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, *Geophys. Res. Lett.*, 39, 109712, <https://doi.org/10.1029/2012GL015607>, 2012.
- Beltrami, H.: Climate from borehole data: Energy fluxes and temperatures since 1500, *Geophys. Res. Lett.*, 29, 26-1–26-4, <https://doi.org/10.1029/2002GL015702>, 2002.
- Beltrami, H. and Bourlon, E.: Ground warming patterns in the Northern Hemisphere during the last five centuries, *Earth Planet. Sc. Lett.*, 227, 169–177, <https://doi.org/10.1016/j.epsl.2004.09.014>, 2004.
- Beltrami, H. and Kellman, L.: An examination of short- and long-term air–ground temperature coupling, *Global Planet. Change*,

- 38, 291–303, [https://doi.org/10.1016/S0921-8181\(03\)00112-7](https://doi.org/10.1016/S0921-8181(03)00112-7), 2003.
- Beltrami, H. and Mareschal, J.-C.: Ground temperature histories for central and eastern Canada from geothermal measurements: Little Ice Age signature, *Geophys. Res. Lett.*, 19, 689–692, <https://doi.org/10.1029/92GL00671>, 1992.
- Beltrami, H. and Mareschal, J.-C.: Resolution of ground temperature histories inverted from borehole temperature data, *Global Planet. Change*, 11, 57–70, [https://doi.org/10.1016/0921-8181\(95\)00002-9](https://doi.org/10.1016/0921-8181(95)00002-9), 1995.
- Beltrami, H. and Taylor, A. E.: Records of climatic change in the Canadian Arctic: towards calibrating oxygen isotope data with geothermal data, *Global Planet. Change*, 11, 127–138, [https://doi.org/10.1016/0921-8181\(95\)00006-2](https://doi.org/10.1016/0921-8181(95)00006-2), 1995.
- Beltrami, H., Smerdon, J. E., Pollack, H. N., and Huang, S.: Continental heat gain in the global climate system, *Geophys. Res. Lett.*, 29, 8-1–8-3, <https://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, *Geophys. Res. Lett.*, 33, 106717, <https://doi.org/10.1029/2006GL025676>, 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, *J. Geophys. Res.-Earth*, 120, 763–778, <https://doi.org/10.1002/2014JF003382>, 2015.
- Bodri, L. and Cermak, V.: Borehole Climatology, Elsevier Science Ltd, Oxford, <https://doi.org/10.1016/B978-008045320-0/50001-6>, 2007.
- Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliessner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, *Nat. Clim. Change*, 2, 417–424, <https://doi.org/10.1038/nclimate1456>, 2012.
- Bullard, E. C. and Schonland, B. F. J.: Heat flow in South Africa, *P. Roy. Soc. Lond. A Mat.*, 173, 474–502, <https://doi.org/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](https://doi.org/10.1007/10201894_62), 1982.
- Chouinard, C., Fortier, R., and Mareschal, J.-C.: Recent climate variations in the subarctic inferred from three borehole temperature profiles in northern Quebec, Canada, *Earth Planetary Sc. Lett.*, 263, 355–369, <https://doi.org/10.1016/j.epsl.2007.09.017>, 2007.
- Cook, B. I., Ault, T. R., and Smerdon, J. E.: Unprecedented 21st century drought risk in the American Southwest and Central Plains, *Sci. Adv.*, 1, e1400082, <https://doi.org/10.1126/sciadv.1400082>, 2015.
- Cuesta-Valero, F. J., García-García, A., Beltrami, H., and Smerdon, J. E.: First assessment of continental energy storage in CMIP5 simulations, *Geophys. Res. Lett.*, 43, 5326–5335, <https://doi.org/10.1002/2016GL068496>, 2016.
- Cuesta-Valero, F. J., García-García, A., Beltrami, H., Zorita, E., and Jaume-Santero, F.: Long-term Surface Temperature (LoST) Database, Figshare, <https://doi.org/10.6084/m9.figshare.8124887.v1>, 2019.
- Demezhko, D. Y. and Gornostaeva, A. A.: Late Pleistocene–Holocene ground surface heat flux changes reconstructed from borehole temperature data (the Urals, Russia), *Clim. Past*, 11, 647–652, <https://doi.org/10.5194/cp-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, *Izv. Atmos. Ocean. Phy.*, 51, 723–736, <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, *Clim. Dynam.*, 46, 427–447, <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., Sect. 9, 741–866, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.020>, 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, *J. Geophys. Res.-Atmos.*, 118, 1139–1150, <https://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, *Environ. Res. Lett.*, 11, 044022, <https://doi.org/10.1088/1748-9326/11/4/044022>, 2016.
- García-García, A., Cuesta-Valero, F. J., Beltrami, H., and Smerdon, J. E.: Characterization of air and ground temperature relationships within the CMIP5 historical and future climate simulations, *J. Geophys. Res.-Atmos.*, 124, 3903–3929, <https://doi.org/10.1029/2018JD030117>, 2019.
- 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, *J. Climate*, 24, 4973–4991, <https://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, *Geophys. Res. Lett.*, 30, 2116, <https://doi.org/10.1029/2003GL018264>, 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, *Geophys. Res. Lett.*, 33, 101703, <https://doi.org/10.1029/2005GL024693>, 2006.
- González-Rouco, J. F., Beltrami, H., Zorita, E., and Stevens, M. B.: Borehole climatology: a discussion based on contributions from climate modeling, *Clim. Past*, 5, 97–127, <https://doi.org/10.5194/cp-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, *J. 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, *Int. J. Climatol.*, 34, 623–642, <https://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, *Geophys. Res. Lett.*, 28, 747–750, <https://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., Sect. 2, 159–254, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.008>, 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, *B. Am. Meteorol. Soc.*, 98, 1841–1856, <https://doi.org/10.1175/BAMS-D-16-0007.1>, 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, *B. Am. Meteorol. Soc.*, 98, 589–602, <https://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, *Sci. Rep.-UK*, 7, 4300, <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, <https://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, *Clim. Past*, 12, 2181–2194, <https://doi.org/10.5194/cp-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, *J. Adv. Model. Earth Sy.*, 5, 422–446, <https://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, *Clim. Past*, 10, 2201–2213, <https://doi.org/10.5194/cp-10-2201-2014>, 2014.
- Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, *Nat. Clim. Change*, 3, 369–373, <https://doi.org/10.1038/nclimate1716>, 2012.
- Knutti, R., Rugenstein, M. A. A., and Hegerl, G. C.: Beyond equilibrium climate sensitivity, *Nat. Geosci.*, 10, 727–736, <https://doi.org/10.1038/ngeo3017>, 2017.
- Kohl, T.: Transient thermal effects below complex topographies, *Tectonophysics*, 306, 311–324, [https://doi.org/10.1016/S0040-1951\(99\)00063-3](https://doi.org/10.1016/S0040-1951(99)00063-3), 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, *P. Natl. Acad. Sci. USA*, 108, 14769–14774, <https://doi.org/10.1073/pnas.1103910108>, 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, *J. Climate*, 26, 1877–1900, <https://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, *J. Climate*, 26, 1085–1111, <https://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, *Ecol. Inform.*, 6, 228–241, <https://doi.org/10.1016/j.ecoinf.2010.12.003>, 2011.
- MacDougall, A. H. and Beltrami, H.: Impact of deforestation on subsurface temperature profiles: implications for the borehole paleoclimate record, *Environ. Res. Lett.*, 12, 074014, <https://doi.org/10.1088/1748-9326/aa7394>, 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, *J. Geophys. Res.-Atmos.*, 115, D12109, <https://doi.org/10.1029/2009JD013170>, 2010.
- MacDougall, A. H., Avis, C. A., and Weaver, A. J.: Significant contribution to climate warming from the permafrost carbon feedback, *Nat. Geosci.*, 5, 719–721, <https://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 Resour. Manag.*, 19, 251–278, <https://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, *Clim. Dynam.*, 6, 135–143, <https://doi.org/10.1007/BF00193525>, 1992.
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González Rouco, J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A.: Information from Paleoclimate Archives, 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., Sect. 5, 383–464, Cambridge University Press, Cambridge, UK and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.013>, 2013.

- 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, *J. Adv. Model. Earth Sy.*, 4, m00A01, <https://doi.org/10.1029/2012MS000154>, 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 Ecol.*, 31, 1063–1075, <https://doi.org/10.1007/s10980-015-0318-x>, 2016.
- McFarlane, N.: Parameterizations: representing key processes in climate models without resolving them, *WIREs Clim. Change*, 2, 482–497, <https://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, *J. Adv. Model. Earth Sy.*, 6, 441–477, <https://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, 613–617, <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, *Agr. Forest Meteorol.*, 92, 211–225, [https://doi.org/10.1016/S0168-1923\(98\)00102-6](https://doi.org/10.1016/S0168-1923(98)00102-6), 1998.
- PAGES 2k-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, *Clim. Past*, 11, 1673–1699, <https://doi.org/10.5194/cp-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, *Clim. Past*, 12, 115–127, <https://doi.org/10.5194/cp-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, *Agr. Forest Meteorol.*, 101, 81–94, [https://doi.org/10.1016/S0168-1923\(99\)00169-0](https://doi.org/10.1016/S0168-1923(99)00169-0), 2000.
- Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Brannon, 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), *Geosci. Model Dev.*, 4, 33–45, <https://doi.org/10.5194/gmd-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, *J. Adv. Model. Earth Sy.*, 6, 141–184, <https://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, *Geosci. Model Dev.*, 10, 3207–3223, <https://doi.org/10.5194/gmd-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., Segsneider, J., Tjiputra, J., Aumont, O., and Romanou, A.: Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment, *Geosci. Model Dev.*, 9, 1827–1851, <https://doi.org/10.5194/gmd-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, *J. Climate*, 25, 4621–4640, <https://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, *J. Climate*, 26, 8597–8615, <https://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, *J. Climate*, 26, 5608–5623, <https://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, *WIREs Clim. Change*, 3, 63–77, <https://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, *J. Geophys. Res.-Sol. Ea.*, 108, 2431, <https://doi.org/10.1029/2002JB002351>, 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>, 2018.
- Stevens, M. B., González-Rouco, J. F., and Beltrami, H.: North American climate of the last millennium: Underground temperatures and model comparison, *J. Geophys. Res.-Earth*, 113, f01008, <https://doi.org/10.1029/2006JF000705>, 2008.
- Stieglitz, M. and Smerdon, J. E.: Characterizing Land–Atmosphere Coupling and the Implications for Subsurface Thermodynamics, *J. Climate*, 20, 21–37, <https://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, *B. Am. Meteorol. Soc.*, 93, 485–498, <https://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, *Palaeogeogr. Palaeoclimatol.*, 43, 237–259, [https://doi.org/10.1016/0031-0182\(83\)90013-5](https://doi.org/10.1016/0031-0182(83)90013-5), 1983.
- Vincenty, T.: Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations, *Surv. Rev.*, 23, 88–93, <https://doi.org/10.1179/sre.1975.23.176.88>, 1975.

- Wang, J. and Bras, R.: Ground heat flux estimated from surface soil temperature, *J. Hydrol.*, 216, 214–226, [https://doi.org/10.1016/S0022-1694\(99\)00008-6](https://doi.org/10.1016/S0022-1694(99)00008-6), 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, *J. Geophys. Res.-Atmos.*, 118, 4326–4347, <https://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, *J. Meteorol. Res.*, 28, 34–56, <https://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/10.3724/SP.J.1248.2013.041>, 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, *J. Meteorol. Soc. Jpn. II*, 90A, 23–64, <https://doi.org/10.2151/jmsj.2012-A02>, 2012.