

Centennial-Scale Temperature Change in Last Millennium Simulations and Proxy-Based Reconstructions

FREDRIK CHARPENTIER LJUNGQVIST,^{a,b,c} QIONG ZHANG,^{d,b} GUDRUN BRATTSTRÖM,^{e,b}
PAUL J. KRUSIC,^{c,d} ANDREA SEIM,^f QIANG LI,^{d,b} QIANG ZHANG,^{d,b} AND ANDERS MOBERG^{d,b}

^a Department of History, Stockholm University, Stockholm, Sweden

^b Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

^c Department of Geography, University of Cambridge, Cambridge, United Kingdom

^d Department of Physical Geography, Stockholm University, Stockholm, Sweden

^e Department of Mathematics, Stockholm University, Stockholm, Sweden

^f Chair of Forest Growth and Dendroecology, Institute of Forest Sciences,
University of Freiburg, Freiburg, Germany

(Manuscript received 14 August 2018, in final form 3 February 2019)


ABSTRACT

Systematic comparisons of proxy-based reconstructions and climate model simulations of past millennium temperature variability offer insights into climate sensitivity and feedback mechanisms, besides allowing model evaluation independently from the period covered by instrumental data. Such simulation–reconstruction comparisons can help to distinguish more skillful models from less skillful ones, which may subsequently help to develop more reliable future projections. This study evaluates the low-frequency simulation–reconstruction agreement within the past millennium through assessing the amplitude of temperature change between the Medieval Climate Anomaly (here, 950–1250 CE) and the Little Ice Age (here, 1450–1850 CE) in PMIP3 model simulations compared to proxy-based local and continental-scale reconstructions. The simulations consistently show a smaller temperature change than the reconstructions for most regions in the Northern Hemisphere, but not in the Southern Hemisphere, as well as a partly different spatial pattern. A cost function analysis assesses how well the various simulations agree with reconstructions. Disregarding spatial correlation, significant differences are seen in the agreement with the local temperature reconstructions between groups of models, but insignificant differences are noted when compared to continental-scale reconstructions. This result points toward a limited possibility to “rank” models by means of their low-frequency temperature variability alone. The systematically lower amplitude of simulated versus reconstructed temperature change indicates either too-small simulated internal variability or that the analyzed models lack some critical forcing or have missing or too-weak feedback mechanisms. We hypothesize that too-cold initial ocean conditions in the models—in combination with too-weak internal variability and slow feedbacks over longer time scales—could account for much of the simulation–reconstruction disagreement.

1. Introduction

Understanding the variability of temperature at local to global scales, over the past one to two millennia, offers insight into natural decadal to centennial-scale change that cannot be discerned from instrumental data. To this end considerable progress has been made

during the past decade using reconstructions from proxy data and climate model simulations (Masson-Delmotte et al. 2013; Smerdon and Pollack 2016; Luterbacher et al. 2016; Christiansen and Ljungqvist 2017; Zhang et al. 2018). Recently made advancements are fourfold: first, the network of temperature-sensitive proxy data has expanded both spatially and temporally (Jones et al. 2009; PAGES 2k Consortium 2013, 2017; Ljungqvist et al. 2012, 2016; Steiger et al. 2018; Tardif et al. 2018).

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Fredrik Charpentier Ljungqvist, fredrik.c.l@historia.su.se



This article is licensed under a Creative Commons Attribution 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1175/JCLI-D-18-0525.1

Second, improved calibration methods with better low-frequency preservation (Tingley et al. 2012; Smerdon and Pollack 2016; Christiansen and Ljungqvist 2017), and pseudoproxy experiments to evaluate the reconstruction methods' performance have been developed (Christiansen et al. 2009; Smerdon 2012; Li and Smerdon 2012; Smerdon et al. 2016; Wang et al. 2014, 2015; Klein and Goosse 2018; Neukom et al. 2018). Third, paleodata assimilation techniques have been developed that constrain model simulations with proxy-based climate estimates (Widmann et al. 2010; Goosse et al. 2012; Goosse 2017; Steiger et al. 2014, 2018; Hakim et al. 2016; Franke et al. 2017). Fourth, paleoclimate model simulations have become more widely available and also improved (Braconnot et al. 2012; Eby et al. 2013; Schmidt et al. 2014a; Harrison et al. 2016), in particular through an improvement of the forcing data used in the simulations (Schmidt et al. 2011, 2012; Jungclaus et al. 2017; see also Steinhilber et al. 2009, 2012; Shapiro et al. 2011; Krivova et al. 2010, 2011; Kaplan et al. 2011; Vieira et al. 2011; Sigl et al. 2015). Moreover, large ensembles have become available (e.g., CESM-LME; Otto-Bliesner et al. 2016).

These advances have facilitated an increasing number of simulation–reconstruction comparison studies (e.g., Fernández-Donado et al. 2013; Phipps et al. 2013; Schurer et al. 2013; PAGES 2k-PMIP3 Group 2015; Ljungqvist et al. 2016; Luterbacher et al. 2016; PAGES Hydro2k Consortium 2017). Simulation–reconstruction comparisons provide an opportunity to evaluate model performance using proxy records as “out of sample” test beds for the models' ability to reproduce both a warmer and colder climate under various forcings and boundary conditions, independent of the models' ability to realistically simulate modern climate conditions (e.g., Moberg 2013). Detection and attribution methods and other types of simulation–reconstruction comparisons have shown that changes in external forcing—in particular orbital, solar, volcanic, and greenhouse gases—are responsible for a large portion of the observed multidecadal to centennial-scale temperature change (e.g., Phipps et al. 2013; Schurer et al. 2013, 2014; Andres and Peltier 2016; Otto-Bliesner et al. 2016). Internal, unforced, climate variability also plays a significant role for multidecadal to centennial-scale temperature variability during the past millennium; it may in particular have been to a large degree responsible for the warmth during portions of medieval times (Goosse et al. 2012; Le et al. 2016; Cheung et al. 2017; Wang et al. 2017, 2018).

Fernández-Donado et al. (2013) compared 26 “last millennium” simulations from eight pre-PMIP3 models with global or hemispheric-scale temperature reconstructions.

They found that the simulations can reproduce the general patterns and timing of temperature variability, although not always the amplitude, during cold periods reasonably well; however, less so during warm periods. In particular, Fernández-Donado et al. (2013) as well as many other studies (e.g., Goosse et al. 2005, 2012; Zorita et al. 2010; Jungclaus et al. 2010; Phipps et al. 2013; Schurer et al. 2013, 2014; Luterbacher et al. 2016) have noted that the simulations fail to reproduce the peak medieval warmth around 900–1100 CE. Instead, they simulate a weaker medieval warming, more than a century after the proxy-reconstructed peak, consistent with high solar and relatively low volcanic forcing at that time. A relatively good agreement between simulated and reconstructed temperatures is found after approximately 1200 CE, when both simulations and proxy-based reconstructions follow the evolution in radiative forcings (Masson-Delmotte et al. 2013; Phipps et al. 2013; Schurer et al. 2013; Goosse 2017; Neukom et al. 2018). Model simulations in general produce smaller centennial-scale variability than proxy-based temperature reconstructions over the last millennium (Ault et al. 2013; Bothe et al. 2013a,b; Phipps et al. 2013; Schurer et al. 2014; Yan et al. 2015; Ljungqvist et al. 2016; Luterbacher et al. 2016; Zhang et al. 2018) and possess a different frequency spectrum (Lovejoy and Varotsos 2016; Dee et al. 2017) with less low-frequency and more high-frequency variability (PAGES 2k-PMIP3 Group 2015; Hartl-Meier et al. 2017).

Simulation–reconstruction comparisons can be carried out for a number of purposes including 1) to assess the performance of model simulations, 2) to explore the mechanisms forcing past climate changes as recorded in proxies, 3) to constrain future climate change projections, and 4) to explore the role of external forcings versus internal causes of past climate variability. Systematic comparisons of proxy-based reconstructions and model simulations can address long-term processes within the climate system and offer important insights into climate sensitivity and feedback mechanisms otherwise inaccessible. The models' performance in simulating low-frequency changes, beyond multidecadal time scales, cannot be evaluated using the short period covered by instrumental data. At the same time, model performance over these longer time scales is of significance for the reliability of long-term projections of anthropogenic global warming. The ability of the models to realistically simulate the low-frequency (centennial scale) temperature changes during the past one to two millennia is of particular interest for testing the models' climate sensitivity and feedback mechanisms

since the boundary conditions (besides anthropogenic forcing) are rather similar to those of the present day.

The centennial-scale temperature change of last millennium simulations with PMIP3 models has hitherto not been systematically benchmarked against reconstructions, in a way that potentially allows a “ranking” of models, as most simulation–reconstruction comparison studies have focused more on interannual to multidecadal scale agreement. Testing the ability to rank models is of interest since the models that better than others capture the amplitude of reconstructed low-frequency temperature change over the past millennium (or other time scales), as well as its spatial signature, are arguably also more likely to provide more reliable long-term climate projections (Schmidt 2010; Braconnot et al. 2012; Schmidt et al. 2014a; Jungclauss et al. 2017). Alternatively, one could argue that those models that more poorly than others simulate reconstructed past climate variations may not be expected to provide the most reliable future projections. However, care must be taken when interpreting results from model ranking exercises because the results are conditional upon the forcing data used to drive paleoclimate simulations. If the forcing data have errors, it will certainly affect the reliability of results from model ranking experiments. If the models systematically underestimate past low-frequency temperature variability, as many previous studies have shown, it may indicate that they have a too-low climate sensitivity, erroneous initial conditions, or wrong forcing configurations or have some missing or too-weak feedback mechanisms over longer time scales. Alternatively, it means that they have a too-small internal variability over longer time scales.

Here, we compare and analyze the reconstruction–model agreement of the low-frequency component over the past millennium climate. We do this by assessing the amplitude of temperature change between 950–1250 CE (henceforth the Medieval Climate Anomaly, MCA) and 1450–1850 CE (henceforth the Little Ice Age, LIA) in the PMIP3 last millennium (850–1850 CE) simulations using a set of 128 local temperature reconstructions and six PAGES 2k Consortium (2013) continental-scale temperature reconstructions. The time intervals of 950–1250 and 1450–1850 CE conform to IPCC definitions (Masson-Delmotte et al. 2013). We use the absolute MCA–LIA difference as a benchmark to assess the agreement between model-simulated and proxy-reconstructed centennial-scale amplitude of temperature change. Thus, we evaluate the models’ performance independently of their properties during the industrial period and without the need of a reference period.

Our objectives of the simulation–reconstruction comparison are as follows:

- 1) To systematically explore how well models simulate the amplitude of the reconstructed MCA–LIA temperature difference and its spatial signature. This may inform us both of their response to external forcing and their ability to simulate internal variability in the low-frequency domain.
- 2) To determine which simulations agree significantly better than others with the reconstructed amplitude of MCA–LIA temperature difference and explore the possibility of ranking the various simulations according to their agreement with the amplitude of this difference.
- 3) As part of our assessment, we investigate to what extent the reconstructed and simulated polar amplification of the MCA–LIA temperature difference agrees. This key aspect of the climate system has been insufficiently explored for the past millennium.

We conclude the study by discussing possible reasons for observed differences in the amplitude of centennial-scale temperature change in simulations and proxy-based reconstructions, and attempt to reconcile observed simulation–reconstruction differences.

2. Data and methods

a. Local temperature-sensitive proxy reconstructions

We assembled a global and millennium-long dataset of 51 annual-mean, 56 summer, and 21 winter temperature-sensitive local (i.e., site-specific) proxy-based reconstructions from the scientific literature. All reconstructions were calibrated to temperature by the original authors and have a resolution of at least two observations per century (Table 1; Fig. 1). Reconstructions with less than annual resolution are required to have their dating constrained by at least one age after 1700 CE (D_1), one age between 950 and 1700 CE (D_2), and one age before 950 CE (D_3) [see Ljungqvist et al. (2016) for further details]. All non-annually resolved local reconstructions are converted to annual time series by simple linear interpolation between the available dates. For the sake of compatibility with the model simulations we let all boreal summer (austral winter) reconstructions, regardless of the exact monthly interval used in calibration, represent June–August (JJA), and all boreal winter (austral summer) reconstructions represent December–February (DJF). Although they may well calibrate to annual-mean temperature, few proxies published as annual temperature reconstructions are in reality likely to be equally temperature sensitive in all 12 months of the

TABLE 1. List of all local reconstructions, sorted from north to south, used in this study with information about site/location, latitude and longitude coordinates, archive type, proxy type, temporal resolution (Res.), seasonality, dating method(s), reference to original publication, temperature difference between 950–1250 and 1450–1850 CE, and the weight assigned to each proxy for the cost function calculations based on their estimated uncertainty interval. The weights are standardized to sum to 1 and may be used directly to compute the cost function for all seasons taken together. Digitized local reconstructions are marked with an asterisk after the reference.

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
Lower Murray Lake	−69.32	81.21	Lake sediment	Mass accumulation rate	Annual	July	Annual lamina thickness	Cook et al. (2009)	0.64	0.006
MSM05/5_712-1	6.77	78.92	Marine sediment	Planktic foraminifers	Multidecadal	JAS	¹⁴ C	Spielhagen et al. (2011)	0.53	0.004
Lomonosovfonna	17.42	78.85	Ice core	$\delta^{18}\text{O}$	Annual to multidecadal	DJF	Glaciological flow model	Divine et al. (2011)	3.06	0.000
Kongressvatnet	13.93	78.02	Marine sediment	Alkenone U_{37}^{K}	Decadal	JJA	¹⁴ C, tephra	D'Andrea et al. (2012)	−0.58	0.002
GISP2	−38.50	72.60	Ice core	$\delta^{18}\text{O}$	Annual	Annual	Annual lamina thickness	Alley (2000)	0.69	0.011
GISP2	−38.50	72.60	Ice core	$\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$	Annual	Annual	Annual lamina thickness	Kobashi et al. (2011)	0.47	0.006
GRIP	−37.38	72.35	Ice core	$\delta^{18}\text{O}$	Annual	Annual	Annual lamina thickness	Vinther et al. (2010)	1.07	0.003
Over Gunnarsfjorden	28.17	71.02	Lake sediment	Pollen	Multidecadal	January ^a	¹⁴ C	Allen et al. (2007)	−3.08	0.000
Over Gunnarsfjorden	28.17	71.02	Lake sediment	Pollen	Multidecadal	July ^b	¹⁴ C	Allen et al. (2007)	1.12	0.001
Indigirka	148.15	70.53	Tree ring	Ring width	Annual	JJA	Annual lamina thickness	Solomina and Alverson (2004) *	0.33	0.032
Avam-Taimyr	93.00	70.00	Tree ring	Ring width	Annual	JJ	Annual lamina thickness	Naurzbaev et al. (2002)	0.37	0.006
Yamal	69.00	69.00	Tree ring	Ring width	Annual	JJA	Annual lamina thickness	Briffa et al. (2013)	0.17	0.034
Finnish Lapland	25.00	69.00	Tree ring	Ring width	Annual	July	Annual lamina thickness	Helama et al. (2010)	0.12	0.010
Kurupa Lake	−154.61	68.35	Lake sediment	Relative absorption-band depth	Multidecadal	JJAS	Annual lamina thickness ^{239,240} Pu, ¹⁴ C	Boldt et al. (2015)	0.22	0.000
Torneträsk	19.80	68.31	Tree ring	Maximum density	Annual	JJA	Annual lamina thickness	Melvin et al. (2013)	0.27	0.009
Northern Scandinavia	25.00	68.00	Tree ring	Maximum density	Annual	JJA	Annual lamina thickness	Esper et al. (2012)	0.25	0.002

TABLE 1. (Continued)

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
MD99-2322	−30.83	67.14	Marine sediment	Diatom	Decadal	August	^{14}C , tephra	Miettinen et al. (2015)	0.08	0.002
Braya Sjø	−50.42	67.00	Lake sediment	Alkenone $\text{U}_{37}^{\text{K}'}$	Multidecadal	JJ	$^{14}\text{C}^{\text{e}}$	von Gunten et al. (2012)	−0.10	0.000
MD95_2011	7.64	66.97	Marine sediment	Diatoms	Decadal	August	^{14}C	Berner et al. (2011)	0.35	0.001
Donard Lake	−61.35	66.66	Lake sediment	Varve thickness	Annual	JJA	Annual lamina thickness	Moore et al. (2001)	0.12	0.005
North Iceland Shelf	−17.22	66.33	Marine sediment	Diatom	Decadal	JJA	^{14}C , tephra	Jiang et al. (2005)	0.44	0.001
North Iceland Shelf	−17.22	66.33	Marine sediment	Diatom	Decadal	DJF	^{14}C , tephra	Jiang et al. (2005)	0.50	0.001
MD99-2275 ^d	−19.30	66.30	Marine sediment	Alkenone $\text{U}_{37}^{\text{K}'}$	Decadal	July	^{14}C , tephra	Sicre et al. (2011)	1.03	0.002
MD99-2275 ^e	−19.30	66.30	Marine sediment	Diatom	Decadal	JAS	^{14}C , tephra	Ran et al. (2011)	0.75	0.002
Screaming Lynx Lake	−145.40	66.07	Lake sediment	Chironomid transfer function	Multidecadal	July	^{210}Pb , ^{14}C	Clegg et al. (2011)	−0.14	0.000
Dye-3	−43.49	65.11	Ice core	$\delta^{18}\text{O}$	Annual	Annual	Annual lamina thickness	Vinther et al. (2010)	1.62	0.001
Iceland	−18.00	65.00	Historical	Documentary	Multidecadal	Annual	Documentary	Bergthórsön (1969)*	0.55	0.002
Lake Pieni-Kauro	30.12	64.28	Lake sediment	Chironomids	Multidecadal	July	^{137}Cs , ^{14}C	Luoto and Helama (2010)	0.38	0.012
Jämtland	14.00	63.00	Tree ring	Density	Annual	AMJJAS	Annual lamina thickness	Zhang et al. (2016)	0.31	0.010
Farewell Lake	−153.63	62.55	Lake sediment	$\delta^{18}\text{O}$	Multidecadal	JJA	^{210}Pb , ^{14}C	Hu et al. (2001)	0.86	0.000
Finnish Lakelands	27.50	62.00	Tree ring	Maximum density	Annual	MJJA	Annual lamina thickness	Helama et al. (2014)	0.36	0.034
Hallet Lake	−146.20	61.50	Lake sediment	Biogenic silica	Multidecadal	JJA	^{14}C	McKay et al. (2008)	−0.31	0.003
RAPID-35-25B	−19.53	61.47	Marine sediment	Mg/Ca	Multidecadal	Annual	^{14}C	Moffa-Sánchez et al. (2014)	0.80	0.001
Moose Lake	−143.61	61.37	Lake sediment	Chironomid transfer function	Multidecadal	July	^{210}Pb , ^{14}C	Clegg et al. (2010)	0.39	0.000
Iceberg Lake ^f	−142.95	60.78	Lake sediment	Varve thickness	Annual	MJ	^{14}C	Loso (2009)	0.07	0.004
Gulf of Alaska ^g	−141.00	59.00	Tree ring	Ring width	Annual	FMAMJJJA	Annual lamina thickness	Wiles et al. (2014)	0.43	0.008
RAPID21-3K ^h	−27.91	57.45	Marine sediment	Alkenone $\text{U}_{37}^{\text{K}'}$	Decadal	MJJA	^{14}C	Sicre et al. (2011)	0.18	0.002
RAPID21-3K	−27.91	57.45	Marine sediment	Diatom	Decadal	August	^{14}C	Miettinen et al. (2012)	−0.39	0.002
Feni Drift	−13.99	55.65	Marine sediment	Mg/Ca	Multidecadal	Annual	^{137}Cs , ^{210}Pb , ^{14}C	Richter et al. (2009)	0.74	0.001

TABLE 1. (Continued)

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
Northern Quebec	−71.00	54.00	Tree ring	Ring width	Annual	July–August	Annual lamina thickness	Gennaretti et al. (2014)	1.108	0.001
Colombia Icefield	−117.15	52.15	Tree ring	$\delta^{13}\text{C}$	Decadal	DJF	Annual lamina thickness	Edwards et al. (2008)	3.71	0.001
Colombia Icefield	−117.15	52.15	Tree ring	Maximum density	Annual	MJJA	Annual lamina thickness	Luckman and Wilson (2005)	0.22	0.020
DeBilt	5.180	52.10	Historical	Documentary	Decadal	DJF	Documentary	van Engelen et al. (2001)	0.50	0.034
DeBilt	5.180	52.10	Historical	Documentary	Decadal	JJA	Documentary	van Engelen et al. (2001)	0.00	0.034
Central England	1.00	52.00	Historical	Documentary	Multidecadal	Annual	Documentary	Lamb (1965)	0.70	0.034
Central England	1.00	52.00	Historical	Documentary	Multidecadal	DJF	Documentary	Lamb (1965)	0.47	0.034
Central England	1.00	52.00	Historical	Documentary	Multidecadal	JA	Documentary	Lamb (1965)	0.82	0.034
Teletskoe Lake	87.61	51.76	Lake sediment	Multiproxy	Annual	Annual	Annual lamina thickness	Kalugin et al. (2009)	0.27	0.001
Onondaga Nuruu	99.08	51.15	Tree ring	Ring width	Annual	JJ	Annual lamina thickness	Davi et al. (2015)	−0.02	0.008
Altai	87.50	50.00	Tree ring	Ring width	Annual	JJA	Annual lamina thickness	Büntgen et al. (2016)	0.31	0.006
Nadas Lake ^b	19.70	47.99	Lake sediment	Pollen	Multidecadal	Annual	^{14}C	Sümegi et al. (2009)	0.87	0.000
Nadas Lake ⁱ	19.70	47.99	Lake sediment	Pollen	Multidecadal	July	^{14}C	Sümegi et al. (2009)	0.80	0.000
Nadas Lake ^j	19.70	47.99	Lake sediment	Pollen	Multidecadal	January	^{14}C	Sümegi et al. (2009)	1.03	0.000
Southeastern Newfoundland	−54.55	47.13	Marine sediment	Alkenone U_{37}^{K}	Multidecadal	Annual	^{210}Pb , ^{14}C	Sicre et al. (2014)	0.11	0.005
Spannagel Cave	11.40	47.05	Speleothem	$\delta^{18}\text{O}$	Decadal	Annual	Th/U	Mangini et al. (2005)	1.09	0.024
Lake Neuchâtel	6.70	46.80	Lake sediment	$\delta^{18}\text{O}$	Multidecadal	Annual	^{14}C	Filippi et al. (1999)*	1.49	0.000
Lötschental	8.00	46.30	Tree ring	Maximum density	Annual	JJAS	Annual lamina thickness	Büntgen et al. (2006)	0.78	0.006
Conroy Lake	−67.88	46.28	Lake sediment	Pollen	Multidecadal	JJA	Annual lamina thickness	Gajewski (1988)	0.57	0.001
Hells Kitchen Lake	−89.42	46.11	Lake sediment	Pollen	Multidecadal	JJA	Annual lamina thickness	Gajewski (1988)	−0.06	0.002

TABLE 1. (Continued)

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
Alpine arc	8.00	46.00	Tree ring	Maximum width	Annual	JJA	Annual lamina thickness	Büntgen et al. (2011)	0.35	0.006
Colle Gnifetti	7.88	45.93	Ice core	Ca ²⁺	Decadal	AMJJAS	Annual lamina thickness, ¹⁴ C	Bohleber et al. (2018)	−0.10	0.030
Lake Anterne	6.47	45.59	Lake sediment	Chironomid assemblages	Multidecadal	July	¹⁴ C	Milliet et al. (2009)	0.44	0.000
Clear Pond	−74.01	43.45	Lake sediment	Pollen	Multidecadal	JJA	Annual lamina thickness	Gajewski (1988)	0.11	0.001
Gulf of Lion	3.28	43.38	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Annual	¹⁴ C	Sicre et al. (2016)	0.21	0.006
Northern Spain	−3.50	42.90	Speleothem	δ ¹³ C	Decadal	Annual	²³⁰ TH	Martin-Chivelet et al. (2011)	0.27	0.034
Jones Lake	−84.56	42.77	Lake sediment	Pollen	Multidecadal	MJJAS	¹⁴ C	Bernabo (1981)	0.09	0.006
Lake Redon	0.77	42.64	Lake sediment	Chrysophyte cysts	Multidecadal	DJFMAM	¹⁴ C	Pla and Catalan (2005)	−0.17	0.001
Lake Sihailongwan	126.36	42.17	Lake sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Summer	Varves, ¹⁴ C	Chu et al. (2012)	0.33	0.001
Qigai Nur	109.50	39.50	Lake sediment	Pollen	Multidecadal	Annual	¹⁴ C	Sun and Feng (2013)	−0.90	0.001
ShiHua Cave	115.56	39.47	Speleothem	Thickness variations	Annual	MJJA	Annual lamina thickness	Tan et al. (2003)	1.01	0.034
Chesapeake Bay	−76.40	39.00	Marine sediment	Mg/Ca	Decadal	MAM	¹⁴ C	Cronin et al. (2003)	0.88	0.000
Sugan Lake	93.90	38.85	Lake sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	JJA	¹⁴ C	He et al. (2013)	1.84	0.000
Qilian Mountains	99.50	38.30	Tree ring	Ring width	Annual	JFMAMJJJA	Annual lamina thickness	Zhang et al. (2014)	−0.09	0.004
Southern Sierra Nevada	−118.90	36.90	Tree ring	Ring width	Annual	JJA	Annual lamina thickness	Graumlich (1993)	0.14	0.023
Eastern Tibetan Plateau	98.50	36.50	Tree ring	Ring width	Annual	Annual	Annual lamina thickness	Liu et al. (2009)*	0.23	0.008
Karakorum Mountains	74.99	36.37	Tree ring	δ ¹³ C	Annual	JJ	Annual lamina thickness	Treydte et al. (2009)	0.25	0.004
Karakorum Mountains	74.99	36.37	Tree ring	Ring width	Annual	Annual	Annual lamina thickness	Esper et al. (2002)	0.47	0.007
North Pacific Gyre ^k	141.80	36.00	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Annual	¹⁴ C	Isono et al. (2009)	0.59	0.001
Southern Colorado Plateau	−111.40	35.20	Tree ring	Ring width	Annual	Annual	Annual lamina thickness	Salzer and Kipfmüller (2005)	−0.01	0.006
Animaqin Mountains	99.07	35.16	Tree ring	Ring width	Annual	AMJ	Annual lamina thickness	Chen et al. (2016)	−0.12	0.034
Kuroshio Extension	145.00	35.00	Speleothem	δ ¹⁸ O	Annual	Annual	²³⁰ TH	McCabe-Glynn et al. (2013)	0.03	0.034

TABLE 1. (Continued)

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
East China	114.00	35.00	Historical	Documentary	Decadal	ONDJFMA	Documentary	Ge et al. (2003)	0.48	0.034
Gahai Lake	102.33	34.24	Lake sediment	Alkenone U ₃₇ ^{K'}	Decadal	JJA	¹⁴ C	He et al. (2013)	2.02	0.001
Seto Inland Sea	132.33	34.18	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	JJAS	¹⁴ C	Kawahata et al. (2017)	−0.15	0.011
Bermuda Rise	−57.63	33.72	Marine sediment	δ ¹⁸ O	Multidecadal	Annual	¹⁴ C	Keigwin (1996)	0.75	0.005
Western Himalaya	76.45	32.50	Tree ring	Ring width	Annual	MJJA	Annual lamina thickness	Yadav et al. (2011)	0.75	0.007
GeoB6008-1	−10.10	30.85	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Annual	¹⁴ C	McGregor et al. (2007)	−0.07	0.007
Yakushima Island	130.30	30.20	Tree ring	δ ¹³ C	Decadal	Annual	Annual lamina thickness	Kitagawa and Matsumoto (1995)*	0.56	0.001
Pigmy basin ^l	−91.42	27.20	Marine sediment	Mg/Ca	Multidecadal	Annual	¹⁴ C	Richey et al. (2007)	1.70	0.000
SO90-56KA /SO90-39KG ⁿ	65.92	24.83	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Annual	Annual lamina thickness, tephra, ¹⁴ C	Doose-Rolinski et al. (2001)	−0.05	0.001
Cores SO90/39KG and SO130/275KL ⁿ	65.92	24.83	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Annual	Varve counting and ¹⁴ C	Böll et al. (2014)	0.35	0.002
ODP Hole 1202B	122.50	24.80	Marine sediment	TEX ₈₆	Decadal	Annual	¹⁴ C	Wu et al. (2012)	2.89	0.000
Northern South China Sea	113.90	19.80	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	Annual	¹⁴ C	Kong et al. (2017)	0.08	0.006
39KG/275KL	68.07	18.04	Marine sediment	Planktic foraminiferal	Decadal	JFM	¹⁴ C	Munz et al. (2015)	−0.02	0.002
Northeastern Caribbean	−66.60	17.88	Marine sediment	Planktonic foraminifera	Multidecadal	ASO	²¹⁰ Pb, ¹³⁷ C, ¹⁴ C	Nyberg et al. (2002)	−0.02	0.001
Northeastern Caribbean	−66.60	17.88	Marine sediment	Planktonic foraminifera	Multidecadal	FMA	²¹⁰ Pb, ¹³⁷ C, ¹⁴ C	Nyberg et al. (2002)	0.86	0.002
GeoB 9501-5	−16.73	16.83	Marine sediment	Mg/Ca	Multidecadal	Annual	¹⁴ C	Kuhnert and Mulitza (2011)	0.52	0.002
GeoB8331-4°P178-15	−29.13	16.72	Marine sediment	Alkenone U ₃₇ ^{K'}	Multidecadal	SONDJFM	¹⁴ C	Leduc et al. (2010)	0.17	0.001
	44.30	11.96	Marine sediment	TEX ₈₆	Multidecadal	Annual	²¹⁰ Pb, ¹⁴ C	Tierney et al. (2015)	0.01	0.001
Cariaco basin ^p	−64.56	10.42	Marine sediment	Mg/Ca	Annual	JJASON	¹⁴ C	Wurtzel et al. (2013)	0.27	0.006
Cariaco basin ^p	−64.56	10.42	Marine sediment	Mg/Ca	Annual	DJFMAM	¹⁴ C	Wurtzel et al. (2013)	0.63	0.016
MD98-2181	126.00	6.30	Marine sediment	Mg/Ca	Multidecadal	Annual	¹⁴ C	Stott et al. (2004)	0.28	0.001
Indonesian Throughflow	119.08	1.40	Marine sediment	Mg/Ca	Multidecadal	Annual	¹⁴ C	Newton et al. (2011)	0.37	0.002

TABLE 1. (Continued)

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
Niño-4 region ^q	−155.00	0.00	Ice core	$\delta^{18}\text{O}$	Decadal	Annual	Annual lamina thickness ^{210}Pb , ^{14}C	Thompson et al. (2013)	0.09	0.001
Indo-Pacific warm pool	119.27	−3.53	Marine sediment	$\delta^{18}\text{O}$	Decadal	Annual		Oppo et al. (2009)	0.61	0.021
TTR17-1_384B ^r	36.09	−4.53	Marine sediment	Alkenone U_{37}^{K}	Multidecadal	Annual	^{210}Pb , ^{14}C	Nieto-Moreno et al. (2013)	0.00	0.001
TTR17-1_384B ^s	36.09	−4.53	Marine sediment	TEX ₈₆	Multidecadal	JJA	^{210}Pb , ^{14}C	Nieto-Moreno et al. (2013)	0.03	0.000
Lake Tanganyika	29.58	−6.33	Lake sediment	TEX ₈₆	Multidecadal	Annual	^{210}Pb , ^{14}C	Tierney et al. (2010)	0.27	0.002
Nevado Illimani	−67.48	−16.39	Ice core	Ice-core ammonium concentration	Annual	Annual	Colcanic time horizons; ^{14}C $^{234}\text{U}/^{238}\text{U}$	Kellerhals et al. (2010)	0.24	0.034
Makapansgat Valley ^t	29.20	−24.01	Speleothem	Color banding	Decadal	Annual		Holmgren et al. (2001)	−0.01	0.014
Laguna Chépica	−70.50	−32.24	Lake sediment	Reflectance spectroscopy	Annual	NDJF	^{210}Pb , ^{137}C , ^{14}C	de Jong et al. (2013)	−0.40	0.007
Rio Alerce ^u	−73.00	−42.00	Tree ring	Ring width	Annual	DJF	Annual lamina thickness	Lara and Villalba (1993)*	0.37	0.022
Mt. Read	148.00	−43.00	Tree ring	Ring width	Annual	NDJFMA	Annual lamina thickness	Cook et al. (2000)	0.03	0.009
Oroko Swamp	170.28	−43.23	Tree ring	Ring width	Annual	JFM	Annual lamina thickness	Cook et al. (2002)	−0.18	0.034
GeoB7186-3 ^v	−75.15	−44.15	Marine sediment	Alkenone U_{37}^{K}	Multidecadal	Annual	^{14}C	Mohtadi et al. (2007)	0.16	0.001
Jacaf Fjord (core CF7-PC33) ^u	−72.97	−44.37	Marine sediment	Alkenone U_{37}^{K}	Multidecadal	DJF	^{14}C	Sepúlveda et al. (2009)	0.48	0.001
Laguna Escondida	−71.82	−45.52	Lake sediment	Biogenic silica flux	Annual	SONDJF	^{210}Pb , ^{137}C , ^{14}C	Elbert et al. (2013)	0.41	0.004
Homestead Scarp	169.10	−52.50	Lake sediment	Pollen	Multidecadal	January	^{14}C	McGlone et al. (2010)	0.13	0.000
Mount Honey	169.10	−52.50	Lake sediment	Pollen	Multidecadal	January	^{14}C	McGlone et al. (2010)	0.20	0.000
James Ross Island	−64.20	−57.69	Ice core	Ice-core δD	Decadal	Annual	Glaciological flow model ^{14}C	Mulvaney et al. (2012)	−0.04	0.021
ODP Site 1098 ^w	−64.21	−64.86	Marine sediment	TEX ₈₆	Multidecadal	Annual		Shevenell et al. (2011)	0.99	0.000
Antarctic Peninsula	−65.00	−67.00	Ice core	$\delta^{18}\text{O}$	Decadal	Annual	Annual lamina thickness	Stenni et al. (2017)	−0.13	0.001
Wilkes Land Coast	120.00	−68.00	Ice core	$\delta^{18}\text{O}$	Decadal	Annual	Annual lamina thickness	Stenni et al. (2017)	0.10	0.003

TABLE 1. (Continued)

Location	Lon (°)	Lat (°)	Archive	Proxy	Res.	Season	Dating	Reference	Temp difference (°C)	Weight
Victoria Land	140.00	−74.00	Ice core	$\delta^{18}\text{O}$	Decadal	Annual	Annual lamina thickness	Stenni et al. (2017)	0.56	0.001
EPICA Dome C	123.40	−75.10	Ice core	Ice-core δD	Multidecadal	Annual	Glaciological flow model	Jouzel et al. (2007)	0.30	0.002
Vostok	108.00	−78.50	Ice core	Ice-core δD	Multidecadal	Annual	Glaciological flow model	Petit et al. (1999)	−0.19	0.001
West Antarctic Ice Sheet	−135.00	−80.00	Ice core	$\delta^{18}\text{O}$	Decadal	Annual	Annual lamina thickness	Stenni et al. (2017)	0.47	0.002
East Antarctic Plateau	60.00	−87.50	Ice core	$\delta^{18}\text{O}$	Decadal	Annual	Annual lamina thickness	Stenni et al. (2017)	0.25	0.001

^a Coldest month.^b Warmest month.^c Dating in [D'Andrea et al. \(2011\)](#).^d Months not explicitly stated; uncertainty estimate from [Rosell-Melé et al. \(2001\)](#).^e Months not explicitly stated.^f Dating in [Loso et al. \(2006\)](#).^g May contain some limitations in fully preserving centennial-scale temperature variability; see [Esper et al. \(2016\)](#).^h Uncertainty estimate from [Rosell-Melé et al. \(2001\)](#).ⁱ Uncertainty inferred from [Davis et al. \(2003\)](#); warmest month.^j Uncertainty inferred from [Davis et al. \(2003\)](#); coldest month.^k Composite of cores MD01-2421; KR02-06 St.A GC; KR02-06 St.A MC. The seasonal interpretation is not explicitly stated by [Isono et al. \(2009\)](#) and is inferred by us from the literature.^l Uncertainty from [Richey et al. \(2011\)](#).^m Uncertainty estimate from [Rosell-Melé et al. \(2001\)](#); dating in [von Rad et al. \(1999\)](#).ⁿ Dating in [von Rad et al. \(1999\)](#).^o Uncertainty estimate from [Rosell-Melé et al. \(2001\)](#); dating in [Herbert and Compton \(2007\)](#); sensitive to the upwelling season in the Southern Benguela which occurs between September and March.^p Dating in [Goñi et al. \(2003\)](#).^q Inferred from $\delta^{18}\text{O}$ ice-core data from Quelccaya Summit Dome Ice Core (13.93°S, 70.83°W).^r Uncertainty estimate from [Rosell-Melé et al. \(2001\)](#).^s Uncertainty estimate from [Kim et al. \(2008\)](#).^t Dating in [Holmgren et al. \(1999\)](#).^u This record may contain some limitations in fully preserving centennial-scale temperature variability; see [Esper et al. \(2016\)](#).^v Uncertainty estimate from [Rosell-Melé et al. \(2001\)](#).^w Dating in [Domack et al. \(2001\)](#).

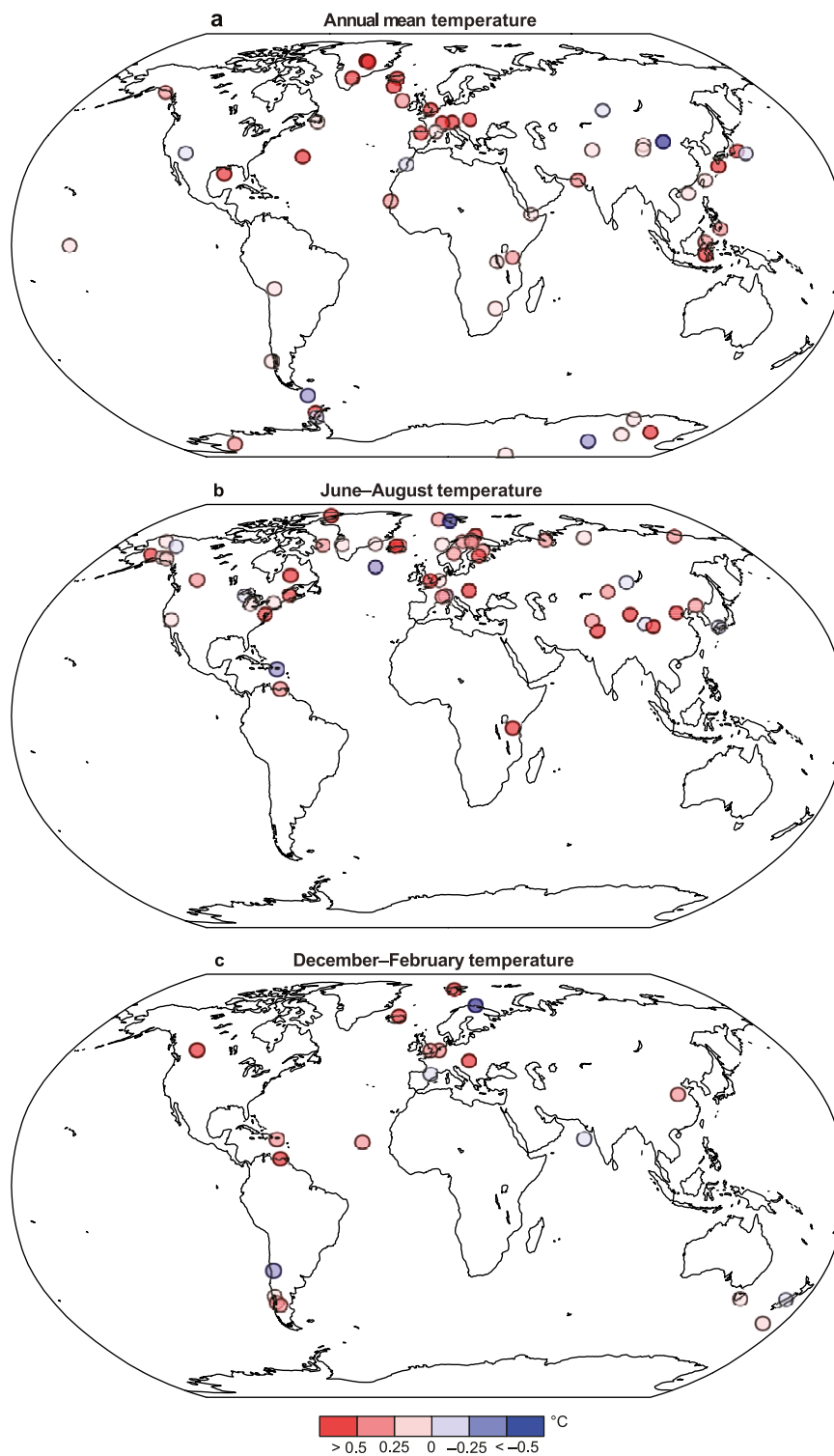


FIG. 1. Temperature difference ($^{\circ}\text{C}$) between the Medieval Climate Anomaly (950–1250 CE) and the Little Ice Age (1450–1850 CE) for local reconstructions reflecting (a) annual-mean temperature (51 records), (b) JJA temperature (56 records), and (c) DJA temperature (21 records).

TABLE 2. The [PAGES 2k Consortium \(2013\)](#) continental-scale temperature reconstructions with information about latitude and longitude coordinates, temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE, standard deviation (SD), variance (V), correction for autocorrelation, and assigned weights. For the Arctic, the updated versions by [McKay and Kaufman \(2014\)](#), for Europe by [Luterbacher et al. \(2016\)](#), and for Antarctica by [Stenni et al. \(2017\)](#) were used.

Region	Coordinates	Temp difference ($^{\circ}\text{C}$)	SD ($^{\circ}\text{C}$)	V	A	$(V \times A)^{-1}$	Weight
Antarctica	60°–90°S, 180°W–180°E	0.37	0.053	0.002 80	7.32	48.80	0.162
Arctic	60°–90°N, 180°W–180°E	0.77	0.028	0.000 76	18.65	70.95	0.236
Asia	23.5°–55°N, 60°–160°E	0.18	0.042	0.001 75	12.44	45.81	0.152
Europe	35°–70°N, 10°W–40°E	0.40	0.033	0.001 06	15.90	59.35	0.197
North America	30°–55°N, 130°–75°W	0.19	0.066	0.004 31	5.96	38.87	0.129
South America	20°–65°S, 80°–30°W	0.09	0.034	0.001 14	23.76	37.01	0.123

year. To ease the comparison with model simulations, we also treat sea surface temperature reconstructions as near-surface temperature indicators. This approach is justifiable given the high correlation between near-surface temperature and sea surface temperature over long time scales (e.g., [Rayner et al. 2003](#)). For the sake of comparability between model data and reconstructions, the simulated values have been made to correspond to the reconstructed ones in both time and space, that is, we use simulated values for the sites where we have proxy data, at the same time points. This means that for sites where the reconstruction has less-than-annual resolution, so will the simulated values used. We then subject these simulated values to the same linear interpolation as the reconstructions, even though actual annually resolved data are of course obtainable for the simulations.

UNCERTAINTY ESTIMATES IN LOCAL RECONSTRUCTIONS

Most local reconstructions used in this study were published with some measure of uncertainty. The nature of this uncertainty (calibration error, measurement error, etc.) is not always given explicitly, nor is it always clear how it is to be understood in statistical terms, for example, whether it is a 95% confidence interval or a standard deviation. Nevertheless, we attempt to make these measures comparable across studies, which is necessary in order to assign weights (see [section 2d](#) and [appendix A](#)) by reinterpreting them as standard deviations SD_{MCA} and SD_{LIA} of the statistical error in the reconstructions. Under the assumption of normality, a confidence interval may be converted into a standard deviation as shown in [Table 3](#), and we have taken the expression “root-mean-square error” to mean a standard deviation. When it was explicitly stated that the uncertainty is the standard deviation (SD) of the average of m values, the uncertainty is re-expressed as SD/\sqrt{m} . However, for many reconstructions assigning an SD value to the annual or interpolated values is qualified guesswork. To proxies with no published estimate of uncertainty

we have assigned the pooled standard deviation¹ of all other reconstructions from the same archive type. The MCA (950–1250 CE) and the LIA (1450–1850 CE) values are represented by 300- and 400-yr averages, so we use SD_{MCA} and SD_{LIA} to calculate the standard deviations of these averages, taking autocorrelation into account (see [section 2d](#) and [appendix A](#) for details).

b. The PAGES 2k continental-scale temperature reconstructions

The [PAGES 2k Consortium \(2013\)](#) produced seven continental-scale reconstructions of various lengths using either a multiproxy network or only tree-ring data: the Arctic ($>60^{\circ}\text{N}$), Europe, Asia, North America, South America, Australasia, and Antarctica ([Table 2](#)). To cover the full study period of 950–1850 CE, we exclude the Australasian reconstruction (1001–2001 CE) and the tree-ring-based reconstruction (1204–1974 CE) for North America. Representing the Arctic, we use the updated [McKay and Kaufman \(2014\)](#) version, for Europe the updated [Luterbacher et al. \(2016\)](#) version, and for Antarctica the updated [Stenni et al. \(2017\)](#) version. Seasonal targets are annual-mean temperature for the Arctic, North America, and Antarctica, JJA for Europe and Asia, and DJF for South America. The number of proxies contributing to each reconstruction and their temporal changes in coverage as well as the proxies’ signal strength among the reconstructions are not homogenous [see Fig. 2 in [PAGES 2k Consortium \(2013\)](#)].

Despite agreeing upon a common set of minimum criteria, the various [PAGES 2k Consortium \(2013\)](#) regional working groups used different criteria for their proxy selection and used different methods for calibration; primarily by scaling the proxy composite to an

¹ That is, we computed the arithmetic mean of the corresponding variances and then obtained the pooled standard deviation by taking the square root (see [Box et al. 2005](#), 184–185).

TABLE 3. Connecting uncertainties u and standard deviations s of the continental-scale temperature reconstructions.

Type of uncertainty	Standard deviation
$2 \times$ standard error	$u/2$
90% confidence interval	$u/1.645$
95% confidence interval	$u/1.96$

instrumental target or by using linear regression-based techniques. The influence of different calibration methods was found, in most cases, to be relatively limited. However, the different proxy selection strategies can arguably result in different strengths of the reconstructed temperature signal and may have an effect on the amount of preserved low-frequency information (e.g., Christiansen and Ljungqvist 2017). In addition, each PAGES2k group calculated their reconstruction's levels of uncertainty differently, all assuming stationarity between temperature and proxy values. We have reinterpreted all these uncertainties so that they should represent SDs (Table 3).

c. PMIP3 last millennium simulations

Ten groups within the Paleoclimate Modeling Intercomparison Project phase 3 (PMIP3; Braconnot et al. 2012) of the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al. 2012) have produced forced-transient simulations covering 850–1850 CE. All simulations follow the PMIP3 protocol regarding

orbital, total solar irradiation, volcanic, and land-use forcings (Schmidt et al. 2011, 2012; Table 4). The surface temperature from the simulations covering the period of 850–1850 CE used in this study are shown in Fig. 2 for the global mean, Northern Hemisphere (NH) mean, and Southern Hemisphere (SH) mean, respectively.

Simulations with the following models were downloaded from the CMIP5 database (<https://esgf-data.dkrz.de/search/cmip5-dkrz/>) and used in this study: BCC_CSM1.1 (BCC hereafter), CCSM4, CSIRO Mk3L1.2 (CSIRO hereafter), FGOALS-s2 (FGOALS hereafter), GISS-E2-R (GISS hereafter), HadCM3, IPSL-CM5A-LR (IPSL hereafter), and MPI-ESM-P (MPI hereafter). The GISS modeling group produced an ensemble with eight members (with different combinations of volcanic, solar and land-use forcings; see Table 4). We selected the three GISS simulations [r11p121 (GISS-R21), r11p124 (GISS-R24), and r11p127 (GISS-R27)] with correctly implemented volcanic forcing and also containing all other forcings as prescribed by Schmidt et al. (2011, 2012). Additionally, we include a recently produced simulation with EC-Earth v3.1 (Hazeleger et al. 2010, 2012; see appendix B). The simulations from the MIROC-ESM and FGOALS-g1 models, which are also available in the CMIP5 database, were discarded in the present study. MIROC-ESM shows a strong warming drift ($\sim 0.09^\circ\text{C century}^{-1}$) caused by a decreasing trend in the Southern Ocean sea

TABLE 4. Description of CMIP5/PMIP3 climate models used for the last millennium simulations and their forcing configuration. When a certain forcing is not specified, a constant “preindustrial” value is used. Orbital forcing in the models follows Berger (1978). Abbreviations in the table for the different forcing datasets are as follows. Solar forcing: Sa = Steinhilber et al. (2009) spliced to Wang et al. (2005); Sb = Vieira and Solanki (2010) spliced to Wang et al. (2005). Volcanic forcing: Va = Gao et al. (2008); Vb = Crowley and Unterman (2013); Vc = Ammann et al. (2007). Well-mixed greenhouse gases concentrations: Ga = Flückiger et al. (1999, 2002); Machida et al. (1995); Gb = Hansen and Sato (2004); Gc = MacFarling Meure et al. (2006); Gd = CO₂ diagnosed by the model. Aerosols: Aa = Lamarque et al. (2010); Ab = Johns et al. (2003). Land use–land cover: La = Pongratz et al. (2009) spliced to Hurtt et al. (2006); Lb = Kaplan et al. (2011).

Model name in the CMIP5 database	Resolution of atmosphere: lon \times lat ($^\circ$), vertical levels	Resolution of ocean: lon \times lat ($^\circ$), vertical levels	Forcings	References
BCC_CSM1.1	$2.8 \times 2.8, 26$	$1 \times 1, 40$	Sb, Va, Gabc, Aa	Xin et al. (2013)
CCSM4	$1.25 \times 0.9, 26$	$1 \times 1, 60$	Sb, Va, Gabc, Aa, La	Gent et al. (2011); Landrum et al. (2013)
CSIRO Mk3L1.2	$5.6 \times 3.2, 18$	$2.8 \times 1.6, 21$	Sa, Vb, Gabc	Rotstayn et al. (2012); Phipps et al. (2011, 2012)
EC-Earth	$1.125 \times 1.125, 62$	$1 \times 1, 46$	Sb, Vc, Gabc, Aa, La	Hazeleger et al. (2010)
FGOALS-s2	$2.81 \times 1.66, 26$	$1 \times 1, 30$	Sa, Vb, Gabc	Bao et al. (2013)
GISS-R21	$2.5 \times 2, 40$	$1 \times 1.25, 32$	Sb, Vb, Gabc, Aa, La	Schmidt et al. (2006)
GISS-R24	$2.5 \times 2, 40$	$1 \times 1.25, 32$	Sb, Vb, Gabc, Aa, La	Schmidt et al. (2006)
GISS-R27	$2.5 \times 2, 40$	$1 \times 1.25, 32$	Sb, Vb, Gabc, Aa, Lb	Schmidt et al. (2006)
HadCM3	$3.75 \times 2.5, 19$	$1.25 \times 1.25, 20$	Sa, Vb, Gacd, Ab, La	Pope et al. (2000); Collins et al. (2001); Schurer et al. (2013)
IPSL-CM5A-LR	$3.75 \times 1.875, 39$	$2 \times 2, 31$	Sb, Vc, Gabc	Dufresne et al. (2013)
MPI-ESM-P	$1.875 \times 1.875, 47$	$1.5 \times 1.5, 40$	Sb, Vb, Gabc, Aa, La	Giorgetta et al. (2013)

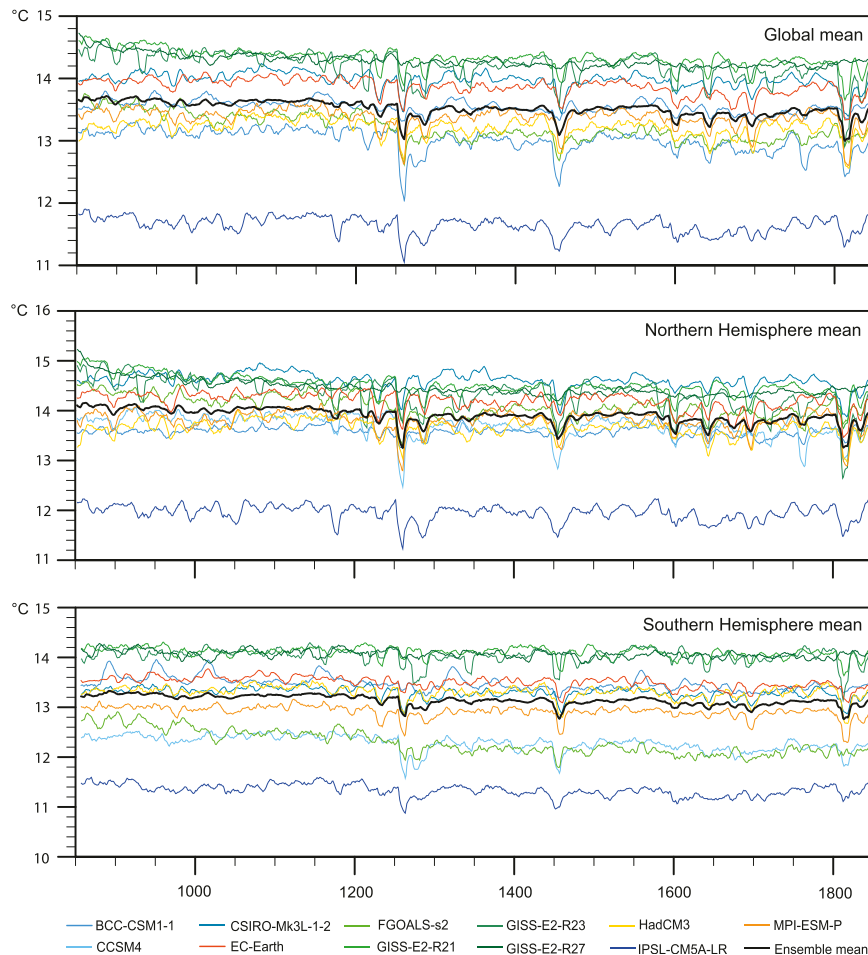


FIG. 2. Surface air temperature evolution ($^{\circ}\text{C}$) over 850–1850 CE in the PMIP3 last millennium simulations for the (top) global mean, (middle) NH, and (bottom) SH. Note that the absolute temperature differs considerably between the individual model simulations and that the large negative temperature “spikes” are forced by large volcanic eruptions.

ice, likely due to an insufficiency of the initialization according to Sueyoshi et al. (2013). FGOALS-gl is excluded because it first starts at 1000 CE. A simple correction is performed for the known drift of the GISS simulations (Schmidt 2012): by subtracting a locally weighted scatterplot smooth (LOWESS) fit (influence of about 600 years) to the GISS preindustrial control run from the annually resolved data of interest (i.e., the gridpoint data for the field evaluation and the relevant time series for the area-averaged assessment).

MODEL DATA PROCESSING

The monthly mean temperature data for 950–1250 CE and 1450–1850 CE were extracted from all the simulations. From the two periods we calculated the annual, the JJA, and DJF means. The difference between the

two periods was then calculated for each seasonal target to represent the amplitude of MCA–LIA temperature change in each grid cell.

Since the models have different resolutions, their data were bilinearly regridded to a common grid size of $1^{\circ} \times 1^{\circ}$ before calculating the multimodel ensemble mean and medians. The SDs of the annual and seasonal values within the periods representing the MCA and the LIA at each new grid cell were calculated and used for a two-tailed Student's t test of mean climate change. Similarly, the Wilcoxon's signed rank test (Wilcoxon 1945) is used for the median.

d. Statistical evaluation of the simulation–reconstruction agreement

The simulation–reconstruction comparison is performed for those areas (e.g., grid cells) for which local or

continental-scale reconstructions are available. A weighted distance between simulations and reconstructions is calculated by means of a cost function (CF). The method was first applied by [Goosse et al. \(2006\)](#) in a data assimilation study to select ensemble members from a large number of simulations that best match a target reconstruction. Another use has been to objectively group simulation data from different models ([Zhang et al. 2010](#)). The CF can also be applied to evaluate the degree of agreement between model results and the corresponding reconstructions. A good-fit measured by the cost-function largely depends on the number of predictors, their uncertainties, and their spatial distribution.

The CF measures the weighted distance between reconstructed and simulated temperature change, and is defined as follows:

$$CF_k = \sqrt{\sum_{i=1}^n w_i (T_i^{\text{model},k} - T_i^{\text{rec}})^2}, \quad (1)$$

where T_i^{rec} is the difference between the reconstructed MCA and LIA temperatures at location i ; $T_i^{\text{model},k}$ is the difference between the MCA and LIA temperatures at location i , simulated by model k ; and w_i is weights, standardized to sum to 1.

In the case of local reconstructions, the CF values will be calculated for each season (annual mean, JJA, or DJF) separately.

For the definition of the weights, see [appendix A](#). Suffice to say here is that reconstructions with large uncertainties receive smaller weights. We also consider an unweighted CF, in which all the weights are equal, that is, $w_i = 1/n$:

$$CF_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_i^{\text{model},k} - T_i^{\text{rec}})^2}. \quad (2)$$

THE COST FUNCTION CALCULATION AS A BLOCK DESIGN

In principle, the CF values could be used to rank the climate models, by means of the statistical distance between simulated temperature reconstructions, in terms of point-wise performance: the lower the CF value, the better the climate model's fit to the reconstructions. However, such numerical assessments are not necessarily reliable: even if all the models were in fact equally good, it is unlikely that any two of the CF values would turn out exactly the same. Hence, one model would appear to be the “best” and another the “worst,” when in such a case all the observed differences between models are due to random variability rather than to any

consistent difference in model performance. This fact points to the need for a significance test to tell when an observed difference between models is meaningful. To this end, consider

$$Y_{ki} = (T_i^{\text{model},k} - T_i^{\text{rec}})^2 \quad (3)$$

measuring the performance of model k at location i . Introducing the notation \bar{Y}_k for the average of all the Y_{ki} for model k , we then have the following relation to the unweighted cost function:

$$CF_k = \sqrt{\bar{Y}_k}.$$

The Y_{ki} may be analyzed as a block design (e.g., [Box et al. 2005](#)), with the locations of the reconstructions as blocks. This means that we only compare Y_{ki} values for different k (models) at the same i (reconstruction location); clearly the Y_{ki} for *different* locations reflect the diverse nature of the locations at least as much as differences in model performance and so should not be allowed to influence the analysis. Since the normality of Y_{ki} is questionable (indeed, they are nonnegative), we perform a nonparametric test, the Friedman test ([Friedman 1937, 1939](#)), complemented by a Friedman–Nemenyi test ([Nemenyi 1963](#)), to decide whether any significant differences exist between models. Both tests are based on ranks but not on any assumption of normality. For each site i the models are ranked so that the model with the smallest Y_{ki} is assigned rank 1, the model with the second smallest Y_{ki} is assigned rank 2, etc. Then the ranks are summed over locations for each model, giving an overall assessment R_k of the performance of model k . The R_k may be used to test the null hypothesis H_0 : the assignment of ranks at the various sites are independent random permutations of the numbers 1, 2, ..., n . This would be the case if the difference between model runs was due to internal variability and random error alone, assuming the time series belonging to the different locations i are independent. In that case all the R_k will be approximately equal. The Friedman test tells us how unequal they need to be in order for H_0 to be rejected. The Friedman–Nemenyi test restricts this comparison to pairs of models, so that we can tell exactly which differences are significant. Yet, in reality, we cannot be certain that neighboring locations are independent, meaning that the effective number of independent sites is smaller than the observed number, so that we have less information. Hence, the p values we present in this context may be too small and should be taken only as rough indicators of which differences between models are meaningful.

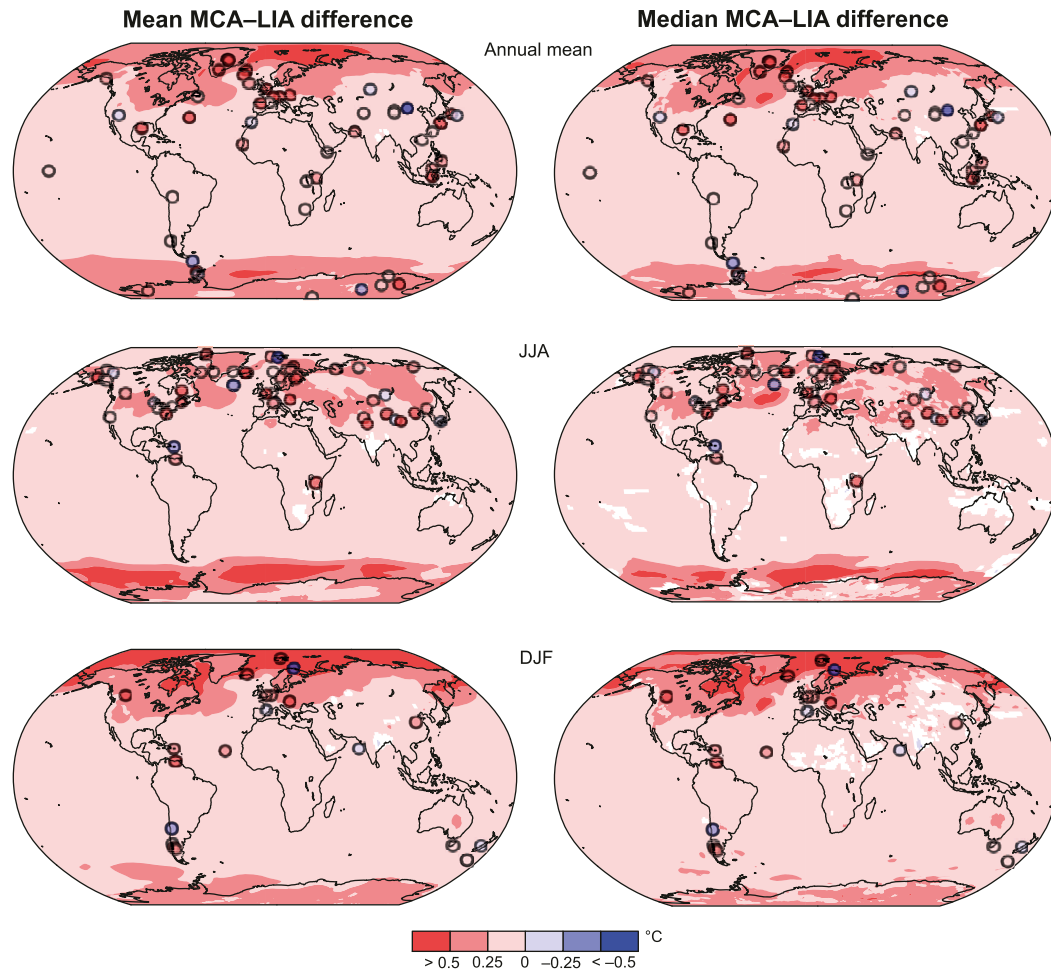


FIG. 3. (left) Ensemble mean and (right) ensemble median near-surface air temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE for (top) annual-mean, (middle) JJA, and (bottom) DJF temperatures, with local reconstructions for the respective season superimposed. White areas show where insignificant ($p > 0.05$) temperature difference between the two periods in the ensemble mean and medians occur. Note the polar amplification in both the ensemble mean and median for all seasons, which is stronger in winter than in summer.

3. Amplitude of temperature change in simulations and local and continental-scale reconstructions

a. Intermodel comparison

The selected simulations from PMIP3 plus EC-Earth show considerable differences in both the amplitude of the MCA–LIA temperature change and its spatial patterns (Figs. 3, 4). Some models show a much larger amplitude of change than others even though similar forcing datasets are used. This is largely explained by differences in the sensitivity to forcings due to differing feedbacks and in ocean heat uptake efficiency (Atwood et al. 2016; Fernández-Donado 2016), with the structural intermodel differences larger than the uncertainties in external forcings (Lehner et al. 2015), and due to the implementation of the forcings (Zanchettin et al. 2016).

When considering either the ensemble mean or median, larger areas show a significant change ($p < 0.05$, two-tailed Student's t test or Wilcoxon rank test) in annual-mean temperature than for JJA or DJF temperatures. A partly similar assessment of MCA–LIA temperature change in the PMIP3 simulations is found in Atwood et al. (2016) though they used 950–1200 CE for the MCA and 1600–1850 CE for the LIA, but the choice of periods has a marginal effect. Our results, and those of Atwood et al. (2016), are consistent with similar analyses of pre-PMIP3 models showing a similar large spread (Fernández-Donado et al. 2013).

Considering the individual simulations (Fig. 4), the three GISS simulations (r11p121, r11p124, and r11p127), EC-Earth, CCSM4, and FGOALS show rather large ($>0.25^{\circ}\text{C}$) and significant temperature changes

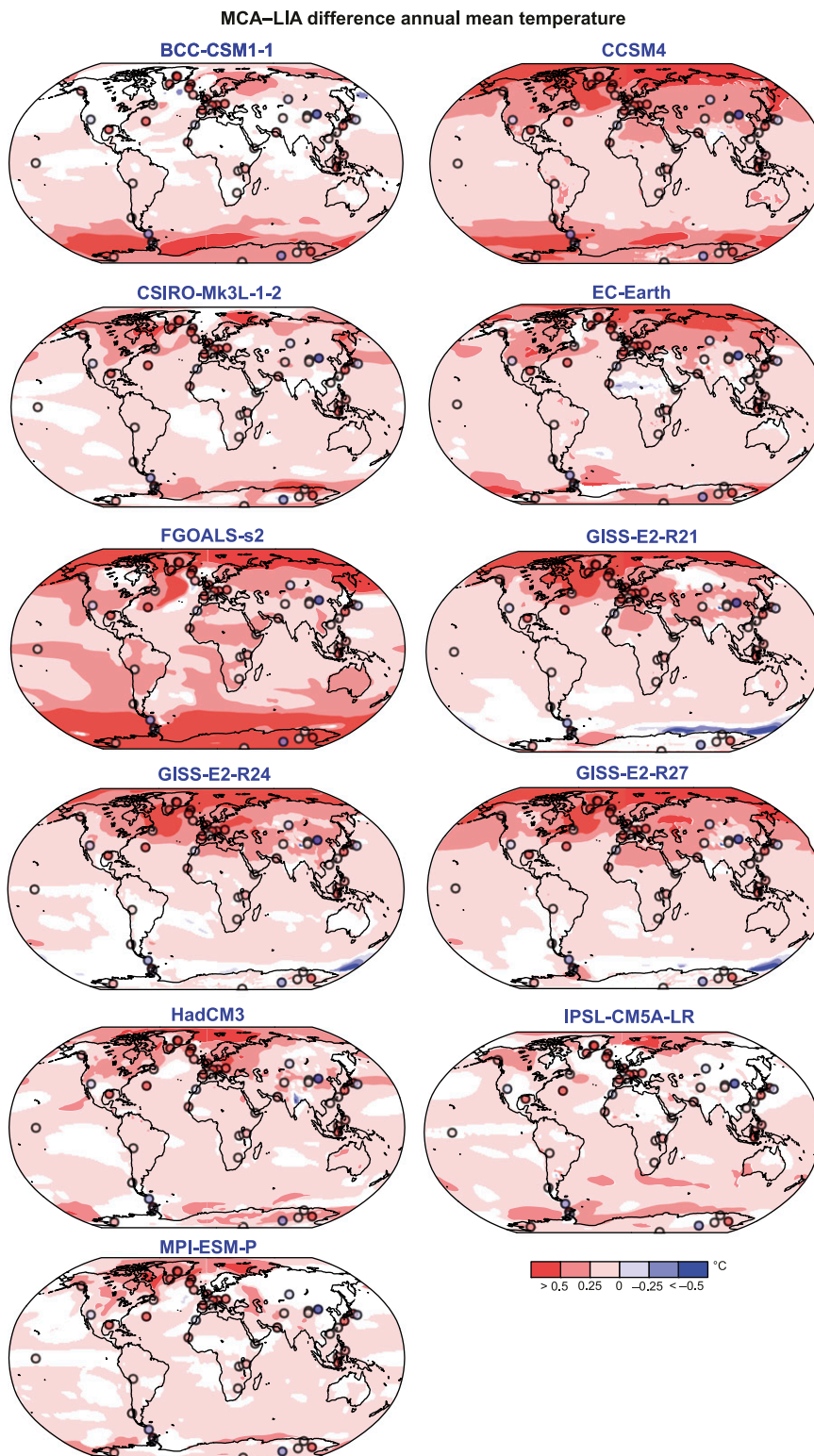


FIG. 4. Near-surface air temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE from the individual climate models for annual-mean, JJA, and DJF temperatures, with local reconstructions for the respective season superimposed. White areas show insignificant ($p > 0.05$) temperature difference between the two periods in the model in question. Note the large differences between the various models.

MCA-LIA difference June–August temperature

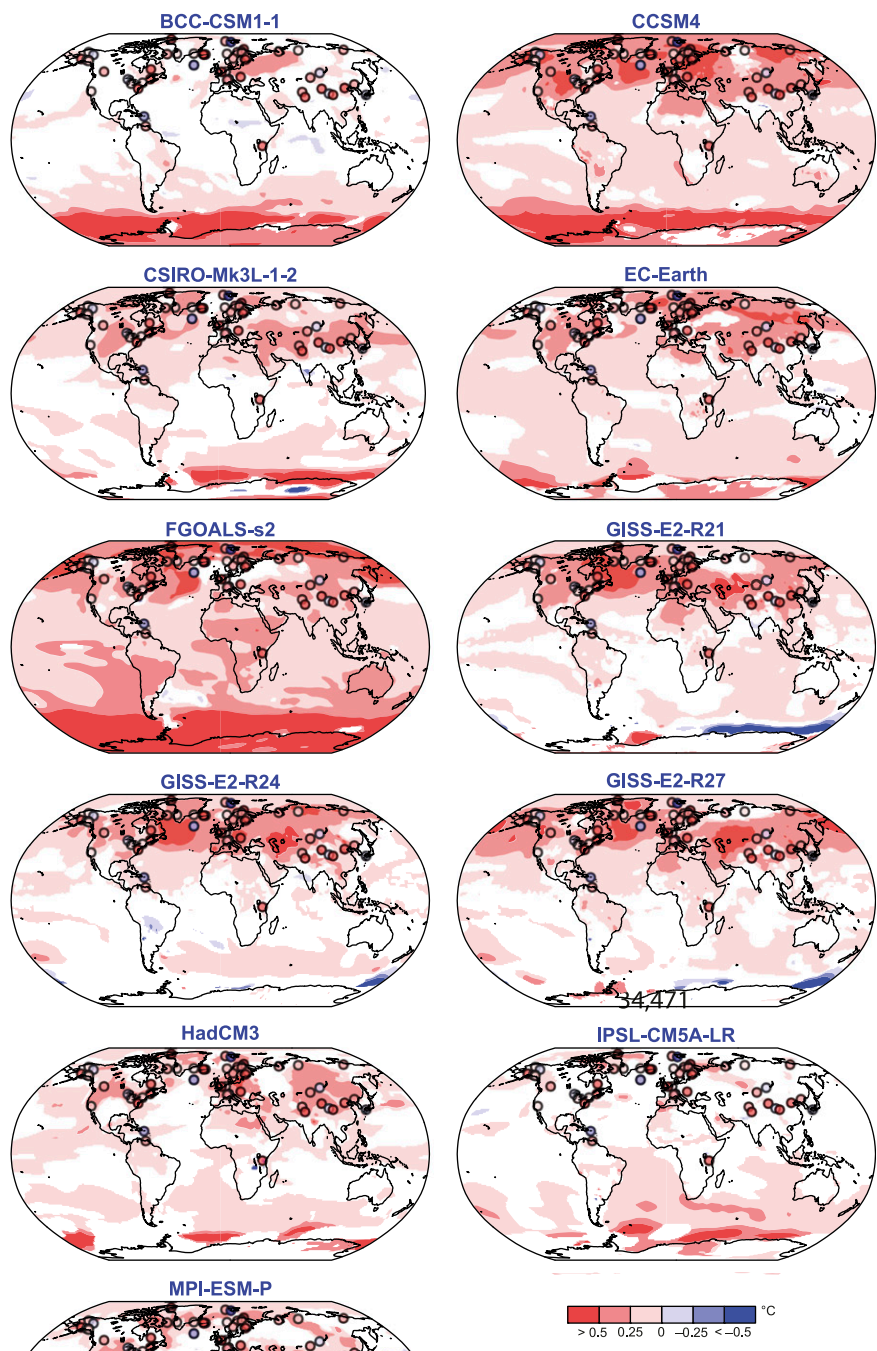


FIG. 4. (Continued)

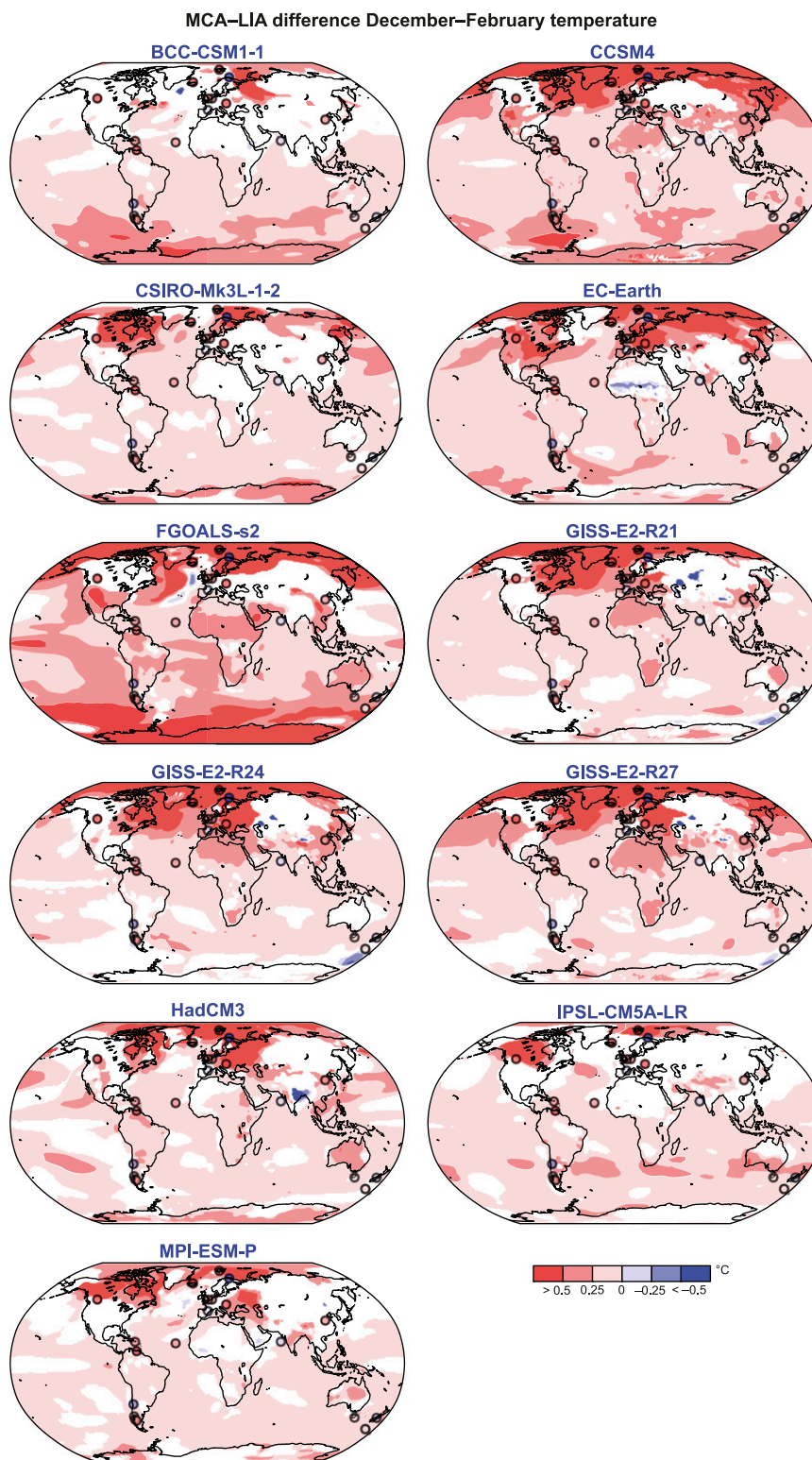


FIG. 4. (Continued)

TABLE 5. Global, NH, and SH temperature differences ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE in local reconstructions and in climate model simulations at locations of local reconstructions by season. Note that reconstruction means and medians are only calculated if at least three local reconstructions for a particular season were available.

CMIP5/PMIP3 model	Global			NH			SH		
	Annual	DJF	JJA	Annual	DJF	JJA	Annual	DJF	JJA
BCC	0.14	0.17	0.09	0.12	0.16	0.09	0.17	0.17	—
CCSM4	0.26	0.37	0.42	0.26	0.42	0.43	0.27	0.27	—
CSIRO	0.14	0.10	0.12	0.14	0.06	0.12	0.14	0.15	—
EC-Earth	0.23	0.31	0.22	0.25	0.39	0.22	0.17	0.17	—
FGOALS	0.34	0.23	0.29	0.27	0.17	0.29	0.49	0.34	—
GISS21	0.17	0.26	0.29	0.19	0.44	0.29	0.12	−0.03	—
GISS24	0.25	0.22	0.31	0.32	0.40	0.32	0.07	−0.08	—
GISS27	0.25	0.30	0.32	0.28	0.49	0.32	0.18	−0.01	—
HadCM3	0.20	0.47	0.27	0.21	0.71	0.33	0.16	0.07	—
IPSL	0.08	0.18	0.05	0.04	0.13	0.04	0.16	0.27	—
MPI	0.10	0.23	0.15	0.08	0.25	0.15	0.14	0.20	—
Mean of model medians	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	—
Median of model medians	0.20	0.23	0.27	0.21	0.39	0.29	0.16	0.17	—
Proxy mean	0.38	0.38	0.41	0.42	0.58	0.41	0.28	0.06	—
Proxy median	0.25	0.30	0.33	0.36	0.46	0.33	0.18	0.08	—
Number of data points	51	21	56	36	13	55	16	8	1

in most areas. BCC, CSIRO, HadCM3, IPSL, and MPI show smaller changes ($<0.25^{\circ}\text{C}$) that are insignificant over large regions, in particular in the tropics. GISS (especially r11p121) is the only model showing a cooling in or near Antarctica, and BCC is the only model showing larger changes in the SH than in the NH. FGOALS, and to a lesser extent CCSM4, stands out by its large temperature change, and is the only model showing changes of $>0.25^{\circ}\text{C}$ in low-latitude regions. EC-Earth also shows a rather large change over the Arctic and the Eurasian continent, most strongly expressed during DJF.

b. Model comparison using local reconstructions

We observe the same general latitudinal pattern in the relative amplitude of MCA–LIA temperature change in the local reconstructions as in the simulations (Table 5; Fig. 5): largest in the NH high latitudes (60° – 90°N) and smallest in the low latitudes (30°N – 30°S). However, the local reconstructions in general show a larger amplitude of MCA–LIA temperature change than the simulations, particularly in the NH (Table 5; Fig. 5). The MCA–LIA difference in the local reconstructions is rather similar across seasons with annual-mean, JJA- and DJF-mean (median) values of 0.38°C (0.25°C), 0.41°C (0.38°C), and 0.38°C (0.30°C), respectively, compared to model means (medians) from corresponding grid cells of 0.19°C (0.18°C), 0.21°C (0.23°C), and 0.27°C (0.17°C), respectively.

Considering local reconstructions of annual-mean temperature, the MCA–LIA difference is consistently larger in the eastern rim of the Pacific, the North Atlantic

and Europe (Figs. 3, 4) compared to in the simulations. A larger annual temperature change in the northern North Atlantic is captured both in most simulations and in the local reconstructions (although the amplitude is larger in absolute terms in the latter). The reconstructed JJA temperature change for locations in Eurasia is consistent with the simulations though generally larger, whereas reconstructions for North America show a less consistent and, in general, smaller change than for Eurasia. Local reconstructions representing DJF temperature are too few, and show an inconsistent signal (likely due to proxy noise), to meaningfully assess the simulation–reconstruction agreement.

For certain seasons and regions, there are a sufficient number of local reconstructions to calculate meaningful continental-scale means and medians of MCA–LIA temperature difference and compare these to the mean and median model values from corresponding grid cells (Table 6). All such continental-scale averages should be interpreted cautiously—they are based on relatively few reconstructions that among themselves show considerable differences. Though the means and medians for the reconstructions, except for Antarctica and South America, exceed the amplitude of change in most models, they are not inconsistent with the two models showing the largest amplitude: CCSM4 and FGOALS.

Could the larger reconstructed Northern Hemisphere MCA–LIA temperature difference, compared to the simulated one, simply be a result of internal variability? It is impossible to assess, in detail, the signature of internal variability in the reconstructed MCA–LIA temperature difference, but for simulations the situation is

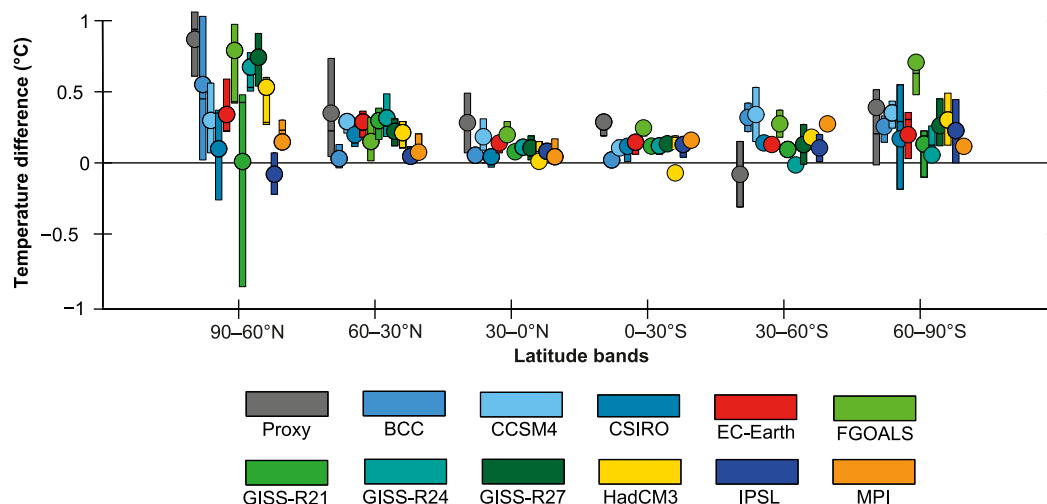


FIG. 5. Temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE for annual-mean temperature in local reconstructions (proxies) and model-simulated data for corresponding grid cells after latitude bands are divided into six 30° segments. The filled circles represent the mean, and the length of the bars represents the quartile range.

more tractable. Ideally, we would use replicates of simulations with the same forcings and the same model to assess internal variability. Failing that, unforced control simulations may be informative, if we are willing to believe that the internal variability of a forced simulation can be reasonably approximated by the variability of an unforced control simulation. The question then becomes whether the observed difference between reconstructed MCA–LIA differences and simulated ones falls within the range of the MCA–LIA differences in a control simulation. We have used the 3000-yr unforced MPI control simulation (Jungclauss et al. 2010; Zhu et al. 2016) to investigate whether this can happen. Within the 3000 years, we can fit 2101 periods of 900 consecutive years. For each one of these, we calculated the temperature difference between the means of the first 300 years (representing MCA) and the

last 400 years (representing LIA) and separated by 200 years.

The large majority of all differences between 300- and 400-yr temperature averages from the control simulation vary by less than 0.1°C for global annual-mean temperature (Fig. 6a), thus explaining only a fraction of the proxy reconstruction–simulation difference at the global-average annual-mean scale. Over the PAGES2k Consortium (2013) regions of Europe and North America, the outer tail of the amplitude of the simulated internal variability (Figs. 6b–g) reaches the size of the difference between the reconstruction and model mean for the same two regions (Tables 6, 7). The Arctic also shows a large simulated internal variability, but its magnitude is less than half the difference between the reconstruction and model mean. For the regions where the forced model simulations do not underestimate the

TABLE 6. Continental-scale means and medians of the MCA–LIA temperature difference ($^{\circ}\text{C}$) in local reconstructions and model simulations from corresponding grid cells for regions with at least four local reconstructions representing the same season.

Geographical area	Season	No. of local reconstructions	Local reconstruction values		Model values	
			Mean	Median	Mean	Median
Arctic ($>60^{\circ}\text{N}$)	Annual mean	6	0.87	0.85	0.35	0.31
Arctic ($>60^{\circ}\text{N}$)	JJA	26	0.34	0.30	0.23	0.25
Antarctica	Annual mean	9	0.31	0.16	0.28	0.29
North America ^a	JJA	18	0.31	0.26	0.19	0.23
Europe	JJA	16	0.45	0.41	0.25	0.25
Asia	JJA	13	0.69	0.42	0.20	0.23
South America	DJA	6	0.18	0.23	0.18	0.17

^a Greenland not included here in North America.

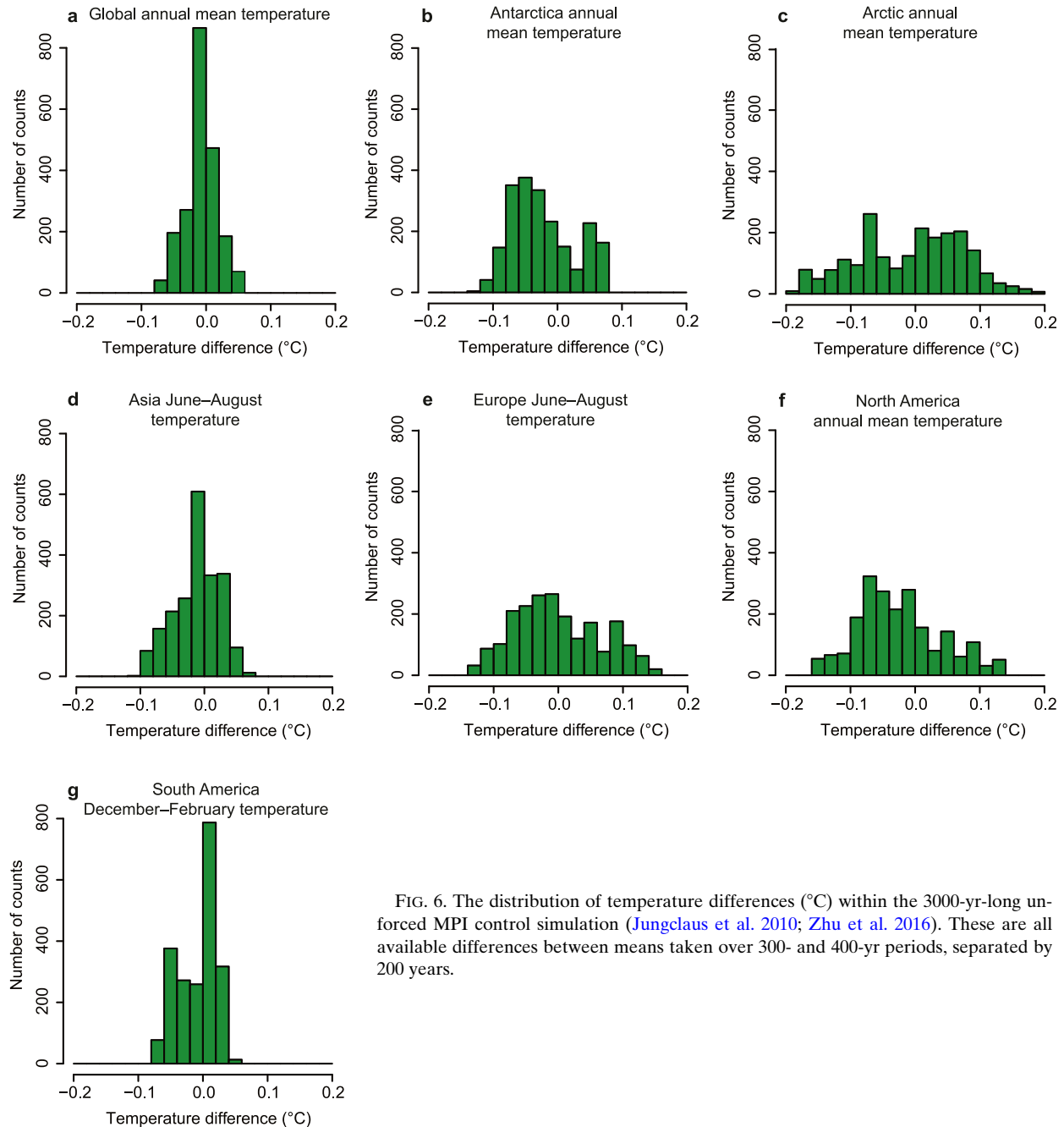


FIG. 6. The distribution of temperature differences ($^{\circ}\text{C}$) within the 3000-yr-long unforced MPI control simulation (Jungclaus et al. 2010; Zhu et al. 2016). These are all available differences between means taken over 300- and 400-yr periods, separated by 200 years.

reconstructed difference—Antarctica, South America and, to a lesser extent, Asia—the simulated internal variability is also relatively small. This implies that internal variability plays a role in those regions where the forced simulations do underestimate the reconstructed MCA–LIA temperature difference. Thus, we cannot exclude the possibility that the different temporal realizations of internal variability between the real world (as evidenced by the reconstructions) and in the model

simulations could account for a substantial fraction of the simulation–reconstruction mismatches.

c. Polar amplification in simulations and the local reconstructions

Polar amplification (or Arctic amplification in the NH) is a commonly used term to denote a larger temperature change in the polar regions compared to other parts of the world. It is often quantified as the ratio

TABLE 7. The temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE in the continental-scale reconstructions and corresponding simulated temperature change over the same PAGES2k geographical region and season. See Table 2 for the latitude and longitude coordinates for each region.

	Antarctica (Annual)	Arctic (Annual)	Asia (JJA)	Europe (JJA)	North America (Annual)	South America (DJF)
BCC	0.53	0.15	0.06	0.18	−0.04	0.21
CCSM4	0.46	0.65	0.25	0.42	0.10	0.27
CSIRO	0.27	0.30	0.21	0.15	0.08	0.11
EC-Earth	0.27	0.47	0.21	0.26	−0.01	0.20
FGOALS	0.89	0.68	0.22	0.17	0.22	0.25
GISS21	0.01	0.47	0.26	0.34	0.03	0.11
GISS24	0.03	0.48	0.20	0.33	0.14	0.11
GISS27	0.05	0.50	0.30	0.26	0.14	0.13
HadCM3	0.19	0.32	0.20	0.27	0.28	0.12
IPSL	0.21	0.21	0.03	0.07	0.15	0.19
MPI	0.18	0.23	0.09	0.13	0.06	0.15
Mean of model mean	0.34	0.39	0.17	0.22	0.11	0.18
Median of model medians	0.31	0.38	0.17	0.22	0.09	0.18
SD models	0.26	0.18	0.09	0.10	0.10	0.06
Reconstructions	0.32	0.77	0.18	0.40	0.17	0.09
Reconstruction–model mean difference	−0.02	0.38	0.01	0.18	0.06	−0.09

(amplification factor) between the temperature change in polar to nonpolar regions (Hind et al. 2016). If we only consider sites with local reconstructions reflecting annual-mean temperature, and by using their median values, the amplification factor associated with the temperature changes between the latitude bands of 60° – 90°N and 0° – 30°N is 3.52. Replacing the medians for local reconstructions by grid cells corresponding from the pool of all models in the calculation, we obtain the amplification factor 2.36. For the SH considering the analogous latitudinal bands (60° – 90°S and 0° – 30°S), the amplification factors using local reconstruction and model medians are 1.11 and 2.19, respectively. Thus, both the local reconstructions and model median agree there is a somewhat smaller polar amplification in the SH than in the NH. The lack of JJA reconstructions outside the NH extratropics, and the limited number of DJF reconstructions, unfortunately precludes a calculation of polar/Arctic amplification for these seasons.

The amplification factor is defined as a ratio, so one should not expect sensible results when its denominator is close to zero, that is, for simulations or reconstructions where no significant MCA–LIA difference has been observed at mid- and low latitudes (cf. Hind et al. 2016). In such cases the uncertainty in the amplification factor may become so large that the factor becomes entirely noninformative: attempts to construct a confidence interval for it yield an “interval” consisting of all real numbers (appendix C; also Fieller 1954; von Luxburg and Franz 2009). Arguably, the notion of an amplification factor is meaningless in this case. For many local

reconstructions and corresponding simulated temperatures, the confidence intervals obtained are noninformative in this sense, in particular for seasonal temperatures. However, for sites with local reconstructions representing annual-mean temperature, the reconstructed amplification factor is significant ($p < 0.05$) for 60° – 90°N versus 0° – 30°N , for 60° – 90°N versus 90°S – 60°N (i.e., Arctic versus non-Arctic), and 60° – 90°N versus 60°S – 60°N (i.e., Arctic versus nonpolar), that is, the interval lies entirely above 1. For the corresponding sites in the climate models, the simulated amplification is significant ($p < 0.05$) for GISS-R24 and GISS-R27 in the NH, and also for the model mean and median. For the SH significant ($p < 0.05$) amplification is seen for CCSM4 and FGOALS, and again for the multimodel mean and median. We note that models with low variability show little or no temperature change in tropical regions, which could explain the noninformativeness of the intervals obtained in these cases.

d. Model comparison using continental-scale reconstructions

Overall, the MCA–LIA temperature amplitude agreement, between proxy reconstructions and model simulations, is slightly better for the continental-scale reconstructions than for the local reconstructions (Tables 7–9). This is partly because of the somewhat smaller amplitude of MCA–LIA temperature difference of the continental-scale reconstructions compared to the local reconstructions. However, a better agreement is mainly evident when considering the mean of the local reconstructions but not when considering the

TABLE 8. The simulated temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE for continental-scale reconstruction locations and the locations of local reconstructions.

	BCC	CCSM4	CSIRO	EC-Earth	FGOALS	GISS21	GISS24	GISS27	HadCM3	IPSL	MPI
Continental scale	0.18	0.36	0.19	0.23	0.41	0.21	0.21	0.23	0.23	0.14	0.14
Local	0.12	0.35	0.13	0.24	0.30	0.24	0.27	0.29	0.27	0.08	0.14
Difference	0.06	0.01	0.06	−0.01	0.11	−0.03	−0.06	−0.06	−0.04	0.06	0.00

median (Table 9). For the Arctic reconstruction, the amplitude of annual-mean temperature change is still twice that of the model mean and median (0.77°C versus 0.39° and 0.38°C). However, CCSM4 and FGOALS show an amplitude almost comparable to the Arctic reconstruction (0.65° and 0.68°C , respectively). The reconstructions for Antarctica and Asia display a similar degree of MCA–LIA temperature difference as the model mean or median, although the various models differ substantially (SD of 0.26°C) for Antarctica (Table 7). The models in general show too-little temperature change in Europe (reconstruction = 0.40°C versus model mean/median = 0.22°C), but CCSM4, GISS21, and GISS24 show simulated values comparable to the reconstructed ones. The models also, but to a lesser extent, underestimate the reconstructed temperature change in North America. In contrast to results for all other regions, all models used for the simulations actually overestimate the amplitude of MCA–LIA temperature change for South America (reconstruction = 0.09°C versus model mean/median = 0.18°C).

4. Cost function analysis of the simulation–reconstruction agreement

For all seasons, the weighted CF values for local simulation–reconstruction comparisons are smaller than the corresponding unweighted values (Fig. 7, Tables 10, 11). This means that locations with local reconstructions having a higher uncertainty (which will be down-weighted in the weighted version) also exhibit a larger difference between model data and reconstructed temperatures than locations of local reconstructions with a lower uncertainty. Our results thus support the recommendation by Schmidt et al. (2014a) that it is important

to have well-characterized estimates of proxy uncertainty in a simulation–reconstruction comparison. The simulated temperatures are in better agreement with local reconstructions if uncertainties in the latter are taken into account, after adjustment for unrealistically small error estimates, see appendix A. However, assigning weights to the different continental-scale reconstructions has only a marginal effect, only slightly affecting the ranking among models, whereas the CF values for the local reconstructions decrease substantially when weights are assigned (Fig. 8, Table 12). This effect could partly be attributed to the weights being far more variable for local than for continental-scale reconstructions. In other words, the weighted CF obtained when using local reconstructions are less similar to the weights in the unweighted CF, which may be viewed as a weighted CF with all its weights equal to $1/n$. The mean and median values for the pool of all simulations produce lower cost function values than the individual simulations. However, mean and median values are not fully comparable to single simulations as much of the internal variability of the individual simulations is averaged out, resulting in a mainly forced temperature component.

The CF value is a measure of how closely a given model reproduces the reconstructed temperatures, and an observed difference between the CF values of two models will suggest a difference in model performance. In section 2d we described a significance test for such observed differences, indicating which of them are meaningful. The most striking differences between models are found when considering all seasonal targets together (p value = 4×10^{-6}). Restricting attention to seasonal targets, we obtain significant ($p < 0.05$) differences for annual-mean temperature (p value = 0.0128) and for JJA (p value = 0.0012). For DJF temperature no significance was found (p value = 0.626). This means that at least some models differ significantly in their agreement with the local reconstructions. The comparison with continental-scale reconstructions yields no significant differences between models (p value = 0.171), which is perhaps not surprising since this test is based on only six observations (i.e., continents) for each model.

TABLE 9. The temperature difference ($^{\circ}\text{C}$) between 950–1250 and 1450–1850 CE for local and continental-scale reconstructions and climate model simulations. The table summarizes data from all localities and for all climate models used in the study.

	Continental scale	Local	Continental scale	Local
	Mean		Median	
Reconstruction	0.32	0.40	0.25	0.30
Model	0.23	0.21	0.20	0.21

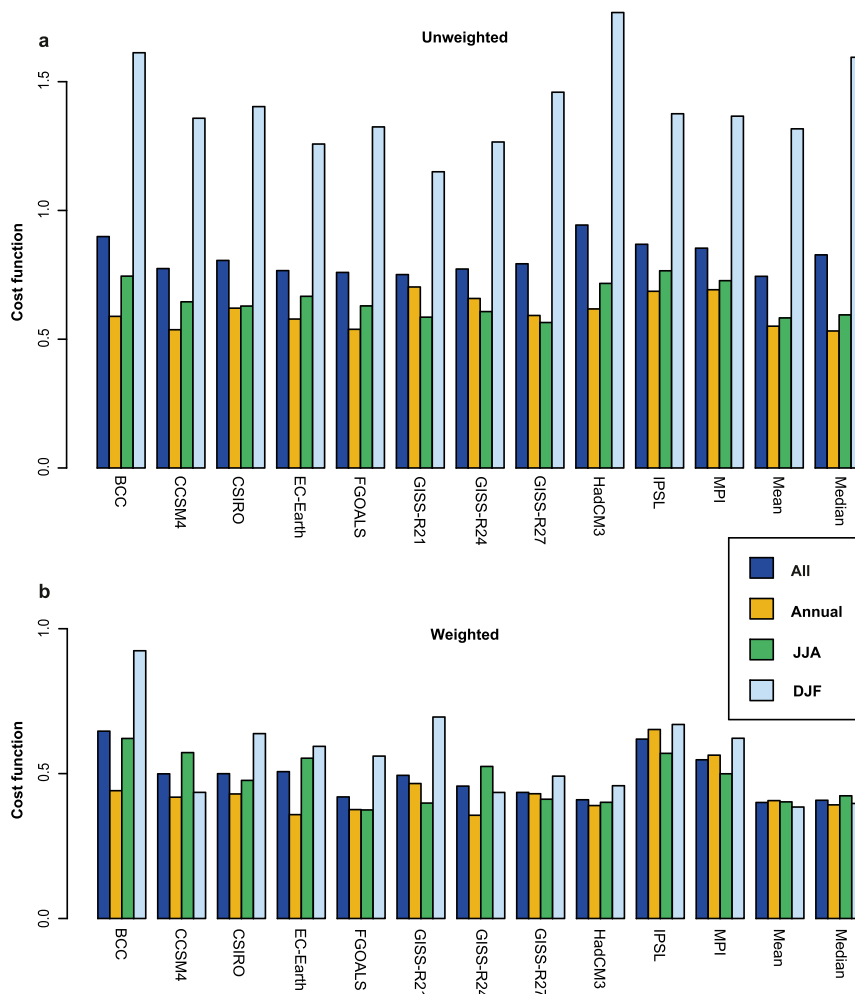


FIG. 7. Cost function values of local reconstructions and model simulations for (a) unweighted and (b) weighted data for summer (green), winter (light blue), and annual-mean temperatures (orange) and for all seasons together (dark blue). The cost function values decrease noticeably when the weights are assigned.

Once a significant difference has been detected, we test *which* models differ significantly using the Friedman–Nemenyi test. Studying the local comparisons, we find significant differences when considering all seasonal targets together, with MPI, IPSL, and BCC differing significantly (by showing a worse agreement) from the mean of all simulations. The results are displayed graphically in Fig. 9. No significant ($p < 0.05$) differences are seen in the continental-scale comparisons. Out of the 11 simulations, HadCM3 and FGOALS show the lowest CF values, that is, best fit to the weighted local reconstructions when considering all seasonal targets together. The lowest agreement is found with MPI, IPSL, and BCC. It is worth noting that the ranking of simulations using local reconstructions versus continental-scale reconstructions results in limited agreement, although

MPI, IPSL, and BCC are among the simulations showing the lowest agreement with both local and continental-scale reconstructions.

5. Discussion and conclusions

After comparing the MCA–LIA temperature difference in the PMIP3 simulations with local and continental-scale temperature reconstructions, we conclude the following:

- 1) Both local and the continental-scale reconstructions on average show a larger amplitude of temperature change than the simulations in the NH, particularly in the high latitudes. Results for the ocean-dominated SH are less conclusive, and even the opposite can be seen for South America (DJF temperatures), that is,

TABLE 10. Unweighted cost functions based on the local reconstructions.

Model	All	Annual	Summer	Winter	Rank: all	Rank: annual	Rank: summer	Rank: winter
BCC	0.90	0.60	0.70	1.60	12	6	12	12
CCSM4	0.80	0.50	0.60	1.40	6	2	8	6
CSIRO	0.80	0.60	0.60	1.40	8	9	6	9
EC-Earth	0.80	0.60	0.70	1.30	4	5	9	2
FGOALS	0.80	0.50	0.60	1.30	3	3	7	5
GISS21	0.80	0.70	0.60	1.20	2	13	3	1
GISS24	0.80	0.70	0.60	1.30	5	10	5	3
GISS27	0.80	0.60	0.60	1.50	7	7	1	10
HadCM3	0.90	0.60	0.70	1.80	13	8	10	13
IPSL	0.90	0.70	0.80	1.40	11	11	13	8
MPI	0.90	0.70	0.70	1.40	10	12	11	7
Mean	0.70	0.60	0.60	1.30	1	4	2	4
Median	0.80	0.50	0.60	1.60	9	1	4	11
Average CF	0.82	0.62	0.66	1.39	—	—	—	—

the models show a larger amplitude of temperature change than the reconstructions there. The best agreement between local reconstructions and model simulations are obtained when comparing the model median (i.e., at the grid cells closest to the proxy data locations) with the medians of the local reconstructions (at a seasonal basis). The medians of the local reconstructions also produce the most similar results to when comparing the model median to the continental-scale reconstructions.

- 2) When comparing simulations with local reconstructions by means of the cost function analysis we detect differences between groups of models, but sites in close proximity are not necessarily independent and hence p values obtained may be too small. For continental-scale reconstructions no significant differences are seen.
- 3) Weighting the local reconstructions by their estimated uncertainty results in better agreement with the simulations in the cost function analysis, whereas
- weighting the continental-scale reconstructions by their uncertainty has no obvious effect.
- 4) It is difficult to evaluate the degree of agreement between reconstructed and simulated Arctic/polar amplification as only some models show a significant amplification while others do not.

The generally larger amplitude of MCA–LIA temperature difference in the NH shown by proxy-based reconstructions compared to simulations is not restricted to particular proxy types, regions or seasons. Nor is it an artifact of geographical bias as the comparison with simulations is conducted only for grid cells corresponding to site locations of the reconstructions (see Table 8). It is unlikely that the proxy-based reconstructions overestimate the amplitude of the MCA–LIA temperature difference. First, most state-of-the-art calibration methods result in an underestimation, rather than an overestimation, of the reconstructed low-frequency temperature amplitude, especially from

TABLE 11. Weighted cost functions based on the local reconstructions (cutoffs: $SD = 0.1$, $d = 0.9$).

Model	All	Annual	Summer	Winter	Rank: all	Rank: annual	Rank: summer	Rank: winter
BCC	0.65	0.44	0.62	0.92	13	10	13	13
CCSM4	0.50	0.42	0.57	0.44	8	7	12	4
CSIRO	0.50	0.43	0.48	0.64	9	8	7	10
EC-Earth	0.51	0.36	0.55	0.59	10	2	10	8
FGOALS	0.42	0.38	0.37	0.56	4	3	1	7
GISS21	0.49	0.47	0.40	0.70	7	11	2	12
GISS24	0.46	0.36	0.52	0.44	6	1	9	3
GISS27	0.44	0.43	0.41	0.49	5	9	5	6
HadCM3	0.41	0.39	0.40	0.46	3	4	3	5
IPSL	0.62	0.65	0.57	0.67	12	13	11	11
MPI	0.55	0.56	0.50	0.62	11	12	8	9
Mean	0.40	0.41	0.40	0.38	1	6	4	1
Median	0.41	0.39	0.42	0.40	2	5	6	2
Average CF	0.50	0.44	0.49	0.59	—	—	—	—

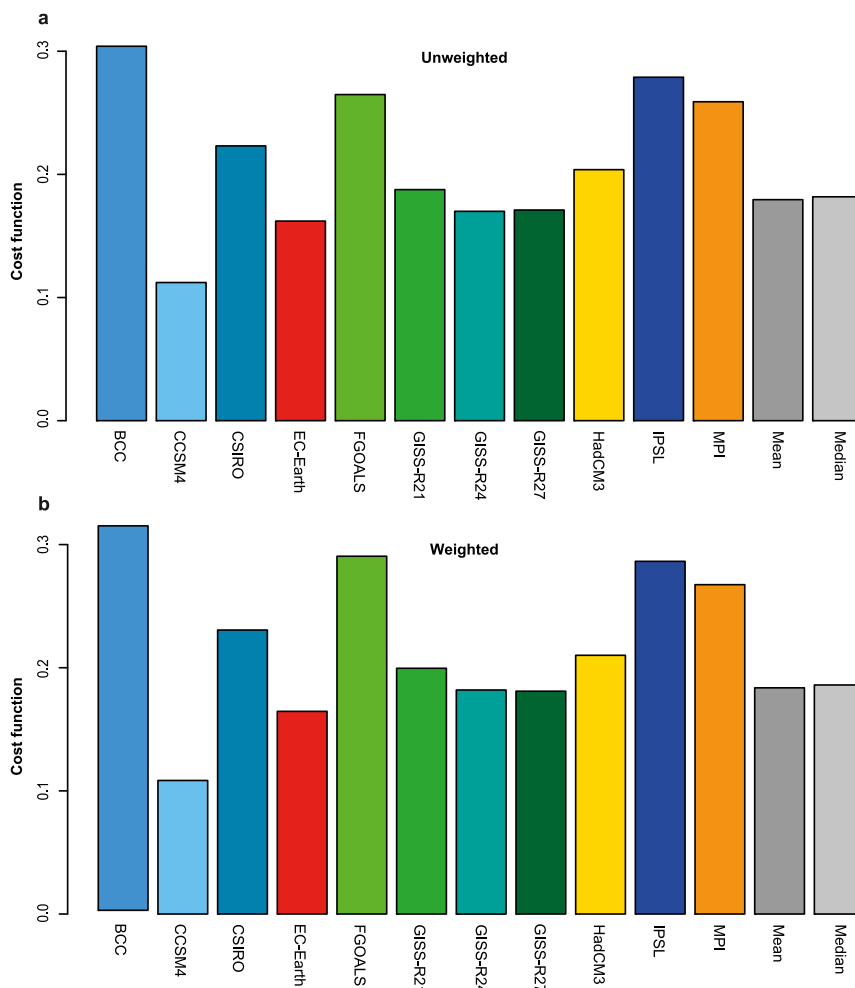


FIG. 8. Cost function values for continental-scale reconstructions and model simulations for (a) unweighted and (b) weighted data.

proxies with a low signal-to-noise ratio (e.g., von Storch et al. 2004; Bürger et al. 2006; Christiansen 2011; Moberg and Brattström 2011; Tingley et al. 2012). Multiproxy reconstructions tend to suffer more from this variance loss than single site reconstructions (Christiansen and Ljungqvist 2017), perhaps explaining the slightly lower amplitude of MCA–LIA temperature difference in the continental-scale reconstructions compared to in the local reconstructions (e.g., Luterbacher et al. 2016). Second, an even larger MCA–LIA amplitude of change is supported by independent estimates of low-frequency temperature change from terrestrial borehole profiles (Huang et al. 2008). However, this result is less conclusive than a large amplitude in temperature difference between the LIA and the present derived from borehole evidence (Pollack and Smerdon 2004). The temperature estimates from boreholes are based on physical modeling, instead of

statistical calibration against time series of instrumental temperature measurements.

a. Uncertainties in the model simulations and in their forcing

In this section, we point out some relevant known issues in model simulations used in this study. To begin, IPSL simulates a $\sim 2^{\circ}\text{C}$ colder global-mean temperature than other models (or the actual temperature). CCSM4 overestimates decadal near-global cooling after large volcanic eruptions (Landrum et al. 2013; Hartl-Meier et al. 2017) (Fig. 2). Such cold biases might have consequences for, in particular, sea ice changes and albedo feedbacks (Miller et al. 2012). In this context, any mean biases in the models that have influences on sea ice changes and albedo will likely influence the simulated equator-to-pole temperature gradient and Arctic/polar amplification. Hence, such biases may in principle affect

TABLE 12. Cost functions based on the continental-scale temperature reconstructions.

Model	Unweighted CF	Unweighted rank	Weighted CF	Weighted rank
BCC	0.30	13	0.31	13
CCSM4	0.11	1	0.11	1
CSIRO	0.22	9	0.23	9
EC-Earth	0.16	2	0.16	2
FGOALS	0.26	11	0.29	12
GISS21	0.19	7	0.20	7
GISS24	0.17	3	0.18	4
GISS27	0.17	4	0.18	3
HadCM3	0.20	8	0.21	8
IPSL	0.28	12	0.29	11
MPI	0.26	10	0.27	10
Mean	0.18	5	0.18	5
Median	0.18	6	0.19	6
Average CF	0.21	—	0.22	—

the degree of agreement between simulations and reconstructions. Unfortunately, as pointed out above, we found it difficult to evaluate the simulation–reconstruction agreement of Arctic/polar amplification. Some of the differences between the models may also be related to the differences in the implemented forcings. However, our research design does not allow us to address the influences of these differences.

The PMIP3 models have previously been found to show little agreement in their response to solar forcing as a result of the different parameterizations used (Le et al. 2016). A choice with a smaller amplitude of solar forcing is in line with present understanding of solar physics (e.g., Gray et al. 2010) as well as with results from detection and attribution studies (Schurer et al. 2013, 2014; Stern and Kaufmann 2014). Still, studies of whether a larger or smaller amplitude of solar forcing variations agrees best with temperature reconstructions have so far been inconclusive (Fernández-Donado et al. 2013; Hind and Moberg 2013; Moberg et al. 2015; Luterbacher et al. 2016).

The PAGES 2k-PMIP3 Group (2015) found the multidecadal temperature response to changes in solar forcing to be larger in the PAGES 2k Consortium (2013) reconstructions for Europe, the Arctic and, to a lesser extent, Asia than suggested by simulations, and Luterbacher et al. (2016) came to similar conclusions

for Europe. This may be due to that the multidecadal solar irradiance change is underestimated by about 40% in the PMIP3 model simulations (Lean 2018). Another possible reason can be the lack of stratospheric chemistry and physics in most of the PMIP3 models that are sensitive to changes in solar irradiance (Mitchell et al. 2015). Chiodo and Polvani (2016) found that the inclusion of an interactive stratospheric chemistry in a climate model can significantly reduce the surface warming resulting from an increase in incoming solar radiation. This behavior can be explained by the so-called stratospheric ozone feedback, which effectively means that stronger solar irradiance will induce more ozone concentration in the stratosphere due to photochemistry (Nowack et al. 2018). Consequently, increased ozone will absorb more solar radiation and result in less solar radiation arriving to the surface, thus reducing the surface temperature response to the increased solar radiation. Unfortunately, none of the PMIP3 models (or EC-Earth) that have been used to perform the “last millennium” transient simulations include stratospheric interactive chemistry; not even the two models that do have a high-top stratosphere: GISS (Schmidt et al. 2014b) and MPI (J. H. Jungclaus 2018, personal communication). This implies that the surface temperature MCA–LIA difference, simulated by the PMIP3 models, may be expected to be larger than if those models actually had included the interactive chemistry. The importance of atmospheric chemistry has been recognized and the Aerosol Chemistry Model Intercomparison Project (AerChemMIP) is endorsed by CMIP6 to quantify the climate effect from aerosols and chemically reactive gases (Collins et al. 2017). All the AerChemMIP models are including an interactive representation of tropospheric aerosols and atmospheric chemistry. Because of the very heavy computation (e.g., up to 5 times the expense for the EC-Earth model) of atmospheric chemistry processes, it is unlikely that any PMIP4 model will include this interactive module in their last millennium transient simulations. Besides the PMIP4 experiments, for modeling group with available computation resource, it would be of interest to perform sensitivity experiments with latest solar irradiance estimation (Lean 2018) and enabled interactive atmospheric

Rank sums

Mean	Median	CSIRO	HadCM3	CCSM4	GISS-R27	GISS-R21	GISS-R24	EC-Earth	FGOALS	MPI	IPSL	BCC
749	799	828	834	866	868	884	917	919	927	975	1026	1056

FIG. 9. Friedman–Nemenyi test for simulated temperatures with all seasons weighted together compared to local reconstructions. The models are ordered by rank sum, and the insignificant ($p > 0.05$) values are connected by a common green horizontal line.

chemistry. Such sensitivity experiments can also be feasibly run merely for the MCA and LIA periods, to test if a better agreement with reconstructions is achievable.

Moreover, [Ineson et al. \(2015\)](#) showed substantial regional climate effects through ultraviolet solar irradiance modulation of the Arctic and North Atlantic Oscillations, which may result in feedbacks not captured by the models. Actually, solar forcing may be more important on multicentennial than shorter time scales (e.g., [Anchukaitis et al. 2017](#); [Pei et al. 2017](#)). However, it appears unlikely that deficiencies in solar forcing are the major reason for the discrepancy between simulations and reconstructions. A main argument for this is that although the amplitude of solar forcing change is uncertain the temporal evolution is well constrained; in particular, peak medieval temperatures occur more than a century *before* the medieval solar maximum (see [section 1](#)).

One possible explanation for why the simulations show a smaller amplitude of MCA–LIA temperature difference than most reconstructions may be residual deep-ocean warmth (from earlier in the Holocene; [Rosenthal et al. 2013, 2017](#)) not considered in the initial conditions of simulations starting at 850 CE. The choice of initial conditions is important: the same climate model, with the same set of forcings, produces different climate evolutions depending on the chosen initial conditions ([González-Rouco et al. 2009](#)). If the oceans are too cold in the model setup, it would exert a significant cold bias during the first centuries of the last millennium simulations and thus result in a biased cold MCA. In line with the recommendations for the PMIP4 protocol ([Jungclauss et al. 2017](#)), it would be preferable for model–data comparison if the simulation starts in the year 0 or runs a transient Holocene simulation, which at the moment is rather challenging for most of the modeling group due to the computation cost. We note that a negative (cooling) drift in the simulations is often corrected, thus making it more difficult to address the possible effect of the oceans being warmer than in equilibrium to forcing.

b. Comparison with existing simulation–reconstruction studies

PMIP3 last millennium simulations have, by some investigators, been found to generally show less variability at multidecadal and longer time scales than reconstructions ([Zorita et al. 2010](#); [Bothe et al. 2013a,b](#); [Fernández-Donado et al. 2013](#); [Schurer et al. 2013](#); [Yan et al. 2015](#)). Assessing this issue more systematically, the [PAGES 2k-PMIP3 group \(2015\)](#) modified this conclusion. Relatively good agreement between simulations and PAGES2k continental-scale reconstructions was

found for the NH, but disagreement in the SH (see also [Neukom et al. 2018](#)). The NH agreement was best at 20–50-yr time scales; many simulations showed less low-frequency but more high-frequency variability than the reconstructions. Conversely, simulations actually showed *more* low-frequency variability than reconstructions for the SH.

Model simulations typically also underestimate the amplitude of change compared to the paleoclimate evidence during earlier warm periods: for example, for the last interglacial period (~130–116 ka) ([Otto-Bliesner et al. 2013](#)), regionally during the mid-Holocene ([Zhang et al. 2010](#); [Braconnot et al. 2012](#); [Harrison et al. 2014](#)), and for the Neoglacial global sea surface cooling trend ([Lohmann et al. 2013](#)). Since orbital forcing, mainly responsible for the climate conditions during these periods, is well constrained, the lack of some important feedback mechanisms in climate models is the likely cause for the reconstruction–simulation discrepancies.

Our finding that it is not possible to rank last millennium simulations by their agreement with continental-scale reconstructions—and that it is only possible to do so, to a limited degree, with local reconstructions—is well in line with previous findings. For example, [Bothe et al. \(2013a,b\)](#) concluded that the inconsistency between simulations and reconstructions at multidecadal and longer time scales renders it difficult to draw firm conclusions regarding model performance and the relative influence of different forcings. Likewise, [Moberg \(2013\)](#) suggested that the temperature changes during the last millennium have been too small to constrain climate sensitivity, and [Fernández-Donado et al. \(2013\)](#) and [Tingley et al. \(2015\)](#) demonstrated the uncertainties and limitations in trying to do so.

c. The importance of internal variability and feedback mechanisms

Not just the models used in PMIP3 studies (see above), but also those used in other studies have been found to typically underestimate the amplitude of unforced variability and/or have deficiencies in their feedback mechanisms on multidecadal and longer time scales ([Cheung et al. 2017](#)). Except for the cooling after large volcanic eruptions, internal variability dominates temperature variability in last millennium PMIP3 simulations ([Lehner et al. 2015](#)). Similarly, [Le et al. \(2016\)](#) found that internal variability plays a larger role, than changes in external forcing, in simulations of temperature variations during the past millennium over the North Atlantic, North Pacific, and Arctic as well as parts of Antarctica and surrounding oceans.

Compared to instrumental data (Laepple and Huybers 2014a; Davey et al. 2002; DelSole 2006) and proxy data (Laepple and Huybers 2014b; Dee et al. 2017), Lovejoy et al. (2013) found models to systematically simulate too-little centennial-scale temperature variability. However, the PAGES2k-PMIP3 Group (2015) found, using spectral analysis, no tendency for a systematic underestimation of low-frequency temperature variability in the simulations, compared to the continental-scale last millennium reconstructions. Rather, they found that results varied among the regions. A plausible explanation for the model–data discrepancies is that the models could have a too-weak internal variability on longer time scales (e.g., Valdes 2011). The models may be too diffusive—in line with model–data discrepancies increasing with time scale (Laepple and Huybers 2014a)—or have a too-weak ocean–atmosphere coupling (Fernández-Donado 2016). Low-frequency internal variability could partly be governed by nonlinear responses from solar and/or volcanic forcings, and include slow feedback mechanisms, deep ocean currents, and/or various biogeochemical processes that are not properly captured by the models.

The role of internal variability is largest on smaller spatial scales (Hawkins and Sutton 2009). Conclusions from the results of experiments using the unforced control simulation show that internal variability is smallest on a global scale and largest in the higher latitudes of the NH (section 3d). The previously reported inability to simulate multidecadal warm periods on regional scales, for instance in Europe during the eighteenth century (Gómez-Navarro et al. 2015), indicates a too-weak internal variability in the models, perhaps in combination with incorrectly specified forcing, model errors, and/or observational errors.

For the long-term cooling trend of the last millennium, orbital forcing may play a larger role than the models suggest. The orbital induced change in insolation (Berger and Loutre 1991) resulted in little change in external forcing globally or annually but regional changes in seasonal forcing, in particular reduced boreal summer insolation, might have caused significant feedbacks including changes in albedo and water vapor (Atwood et al. 2016). The inability of mid-Holocene simulations (Braconnot et al. 2012; Hargreaves et al. 2013; Lohmann et al. 2013; Harrison et al. 2014; Liu et al. 2014) to fully capture the warm mid-Holocene conditions suggests that some feedbacks from orbital forcing can be underestimated or unaccounted for.

We find the largest simulated MCA–LIA temperature difference over the high-latitude North Atlantic section

in agreement with local reconstructions (Figs. 3, 4). This is primarily a result of changes in sea ice coverage (Landrum et al. 2013), highlighting the importance of this feedback. The larger reconstructed MCA–LIA temperature difference farther south in the NH suggests some of the amplifying real-world mechanisms at work north of $\sim 60^\circ\text{N}$ also influence the climate much farther south. We propose that this may be due to a stronger effect of seasonal orbital forcing, changes in ocean currents, sea ice, vegetation, and water vapor and their feedbacks or any combination thereof. Identifying these mechanisms should be a priority in future studies as they may also be informative with regard to amplification mechanisms under ongoing anthropogenic climate change.

d. Concluding remarks

We have demonstrated, within the PMIP3 last millennium simulations, the existence of both large inter-model spread in simulating the amplitude of MCA–LIA temperature change and that reconstructions based on proxy data from the NH systematically show a larger amplitude of this change than simulations, though not in the SH. The larger MCA–LIA temperature difference in the reconstructions from the NH compared to the simulations suggest either that the temperature change in the simulations arose mainly from internal variability, or that the models used for the transient simulations fail to represent some feedback mechanisms correctly or even lack some critical forcing completely. If it is indeed deficits in the models' feedback mechanisms that are responsible, it would have implications for the skill of future climate projections, possibly implying that the models have too-small climate sensitivity [as, e.g., concluded for Europe by Luterbacher et al. (2016)]. From the assessment of an unforced control simulation we cannot exclude the possibility that the difference in simulated and reconstructed MCA–LIA temperature amplitude is the result of internal variability at the outer tails of what is possible. As the simulated and real-world (reconstructed) internal variability will never show perfect temporal agreement—even if the models would simulate the correct amplitude of internal variability—the models will never simulate the full range of reconstructed MCA–LIA temperature difference if internal variability played an important role for this temperature difference.

As models that more accurately simulate past climate may be expected to provide more reliable future projections, we tested the ability to rank model simulations according to their agreement with the reconstructed MCA–LIA temperature difference. Recall that we used the nonparametric Friedman–Nemenyi

test to decide whether any significant differences exist between pairs of models. Also, recall that our calculated p values may be too small because we cannot be certain that neighboring locations are independent. Thus, our significance tests should be taken only as rough indicators of which differences between models are meaningful. Bearing this in mind, we found it, to some degree, possible to detect statistically significant ($p < 0.05$) differences in how well various model simulations agreed with local reconstructions when data for summer, winter and annual-mean temperatures are weighted together. On the other hand, no statistically significant differences could be found among the various simulations' agreement with continental-scale reconstructions. These results indicate that the differences between the temperature data generated by the different climate models are small in relation to the uncertainties and heterogeneities in the simulations as well as in the local and continental-scale reconstructions. However, if similar reconstruction–model comparisons would also be performed for periods with large amplitudes in temperature change—for example, the Last Glacial Maximum and the mid-Holocene—the results obtained may be more informative for evaluating model performances.

To the knowledge of the authors, it has never been extensively discussed before that too-cold initial ocean conditions in the climate models' settings may to a large extent be responsible for the smaller simulated than reconstructed MCA–LIA temperature difference. We argue that in order to test this hypothesis, and to accurately simulate the climate of the second millennium CE, it is necessary to also simulate the climate of the full first millennium CE as the ocean heat content may not be in equilibrium with the forcing at the start of simulations. This will now be done by some PMIP4 modeling groups (Jungclauss et al. 2017).

In conclusion, we propose three possibilities, or any combination thereof, to explain the smaller amplitude of temperature change in the model simulations compared to reconstructions: 1) too-cold initial ocean conditions in the climate models' settings as discussed above, 2) too-weak dynamics of internal variability in the models over longer time scales and/or simply the mismatch in timing of real-world (reconstructed) and simulated internal variability, and 3) some missing or too-weak feedback mechanisms in the simulated climate systems. While points 1 and 2 would have limited implications for future projections, point 3 would imply that the models might have an erroneous climate sensitivity over longer time scales. It is challenging to address these questions, but we expect two of the experimental designs outlined in Jungclauss et al. (2017)

for future PMIP4 last millennium simulations to be particularly informative in this regard: 1) starting simulation in year 0 or, preferably, near the beginning of the Holocene (e.g., about 10 000 years ago) as it would allow testing the hypothesis about too-cold initial ocean conditions, and 2) running simulations with time series of both low and high solar forcing estimates to test uncertainties of this potentially important, but still poorly constrained, forcing. In both cases, a number of different ensembles from the same model, also run with other forcing alternatives, would be needed to robustly assess the influence of individual forcings. Finally, systematic comparisons with comprehensive sets of quality-assessed paleoclimate reconstructions should be made by using appropriate methods that allow the drawing of conclusions in statistical terms.

Acknowledgments. This study was partly funded by the Swedish Research Council (Grant 2012-5924, “A statistical framework for comparing paleoclimate data and climate model simulations”). Q.Z. was also partly funded by the Swedish Research Council for the Swedish–French project “Greenland in a warming Arctic” (GIWA, Grant 2014-06578). F.C.L. was partly supported by the Royal Swedish Academy of Letters, History and Antiquities and the Bank of Sweden Tercentenary Foundation. A.S. was supported by the German Research Foundation (DFG, SE 2802/1-1). The EC-Earth model “last millennium” simulation was performed with resources provided by Cray XC30 HPC systems at the European Centre for Medium-Range Weather Forecasts (ECMWF), and model data analysis is performed on the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre (NSC). Dr. Hanna S. Sundqvist made valuable contributions in the early stages of this project. We thank the three reviewers whose useful comments helped to improve this article.

APPENDIX A

Uncertainty Estimates for the Cost Function (CF) Calculations

In the calculation of CF it is reasonable to let the weights be inversely proportional to the variances based on the known or estimated uncertainties of the local and continental-scale temperature reconstructions (see “Uncertainty estimates in local reconstructions” in section 2a). We calculated averages of 300 years for the MCA and of 400 years for the LIA of reconstructed temperatures except for North America where we consider averages of only 10 years for the MCA and 13 years for the LIA.

Having converted the published uncertainties in temperatures for MCA and LIA into standard deviations SD_{MCA} and SD_{LIA} (see Table 3), we can estimate the variance of the (MCA–LIA) difference. We denote the number of observations in each average by n ; thus, for North America $n = 10/13$, and for all other regional reconstructions $n = 300/400$. If the observations are

independent, then the variance of the difference can be estimated simply as

$$V = \frac{SD_{MCA}^2}{n} + \frac{SD_{LIA}^2}{n}. \quad (A1)$$

It can be shown (e.g., Wei 1989) that the variance adjusted for autocorrelation is

$$\begin{aligned} \text{Var}(\bar{X}_{MCA} - \bar{X}_{LIA}) &= \text{Var}(\bar{X}_{MCA}) + \text{Var}(\bar{X}_{LIA}) = \left(\frac{SD_{MCA}^2}{n} + \frac{SD_{LIA}^2}{n} \right) \left[1 + 2 \sum_{f=1}^{n-1} \left(1 - \frac{f}{n} \right) r(f) \right] \\ &= V \times \left[1 + 2 \sum_{f=1}^{n-1} \left(1 - \frac{f}{n} \right) r(f) \right], \end{aligned} \quad (A2)$$

where $r(f)$ is the estimated autocorrelation function (ACF) at lag f for the time series of reconstructed temperatures. The first equality holds because MCA and LIA are sufficiently far apart to be considered independent, even in the presence of autocorrelation. The standard deviations SD_{MCA} and SD_{LIA} were estimated above using published uncertainty estimates. For the autocorrelation no such information exists, and we are forced to try to extract it from the data itself. A linear trend in the data will automatically produce large autocorrelations at all lags and will also violate the stationarity assumption implicit in Eq. (A2). Hence, it makes sense to remove the trend and calculate the ACF on the residuals. Another reason for doing this is that what we really want is the autocorrelation in the noise, not in the temperature signal. If we introduce the notation

$$A = 1 + 2 \sum_{f=1}^{n-1} \left(1 - \frac{f}{n} \right) r(f) \quad (A3)$$

for the correction factor, then the variance corrected for autocorrelation is $V \times A$.^{A1} The weight assigned to each continental-scale temperature reconstruction is proportional to $(V \times A)^{-1}$. Finally, the weights are standardized to sum to 1. The results are summarized in Table 2. For the continental-scale reconstructions, this method of assigning weights is unproblematic, but applying it mechanically to data using local proxies produces unreasonable results: in some cases, a single reconstruction receives well over half the weight, which is clearly not desirable. This is partly due to some

records having an unrealistically low uncertainty estimate, which inflates the weight. Hence, we impose a cutoff on the standard deviations of 0.1. In addition, we set the upper 10% of the weights equal to the upper 10% quantile, thus guarding against any one reconstruction dominating. After these adjustments, we again standardize the weights to sum to 1.

APPENDIX B

The EC-Earth Climate Model “Last Millennium” (850–1850 CE) Simulation

EC-Earth is developed by a consortium of European research institutions (Hazeleger et al. 2010, 2012). The atmospheric component of EC-Earth is based on Integrated Forecasting System (IFS), which is developed at the European Centre for Medium-Range Weather Forecasts (ECMWF), and the ocean component is based on Nucleus for European Modeling of the Ocean (NEMO) (Madec 2008), including a sea ice model LIM3 (Bouillon et al. 2009). The goal of EC-Earth is to build a fully coupled atmosphere–ocean–land–biosphere model usable for seasonal to decadal climate prediction and climate projections (Hazeleger et al. 2010). A number of model integrations such as “historic” simulations (1850–2005 CE) and scenarios of future projections have been conducted using EC-Earth, version 2.3 (v2.3), to contribute to the CMIP5. However, EC-Earth v2.3 has not previously run any PMIP3 simulations except one 50-yr mid-Holocene simulation. We have implemented several components that are necessary for performing paleoclimate simulations such as orbital forcing and land–ice surface processes in EC-Earth v3.1, and performed several PMIP3 simulations with relatively higher resolution than the other PMIP3 models: T159L62 for the atmosphere model (IFS), ORCA1L46, and 46 vertical

^{A1} For reasons of numerical stability, we do not, in fact, let the sum in the correction for autocorrelation run all the way up to $n - 1$, but only to the smallest integer $\geq 10 \log^{10}(n)$, if this number is less than $n - 1$.

levels for the ocean model (NEMO) (Table 4). This might be one reason that mid-Holocene simulations show good agreement with proxy-based reconstructions (Muschitiello et al. 2015; Pausata et al. 2016). The last millennium experiment with EC-Earth v3.1 has applied the recommended forcings following PMIP3 protocol (<https://wiki.lscce.ipsl.fr/pmip3/doku.php/pmip3:design:lm:final>; Table 4), and the initial condition starts from an equilibrium state at 850 CE after 300 years of spinup. The seasonal and latitudinal distribution of the orbital modulation of insolation is calculated following Berger (1978). Variations in total solar irradiance follow Vieira and Solanki (2010) from 850 to 1609 CE, and from 1610 CE Wang et al. (2005). The volcanic aerosol forcing follows Crowley and Unterman (2013). Changes in greenhouse gas forcing derive from Hansen and Sato (2004) and MacFarling Meure et al. (2006). Land use changes are taken from Pongratz et al. (2008). The EC-Earth last millennium simulation, as used in this article, is available for download from http://model.natgeo.su.se/past1000/tas_EC-Earth3.1_mon.nc.

APPENDIX C

Confidence Interval Calculation for the Arctic Amplification Ratios

Our confidence intervals for the Arctic amplification ratios (Hind et al. 2016) have been constructed using Fieller's method (Fieller 1954; von Luxburg and Franz 2009). We consider two independent normally distributed samples, x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m with arithmetic means \bar{x} and \bar{y} ; think of them as the mid- and low-latitude temperature differences and polar temperature differences respectively. The Arctic amplification is then estimated as the ratio \bar{y}/\bar{x} . In the case of two independent samples, a $(1 - \alpha)$ -confidence interval takes the following form when the denominator \bar{x} is significantly $\neq 0$:

$$I_{\text{Fieller}} = \frac{\bar{x}\bar{y} \pm \sqrt{\bar{x}^2\bar{y}^2 - \bar{x}^T\bar{y}^T}}{\bar{x}^T}, \quad (\text{C1})$$

where $\bar{x}^T = \bar{x}^2 - z_{\alpha/2}^2 \sigma_{\bar{x}}^2$ and $\bar{y}^T = \bar{y}^2 - z_{\alpha/2}^2 \sigma_{\bar{y}}^2$, and $z_{\alpha/2}$ is the upper $\alpha/2$ normal quantile. Hence, \bar{x} is significantly $\neq 0$ precisely when $\bar{x}^T > 0$. In this case, $\bar{x}^2\bar{y}^2 > \bar{x}^T\bar{y}^T$, so the expression under the root sign is positive and we obtain an ordinary, bounded confidence interval. On the other hand, if \bar{x} is not significantly $\neq 0$, then the Fieller confidence interval is either the entire real line, or an “inverted” confidence interval which contains everything *except* the numbers between the two limits given by the formula for I_{Fieller} above.

REFERENCES

- Allen, J. R. M., A. J. Long, C. J. Ottley, D. G. Pearson, and B. Huntley, 2007: Holocene climate variability in northernmost Europe. *Quat. Sci. Rev.*, **26**, 1432–1453, <https://doi.org/10.1016/j.quascirev.2007.02.009>.
- Alley, R. B., 2000: The Younger Dryas cold interval as viewed from central Greenland. *Quat. Sci. Rev.*, **19**, 213–226, [https://doi.org/10.1016/S0277-3791\(99\)00062-1](https://doi.org/10.1016/S0277-3791(99)00062-1).
- Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas, 2007: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model. *Proc. Natl. Acad. Sci. USA*, **104**, 3713–3718, <https://doi.org/10.1073/pnas.0605064103>.
- Anchukaitis, K. J., and Coauthors, 2017: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions. *Quat. Sci. Rev.*, **163**, 1–22, <https://doi.org/10.1016/j.quascirev.2017.02.020>.
- Andres, H. J., and W. R. Peltier, 2016: Regional influences of natural external forcings on the transition from the Medieval Climate Anomaly to the Little Ice Age. *J. Climate*, **29**, 5779–5800, <https://doi.org/10.1175/JCLI-D-15-0599.1>.
- Atwood, A. R., E. Wu, D. M. W. Frierson, and D. S. Battisti, 2016: Quantifying climate forcings and feedbacks over the last millennium in the CMIP5–PMIP3 models. *J. Climate*, **29**, 1161–1178, <https://doi.org/10.1175/JCLI-D-15-0063.1>.
- Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. S. George, B. Otto-Bliesner, C. A. Woodhouse, and C. Deser, 2013: The continuum of hydroclimate variability in western North America during the last millennium. *J. Climate*, **26**, 5863–5878, <https://doi.org/10.1175/JCLI-D-11-00732.1>.
- Bao, Q., and Coauthors, 2013: The flexible global ocean-atmosphere-land system model, spectral version 2: FGOALS-s2. *Adv. Atmos. Sci.*, **30**, 561–576, <https://doi.org/10.1007/s00376-012-2113-9>.
- Berger, A. L., 1978: Long-term variations of daily insolation and Quaternary climatic changes. *J. Atmos. Sci.*, **35**, 2362–2367, [https://doi.org/10.1175/1520-0469\(1978\)035<2362:LTVODI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2).
- , and M. F. Loutre, 1991: Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.*, **10**, 297–317, [https://doi.org/10.1016/0277-3791\(91\)90033-Q](https://doi.org/10.1016/0277-3791(91)90033-Q).
- Berghórrsson, P., 1969: An estimate of drift ice and temperature in Iceland in 1000 years. *Jökull*, **19**, 94–101.
- Bernabo, J. C., 1981: Quantitative estimates of temperature changes over the last 2700 years in Michigan based on pollen data. *Quat. Res.*, **15**, 143–159, [https://doi.org/10.1016/0033-5894\(81\)90101-0](https://doi.org/10.1016/0033-5894(81)90101-0).
- Berner, K. S., N. Koç, F. Godtliebsen, and D. Divine, 2011: Holocene climate variability of the Norwegian Atlantic Current during high and low solar insolation forcing. *Paleoceanography*, **26**, PA2220, <https://doi.org/10.1029/2010PA002002>.
- Bohleber, P., T. Erhardt, N. Spaulding, H. Hoffmann, H. Fischer, and P. Mayewski, 2018: Temperature and mineral dust variability recorded in two low-accumulation Alpine ice cores over the last millennium. *Climate Past*, **14**, 21–37, <https://doi.org/10.5194/cp-14-21-2018>.
- Boldt, B. R., D. S. Kaufman, N. P. McKay, and J. P. Briner, 2015: Holocene summer temperature reconstruction from sedimentary chlorophyll content, with treatment of age uncertainties, Kurupa Lake, Arctic Alaska. *Holocene*, **25**, 641–650, <https://doi.org/10.1177/0959683614565929>.
- Böll, A., and Coauthors, 2014: Late Holocene primary productivity and sea surface temperature variations in the northeastern Arabian Sea: Implications for winter monsoon variability. *Paleoceanography*, **29**, 778–794, <https://doi.org/10.1002/2013PA002579>.

- Bothe, O., J. H. Jungelaus, and D. Zanchettin, 2013a: Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble. *Climate Past*, **9**, 2471–2487, <https://doi.org/10.5194/cp-9-2471-2013>.
- , —, —, and E. Zorita, 2013b: Climate of the last millennium: Ensemble consistency of simulations and reconstructions. *Climate Past*, **9**, 1089–1110, <https://doi.org/10.5194/cp-9-1089-2013>.
- Bouillon, S., M. A. Morales Maqueda, V. Legat, and T. Fichefet, 2009: An elastic–viscous–plastic sea ice model formulated on Arakawa B and C grids. *Ocean Modell.*, **27**, 174–184, <https://doi.org/10.1016/j.ocemod.2009.01.004>.
- Box, G. E. P., J. S. Hunter, and W. G. Hunter, 2005: *Statistics for Experimenters*. 2nd ed. Wiley, 633 pp.
- Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao, 2012: Evaluation of climate models using palaeoclimatic data. *Nat. Climate Change*, **2**, 417–424, <https://doi.org/10.1038/nclimate1456>.
- Briffa, K. R., T. M. Melvin, T. J. Osborn, R. M. Hantemirov, A. V. Kirdyanov, V. S. Mazepa, S. G. Shiyatov, and J. Esper, 2013: Reassessing the evidence for tree-growth and inferred temperature change during the Common Era in Yamalia, north-west Siberia. *Quat. Sci. Rev.*, **72**, 83–107, <https://doi.org/10.1016/j.quascirev.2013.04.008>.
- Büntgen, U., and Coauthors, 2011: 2500 years of European climate variability and human susceptibility. *Science*, **331**, 578–582, <https://doi.org/10.1126/science.1197175>.
- , and Coauthors, 2016: Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nat. Geosci.*, **9**, 231–236, <https://doi.org/10.1038/ngeo2652>.
- , D. C. Frank, D. Nievergelt, and J. Esper, 2006: Summer temperature variations in the European Alps, A.D. 755–2004. *J. Climate*, **19**, 5606–5623, <https://doi.org/10.1175/JCLI3917.1>.
- Bürger, G., I. Fast, and U. Cubasch, 2006: Climate reconstruction by regression—32 variations on a theme. *Tellus*, **58A**, 227–235, <https://doi.org/10.1111/j.1600-0870.2006.00164.x>.
- Chen, F., Y. Zhang, X. Shao, M. Q. Li, and Z. Y. Yin, 2016: A 2000-year temperature reconstruction in the Animaqin Mountains of the Tibet Plateau, China. *Holocene*, **26**, 1904–1913, <https://doi.org/10.1177/0959683616646187>.
- Cheung, A. H., M. E. Mann, B. A. Steinman, L. M. Frankcombe, M. H. England, and S. K. Miller, 2017: Comparison of low frequency internal climate variability in CMIP5 models and observations. *J. Climate*, **30**, 4763–4776, <https://doi.org/10.1175/JCLI-D-16-0712.1>.
- Chiodo, G., and L. M. Polvani, 2016: Reduction of climate sensitivity to solar forcing due to stratospheric ozone feedback. *J. Climate*, **29**, 4651–4663, <https://doi.org/10.1175/JCLI-D-15-0721.1>.
- Christiansen, B., 2011: Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided? *J. Climate*, **24**, 674–692, <https://doi.org/10.1175/2010JCLI3646.1>.
- , and F. C. Ljungqvist, 2017: Challenges and perspectives for large-scale temperature reconstructions of the past two millennia. *Rev. Geophys.*, **55**, 40–96, <https://doi.org/10.1002/2016RG000521>.
- , T. Schmith, and P. Thejll, 2009: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *J. Climate*, **22**, 951–976, <https://doi.org/10.1175/2008JCLI2301.1>.
- Chu, G., and Coauthors, 2012: Seasonal temperature variability during the past 1600 years recorded in historical documents and varved lake sediment profiles from Northeastern China. *Holocene*, **22**, 785–792, <https://doi.org/10.1177/0959683611430413>.
- Clegg, B. F., G. H. Clarke, M. L. Chipman, M. Chou, I. R. Walker, W. Tinner, and F. S. Hu, 2010: Six millennia of summer temperature variation based on midge analysis of lake sediments from Alaska. *Quat. Sci. Rev.*, **29**, 3308–3316, <https://doi.org/10.1016/j.quascirev.2010.08.001>.
- , R. Kelly, G. H. Clarke, I. R. Walker, and F. S. Hu, 2011: Nonlinear response of summer temperature to Holocene insolation forcing in Alaska. *Proc. Natl. Acad. Sci. USA*, **108**, 19 299–19 304, <https://doi.org/10.1073/pnas.1110913108>.
- Collins, W. J., and Coauthors, 2017: AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geosci. Model Dev.*, **10**, 585–607, <https://doi.org/10.5194/gmd-10-585-2017>.
- Collins, M., S. F. B. Tett, and C. Cooper, 2001: The internal climate variability of a HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **17**, 61–81, <https://doi.org/10.1007/s003820000094>.
- Cook, E. R., B. M. Buckley, R. D. D’Arrigo, and M. J. Peterson, 2000: Warm-season temperatures since 1600 BC reconstructed from Tasmanian tree rings and their relationship to large-scale sea surface temperature anomalies. *Climate Dyn.*, **16**, 79–91, <https://doi.org/10.1007/s003820050006>.
- , J. G. Palmer, and R. D. D’Arrigo, 2002: Evidence for a ‘Medieval Warm Period’ in a 1,100 year tree-ring reconstruction of past austral summer temperatures in New Zealand. *Geophys. Res. Lett.*, **29**, 1667, <https://doi.org/10.1029/2001GL014580>.
- Cook, T. L., R. S. Bradley, J. S. Stoner, and P. Francus, 2009: Five thousand years of sediment transfer in a high Arctic watershed recorded in annually laminated sediments from Lower Murray Lake, Ellesmere Island, Nunavut, Canada. *J. Paleolimnol.*, **41**, 77–94, <https://doi.org/10.1007/s10933-008-9252-0>.
- Cronin, T. M., G. S. Dwyer, T. Kamiya, S. Schwede, and D. A. Willard, 2003: Medieval Warm Period, Little Ice Age and 20th century temperature variability from Chesapeake Bay. *Global Planet. Change*, **36**, 17–29, [https://doi.org/10.1016/S0921-8181\(02\)00161-3](https://doi.org/10.1016/S0921-8181(02)00161-3).
- Crowley, T., and M. B. Unterman, 2013: Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth Syst. Sci. Data*, **5**, 187–197, <https://doi.org/10.5194/essd-5-187-2013>.
- D’Andrea, W. J., Y. Huang, S. C. Fritz, and N. J. Anderson, 2011: Abrupt Holocene climate change as an important factor for human migration in West Greenland. *Proc. Natl. Acad. Sci. USA*, **108**, 9765–9769, <https://doi.org/10.1073/pnas.1101708108>.
- , D. Vaillencourt, N. L. Balascio, A. Werner, S. Roof, M. Retelle, and R. S. Bradley, 2012: Mild Little Ice Age and unprecedented recent warmth in an 1800 year lake sediment record from Svalbard. *Geology*, **40**, 1007–1010, <https://doi.org/10.1130/G33365.1>.
- Davey, M., and Coauthors, 2002: STOIC: A study of coupled model climatology and variability in tropical ocean regions. *Climate Dyn.*, **18**, 403–420, <https://doi.org/10.1007/s00382-001-0188-6>.
- Davi, N. K., and Coauthors, 2015: A long-term context (931–2005 C.E.) for rapid warming over Central Asia. *Quat. Sci. Rev.*, **121**, 89–97, <https://doi.org/10.1016/j.quascirev.2015.05.020>.
- Davis, B. A. S., S. Brewer, A. C. Stevenson, and J. Guiot, 2003: The temperature of Europe during the Holocene reconstructed from pollen data. *Quat. Sci. Rev.*, **22**, 1701–1716, [https://doi.org/10.1016/S0277-3791\(03\)00173-2](https://doi.org/10.1016/S0277-3791(03)00173-2).
- de Jong, R., L. von Gunten, A. Maldonado, and M. Grosjean, 2013: Late Holocene summer temperatures in the central Andes reconstructed from the sediments of high-elevation Laguna Chepical, Chile (32° S). *Climate Past*, **9**, 1921–1932, <https://doi.org/10.5194/cp-9-1921-2013>.

- Dee, S. G., L. A. Parsons, G. R. Loope, J. T. Overpeck, T. R. Ault, and J. Emile-Geay, 2017: Improved spectral comparisons of paleoclimate models and observations via proxy system modeling: Implications for multi-decadal variability. *Earth Planet. Sci. Lett.*, **476**, 34–46, <https://doi.org/10.1016/j.epsl.2017.07.036>.
- DelSole, T., 2006: Low-frequency variations of surface temperature in observations and simulations. *J. Climate*, **19**, 4487–4507, <https://doi.org/10.1175/JCLI3879.1>.
- Divine, D., and Coauthors, 2011: Thousand years of winter surface air temperature variations in Svalbard and northern Norway reconstructed from ice core data. *Polar Res.*, **30**, 7379, <https://doi.org/10.3402/polar.v30i0.7379>.
- Domack, E., A. Leventer, R. Dunbar, F. Taylor, S. Brachfield, and C. Sjunneskog, 2001: Chronology of the Palmer Deep site, Antarctic Peninsula: A Holocene palaeoenvironmental reference for the circum-Antarctic. *Holocene*, **11**, 1–9, <https://doi.org/10.1191/095968301673881493>.
- Doose-Rolinski, H., U. Rogalla, G. Scheeder, A. Lücke, and U. von Rad, 2001: High resolution temperature and evaporation changes during the late Holocene in the northeastern Arabian Sea. *Paleoceanography*, **16**, 358–367, <https://doi.org/10.1029/2000PA000511>.
- Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth system model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, <https://doi.org/10.1007/s00382-012-1636-1>.
- Eby, M., and Coauthors, 2013: Historical and idealized climate model experiments: an intercomparison of Earth system models of intermediate complexity. *Climate Past*, **9**, 1111–1140, <https://doi.org/10.5194/cp-9-1111-2013>.
- Edwards, T. W. D., S. J. Birks, B. H. Luckman, and G. M. MacDonald, 2008: Climatic and hydrologic variability during the past millennium in the eastern Rocky Mountains and northern Great Plains of western Canada. *Quat. Res.*, **70**, 188–197, <https://doi.org/10.1016/j.yqres.2008.04.013>.
- Elbert, J., and Coauthors, 2013: Late Holocene air temperature variability reconstructed from the sediments of Laguna Escondida, Patagonia, Chile (45°3'S). *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **369**, 482–492, <https://doi.org/10.1016/j.palaeo.2012.11.013>.
- Esper, J., F. H. Schweingruber, and M. Winiger, 2002: 1300 years of climatic history for Western Central Asia inferred from tree-rings. *Holocene*, **12**, 267–277, <https://doi.org/10.1191/0959683602hl543rp>.
- , and Coauthors, 2012: Orbital forcing of tree-ring data. *Nat. Climate Change*, **2**, 862–866, <https://doi.org/10.1038/nclimate1589>.
- , and Coauthors, 2016: Ranking of tree-ring based temperature reconstructions of the past millennium. *Quat. Sci. Rev.*, **145**, 134–151, <https://doi.org/10.1016/j.quascirev.2016.05.009>.
- Fernández-Donado, L., and Coauthors, 2013: Temperature response to external forcing in simulations and reconstructions of the last millennium. *Climate Past*, **9**, 393–421, <https://doi.org/10.5194/cp-9-393-2013>.
- , 2016: Forced and internal variability in temperature simulations and reconstructions of the Common Era. Ph.D. dissertation, Universidad Complutense de Madrid, 167 pp., <https://eprints.ucm.es/35730/>.
- Fieller, E. C., 1954: Some problems in interval estimation. *J. Roy. Stat. Soc.*, **16B**, 175–185, <https://doi.org/10.1111/j.2517-6161.1954.tb00159.x>.
- Filippi, M. L., P. Lambert, J. Hunziker, B. Kubler, and S. Bernasconi, 1999: Climatic and anthropogenic influence on the stable isotope record from bulk carbonates and ostracodes in Lake Neuchâtel, Switzerland, during the last two millennia. *J. Paleolimnol.*, **21**, 19–34, <https://doi.org/10.1023/A:1008005622256>.
- Flückiger, J., and Coauthors, 1999: Variations in atmospheric N₂O concentration during abrupt climatic changes. *Science*, **285**, 227–230, <https://doi.org/10.1126/science.285.5425.227>.
- , and Coauthors, 2002: High-resolution Holocene N₂O ice core record and its relationship with CH₄ and CO₂. *Global Biogeochem. Cycles*, **16**, <https://doi.org/10.1029/2001GB001417>.
- Franke, J., S. Brönnimann, J. Bhend, and Y. Brugnara, 2017: A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations. *Sci. Data*, **4**, 170076, <https://doi.org/10.1038/sdata.2017.76>.
- Friedman, M., 1937: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Stat. Assoc.*, **32**, 675–701, <https://doi.org/10.1080/01621459.1937.10503522>.
- , 1939: A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Stat. Assoc.*, **34**, 109, <https://doi.org/10.2307/2279169>.
- Gajewski, K., 1988: Late Holocene climate changes in eastern North America estimated from pollen data. *Quat. Res.*, **29**, 255–262, [https://doi.org/10.1016/0033-5894\(88\)90034-8](https://doi.org/10.1016/0033-5894(88)90034-8).
- Gao, C., A. Robock, and C. Ammann, 2008: Volcanic forcing of climate over the last 1500 years: An improved ice core-based index for climate models. *J. Geophys. Res.*, **113**, D23111, <https://doi.org/10.1029/2008JD010239>.
- Ge, Q., J. Zheng, X. Fang, Z. Man, X. Zhang, P. Zhang, and W.-C. Wang, 2003: Winter half-year temperature reconstruction for the middle and lower reaches of the Yellow River and Yangtze River, China, during the past 2000 years. *Holocene*, **13**, 933–940, <https://doi.org/10.1191/0959683603hl680rr>.
- Gennaretti, F., D. Arseneault, A. Nicault, L. Perreault, and Y. Bégin, 2014: Volcano-induced regime shifts in millennial tree-ring chronologies from northeastern North America. *Proc. Natl. Acad. Sci. USA*, **111**, 10 077–10 082, <https://doi.org/10.1073/pnas.1324220111>.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, <https://doi.org/10.1175/2011JCLI4083.1>.
- Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, <https://doi.org/10.1002/jame.20038>.
- Gómez-Navarro, J. J., and Coauthors, 2015: A regional climate palaeosimulation for Europe in the period 1500–1990—Part 2: Shortcomings and strengths of models and reconstructions. *Climate Past*, **11**, 1077–1095, <https://doi.org/10.5194/cp-11-1077-2015>.
- Goñi, M. A., and Coauthors, 2003: Biogenic fluxes in the Cariaco Basin: A combined study of sinking particulates and underlying sediments. *Deep-Sea Res.*, **50**, 781–807, [https://doi.org/10.1016/S0967-0637\(03\)00060-8](https://doi.org/10.1016/S0967-0637(03)00060-8).
- González-Rouco, J. F., H. Beltrami, E. Zorita, and M. B. Stevens, 2009: Borehole climatology: A discussion based on contributions from climate modelling. *Climate Past*, **5**, 97–127, <https://doi.org/10.5194/cp-5-97-2009>.
- Goosse, H., and Coauthors, 2005: Modelling the climate of the last millennium: What causes the differences between simulations. *Geophys. Res. Lett.*, **32**, L06710, <https://doi.org/10.1029/2005GL022368>.

- , 2017: Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries. *Climate Dyn.*, **48**, 1483–1501, <https://doi.org/10.1007/s00382-016-3154-z>.
- , H. Renssen, A. Timmermann, R. S. Bradley, and M. E. Mann, 2006: Using palaeoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Climate Dyn.*, **27**, 165–184, <https://doi.org/10.1007/s00382-006-0128-6>.
- , E. Cresspin, S. Dubinkina, M.-F. Loutre, M. E. Mann, H. Renssen, Y. Sallaz-Damaz, and D. Shindell, 2012: The role of forcing and internal dynamics in explaining the “Medieval Climate Anomaly.” *Climate Dyn.*, **39**, 2847–2866, <https://doi.org/10.1007/s00382-012-1297-0>.
- Graumlich, L. J., 1993: A 1000-year record of temperature and precipitation in the Sierra Nevada. *Quat. Res.*, **39**, 249–255, <https://doi.org/10.1006/qres.1993.1029>.
- Gray, L. J., and Coauthors, 2010: Solar influences on climate. *Rev. Geophys.*, **48**, RG4001, <https://doi.org/10.1029/2009RG000282>.
- Hakim, G. J., and Coauthors, 2016: The last millennium climate reanalysis project: Framework and first results. *J. Geophys. Res. Atmos.*, **121**, 6745–6764, <https://doi.org/10.1002/2016JD024751>.
- Hansen, J., and M. Sato, 2004: Greenhouse gas growth rates. *Proc. Natl. Acad. Sci. USA*, **101**, 16 109–16 114, <https://doi.org/10.1073/pnas.0406982101>.
- Hargreaves, J. C., J. D. Annan, R. Ohgaito, A. Paul, and A. Abe-Ouchi, 2013: Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene. *Climate Past*, **9**, 811–823, <https://doi.org/10.5194/cp-9-811-2013>.
- Harrison, S., and Coauthors, 2014: Climate model benchmarking with glacial and mid-Holocene climates. *Climate Dyn.*, **43**, 671–688, <https://doi.org/10.1007/s00382-013-1922-6>.
- , P. J. Bartlein, and I. C. Prentice, 2016: What have we learnt from palaeoclimate simulations? *J. Quat. Sci.*, **31**, 363–385, <https://doi.org/10.1002/jqs.2842>.
- Hartl-Meier, C. T. M., and Coauthors, 2017: Millennial-scale temperature coherence in proxy reconstructions and climate models. *Geophys. Res. Lett.*, **44**, 9458–9469, <https://doi.org/10.1002/2017GL073239>.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, <https://doi.org/10.1175/2009BAMS2607.1>.
- Hazeleger, W., and Coauthors, 2010: EC-Earth: A seamless earth-system prediction approach in action. *Bull. Amer. Meteor. Soc.*, **91**, 1357–1363, <https://doi.org/10.1175/2010BAMS2877.1>.
- , and Coauthors, 2012: EC-Earth V2.2: Description and validation of a new seamless earth system prediction model. *Climate Dyn.*, **39**, 2611–2629, <https://doi.org/10.1007/s00382-011-1228-5>.
- He, Y.-X., and Coauthors, 2013: Solar influenced late Holocene temperature changes on the northern Tibetan Plateau. *Chin. Sci. Bull.*, **58**, 1053–1059, <https://doi.org/10.1007/s11434-012-5619-8>.
- Helama, S., M. M. Fauria, K. Mielikäinen, M. Timonen, and M. Eronen, 2010: Sub-Milankovitch solar forcing of past climates: Mid and late Holocene perspectives. *Geol. Soc. Amer. Bull.*, **122**, 1981–1988, <https://doi.org/10.1130/B30088.1>.
- , M. Vartiainen, J. Holopainen, H. M. Mäkelä, T. Kolström, and J. Meriläinen, 2014: A palaeotemperature record for the Finnish Lakeland based on microdensitometric variations in tree rings. *Geochronometria*, **41**, 265–277, <https://doi.org/10.2478/s13386-013-0163-0>.
- Herbert, C. T., and J. S. Compton, 2007: Geochronology of Holocene sediments on the western margin of South Africa. *S. Afr. J. Geol.*, **110**, 327–338, <https://doi.org/10.2113/gssajg.110.2-3.327>.
- Hind, A., and A. Moberg, 2013: Past millennial solar forcing magnitude. *Climate Dyn.*, **41**, 2527–2537, <https://doi.org/10.1007/s00382-012-1526-6>.
- , Q. Zhang, and G. Brattström, 2016: Problems encountered when defining Arctic amplification as a ratio. *Sci. Rep.*, **6**, 30 469, <https://doi.org/10.1038/srep30469>.
- Holmgren K., and et al, 1999: A 3000-year high-resolution record of palaeoclimate for north-eastern South Africa. *Holocene*, **9**, 295–309, <https://doi.org/10.1191/095968399672625464>.
- Holmgren, K., P. D. Tyson, A. Moberg, and O. Svanered, 2001: A preliminary 3000-year regional temperature reconstruction for South Africa. *S. Afr. J. Sci.*, **97**, 49–51.
- Hu, F. S., E. Ito, T. A. Brown, B. B. Curry, and D. R. Engstrom, 2001: Pronounced climatic variations in Alaska during the last two millennia. *Proc. Natl. Acad. Sci. USA*, **98**, 10 552–10 556, <https://doi.org/10.1073/pnas.181333798>.
- Huang, S. P., H. N. Pollack, and P. Y. Shen, 2008: A late Quaternary climate reconstruction based on borehole heat flux data, borehole temperature data, and the instrumental record. *Geophys. Res. Lett.*, **35**, L13703, <https://doi.org/10.1029/2008GL034187>.
- Hurt, G. C., and Coauthors, 2006: The underpinnings of land-use history: Three centuries of global gridded landuse transitions, wood harvest activity, and resulting secondary lands. *Global Change Biol.*, **12**, 1208–1229, <https://doi.org/10.1111/j.1365-2486.2006.01150.x>.
- Ineson, S., and Coauthors, 2015: Regional climate impacts of a possible future grand solar minimum. *Nat. Commun.*, **6**, 7535, <https://doi.org/10.1038/ncomms8535>.
- Isono, D., M. Yamamoto, T. Irino, T. Oba, M. Murayama, T. Nakamura, and H. Kawahata, 2009: The 1500-year climate oscillation in the midlatitude North Pacific during the Holocene. *Geology*, **37**, 591–594, <https://doi.org/10.1130/G25667A.1>.
- Jiang, J., M. Eiriksson, K. L. Schultz, K.-L. Knudsen, and M.-S. Seidenkrantz, 2005: Evidence for solar forcing of sea surface temperature on the North Icelandic Shelf during the late Holocene. *Geology*, **33**, 73–76, <https://doi.org/10.1130/G21130.1>.
- Johns, T. C., and Coauthors, 2003: Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Climate Dyn.*, **20**, 583–612, <https://doi.org/10.1007/s00382-002-0296-y>.
- Jones, P., and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. *Holocene*, **19**, 3–49, <https://doi.org/10.1177/0959683608098952>.
- Jouzel, J., and Coauthors, 2007: Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, **317**, 793–796, <https://doi.org/10.1126/science.1141038>.
- Jungclauss, J. H., and Coauthors, 2010: Climate and carbon-cycle variability over the last millennium. *Climate Past*, **6**, 723–737, <https://doi.org/10.5194/cp-6-723-2010>.
- , and Coauthors, 2017: The PMIP4 contribution to CMIP6 – Part 3: the last millennium, scientific objective, and experimental design for the PMIP4 *past1000* simulations. *Geosci. Model Dev.*, **10**, <https://doi.org/10.5194/gmd-10-4005-2017>.
- Kalugin, I. A., A. V. Daryin, and V. V. Babich, 2009: Reconstruction of annual air temperatures for three thousand years in Altai region by lithological and geochemical indicators in Teletskoe Lake sediments. *Dokl. Earth Sci.*, **426**, 681–684, <https://doi.org/10.1134/S1028334X09040382>.

- Kaplan, J. O., K. M. Krumhardt, E. C. Ellis, W. F. Ruddiman, C. Lemmen, and K. Klein Goldewijk, 2011: Holocene carbon emissions as a result of anthropogenic land cover change. *Holocene*, **21**, 775–791, <https://doi.org/10.1177/0959683610386983>.
- Kawahata, H., and Coauthors, 2017: Climatic change and its influence on human society in western Japan during the Holocene. *Quat. Int.*, **440**, 102–117, <https://doi.org/10.1016/j.quaint.2016.04.013>.
- Keigwin, L. D., 1996: The Little Ice Age and Medieval Warm Period in the Sargasso Sea. *Science*, **274**, 1504–1508, <https://doi.org/10.1126/science.274.5292.1504>.
- Kellerhals, T., S. Brüttsch, M. Sigl, S. Knüsel, H. W. Gäggeler, and M. Schwikowski, 2010: Ammonium concentration in ice cores: A new proxy for regional temperature reconstruction? *J. Geophys. Res.*, **115**, D16123, <https://doi.org/10.1029/2009JD012603>.
- Kim, J. H., S. Schouten, E. C. Hopmans, B. Donner, and J. S. Sinninghe Damsté, 2008: Global sediment core-top calibration of the TEX₈₆ paleothermometer in the ocean. *Geochim. Cosmochim. Acta*, **72**, 1154–1173, <https://doi.org/10.1016/j.gca.2007.12.010>.
- Kitagawa, H., and E. Matsumoto, 1995: Climatic implications of $\delta^{13}\text{C}$ variations in a Japanese cedar (*Cryptomeria japonica*) during the last two millennia. *Geophys. Res. Lett.*, **22**, 2155–2158, <https://doi.org/10.1029/95GL02066>.
- Klein, F., and H. Goosse, 2018: Reconstructing East African rainfall and Indian Ocean sea surface temperatures over the last centuries using data assimilation. *Climate Dyn.*, **50**, 3909–3929, <https://doi.org/10.1007/s00382-017-3853-0>.
- Kobashi, T., K. Kawamura, J. P. Severinghaus, J.-M. Barnola, T. Nakaegawa, B. M. Vinther, S. J. Johnsen, and J. E. Box, 2011: High variability of Greenland surface temperature over the past 4000 years estimated from trapped air in an ice core. *Geophys. Res. Lett.*, **38**, L21501, <https://doi.org/10.1029/2011GL049444>.
- Kong, D., G. Wei, M. T. Chen, S. Peng, and Z. Liu, 2017: Northern South China Sea SST changes over the last two millennia and possible linkage with solar irradiance. *Quat. Int.*, **459**, 29–34, <https://doi.org/10.1016/j.quaint.2017.10.001>.
- Krivova, N. A., L. E. A. Vieira, and S. K. Solanki, 2010: Reconstruction of solar spectral irradiance since the Maunder minimum. *J. Geophys. Res.*, **115**, A12112, <https://doi.org/10.1029/2010JA015431>.
- , S. K. Solanki, and Y. C. Unruh, 2011: Towards a long-term record of solar total and spectral irradiance. *J. Atmos. Sol.-Terr. Phys.*, **73**, 223–234, <https://doi.org/10.1016/j.jastp.2009.11.013>.
- Kuhnert, H., and S. Mulitza, 2011: Multidecadal variability and late medieval cooling of near-coastal sea surface temperatures in the eastern tropical North Atlantic. *Paleoceanography*, **26**, PA4224, <https://doi.org/10.1029/2011PA002130>.
- Laepple, T., and P. H. Huybers, 2014a: Global and regional variability in marine surface temperatures. *Geophys. Res. Lett.*, **41**, 2528–2534, <https://doi.org/10.1002/2014GL059345>.
- , and —, 2014b: Ocean surface temperature variability: Large model–data differences at decadal and longer periods. *Proc. Natl. Acad. Sci. USA*, **111**, 16 682–16 687, <https://doi.org/10.1073/pnas.1412077111>.
- Lamarque, J.-F., and Coauthors, 2010: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos. Chem. Phys.*, **10**, 7017–7039, <https://doi.org/10.5194/acp-10-7017-2010>.
- Lamb, H. H., 1965: The early Medieval warm epoch and its sequel. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **1**, 13–37, [https://doi.org/10.1016/0031-0182\(65\)90004-0](https://doi.org/10.1016/0031-0182(65)90004-0).
- Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, and H. Teng, 2013: Last millennium climate and its variability in CCSM4. *J. Climate*, **26**, 1085–1111, <https://doi.org/10.1175/JCLI-D-11-00326.1>.
- Lara, A., and R. Villalba, 1993: A 3620-year temperature record from *Fitzroya Cupressoides* tree rings in southern South America. *Science*, **260**, 1104–1106, <https://doi.org/10.1126/science.260.5111.1104>.
- Le, T., J. Sjolte, and R. Muscheler, 2016: The influence of external forcing on subdecadal variability of regional surface temperature in CMIP5 simulations of the last millennium. *J. Geophys. Res.*, **121**, 1671–1682, <https://doi.org/10.1002/2015JD024423>.
- Lean, J. L., 2018: Estimating solar irradiance since 850 CE. *Earth Space Sci.*, **5**, 133–149, <https://doi.org/10.1002/2017EA000357>.
- Leduc, G., R. Schneider, J.-H. Kim, and G. Lohmann, 2010: Holocene and Eemian sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry. *Quat. Sci. Rev.*, **29**, 989–1004, <https://doi.org/10.1016/j.quascirev.2010.01.004>.
- Lehner, F., F. Joos, C. C. Raible, J. Mignot, A. Born, K. M. Keller, and T. F. Stocker, 2015: Climate and carbon cycle dynamics in a CESM simulation from 850–2100 CE. *Earth Syst. Dyn.*, **6**, 411–434, <https://doi.org/10.5194/esd-6-411-2015>.
- Lohmann, G., M. Pfeiffer, T. Laepple, G. Leduc, and J.-H. Kim, 2013: A model–data comparison of the Holocene global sea surface temperature evolution. *Climate Past*, **9**, 1807–1839, <https://doi.org/10.5194/cp-9-1807-2013>.
- Li, B., and J. Smerdon, 2012: Defining spatial comparison metrics for evaluation of paleoclimatic field reconstructions of the Common Era. *Environmetrics*, **23**, 394–406, <https://doi.org/10.1002/env.2142>.
- Liu, Y., and Coauthors, 2009: Annual temperatures during the last 2485 years in the Eastern Tibetan Plateau inferred from tree rings. *Sci. China D*, **52**, 348–359, <https://doi.org/10.1007/s11430-009-0025-z>.
- Liu, Z., and Coauthors, 2014: The Holocene temperature conundrum. *Proc. Natl. Acad. Sci. USA*, **111**, E3501–E3505, <https://doi.org/10.1073/pnas.1407229111>.
- Ljungqvist, F. C., P. J. Krusic, G. Brattström, and H. S. Sundqvist, 2012: Northern Hemisphere temperature patterns in the last 12 centuries. *Climate Past*, **8**, 227–249, <https://doi.org/10.5194/cp-8-227-2012>.
- , —, H. S. Sundqvist, E. Zorita, G. Brattström, and D. Frank, 2016: Northern Hemisphere hydroclimate variability over the past twelve centuries. *Nature*, **532**, 94–98, <https://doi.org/10.1038/nature17418>.
- Loso, M. G., 2009: Summer temperatures during the Medieval Warm Period and Little Ice Age inferred from varved proglacial lake sediments in southern Alaska. *J. Paleolimnol.*, **41**, 117–128, <https://doi.org/10.1007/s10933-008-9264-9>.
- , R. S. Anderson, S. P. Anderson, and P. J. Reimer, 2006: A 1500-year record of temperature and glacial response inferred from varved Iceberg Lake, southcentral Alaska. *Quat. Res.*, **66**, 12–24, <https://doi.org/10.1016/j.yqres.2005.11.007>.
- Lovejoy, S., and C. Varotsos, 2016: Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings. *Earth Syst. Dyn.*, **7**, 133–150, <https://doi.org/10.5194/esd-7-133-2016>.
- , D. Schertzer, and D. Varon, 2013: Do GCM's predict the climate... or macroweather? *Earth Syst. Dyn.*, **4**, 439–454, <https://doi.org/10.5194/esd-4-439-2013>.

- Luckman, B. H., and R. J. S. Wilson, 2005: Summer temperatures in the Canadian Rockies during the last millennium: A revised record. *Climate Dyn.*, **24**, 131–144, <https://doi.org/10.1007/s00382-004-0511-0>.
- Luoto, T. P., and S. Helama, 2010: Palaeoclimatological and palaeolimnological records from fossil midges and tree-rings: The role of the North Atlantic Oscillation in eastern Finland through the Medieval Climate Anomaly and Little Ice Age. *Quat. Sci. Rev.*, **29**, 2411–2423, <https://doi.org/10.1016/j.quascirev.2010.06.015>.
- Luterbacher, J., and Coauthors, 2016: European summer temperatures since Roman times. *Environ. Res. Lett.*, **11**, 024001, <https://doi.org/10.1088/1748-9326/11/2/024001>.
- MacFarling Meure, C., D. Etheridge, C. Trudinger, P. Steele, R. Langenfelds, T. van Ommen, A. Smith, and J. Elkins, 2006: Law dome CO₂, CH₄ and N₂O ice core records extended to 2000 years BP. *Geophys. Res. Lett.*, **33**, L14810, <https://doi.org/10.1029/2006GL026152>.
- Machida, T., T. Nakazawa, Y. Fujii, S. Aoki, and O. Watanabe, 1995: Increase in the atmospheric nitrous oxide concentration during the last 250 years. *Geophys. Res. Lett.*, **22**, 2921–2924, <https://doi.org/10.1029/95GL02822>.
- Madec, G., 2008: NEMO ocean engine. Note du Pôle de Modélisation de l'Institut Pierre-Simon Laplace 27, 209 pp.
- Mangini, A., C. Spötl, and P. Verdes, 2005: Reconstruction of temperature in the Central Alps during the past 2000 yr from a $\delta^{18}\text{O}$ stalagmite record. *Earth Planet. Sci. Lett.*, **235**, 741–751, <https://doi.org/10.1016/j.epsl.2005.05.010>.
- Martín-Chivelet, J., M. B. Muñoz-García, R. L. Edwards, M. J. Turrero, and A. I. Ortega, 2011: Land surface temperature changes in Northern Iberia since 4000 yr BP, based on $\delta^{13}\text{C}$ of speleothems. *Global Planet. Change*, **77**, 1–12, <https://doi.org/10.1016/j.gloplacha.2011.02.002>.
- Masson-Delmotte, V., and Coauthors, 2013: Information from paleoclimate archives. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 383–464.
- McCabe-Glynn, S., K. R. Johnson, C. Strong, M. Berkelhammer, A. Sinha, H. Cheng, and R. L. Edwards, 2013: Variable North Pacific influence on drought in southwestern North America since AD 854. *Nat. Geosci.*, **6**, 617–621, <https://doi.org/10.1038/ngeo1862>.
- McGlone, M. S., C. S. M. Turney, J. M. Wilmshurst, and K. Pahnke, 2010: Divergent trends in land and ocean temperature in the Southern Ocean over the past 18,000 years. *Nat. Geosci.*, **3**, 622–626, <https://doi.org/10.1038/ngeo931>.
- McGregor, H. V., M. Dima, H. W. Fischer, and S. Mülitzal, 2007: Rapid 20th-century increase in coastal upwelling off northwest Africa. *Science*, **315**, 637–639, <https://doi.org/10.1126/science.1134839>.
- McKay, N. P., and D. S. Kaufman, 2014: An extended Arctic proxy temperature database for the past 2,000 years. *Sci. Data*, **1**, 140026, <https://doi.org/10.1038/sdata.2014.26>.
- , —, and N. Michelutti, 2008: Biogenic silica concentration as a high-resolution, quantitative temperature proxy at Hallet Lake, south-central Alaska. *Geophys. Res. Lett.*, **35**, L05709, <https://doi.org/10.1029/2007GL032876>.
- Melvin, T. M., H. Grudd, and K. R. Briffa, 2013: Potential bias in “updating” tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Torneträsk density and ring-width data. *Holocene*, **23**, 364–373, <https://doi.org/10.1177/0959683612460791>.
- Miettinen, A., D. Divine, N. Koç, F. Godtliessen, and I. R. Hall, 2012: Multicentennial variability of the sea surface temperature gradient across the subpolar North Atlantic over the last 2.8 kyr. *J. Climate*, **25**, 4205–4219, <https://doi.org/10.1175/JCLI-D-11-00581.1>.
- , —, K. Husum, N. Koç, and A. E. Jennings, 2015: Exceptional ocean surface conditions on the SE Greenland shelf during the Medieval Climate Anomaly. *Paleoceanography*, **30**, 1657–1674, <https://doi.org/10.1002/2015PA002849>.
- Miller, G. H., and Coauthors, 2012: Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks. *Geophys. Res. Lett.*, **39**, L02708, <https://doi.org/10.1029/2011GL050168>.
- Millet, L., F. Arnaud, O. Heiri, M. Magny, V. Verneaux, and M. Desmet, 2009: Late-Holocene summer temperature reconstruction from chironomid assemblages of Lake Anterne, northern French Alps. *Holocene*, **19**, 317–328, <https://doi.org/10.1177/0959683608100576>.
- Mitchell, D. M., and Coauthors, 2015: Solar signals in CMIP-5 simulations: The stratospheric pathway. *Quart. J. Roy. Meteor. Soc.*, **141**, 2390–2403, <https://doi.org/10.1002/qj.2530>.
- Moberg, A., 2013: Comparisons of simulated and observed Northern Hemisphere temperature variations during the past millennium – Selected lessons learned and problems encountered. *Tellus*, **65B**, 19921, <https://doi.org/10.3402/tellusb.v65i0.19921>.
- , and G. Brattström, 2011: Prediction intervals for climate reconstructions with autocorrelated noise—An analysis of ordinary least squares and measurement error methods. *Paleogeogr. Palaeoclimatol. Palaeoecol.*, **308**, 313–329, <https://doi.org/10.1016/j.palaeo.2011.05.035>.
- , R. Sundberg, H. Grudd, and A. Hind, 2015: Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium – Part 3: Practical considerations, relaxed assumptions, and using tree-ring data to address the amplitude of solar forcing. *Climate Past*, **11**, 425–448, <https://doi.org/10.5194/cp-11-425-2015>.
- Moffa-Sánchez, P., I. R. Hall, S. Barker, D. J. R. Thornalley, and I. Yashayaev, 2014: Surface changes in the eastern Labrador Sea around the onset of the Little Ice Age. *Paleoceanography*, **29**, 160–175, <https://doi.org/10.1002/2013PA002523>.
- Mohtadi, M., O. E. Romero, J. Kaiser, and D. Hebbeln, 2007: Cooling of the southern high latitudes during the Medieval Period and its effect on ENSO. *Quat. Sci. Rev.*, **26**, 1055–1066, <https://doi.org/10.1016/j.quascirev.2006.12.008>.
- Moore, J. J., K. A. Hughen, G. H. Miller, and J. T. Overpeck, 2001: Little Ice Age recorded in summer temperature reconstruction from varved sediments of Donard Lake, Baffin Island, Canada. *J. Paleolimnol.*, **25**, 503–517, <https://doi.org/10.1023/A:1011181301514>.
- Mulvaney, R., and Coauthors, 2012: Recent Antarctic Peninsula warming relative to Holocene climate and ice-shelf history. *Nature*, **489**, 141–144, <https://doi.org/10.1038/nature11391>.
- Munz, P. M., and Coauthors, 2015: Decadal-resolution record of winter monsoon intensity over the last two millennia from planktic foraminiferal assemblages in the northeastern Arabian Sea. *Holocene*, **25**, 1756–1771, <https://doi.org/10.1177/0959683615591357>.
- Muschitiello, F., Q. Zhang, H. S. Sundqvist, F. J. Davies, and H. Renssen, 2015: Arctic climate response to the termination of the African Humid Period. *Quat. Sci. Rev.*, **125**, 91–97, <https://doi.org/10.1016/j.quascirev.2015.08.012>.
- Naurzbaev, M. M., E. A. Vaganov, O. V. Sidorova, and F. H. Schweingruber, 2002: Summer temperatures in eastern Taimyr inferred from a 2427-year late-Holocene tree-ring

- chronology and earlier floating series. *Holocene*, **12**, 727–736, <https://doi.org/10.1191/0959683602hl586rp>.
- Nemenyi, P., 1963: Distribution-free multiple comparisons. Ph.D. dissertation, Princeton University, 254 pp.
- Neukom, R., A. P. Schurer, N. J. Steiger, and G. C. Hegerl, 2018: Possible causes of data model discrepancy in the temperature history of the last millennium. *Sci. Rep.*, **8**, 7572, <https://doi.org/10.1038/s41598-018-25862-2>.
- Newton, A., R. Thunell, and L. Stott, 2011: Changes in the Indonesian Throughflow during the past 2000 yr. *Geology*, **39**, 63–66, <https://doi.org/10.1130/G31421.1>.
- Nieto-Moreno, V., F. Martínez-Ruiz, S. Giral, D. Gallego-Torres, J. García-Orellana, P. Masqué, and M. Ortega-Huertas, 2013: Climate imprints during the ‘Medieval Climate Anomaly’ and the ‘Little Ice Age’ in marine records from the Alboran Sea basin. *Holocene*, **23**, 1227–1237, <https://doi.org/10.1177/0959683613484613>.
- Nowack, P. J., N. L. Abraham, P. Braesicke, and J. A. Pyle, 2018: The impact of stratospheric ozone feedbacks on climate sensitivity estimates. *J. Geophys. Res. Atmos.*, **123**, 4630–4641, <https://doi.org/10.1002/2017JD027943>.
- Nyberg, J., B. A. Malmgren, A. Kuijpers, and A. Winter, 2002: A centennial-scale variability of tropical North Atlantic surface hydrography during the late Holocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **183**, 25–41, [https://doi.org/10.1016/S0031-0182\(01\)00446-1](https://doi.org/10.1016/S0031-0182(01)00446-1).
- Oppo, D. W., Y. Rosenthal, and B. K. Linsley, 2009: 2,000-year-long temperature and hydrology reconstructions from the Indo-Pacific warm pool. *Nature*, **460**, 1113–1116, <https://doi.org/10.1038/nature08233>.
- Otto-Bliesner, B. L., N. Rosenbloom, E. J. Stone, N. P. McKay, D. J. Lunt, E. C. Brady, and J. T. Overpeck, 2013: How warm was the last interglacial? New model–data comparisons. *Philos. Trans. Roy. Soc. London*, **371A**, 20130097, <https://doi.org/10.1098/rsta.2013.0097>.
- Otto-Bliesner, B. L., and Coauthors, 2016: Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **97**, 735–754, <https://doi.org/10.1175/BAMS-D-14-00233.1>.
- PAGES2k-PMIP3 Group, 2015: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Climate Past*, **11**, 1673–1699, <https://doi.org/10.5194/cp-11-1673-2015>.
- PAGES 2k Consortium, 2013: Continental-scale temperature variability during the last two millennia. *Nat. Geosci.*, **6**, 339–346, <https://doi.org/10.1038/ngeo1797>.
- , 2017: A global multiproxy database for temperature reconstructions of the Common Era. *Sci. Data*, **4**, 170088, <https://doi.org/10.1038/sdata.2017.88>.
- PAGES Hydro2k Consortium, 2017: Comparing proxy and model estimates of hydroclimate variability and change over the Common Era. *Climate Past*, **13**, 1851–1900, <https://doi.org/10.5194/cp-13-1851-2017>.
- Pausata, F. S. R., G. Messori, and Q. Zhang, 2016: Impacts of dust reduction on the northward expansion of the African monsoon during the Green Sahara period. *Earth Planet. Sci. Lett.*, **434**, 298–307, <https://doi.org/10.1016/j.epsl.2015.11.049>.
- Pei, Q., D. D. Zhang, J. Li, and H. F. Lee, 2017: Proxy-based Northern Hemisphere temperature reconstruction for the mid-to-late Holocene. *Theor. Appl. Climatol.*, **130**, 1043–1053, <https://doi.org/10.1007/s00704-016-1932-5>.
- Petit, J. R., and Coauthors, 1999: Climate and atmospheric history of the past 420,000 years from the Vostok Ice Core, Antarctica. *Nature*, **399**, 429–436, <https://doi.org/10.1038/20859>.
- Phipps, S. J., and Coauthors, 2013: Paleoclimate data–model comparison and the role of climate forcings over the past 1500 years. *J. Climate*, **26**, 6915–6936, <https://doi.org/10.1175/JCLI-D-12-00108.1>.
- , L. D. Rotstayn, H. B. Gordon, J. L. Roberts, A. C. Hirst, and W. F. Budd, 2011: The CSIRO Mk3L climate system model version 1.0—Part 1.0: Description and evaluation. *Geosci. Model Dev.*, **4**, 483–509, <https://doi.org/10.5194/gmd-4-483-2011>.
- , —, —, —, —, and —, 2012: The CSIRO Mk3L climate system model version 1.0—Part 2: Response to external forcings. *Geosci. Model Dev.*, **5**, 649–682, <https://doi.org/10.5194/gmd-5-649-2012>.
- Pla, S., and J. Catalan, 2005: Chrysophyte cysts from lake sediments reveal the submillennial winter/spring climate variability in the northwestern Mediterranean region throughout the Holocene. *Climate Dyn.*, **24**, 263–278, <https://doi.org/10.1007/s00382-004-0482-1>.
- Pollack, H. N., and J. E. Smerdon, 2004: Borehole climate reconstructions: Spatial structure and hemispheric averages. *J. Geophys. Res.*, **109**, D11106, <https://doi.org/10.1029/2003JD004163>.
- Pongratz, J., C. H. Reick, T. Raddatz, and M. Claussen, 2008: A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochem. Cycles*, **22**, GB3018, <https://doi.org/10.1029/2007GB003153>.
- , —, —, and —, 2009: Effects of anthropogenic land cover change on the carbon cycle of the last millennium. *Global Biogeochem. Cycles*, **23**, GB4001, <https://doi.org/10.1029/2009GB003488>.
- Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dyn.*, **16**, 123–146, <https://doi.org/10.1007/s003820050009>.
- Ran, L., H. Jiang, K. L. Knudsen, and J. Eiríksson, 2011: Diatom-based reconstruction of palaeoceanographic changes on the North Icelandic shelf during the last millennium. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **302**, 109–119, <https://doi.org/10.1016/j.palaeo.2010.02.001>.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Richey, J. N., R. Z. Poore, B. P. Flower, and T. M. Quinn, 2007: 1400 yr multi-proxy record of climate variability from the northern Gulf of Mexico. *Geology*, **35**, 423–426, <https://doi.org/10.1130/G23507A.1>.
- , D. J. Hollander, B. P. Flower, and T. I. Eglinton, 2011: Merging late Holocene molecular organic and foraminiferal-based geochemical records of sea surface temperature in the Gulf of Mexico. *Paleoceanography*, **26**, PA1209, <https://doi.org/10.1029/2010PA002000>.
- Richter, T. O., F. J. C. Peeters, and T. C. E. van Weering, 2009: Late Holocene (0–2.4 ka BP) surface water temperature and salinity variability, Feni Drift, NE Atlantic Ocean. *Quat. Sci. Rev.*, **28**, 1941–1955, <https://doi.org/10.1016/j.quascirev.2009.04.008>.
- Rosell-Melé, A., and Coauthors, 2001: Precision of the current methods to measure the alkenone proxy U_{37}^K and absolute alkenone abundance in sediments: Results of an interlaboratory comparison study. *Geochem. Geophys. Geosyst.*, **2**, 1046, <https://doi.org/10.1029/2000GC000141>.

- Rosenthal, Y., B. K. Linsley, and D. W. Oppo, 2013: Pacific Ocean heat content during the past 10,000 years. *Science*, **342**, 617–621, <https://doi.org/10.1126/science.1240837>.
- , J. Kalansky, A. Morley, and B. Linsley, 2017: A paleo-perspective on ocean heat content: Lessons from the Holocene and Common Era. *Quat. Sci. Rev.*, **155**, 1–12, <https://doi.org/10.1016/j.quascirev.2016.10.017>.
- Rotstayn, L. D., S. J. Jeffrey, M. A. Collier, S. M. Dravitzki, A. C. Hirst, J. I. Syktus, and K. K. Wong, 2012: Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations. *Atmos. Chem. Phys.*, **12**, 6377–6404, <https://doi.org/10.5194/acp-12-6377-2012>.
- Salzer, M. W., and K. F. Kipfmüller, 2005: Reconstructed temperature and precipitation on a millennial timescale from tree-rings in the Southern Colorado Plateau, U.S.A. *Climatic Change*, **70**, 465–487, <https://doi.org/10.1007/s10584-005-5922-3>.
- Schmidt, R., and Coauthors, 2006: Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *J. Climate*, **19**, 153–192, <https://doi.org/10.1175/JCLI3612.1>.
- , 2010: Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change. *J. Quat. Sci.*, **25**, 79–87, <https://doi.org/10.1002/jqs.1314>.
- , and Coauthors, 2011: 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>.
- , 2012: Interactive comment on “Constraining the temperature history of the past millennium using early instrumental observations” by P. Brohan et al. *Climate Past Discuss.*, **8**, C393–C398.
- , and Coauthors, 2012: Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1). *Geosci. Model Dev.*, **5**, 185–191, <https://doi.org/10.5194/gmd-5-185-2012>.
- , and Coauthors, 2014a: Using palaeo-climate comparisons to constrain future projections in CMIP5. *Climate Past*, **10**, 221–250, <https://doi.org/10.5194/cp-10-221-2014>.
- , and Coauthors, 2014b: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, **6**, 141–184, <https://doi.org/10.1002/2013MS000265>.
- Schurer, A. P., G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps, 2013: Separating forced from chaotic climate variability over the past millennium. *J. Climate*, **26**, 6954–6973, <https://doi.org/10.1175/JCLI-D-12-00826.1>.
- , S. F. B. Tett, and G. C. Hegerl, 2014: Small influences of solar variability on climate over the past millennium. *Nat. Geosci.*, **7**, 104–108, <https://doi.org/10.1038/ngeo2040>.
- Sepúlveda, J., S. Pantoja, K. A. Huguen, S. Bertrand, D. Figueroa, T. Leon, N. J. Drenzek, and C. Lange, 2009: Late Holocene sea-surface temperature and precipitation variability in northern Patagonia, Chile (Jacaf Fjord, 44°S). *Quat. Res.*, **72**, 400–409, <https://doi.org/10.1016/j.yqres.2009.06.010>.
- Shapiro, A. I., W. Schmutz, E. Rozanov, M. Schoell, M. Haberleiter, A. V. Shapiro, and S. Nyeki, 2011: A new approach to long-term reconstruction of the solar irradiance leads to large historical solar forcing. *Astron. Astrophys.*, **529**, A67, <https://doi.org/10.1051/0004-6361/201016173>.
- Shevenell, A., A. Ingalls, E. Domack, and C. Kelly, 2011: Holocene Southern Ocean surface temperature variability west of the Antarctic Peninsula. *Nature*, **470**, 250–254, <https://doi.org/10.1038/nature09751>.
- Sicre, M.-A., and Coauthors, 2011: Sea surface temperature variability in the subpolar Atlantic over the last two millennia. *Paleoceanography*, **26**, PA4218, <https://doi.org/10.1029/2011PA002169>.
- , K. Weckström, and M.-S. Seidenkrantz, 2014: Labrador Current variability over the last 2000 years. *Earth Planet. Sci. Lett.*, **400**, 26–32, <https://doi.org/10.1016/j.epsl.2014.05.016>.
- , B. Jalali, B. Martrat, S. Schmidt, M.-A. Bassetti, and N. Kallel, 2016: Sea surface temperature variability in the North Western Mediterranean Sea (Gulf of Lion) during the Common Era. *Earth Planet. Sci. Lett.*, **456**, 124–133, <https://doi.org/10.1016/j.epsl.2016.09.032>.
- Sigl, M., and Coauthors, 2015: Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, **523**, 543–549, <https://doi.org/10.1038/nature14565>.
- Smerdon, J. E., 2012: Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *Wiley Interdiscip. Rev.: Climate Change*, **3**, 63–77, <https://doi.org/10.1002/wcc.149>.
- , and H. N. Pollack, 2016: Reconstructing Earth’s surface temperature over the past 2000 years: the science behind the headlines. *Wiley Interdiscip. Rev. Climate Change*, **7**, 746–771, <https://doi.org/10.1002/wcc.418>.
- , S. Coats, and T. R. Ault, 2016: Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era. *Climate Dyn.*, **46**, 1921–1942, <https://doi.org/10.1007/s00382-015-2684-0>.
- Solomina, O., and K. Alverson, 2004: High latitude Eurasian palaeoenvironments: Introduction and synthesis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, **209**, 1–18, <https://doi.org/10.1016/j.palaeo.2004.02.027>.
- Spielhagen, R. F., and Coauthors, 2011: Enhanced modern heat transfer to the Arctic by warm Atlantic water. *Science*, **331**, 450–453, <https://doi.org/10.1126/science.1197397>.
- Steiger, N. J., G. J. Hakim, E. J. Steig, D. S. Battisti, and G. H. Roe, 2014: Assimilation of time-averaged pseudoproxies for climate reconstruction. *J. Climate*, **27**, 426–441, <https://doi.org/10.1175/JCLI-D-12-00693.1>.
- , J. E. Smerdon, E. R. Cook, and B. I. Cook, 2018: A reconstruction of global hydroclimate and dynamical variables over the Common Era. *Sci. Data*, **5**, 180086, <https://doi.org/10.1038/sdata.2018.86>.
- Steinhilber, F., J. Beer, and C. Frohlich, 2009: Total solar irradiance during the Holocene. *Geophys. Res. Lett.*, **36**, L19704, <https://doi.org/10.1029/2009GL040142>.
- , and Coauthors, 2012: 9,400 years of cosmic radiation and solar activity from ice cores and tree rings. *Proc. Natl. Acad. Sci. USA*, **109**, 5967–5971, <https://doi.org/10.1073/pnas.1118965109>.
- Stenni, B., and Coauthors, 2017: Antarctic climate variability on regional and continental scales over the last 2000 years. *Climate Past*, **13**, 1609–1634, <https://doi.org/10.5194/cp-13-1609-2017>.
- Stern, D. I., and R. K. Kaufmann, 2014: Anthropogenic and natural causes of climate change. *Climatic Change*, **122**, 257–269, <https://doi.org/10.1007/s10584-013-1007-x>.
- Stott, L., K. Cannariato, R. Thunell, G. H. Haug, A. Koutavas, and S. Lund, 2004: Decline of surface temperature and salinity in the western tropical Pacific Ocean in the Holocene epoch. *Nature*, **431**, 56–59, <https://doi.org/10.1038/nature02903>.
- Sueyoshi, T., and Coauthors, 2013: Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM. *Geosci. Model Dev.*, **6**, 819–836, <https://doi.org/10.5194/gmd-6-819-2013>.
- Sümeği, P., G. Jakab, P. Majkut, T. Töröcsik, and Cs. Zatykó, 2009: Middle Age palaeoecological and palaeoclimatological reconstruction in the Carpathian Sea. *Idojaras*, **113**, 265–298.

- Sun, A. Z., and Z. D. Feng, 2013: Holocene climatic reconstructions from the fossil pollen record at Qigai Nuur in the southern Mongolian Plateau. *Holocene*, **23**, 1391–1402, <https://doi.org/10.1177/0959683613489581>.
- Tan, M., T. Liu, J. Hou, X. Qin, H. Zhang, and T. Li, 2003: Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature. *Geophys. Res. Lett.*, **30**, 1617, <https://doi.org/10.1029/2003GL017352>.
- Tardif, R., and Coauthors, 2018: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate Past Discuss.*, <https://doi.org/10.5194/cp-2018-120>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Thompson, L. G., E. Mosley-Thompson, M. E. Davis, V. S. Zagorodnov, I. M. Howat, V. N. Mikhailenko, and P.-N. Lin, 2013: Annually resolved ice core records of tropical climate variability over the past ~1800 years. *Science*, **340**, 945–950, <https://doi.org/10.1126/science.1234210>.
- Tierney, J. E., M. T. Mayes, N. Meyer, C. Johnson, P. W. Swarzenski, A. S. Cohen, and J. M. Russell, 2010: Late-twentieth-century warming in Lake Tanganyika unprecedented since AD 500. *Nat. Geosci.*, **3**, 422–425, <https://doi.org/10.1038/ngeo865>.
- , C. C. Ummenhofer, and P. B. deMenocal, 2015: Past and future rainfall in the Horn of Africa. *Sci. Adv.*, **1**, e1500682, <https://doi.org/10.1126/sciadv.1500682>.
- Tingley, M. P., P. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam, 2012: Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.*, **35**, 1–22, <https://doi.org/10.1016/j.quascirev.2012.01.012>.
- , —, —, —, —, and —, 2015: Discriminating between GCM forcing configurations using Bayesian reconstructions of Late-Holocene temperatures. *J. Climate*, **28**, 8264–8281, <https://doi.org/10.1175/JCLI-D-15-0208.1>.
- Treydte, K. S., D. C. Frank, M. Saurer, G. Helle, G. H. Schleser, and J. Esper, 2009: Impact of climate and CO₂ on a millennium-long tree-ring carbon isotope record. *Geochim. Cosmochim. Acta*, **73**, 4635–4647, <https://doi.org/10.1016/j.gca.2009.05.057>.
- Valdes, P., 2011: Built for stability. *Nat. Geosci.*, **4**, 414–416, <https://doi.org/10.1038/ngeo1200>.
- van Engelen, A. F. V., J. Buisman, and F. Ijnsen, 2001: A millennium of weather, winds and water in the Low Countries. *History and Climate: Memories of the Future?* P. D. Jones, Eds., Kluwer Academic, 101–124.
- Vieira, L. E. A., and S. K. Solanki, 2010: Evolution of the solar magnetic flux on time scales of years to millennia. *Astron. Astrophys.*, **509**, A100, <https://doi.org/10.1051/0004-6361/200913276>.
- , —, N. A. Krivova, and I. Usoskin, 2011: Evolution of the solar irradiance during the Holocene. *Astron. Astrophys.*, **531**, A6, <https://doi.org/10.1051/0004-6361/201015843>.
- Vinther, B. M., P. D. Jones, K. R. Briffa, H. B. Clausen, K. K. Andersen, D. Dahl-Jensen, and S. J. Johnsen, 2010: Climatic signals in multiple highly resolved stable isotope records from Greenland. *Quat. Sci. Rev.*, **29**, 522–538, <https://doi.org/10.1016/j.quascirev.2009.11.002>.
- von Gunten, L., W. J. D'Andrea, R. S. Bradley, and Y. Huang, 2012: Proxy-to-proxy calibration: Increasing the temporal resolution of quantitative climate reconstructions. *Sci. Rep.*, **2**, 609, <https://doi.org/10.1038/srep00609>.
- von Luxburg, U., and V. H. Franz, 2009: A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations and bootstrap. *Stat. Sin.*, **19**, 1095–1117.
- von Rad, U., H. Schulz, V. Riech, M. den Dulk, U. Berner, and F. Sirocko, 1999: Multiple monsoon-controlled breakdown of oxygen-minimum conditions during the past 30,000 years documented in laminated sediments off Pakistan. *Palaeogeogr. Climat. Ecol.*, **152**, 129–161, [https://doi.org/10.1016/S0031-0182\(99\)00042-5](https://doi.org/10.1016/S0031-0182(99)00042-5).
- von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett, 2004: Reconstructing past climate from noisy data. *Science*, **306**, 679–682, <https://doi.org/10.1126/science.1096109>.
- Wang, Y. M., J. L. Lean, and N. R. Sheeley, 2005: Modeling the sun's magnetic field and irradiance since 1713. *Astrophys. J.*, **625**, 522–538, <https://doi.org/10.1086/429689>.
- Wang, J., J. Emile-Geay, D. Guillot, J. Smerdon, and B. Rajaratnam, 2014: Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Climate Past*, **10**, 1–19, <https://doi.org/10.5194/cp-10-1-2014>.
- , —, —, N. P. McKay, and B. Rajaratnam, 2015: Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation. *Geophys. Res. Lett.*, **42**, 7162–7170, <https://doi.org/10.1002/2015GL065265>.
- Wang, J.-L., B. Yang, F. C. Ljungqvist, J. Luterbacher, T. J. Osborn, K. R. Briffa, and E. Zorita, 2017: Internal and external forcing of multidecadal Atlantic climate variability over the past 1,200 years. *Nat. Geosci.*, **10**, 512–517, <https://doi.org/10.1038/ngeo2962>.
- Wang, J., B. Yang, T. J. Osborn, F. C. Ljungqvist, H. Zhang, and J. Luterbacher, 2018: Causes of East Asian temperature multidecadal variability since 850 CE. *Geophys. Res. Lett.*, **45**, 13 485–13 494, <https://doi.org/10.1029/2018GL080725>.
- Wei, W. W. S., 1989: *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley, 478 pp.
- Widmann, M., H. Goosse, G. van der Schrier, R. Schnur, and J. Barkmeijer, 2010: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate Past*, **6**, 627–644, <https://doi.org/10.5194/cp-6-627-2010>.
- Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biom. Bull.*, **1**, 80–83, <https://doi.org/10.2307/3001968>.
- Wiles, G. C., R. D. D'Arrigo, D. Barclay, R. Wilson, S. K. Jarvis, L. Vargo, and D. Frank, 2014: Surface air temperature variability for the Gulf of Alaska over the past 1200 years. *Holocene*, **24**, 198–208, <https://doi.org/10.1177/0959683613516815>.
- Wu, W., W. Tan, L. Zhou, H. Yang, and Y. Xu, 2012: Sea surface temperature variability in the southern Okinawa Trough during last 2700 years. *Geophys. Res. Lett.*, **39**, L14705, <https://doi.org/10.1029/2012GL052749>.
- Wurtzel, J. B., D. E. Black, R. C. Thunell, L. C. Peterson, E. J. Tappa, and S. Rahman, 2013: Mechanisms of southern Caribbean SST variability over the last two millennia. *Geophys. Res. Lett.*, **40**, 5954–5958, <https://doi.org/10.1002/2013GL058458>.
- Xin, X., T. Wu, and J. Zhang, 2013: Introduction of CMIP5 experiments carried out with the climate system models of Beijing Climate Center (in Chinese). *Adv. Clim. Chang. Res.*, **4**, 41–49, <https://doi.org/10.3724/SP.J.1248.2013.00041>.
- Yadav, R. R., A. Braeuning, and J. Singh, 2011: Tree ring inferred summer temperature variations over the last millennium in western Himalaya, India. *Climate Dyn.*, **36**, 1545–1554, <https://doi.org/10.1007/s00382-009-0719-0>.
- Yan, Q., Z. Zhang, H. Wang, and D. Jiang, 2015: Simulated warm periods of climate over China during the last two millennia:

- The Sui-Tang warm period versus the Song-Yuan warm period. *J. Geophys. Res. Atmos.*, **120**, 2229–2241, <https://doi.org/10.1002/2014JD022941>.
- Zanchettin, D., and Coauthors, 2016: The Model Intercomparison Project on the climatic response to volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6. *Geosci. Model Dev.*, **9**, 2701–2719, <https://doi.org/10.5194/gmd-9-2701-2016>.
- Zhang, H., and Coauthors, 2018: East Asian warm season temperature variations over the past two millennia. *Nat. Commun.*, **8**, 7702.
- Zhang, Q., H. S. Sundqvist, A. Moberg, H. Körnich, J. Nilsson, and K. Holmgren, 2010: Climate change between the mid and late Holocene in northern high latitudes—Part 2: Model–data comparisons. *Climate Past*, **6**, 609–626, <https://doi.org/10.5194/cp-6-609-2010>.
- Zhang, Y., X. M. Shao, Z.-Y. Yin, and Y. Wang, 2014: Millennial minimum temperature variations in the Qilian Mountains, China: Evidence from tree rings. *Climate Past*, **10**, 1763–1778, <https://doi.org/10.5194/cp-10-1763-2014>.
- Zhang, P., H. W. Linderholm, B. E. Gunnarson, J. Björklund, and D. Chen, 2016: 1200 years of warm-season temperature variability in central Scandinavia inferred from tree-ring density. *Climate Past*, **12**, 1297–1312, <https://doi.org/10.5194/cp-12-1297-2016>.
- Zhu, X., J. Bye, K. Fraedrich, and I. Bordi, 2016: Statistical structure of intrinsic climate variability under global warming. *J. Climate*, **29**, 5935–5947, <https://doi.org/10.1175/JCLI-D-15-0505.1>.
- Zorita, E., and Coauthors, 2010: European temperature records of the past five centuries based on documentary/instrumental information compared to climate simulations. *Climatic Change*, **101**, 143–168, <https://doi.org/10.1007/s10584-010-9824-7>.