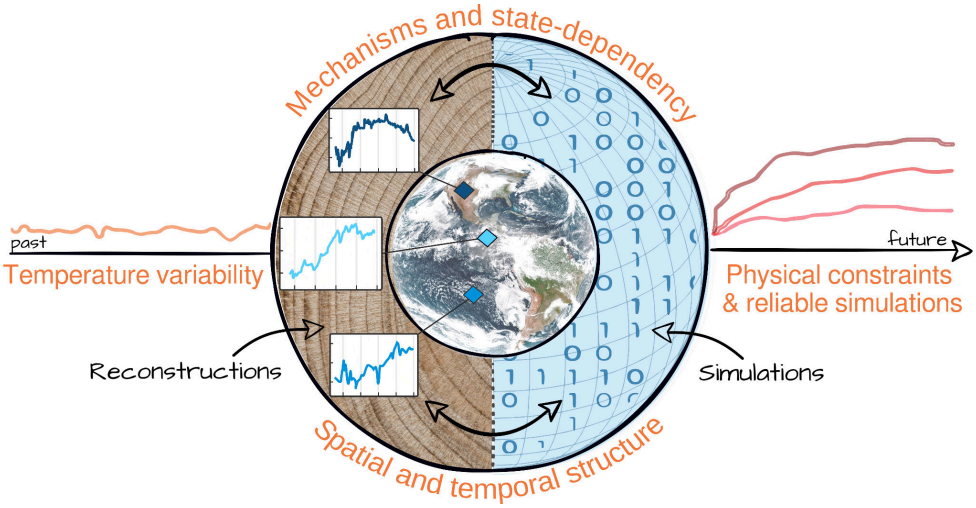
Dissertation

Structure and Causes of Variability in Earth's Surface Air Temperature at the Local to Global Level



Beatrice Ellerhoff

Structure and Causes of Variability in Earth's Surface Air Temperature at the Local to Global Level

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Beatrice Marie Ellerhoff geb. Latz
aus Saarbrücken

Tübingen
2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 19.05.2023

Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin:	Prof. Dr. Kira Rehfeld
2. Berichterstatter:	Prof. Dr. Michel Crucifix

Acknowledgments

I wish to express my profound thanks to Kira Rehfeld for the opportunity to pursue a Ph.D. in her group. Kira, you encouraged me to get involved in climate research and always supported me to develop ideas and realize projects in your group. Thank you for your invaluable advice as a supervisor and our countless scientific discussions.

This journey would have been a different one without the SPACY people. One of my favorite jazz pianists, Herbie Hancock, said: "It is people's hearts that move the age." I am more than grateful for the group's refreshing, motivating, and enriching approach to the daily work and the biggest questions of our time. I wish SPACY to always keep this spirit. Thank you for the fantastic times we spent together and the friendships we made.

Special thanks to Elisa Ziegler, Janica Bühler, Maybritt Schillinger, Markus Maisch, Mathurin Choblet, Moritz Adam, and Nils Weitzel for their proofreading and comments on this thesis. Moreover, I thank Stephan Sprinz for performing numerical experiments with the ClimBayes software package for the analysis of economic damages.

I also thank Uwe Mikolajewicz and Marie-Luise Kapsch for permission to use the latest MPI-ESM simulation outputs and for the opportunity to discuss the data with them.

I am grateful to my co-supervisor Thomas Gasenzer for many inspiring discussions throughout the past years.

I thank the Heinrich-Böll-Stiftung, the Heidelberg Graduate School for Physics and the PalMod project for the financial support of this thesis. I am grateful for discussions within CVAS, a working group of the Past Global Changes (PAGES) project.

Words cannot express my gratitude to my friends and family, especially Tim, Mum, Simon, Grandma, and Grandpa for all the years of loving support. I dedicate this work to the memory of my father, who remains the greatest source of inspiration to me.

Table of Contents

List of abbreviations

AMOC	Atlantic Meridional Overturning Circulation
AOD	Aerosol Optical Depth
AOGCM	Atmosphere-Ocean General Circulation Model
BP	Before Present (1950 CE)
CMIP	Coupled Model Intercomparison Project
CE	Common Era
DO	Dansgaard Oeschger
EBM	Energy Balance Model
ECS	Equilibrium Climate Sensitivity
EMIC	Earth System Model of Intermediate Complexity
ENSO	El Niño–Southern Oscillation
ESM	Earth System Model
GMST	Global Mean Surface Temperature
HadCM	Hadley Centre Coupled Model
IAM	Integrated Assessment Model
kyrs	kiloyears
K	Kelvin
LGM	Last Glacial Maximum
MPI-ESM	Max Planck Institute for Meteorology Earth System Model
OU	Ornstein Uhlenbeck
PMIP	Paleoclimate Modelling Intercomparison Project
PSD	Power Spectral Density
SSP	Shared Socioeconomic Pathway
yrs	years

List of Figures

The figures are own work if not stated otherwise.

Figure 2 a, 3, 5, and 8 (same as the one on the thesis cover) use icons and pictures under free-to-use copyrights from Unsplash (https://unsplash.com), Wikimedia Commons (https://creativecommons.org/licenses/by-sa/4.0), and Flaticon (https://flaticon.com).

List of Tables

1 Summary

Fluctuations in Earth's surface air temperature determine the climatic conditions for life on our planet and influence, among other factors, the occurrence of extreme climatic events. Therefore, a precise understanding of this variability in temperature is paramount to society, politics, and the economy. However, many characteristics and impacts of temperature variability are still uncertain, especially at the local spatial scale and on decadal to multicentennial timescales. Moreover, the relative contribution of climate system-inherent and external drivers of variability needs to be better constrained. Finally, how temperature variability changes with the mean climate state and, thus, under current and future anthropogenic warming remains to be clarified. These uncertainties affect long-term planning, for example, concerning mitigation and adaptation strategies.

In three publications, this dissertation examines the statistical properties of temperature variability as a function of the underlying spatiotemporal scales, external drivers, and mean climate state. Spectral methods are combined with time series analysis, conceptual modeling, and Bayesian inference to quantify temperature variability from climate model simulations and paleoclimate records. The results confirm overall confidence in the simulated global temperature variability. Climate models, however, respond more strongly to external forcing and show fewer internal fluctuations at decadal scales than paleoclimate reconstructions of global temperature. At the local level, there are significant data-model mismatches in temperature variance and correlation properties over decadal and longer timescales. Improved representation of natural forcing in climate model simulations can partially offset these differences. By integrating the response to volcanic eruptions, sea ice dynamics plays an essential role in amplifying local multidecadal variability. Decreasing sea ice extent attenuates local variability under warming, especially in high latitudes. Overall, global variability tends to be dominated by external forcing, while local variability is primarily caused by state-dependent internal variations and remains subject to substantial uncertainties.

This work extends these findings with complementary investigations. A new estimate of the power spectral density for global temperature beyond the last millennia complements the analyses. The comparison of transient climate model experiments covering the last 27 thousand years using prescribed or interactively coupled ice sheets reveals benefits from dynamic simulation of ice sheet feedbacks for representing millennial-scale variability. Further studies analyze the potential impacts of uncertainties in the simulated variability on projections of extreme climate events, attribution studies, and risk assessments. Missing interannual to millennial temperature variability in simulations could lead to underestimating future climate impacts.

Limitations of the presented analyses arise from uncertainties of the data and the assumption that weakly-stationary stochastic processes sufficiently describe the considered temperature time series. Since this work focuses on past and present climates, follow-up studies are needed to test the robustness of the findings in future projections, particularly against the influence of non-linear processes.

Altogether, the insights gained on the timescale-, forcing-, and state-dependent statistics of local to global temperatures open new possibilities for improving the analysis and understanding of temperature variations, as well as the ability of models to simulate climate variability. Despite remaining uncertainties in model simulations and reconstructions, the reliable simulation of future local temperature variability requires improved representations of ice sheet feedbacks and natural forcing in climate model experiments. This will contribute to a reliable assessment of future climate-related risks and help inform long-term planning, mitigation, and adaptation efforts.

Kurzfassung

Die Schwankungen der Oberflächenlufttemperatur bestimmen die klimatischen Bedingungen für das Leben auf unserem Planeten. Durch diese Temperaturvariabilität werden etwa extreme Klimaereignisse beeinflusst. Temperaturvariabilität zu verstehen ist daher für Gesellschaft, Politik und Wirtschaft unerlässlich. Besonders auf lokaler Ebene und über Zeitskalen von Jahren bis Jahrtausenden ist Temperaturvariabilität noch unzureichend verstanden. Es ist notwendig, Diskrepanzen zwischen simulierter und rekonstruierter Variabilität zu klären. Zudem gilt es die Anteile jener Fluktuationen, die im Klimasystem intern generiert oder extern angetrieben werden, zu bestimmen. Noch ist zu wenig über den Einfluss der durch den Menschen verursachten Erwärmung auf die Klimavariabilität bekannt. Diese Unsicherheiten erschweren die langfristige sozioökonomische Planung und Entwicklung von Mitigations- und Adaptationsstrategien.

Diese Dissertation untersucht in drei Publikationen die Temperaturvariabilität unter Berücksichtigung der räumlichen und zeitlichen Skalen, der natürlichen Antriebe und des mittleren Klimazustands. Zeitreihenanalysen, konzeptionelle Modellierung und Bayes'sche Verfahren quantifizieren die Variabilität aus Paläoklimarekonstruktionen und Klimamodellsimulationen. Die Ergebnisse zeigen, dass Klimamodelle die globale Temperaturvariabilität zuverlässig simulieren. Auf dekadischen Zeitskalen sind geringe Abweichungen zu finden, da Modelle stärker auf Vulkanausbrüche reagieren. Auf lokaler Ebene und Zeitskalen länger als zehn Jahren zeigen sich signifikante Unterschiede in der Varianz und den Korrelationseigenschaften simulierter und rekonstruierter Temperaturen. Die Einbeziehung natürlicher Einflüsse auf die Strahlungsbilanz in Modellen kann diese Diskrepanz verringern. Meereis ist dabei wichtig, um die Reaktion des Oberflächenklimas auf Vulkanausbrüche zu verstärken. In wärmeren Klimazuständen führen geringere Meereis-Rückkopplungseffekte zu weniger Variabilität. Zusammenfassend wird die globale Temperatur hauptsächlich durch externe Antriebe bestimmt, während lokal interne zustandsabhängige Schwankungen dominieren und noch große Unsicherheiten auftreten.

Diese Arbeit ergänzt die Erkenntnisse der vorgestellten Publikationen durch weitere Untersuchungen. Dabei wird die Temperaturvariabilität aus Klimamodellsimulationen der letzten 27 tausend Jahre mit interaktiv gekoppelten Eisschilden analysiert. Die dynamische Simulation von Rückkopplungen in Reaktion auf Schmelzwasser aus Eisschilden ermöglicht eine bessere Darstellung der Variabilität. Allerdings reichen diese Verbesserungen nicht aus, um die Diskrepanz zwischen simulierten und rekonstruierten lokalen Temperaturen vollständig aufzuheben. Limitierungen der simulierten Variabilität kann sich auf die statistische Vorhersage extremer Klimaereignisse sowie Attributionsstudien auswirken. Des Weiteren sollte die Abschätzung ökonomischer Schäden durch Klimaveränderungen mögliche Variabilitätseffekte berücksichtigen.

Limitierungen entstehen primär durch Datenunsicherheiten und die Annahme, dass schwachstationäre stochastische Prozesse die Temperaturveränderungen hinreichend beschreiben. Da diese Arbeit den Fokus auf das vergangene und gegenwärtige Klima legt, sollten Folgestudien die vorgestellten Methoden und Ergebnisse in Simulationen zukünftiger Szenarien testen.

Die gewonnenen Erkenntnisse eröffnen neue Möglichkeiten, um das Verständnis von Klimaschwankungen zu verbessern und die Fähigkeiten von Modellen zur Simulation von Temperaturvariabilität zu testen. Ungeachtet verbleibender Unsicherheiten erfordern Projektionen lokaler Temperaturen eine verbesserte Darstellung von Eisschild-Rückkopplungseffekten und natürlichen Antrieben. Die dargelegten Ergebnisse werden daher zu einer zuverlässigen Simulation künftiger Klimarisiken beitragen und langfristige Mitigations- und Anpassungsmaßnahmen wissenschaftlich unterstützen.

2 List of publications

P1: B. Ellerhoff and K. Rehfeld. "Probing the Timescale Dependency of Local and Global Variations in Surface Air Temperature from Climate Simulations and Reconstructions of the Last Millennia". *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 104.6 (2021). doi: 10.1103/PhysRevE.104.064136.

P2: B. Ellerhoff, M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime, and K. Rehfeld. "Contrasting State-Dependent Effects of Natural Forcing on Global and Local Climate Variability". *Geophysical Research Letters* 49.10 (2022). doi: 10.1029/2022GL098335.

P3: M. Schillinger, B. Ellerhoff, R. Scheichl, and K. Rehfeld. "Separating internal and externally-forced contributions to global temperature variability using a Bayesian stochastic energy balance framework." *Chaos* 32, 113146. (2022). doi: 10.1063/5.0106123

The above publications Ellerhoff & Rehfeld (2021), Ellerhoff et al. (2022), and Schillinger et al. (2022a) are hereafter referred to as **P1, P2**, and **P3**, respectively. The publications and their supplementary materials are printed in the Appendix of this thesis (App. F-App. H).

3 Data and code availability

Data and code to reproduce the analysis of **P1-P3** are available at `https://github.com/paleovar`. The developed software packages and snapshots of the code have been published under Zenodo identifiers. The specific links are given in the data availability statements of **P1-P3**. In addition, data and code to reproduce the figures of this thesis' introduction, discussion, and appendix are published under restricted access at Zenodo (`https://zenodo.org/record/7543184`). The repository is confidential and will be made openly available upon publication of this dissertation.

4 Declaration of author contributions

Here, I declare my contributions and those of my co-authors to **P1–P3**. The authors are abbreviated by the first letter of their first and last names.

P1: B. Ellerhoff and K. Rehfeld (2021)

I carried out the analysis, created the figures, and wrote the paper, supervised by KR. Together, we designed the study and finalized the revisions.

P2: B. Ellerhoff, M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime, and K. Rehfeld (2022)

I carried out the analysis, created the figures, and wrote the paper, supervised by KR. I built on previous research and code developed by MK during his Master's thesis. KR performed the simulations with the HadCM3 climate model, with help from MH and LS. Based on the initial clean-up by MK and KR, I completed the post-processing of the runs and prepared them for publication. I ran the TransEBM simulations with help from EZ, who implemented the TransEBM model (Ziegler & Rehfeld, 2021). All authors contributed to revisions of the manuscript. MK, KR, and I designed the study.

P3: M. Schillinger, B. Ellerhoff, R. Scheichl, and K. Rehfeld (2022a)

MS and I contributed equally to this work. We carried out the analysis, created the figures, and wrote the paper, supervised by KR and RS. MS led the development of the ClimBayes package (Schillinger et al., 2022b). I led the variability analysis based on spectral methods. The project was based on MS's master thesis, supervised by RS, KR, and myself. All authors designed the study and contributed to revisions of the manuscript.

5 Introduction

Throughout human history, the desire to understand nature and its laws has led to countless breakthroughs in physics and related disciplines. Observations of the Earth's dynamic system have repeatedly given new impetus to climate research, prompting the field to find explanations for phenomena as diverse as glacial cycles, the greenhouse effect, and ozone depletion. Today, human impacts on the climate and environment threaten life on Earth. Among others, the occurrence of extreme weather and climate events is noticeably changing (Fischer & Knutti, 2015). Unabated global warming is affecting human health, food security, and entire ecosystems (IPCC, 2022b). Consequently, climate research is driven by the need to estimate and constrain possible future climate pathways based on information about past and present changes. At the same time, the time window to act and protect human well-being and our planet's health is rapidly closing (IPCC, 2021b).

Climate model simulations of global mean temperature have quantified and attributed anthropogenic warming (IPCC, 2021b; Manabe & Wetherald, 1967), paving the way for effective mitigation strategies. Besides changes in the mean, climate variability also comprises changes in the variance and higher-order statistics (IPCC, 2022a). Predicting air temperature fluctuations at the Earth's surface, here temperature variability, is at least equally important as predictions of the mean (Katz & Brown, 1992). In particular, temperature variability influences the climatic conditions at the local level, including the probability of droughts and extreme heat (Schär et al., 2004). Therefore, risk assessment and adaptation strategies depend on accurate information about local temperature variability.

Klaus Hasselmann, honored with the 2021 Nobel Prize in Physics, was one of the first to recognize variability as a prominent feature of global temperature. He attributed parts of the variability to random effects from high-frequency ("weather") components of the Earth system (Hasselmann, 1976). His work enabled the unambiguous separation of natural forced and unforced variations from anthropogenic effects by so-called "fingerprinting" approaches (Hegerl et al., 1996). Despite these and other milestones in climate science, the range of future climate variability is still subject to uncertainties (IPCC, 2021b). More knowledge about temperature variability is required, especially regarding local temperatures at timescales of a few years to millennia (Franzke et al., 2020).

Figure 1 illustrates the "knowledge gap" on temperature variability by showing three profoundly different concepts of temperature variability across scales (Huybers & Curry, 2006; Lovejoy, 2015; Mitchell, 1976). The concepts correspond to analytical and data-based estimates of the power spectral density (PSD) $S(\tau)$. The PSD quantifies the contributions of fluctuations to the total variance as a function of the underlying timescale τ (details in Section 5.2.2). The first estimate of the PSD, or "spectrum", of global temperature was a detailed and pioneering sketch by Mitchell (1976) (Fig. 1 a). His seminal paper shaped the early understanding of this spectrum as the composite of multiple (pseudo-)oscillatory variations, giving rise to spectral peaks.

In contrast to Mitchell's "foreground" view, Huybers & Curry (2006) drew attention to the "background continuum" of the spectrum (Fig. 1 b). Parts of the fitted spectral continuum follow a power law

$$S(\tau) \sim \tau^{\beta} ,$$

with spectral exponent β and scaling interval $\tau \in [\tau_1, \tau_2]$. The temporal scaling suggests that the temperature fluctuations can be described by the same stochastic process and, thus, statistical properties within a scaling interval. The concept of temporal scaling stimulated numerous studies (Fraedrich et al., 2004; Fredriksen & Rypdal, 2017; Nilsen et al., 2016), however, its underlying mechanisms are still debated (Franzke et al., 2020).

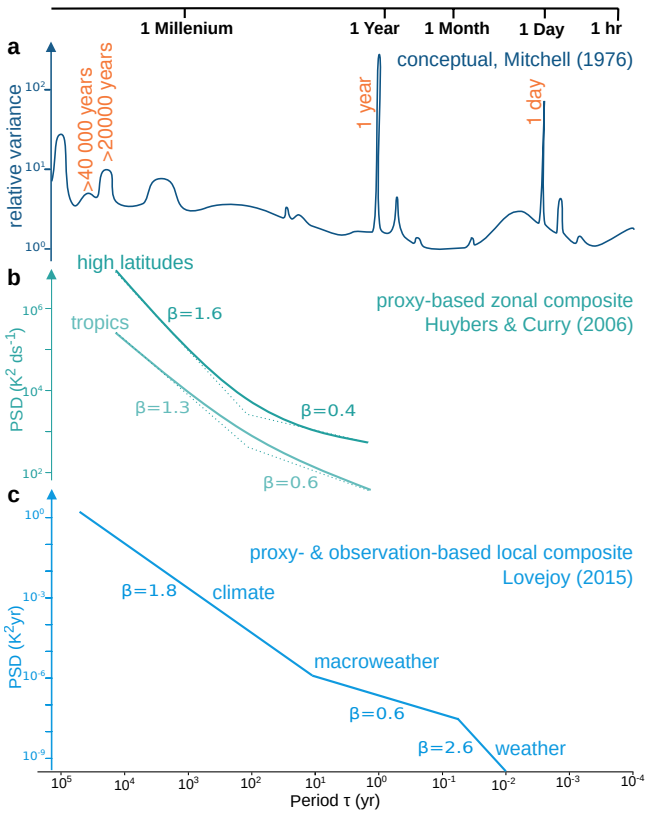


Figure 1: Conceptual illustrations of estimated power spectral densities (PSD) for surface air temperature and scaling coefficients β prior to this dissertation thesis. **a** Sketched temperature spectrum with highlighted peaks at the Milankovitch, annual, and diurnal cycles (Mitchell, 1976). **b** Fitted PSD and power-law scaling β to composite spectra from tropical and high latitude temperature data (Huybers & Curry, 2006). **c** PSD and scaling laws fitted to a spectral composite of temperature data from different locations (Lovejoy, 2015). The so-called weather, macroweather, and climate regimes (Lovejoy, 2015) are highlighted. All lines were redrawn from the original figures.

One reason for the remaining uncertainties is unequivocal evidence for scale breaks. The latter are periods at which the scaling behavior changes (Nilsen et al., 2016). Huybers & Curry (2006) suggested a change in scaling behavior at the centennial scale for tropical and high-latitude zonal mean spectra based on paleoclimate and instrumental records (Fig. 1 b). However, this selection of data supporting spectral estimates is limited to a few records. The PSD estimate by Lovejoy (2015) proposes different regimes with distinct temporal scaling behavior (Fig. 1 c), but the records seem unrelated because they refer to different locations. The use of different methods for quantifying spectral estimates, such as relative magnitudes (Fig. 1 a) or normalization techniques (Fig. 1 b), complicate the comparison between spectral estimates and scaling behavior (Hébert et al., 2021).

The guiding aim of this thesis is to improve the understanding of the structure and causes of temperature variability by presenting and analyzing new estimates of the temperature spectrum. To this end, this work quantifies and contrasts the *timescale dependency* of temperature fluctuations across scales from paleoclimate records, observational data, and model simulations of varying complexity. To identify potential variability mechanisms, the presented studies also examine the *forcing dependency* of temperature variability, that is, the extent to which fluctuations are externally driven or internally generated. As the amplitude of temperature variations does not only vary with timescale but also with the mean climate (Rehfeld et al., 2018), the *state dependency* of forced and unforced temperature variability from the last glacial to the present day is assessed. The improved understanding of local to global temperature variability can help to identify the *mechanisms and potential impacts* of temperature variability and to test the ability of climate models to simulate temperature variability with sufficient accuracy.

This introductory chapter reviews the physics-based concepts relevant to the pursued research strategy. Section 5.1 defines the climate system, its components, and characteristic timescales. A brief history of Earth's surface temperature (Sec. 5.1.1) reviews paleoclimate reconstructions as a data source for past temperature variability. The notion of radiative equilibrium and feedback is

introduced to explain the state space of climate (Sec. 5.1.2) and emerging dynamical phenomena, such as abrupt transitions. Section 5.2 presents the methods used in this work. These include simple to comprehensive climate models (Sec. 5.2.1), spectral analysis (Sec. 5.2.2), and scale invariance (Sec. 5.2.3). Section 5.3 elaborates the existing knowledge gaps on temperature variability and motivates the guiding questions of this dissertation. All details about the numerical implementations, statistical framework, uncertainty analysis, and specific datasets are given in the data and methods parts of **P1-P3** (App. F - App. H).

5.1 Variability as an inherent property of Earth's dynamic system

Changes in Earth's climate result from the dynamics of climate subsystems, including the atmosphere, hydrosphere, cryosphere, lithosphere, and biosphere (IPCC, 2022a), as well as external forcing from natural and anthropogenic sources (Fig. 2 a). Natural forcing corresponds to orbital, solar, volcanic, and natural greenhouse gas perturbations to Earth's radiative balance. Anthropogenic forcing relates to human activities and mainly stems from land cover changes and greenhouse gas emissions. Both the external forcing and climate subsystems operate at distinct characteristic timescales (Fig. 2 b). The transfer of mass, energy, and momentum intimately connects these components. As a result, climate variability is closely bound to the physical properties of the climate system.

To first order, climate variability can be separated into external and internal variability. Here, "external" refers to the forced component of variability, whereas "internal" describes the intrinsic chaotic component of variability. Natural variability describes internal and external variations without anthropogenic influences. Statistical descriptions of climate variability range from monthly to geological periods on the order of millions of years (Fig. 2 b). In contrast to meteorological studies, climate variability typically neglects details of the instantaneous atmospheric state, including day-to-day and synoptic changes. Nevertheless, climate statistics capture these fluctuations in the mean and higher-order moments over many individual realizations of fast and chaotic weather dynamics. The wide range of timescales is linked to a multitude of spatial scales, from kilometer-wide weather systems to planetary waves spanning large parts of the Earth.

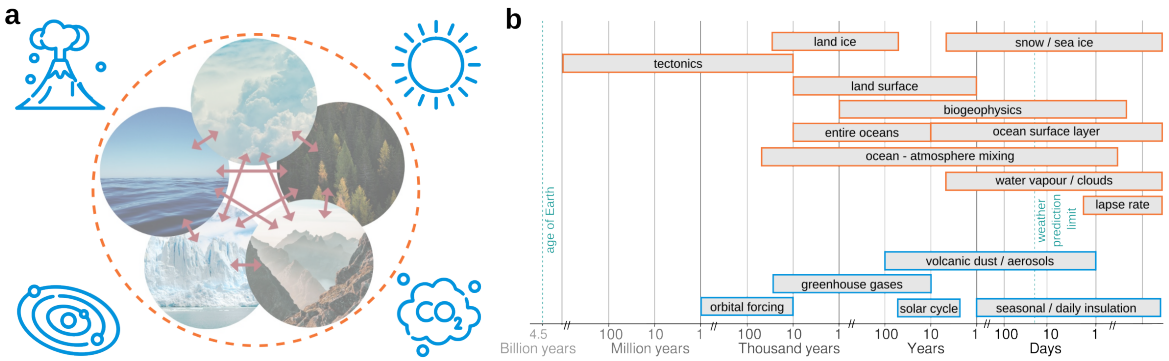


Figure 2: a Main components (hydrosphere, atmosphere, biosphere, lithosphere, and cryosphere) and drivers (volcanic aerosols, solar irradiance, greenhouse gas emissions, and orbital forcing) of Earth's climate system. **b** Selection of approximate characteristic timescales for internal variations (orange) and external forcing (blue) (PALAEOSENS Project Members, 2012; Peixoto & Oort, 1992; Rohling et al., 2018).

5.1.1 From paleoclimate reconstructions to future scenarios

Earth's temperature variations provide insights into our planet's history. Instrumental measurements of local surface and near-surface temperature have been available since about industrial times. The longest instrumental observation is the Hadley Centre Central England Temperature (HadCET), starting 1659 CE (Parker et al., 1992). Robust statistical estimates of the global temperature are avail-

able after approximately 1850 CE (Morice et al., 2021), whereby global temperature commonly refers to the globally-averaged surface air temperature.

Paleoclimate reconstructions

Paleoclimate research based on climate proxies is needed to resolve variability prior to the last centuries (IPCC, 2021b). Climate proxies preserve past physical or biogeochemical characteristics of particular, often seasonal, climate parameters such as summer temperatures (Evans et al., 2013). Examples are stable isotopes in ice cores and corals, growth rates in tree rings, pollen species in lake sediments, and foraminifera in ocean sediments. Their biological and chemical composition, growth or sedimentation rate relate to past climate conditions. Therefore, proxy records allow for reconstructions of past temperatures based on calibration methods. Nevertheless, there are fundamental challenges in separating climatic from non-climatic information (Lücke et al., 2019; Reschke et al., 2019) as well as externally forced from internally generated fluctuations. Proxy-based reconstructions are subject to numerous uncertainties in the agglomeration of the climatic signal, the transformation to temperature, and the statistical interpretation. There is a shift from analyzing individual proxies to synthesizing records in large data projects. These compilations have improved the availability, documentation, and comparability of paleoclimate reconstructions (Comas-Bru et al., 2020; Marcott et al., 2013) and analysis methods (Kaufman et al., 2020; PAGES2k Consortium, 2017). Moreover, combining multiple proxy records with statistical methods and climate field reconstructions allows for reliable hemispheric and global estimates (Kaufman et al., 2020; Lisiecki & Raymo, 2005; PAGES 2k Consortium, 2019; Snyder, 2016).

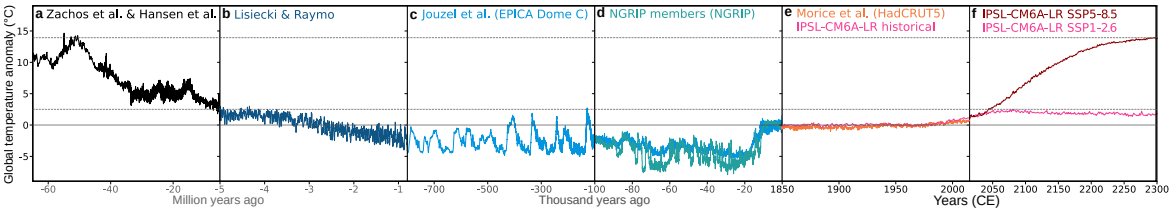


Figure 3: Patchwork of paleoclimate reconstructions (Hansen et al., 2013; Jouzel et al., 2007; Lisiecki & Raymo, 2005; North Greenland Ice Core Project members, 2004; Zachos et al., 2008), historical observations (Morice et al., 2021), and model simulation (Boucher et al., 2020) of the global temperature from 66 million years ago to 2300 CE. Anomalies are given with respect to 1960-1989 CE. Dashed lines correspond to the maximum global annual mean temperature of two scenarios, SSP1-2.6 and SSP5-8.5, as simulated by the IPSL model (Boucher et al., 2020). The figure was adapted from Wikimedia Commons (Fergus, 2014) using code from Elisa Ziegler (Ziegler, 2022).

Past to future temperatures on Earth

Paleoclimate reconstructions of global mean temperature present a wealth of variability phenomena over the last million years (Fig. 3). One important feature of past climate on Earth is the Paleocene–Eocene thermal maximum (PETM), an early warm period approximately 55 million years ago (McInerney & Wing, 2011) (Fig. 3 a). Back then, global temperature was more than 10 °C degrees higher than today. Some warming scenarios project similarly high temperatures within the next centuries (Boucher et al., 2020). Although life on Earth was quite different from today, research on the PETM can help understand the impacts of anthropogenic warming on flora and fauna (McInerney & Wing, 2011; Wing et al., 2005).

Other characteristics of past temperatures are pseudo-oscillatory switches between warm and cold periods, so-called glacial and interglacial states (Fig. 3 b and c). Climate research has greatly benefited from studying such glacial-interglacial cycles and their underlying physics (Crucifix, 2012; Jouzel et al., 2007; Shakun & Carlson, 2010). In his seminal work, Milankovitch (1941) attributed the changes between glacial and interglacial periods to seasonal and latitudinal-dependent changes in

solar insolation from slow periodic variations in Earth's orbit. However, there is still a debate about the role of internal mechanisms and external causes in the occurrence of these transitions (Shackleton, 2000; Willeit et al., 2019). While Milankovitch's theory explains the presence of 41 kiloyears (kyrs) long periods, different hypotheses exist for the change in cyclicity to periods of approximates 100 kyrs around 650 to 1250 kyrs ago, referred to as the "100,000-year problem" (Clark et al., 2006).

Moreover, the study of glacial-interglacial cycles has revealed sudden temperature changes (Fig. 3 d), primarily associated with so-called Dansgaard–Oeschger (DO) (Dansgaard et al., 1993) and Heinrich events (Heinrich, 1988). These involve self-amplifying non-linear processes and critical transitions in ocean circulations, such as the Atlantic Meridional Overturning Circulation (AMOC) (Henry et al., 2016). The most recent deglaciation, called "last deglaciation", describes the transition from the Last Glacial Maximum (LGM, approx. 25-17 kyrs ago) to the Holocene (11 750 before present (BP) to today). The latter is a particularly stable climate period, showing only small global temperature variations. The temperature increase from the LGM to the Holocene is of the same magnitude as that of warming scenarios by 2100 CE (IPCC, 2021b; Shakun & Carlson, 2010). Therefore, the comparison of variability in different mean climates (Rehfeld et al., 2018) can guide the validation of climate model simulations across climate states and improve the accuracy of projections (Rehfeld et al., 2020).

Human interventions in the climate are determining the near-term global temperature evolution (IPCC, 2022b). Figure 3 e depicts emission-driven simulations for two distinct future scenarios. The scenarios correspond to shared socioeconomic pathways (SSPs), providing greenhouse gas emissions based on projected socioeconomic changes and climate policies up to 2100 CE, and their extensions to 2300 CE. Reaching the Paris target to limit warming to +2 °C compared to pre-industrial times, preferably +1.5 °C, by 2100 CE is still achievable (IPCC, 2022b). The SSP1-1.9 and SSP1-2.6 scenarios are projected to yield a global temperature increase below +1.5 °C and +2 °C, respectively. Conversely, sustained high emission scenarios, such as SSP5-8.5, would drive the climate within a few decades into a range of temperatures unseen over the past million years (Fig. 3 e). While there is strong confidence in the global response to greenhouse gas emissions, uncertainties about future projections remain (Sherwood et al., 2020), especially on local to regional scales (IPCC, 2021b). These uncertainties affect the assessment of climate-related risks for society and ecosystems and, thus, require an improved understanding of local temperature variability.

5.1.2 From nature's laws to describing temperature variability

To characterize temperature fluctuations, it is essential to describe temperature as a quantity of the dynamical climate system and its governing equations. Depending on the physical systems, different mathematical formulations of temperature exist. Microscopically, for example, temperature can describe the mean translational kinetic energy of particles with velocity vector $\vec{v}$ following the Maxwell-Boltzmann law $T \propto \sqrt{\langle \vec{v} \cdot \vec{v} \rangle}$. Macroscopically, temperature is a state variable of a physical system that obeys the laws of thermodynamics. For example, the ideal gas law defines the temperature as a function of pressure p and volume V via $T \propto pV$, given that the total number of molecules is fixed.

Energy balance and equilibrium solutions

In climate research, temperature relates to nature's laws of energy conservation and thermodynamics. It is the key quantity to describe the Earth's climate system's physical state and statistical properties (Peixoto & Oort, 1992). A zero-dimensional, so-called Budyko-Sellers type energy balance model (EBMs) (Budyko, 1969; Sellers, 1969), serves as a simple approximation to the global temperature from

the balance between incoming and outgoing radiative energy

$$C\frac{dT}{dt} = R_{in} - R_{out}\,, \tag{1}$$

where T is the time-dependent temperature evolution. C is a positive constant corresponding to Earth's heat capacity. The incoming short-wave and outgoing long-wave radiation are functions of T, typically written as:

$$R_{in} = (1 - \alpha(T))\, S_0$$
$$R_{out} = \sigma\epsilon(T)T^4$$

The absorbed radiation R_{in} equals the incoming radiation $S_0 = Q_0/4$ from solar insolation Q_0 minus the reflected one $\alpha(T)S_0$ due to albedo effects $\alpha(T)$. The effective emissivity $\epsilon(T)$ describes the back radiation relative to that of a perfect black body radiator, and σ is the Stefan-Boltzmann constant. More (less) incoming than outgoing energy heats (cools) the Earth's surface.

To find steady solutions to the energy balance (1), it is possible to formulate a variational principle (Benzi et al., 1983; Ghil, 1976; North et al., 1979). Accordingly, the steady solutions of a differential equation $f(T)$ correspond to the extrema of a pseudo-potential V via

$$f(T) = -\frac{dV}{dT} \stackrel{!}{=} 0 \quad \text{and} \quad V(T) = -\int_{-\infty}^{\infty} f(T)\, dT\,.$$

Here, the differential equation reads $f(T) = C\frac{dT}{dt}$ and a steady solution $f(T_0) = 0$ gives rise to a so-called "fixed point" T_0. At such fixed points, the temperature evolution T can be approximated after time-dependent perturbation ΔT using a Taylor series

$$f(T) = f(T_0) + f'(T_0)\Delta T + \mathcal{O}((\Delta T^2)) = f'(T_0)\Delta T + \mathcal{O}((\Delta T^2))\,,$$

with $f'(T_0)$ denoting the time derivative of $f(T)$ at T_0. For sufficiently small fluctuations ΔT at the start of the perturbation, higher-order fluctuations $\mathcal{O}((\Delta T^2))$ are negligible. As a result, the approximate time-evolution for the perturbation ΔT reads

$$C\frac{d\Delta T}{dt} \approx f'(T_0)\Delta T.$$

The solution to this equation is an exponential function $\Delta T(t) = \Delta T(t_0)\exp(\lambda_f t)$ (Ghil, 1976) with initial perturbation $\Delta T(t_0)$ and feedback parameter $\lambda_f = f'(T_0)/C$. The sign of λ_f determines whether the perturbations grow or decay exponentially over a period of time, and relates to the concavity of the pseudo-potential $V(T)$ at T_0. The fixed point is linearly stable (unstable) for negative (positive) feedback. If $\lambda_f = 0$, the fixed point is marginally stable, and higher-order fluctuations must be considered.

Bifurcation and abrupt transitions

In case multiple fixed points exist in a system, transitions between them can occur, which is described in the mathematical framework of bifurcations. Considering the global temperature evolution as formulated in Equation (1), the number of possible states depends on the parametrization $\alpha(T)$ and $\epsilon(T)$. Typical parametrizations share the property that, as temperature increases, the negative black body feedback partially counterbalances the positive feedback from the planetary albedo (Budyko, 1969; Sellers, 1969). A corresponding potential $V(T)$ can be modeled by a double-well potential

$$V_\nu(T) = \frac{T^4}{4} - \frac{T^2}{2} - \nu T\,,$$

with control parameter ν (Fig. 4 a) (Boers et al., 2022; Ditlevsen & Ashwin, 2018; Ghil & Lucarini, 2020). Here, the control parameter ν describes the difference between R_{in} and R_{out} and, thus, the global temperature.

For control parameters $|\nu| \geq \nu_c$ above a critical value ν_c, the system has only one fixed point (Fig. 4 a, blue curve). For $-\nu_c < \nu < \nu_c$, the dynamical system exhibits two stable solutions (Fig. 4 a, green curve), described by the term "bistability". For example, the two stable states correspond to the existence of a cold ("deep-freeze") and warm, present-day climate state (Budyko, 1969; Sellers, 1969). The bifurcation diagram (Fig. 4 b) describes this change in the number of fixed points and visualizes its dependence on ν.

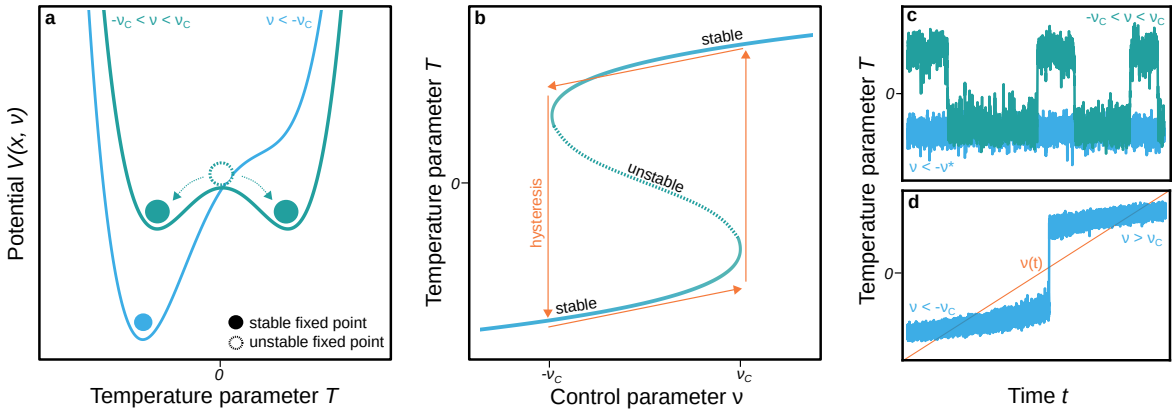


Figure 4: Potential function (**a**), bifurcation diagram (**b**), and time series examples (**c** and **d**) of a dynamical system with bistability as a function of the control parameter ν. The illustration and underlying mathematical description correspond to global temperature bistability (Boers et al., 2022; Ditlevsen & Ashwin, 2018; Ditlevsen & Johnsen, 2010; Ghil & Lucarini, 2020). Dashed lines indicate unstable fixed points and branches. Arrows in the bifurcation diagram (**b**) correspond to hysteresis jumps from one stable branch to the other. The change in the system state is numerically simulated under (**c**) a noise-induced transition, corresponding to the motion of a particle in the double-well potential, and (**d**) a bifurcation-induced transition using a time-dependent linear ramp $\nu(t)$.

For a certain value of ν, the system can abruptly jump from one to the other fixed point. This is because stable solutions can coexist for the same ν (Fig. 4 b). Stochastic noise, such as random internal fluctuations, can drive the transition from one to another fixed point (Fig. 4 c). Gradual changes in ν can also cause a transition (Fig. 4 d). After such a transition, only bifurcation or closing the hysteresis loop (Fig. 4 b) retains the initial state (Boers et al., 2022).

In addition to global temperature, many studies have found multistability for other climate parameters. In this context, the concept of so-called "tipping points" (Lenton et al., 2008; Lenton & Schellnhuber, 2007) has recently gained much attention (Armstrong McKay et al., 2022; Lenton et al., 2019; Wunderling et al., 2021). "Tipping" describes transition phenomena that are triggered above a critical threshold and different types of tipping have been classified (Ashwin et al., 2012). Its notion of an accelerated and irreversible change is controversial but stresses the risk that small fluctuations can cause a pronounced change in the climate system (Crucifix & Annan, 2019). Detecting and predicting tipping points might be complicated by incomplete knowledge of long-term variability and stochastic dynamics in climate subsystems. This highlights the role of precise knowledge of the structure and causes of variability in assessing climate-related risks.

Climate sensitivity

As a central quantity in climate research, the equilibrium climate sensitivity (ECS) quantifies the long-term response to external forcing. The term "equilibrium" refers to some equilibrated slower dynamics (von der Heydt et al., 2021), independent of fast fluctuations such as weather. Building on this timescale separation, the expected long-term change in global temperature ΔT after response to radiative forcing ΔR can be described by a general equilibrium sensitivity (Ghil & Lucarini,

2020)

$$S_E = \frac{\Delta T}{\Delta R}\,.$$

The ECS is then defined as the global warming ECS $= S_E\,\Delta R_{2x\mathrm{CO}_2}$ corresponding to the radiative perturbation $\Delta R_{2x\mathrm{CO}_2}$ of doubling the atmospheric CO_2 concentration (IPCC, 2022a) and after reaching a new equilibrium state. Current estimates suggest that ECS = 3 K with a likely range of 2.5 - 4 K (IPCC, 2021b).

Considering anomalies ΔT with respect to some equilibrium stat T_0, a new equilibrium $T = T_0 + \Delta T$ is reached when $R(\Delta T) = R_{out}(\Delta T) - R_{in}(\Delta T) = 0$ Wm^{-2} holds. Here, the radiative forcing $R(T_0) = 0$ Wm^{-2} is set to zero for simplicity. The change $\Delta R = -R(\Delta T)$ in response to external radiative forcing reads

$$\Delta R = -\frac{dR}{dT}\bigg|_{T=T_0} \Delta T + \mathcal{O}\left((\Delta T)^2\right)$$

Assuming that R_{in} stays constant, the reference sensitivity S_E for the radiative balance (1) is solely determined by the Planck feedback

$$S_E = -\frac{\partial R}{\partial T}\bigg|_{T=T_0} = 4\sigma\epsilon(T_0)T^3\,, \tag{2}$$

which defines the reference feedback parameter $\lambda_0(T_0) = 1/S_E$. However, R can depend on various other feedback processes $\chi_i(T)$, that contribute to the radiative adjustment

$$\frac{dR}{dT} = \frac{\partial R}{\partial T} + \sum_i \frac{\partial R}{\partial \chi_i}\frac{d\chi_i}{dT}\,.$$

Neglecting higher-order fluctuations, the change in temperature ΔT corresponding to a change in radiation ΔR reads

$$\Delta T \approx \frac{\lambda_0(T_0)}{1 - \sum_i f_i(T_0)}\Delta R \quad \text{with} \quad f_i(T_0) = \lambda_0(T_0)\frac{\partial R}{\partial \chi_i}\frac{d\chi_i}{dT}\bigg|_{T=T_0}\,.$$

The ECS found in observations and climate model simulations is typically higher than that expected from (2) due to feedbacks $\frac{\partial R}{\partial \chi_i}\frac{d\chi_i}{dT}$ and non-linear processes $\mathcal{O}\left((\Delta T)^2\right)$. Causes for the large inter-model spread in ECS estimates are uncertainties about the role of non-linear feedback processes (Ceppi et al., 2017; Knutti & Rugenstein, 2015; Lovejoy & Schertzer, 2012; Rypdal & Rypdal, 2016), forced (Forster, 2016), and internal variability (Huber et al., 2014; Olson et al., 2013). Thus, ECS estimates will directly benefit from studying forced versus unforced variability and potential feedback mechanisms contributing to climate variability.

5.2 Concepts of temperature variability research

Temperature data to study climate variability can be obtained from direct instrumental measurements, reanalysis, paleoclimate reconstructions, and simulations with climate models of varying complexity. These datasets vary in spatiotemporal resolution and are subject to inevitable uncertainties, requiring statistical and physics-based analysis. This chapter introduces climate models of various complexity and reviews the time series analysis and spectral methods used in this work.

5.2.1 Climate model simulations

Climate model simulations are crucial to project the future climate and aid in understanding climate observations and reconstruction. They represent the physical, chemical, and biological processes of the climate system, their interactions, and responses to external drivers (IPCC, 2022a) (Fig. 5). In

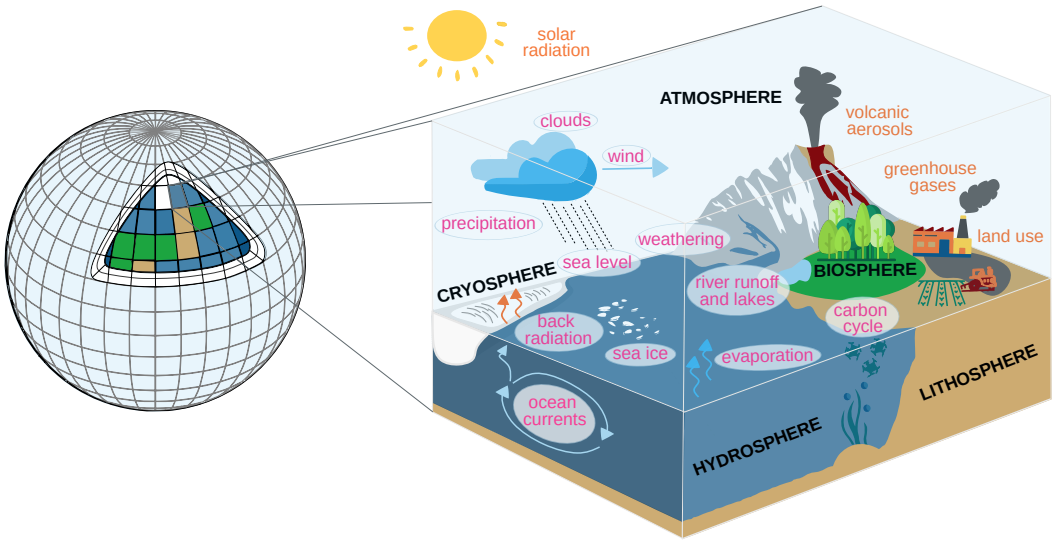


Figure 5: Simplified architecture of complex climate models, such as ESMs. The model represents the main components (black), drivers (orange), and processes (pink) by solving their governing equations on a three-dimensional grid of cells. The figure is inspired by the schematic of the UVic ESCM 2.10 climate model (Mengis et al., 2020).

addition, the models account for the conservation of momentum, mass, and energy. To this end, they employ physics-based laws such as the first law of thermodynamics, the Stefan-Boltzmann law, and the Navier-Stokes equations of fluid motion. The numerical framework of climate models enforces these constraints by differential equations. Physical or empirical laws parametrize sub-grid processes. Running model experiments corresponds to computing the discretized and approximate solutions to the implemented equations. As such, climate simulations depend on the choice of boundary and initial conditions. Boundary conditions include Earth system properties such as orography, land-sea mask, greenhouse gas concentrations, bathymetry, vegetation distribution, and orbital parameters. They can vary in time by interactive modeling or prescribed values. Initial conditions describe the initialized state of the climate system's components, such as the oceans and atmosphere. The initial state corresponds to an equilibrium reached after a spin-up run.

Models differ in their number of degrees of freedom, spatiotemporal resolution, and the extent to which processes are parametrized or directly simulated (Fig. 6). Earth system models (ESMs) are the highest, most comprehensive class of climate models. They increasingly include interactive chemistry and biology, allowing for enhanced representation of the carbon cycle, atmospheric chemistry, ice sheets, and dynamic vegetation (IPCC, 2022a). As such, they account for more interactions within the Earth system than their predecessors, so-called atmosphere-ocean general circulation models (AOGCMs). Earth system models of intermediate complexity (EMICs) resolve climate processes typically at a lower resolution and use more idealized representations (IPCC, 2022a) compared to AOGCMs.

Stochastic energy balance models

Comprehensive climate models require high programming efforts, are computationally expensive, and produce large amounts of data, especially for high-resolution and long simulations. Despite the complexity of the climate system, conceptual (also called "simple") energy balance models (EBMs) (Fig. 6) provide an easy-to-interpret and computationally cheap approach to model the climate's main characteristics. The simplicity of conceptual climate models is not a deficiency but helps explain global warming, climate dynamics, and the output of more complex models. For example, they laid the basis for attributing anthropogenic warming to greenhouse gas emission (Hegerl et al., 1996), investigating climate sensitivity (Ghil, 1984), and Glacial/Interglacial cycles (Dortmans et al., 2019;

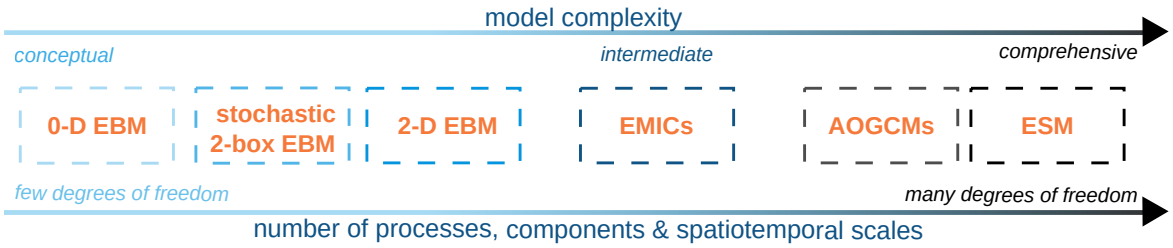


Figure 6: Conceptual ranking of the complexity of climate models used in this work. The model typology ranges from conceptual energy balance models (EBMs) and Earth system models of intermediate complexity (EMICs) to highly comprehensive atmosphere-ocean general circulation models (AOGCMs) and Earth system models (ESMs). The dashed lines indicate that the distinction between the individual models and their positioning on the scale of increasing degrees of freedom can overlap.

North et al., 1983). Simple climate models have also been used for future projections of global and regional temperatures (Hébert & Lovejoy, 2018; Myrvoll-Nilsen et al., 2020). In addition to analytical exercises, simple climate models can emulate the output of more complex model simulations (Meinshausen et al., 2011). Emulators "learn" another model's properties by fitting the free parameters of the emulator to the target data or by using machine learning methods. Because of their flexibility and efficiency, emulators can help explore a wide range of parameters and scenarios.

Assuming a linear and stationary temperature response to external forcing $F(t)$ to the energy balance (1), the global temperature evolution can be approximated by

$$C\frac{d}{dt}T(t) = -\lambda T(t) + F(t)\,,$$

with feedback parameter λ (Budyko, 1969; Sellers, 1969). Hasselmann (1976) extended this approach to a stochastic model, thereby initiating the field of stochastic climate modeling. Stochastic models are motivated by the fundamental question of how random long-term fluctuations in the climate systems can emerge from the chaotic short-term dynamics of the atmosphere. Hasselmann's seminal approach relies on the distinction between deterministic and random forcings:

$$C\frac{d}{dt}T(t) = -\lambda T(t) + F(t) + \xi(t) \tag{3}$$

The process $\xi(t) = \sigma_W \epsilon(t)$ with standard deviation σ_W is typically assumed as a white noise process $\epsilon(t)$ and gives rise to stochastic dynamics, motivated by random "weather forcing". The solution to Equation (3) is an Itô-integral over the Wiener process $W(s)$

$$T(t) = \int_{-\infty}^{t} R(t-s)\frac{1}{C}F(s)ds + \int_{-\infty}^{t} R(t-s)\frac{\sigma_W}{C}dW(s)\,, \tag{4}$$

with response function $R(t)$ (**P3**). The second term in Equation (4) corresponds to an Ornstein-Uhlenbeck (OU) process and explains long-term temperature variability as the integral response to short-term stochastic fluctuations (Hasselmann, 1976; Rypdal & Rypdal, 2014). In **P3** (Schillinger et al., 2022a), we combine a multibox extension of the Hasselmann EBM (4) with Bayesian inference to emulate the statistics of forced and unforced variability in state-of-the-art complex climate models.

5.2.2 Time series and spectral analysis

Understanding mechanisms and making predictions of climate variability requires time series analysis. A time series $x(t)$ represents a sequence of data points at consecutive points in time. To describe their statistical properties, time series are typically considered to be realizations of mathematical models. Since random fluctuations are ubiquitous in nature, stochastic processes are particularly

suited to describe time series of climate parameters (Chatfield, 2003). A stochastic process $X(t)$ is an ordered set of random variables. Here, the main statistical measures of stochastic processes are presented in the framework of real-valued processes that are continuous in time ($-\infty < t < \infty$, if not stated otherwise).

Autocovariance and spectral analysis of stochastic processes

The autocovariance gives the covariance of $X(t_1)$ and $X(t_2)$ at times t_1 and t_2

$$\gamma(t_1, t_2) = E\left[\left(X(t_1) - \mu(t_1)\right)\left(X(t_2) - \mu(t_2)\right)\right] ,$$

with mean $\mu(t) = E\left[X(t)\right]$. If the autocovariance and mean function are time-independent, the process is weakly stationary, and the autocovariance depends solely on the lag $h = t_2 - t_1$:

$$\gamma(h) = E\left[\left(X(t) - \mu\right)\left(X(t+h) - \mu\right)\right]$$

For zero-mean processes, $\gamma(h = 0) = E\left[X^2\right] = \sigma^2$ denotes the variance. Strict stationarity additionally requires the temporal independence of all higher-order moments.

Spectral analysis plays a key role in quantifying temperature fluctuations across timescales. The PSD quantifies the contribution of fluctuations, related to oscillations with certain frequencies, to the total variance. For a weakly-stationary stochastic process, the PSD is defined by the Fourier transform $\mathcal{F}$ of the autocovariance (Chatfield, 2003; Khintchine, 1934; Wiener, 1930)

$$S(\omega) = \mathcal{F}\{\gamma(h)\}(\omega) . \tag{5}$$

Practically, the PSD is often calculated from the Fourier transform of the process, that is $\hat{X}(\omega) = \mathcal{F}\{X(t)\}(\omega)$. To avoid convergence problems (Ditlevsen et al., 2020) (App. A), it is assumed that the stochastic process is ergodic and zero outside an interval $t \in [-T/2, T/2]$, here denoted by $X_T(t)$. Following the proof from Equation (14) in Appendix A, the PSD can be expressed by means of the squared Fourier transform

$$S(\omega) = \lim_{T\to\infty} \frac{1}{T} |\hat{X}_T(\omega)|^2 \tag{6}$$

in the limit of $T \to \infty$.

Moreover, the PSD relates to the variance $\gamma(0) = E[X^2]$ of the stochastic process via

$$E[X^2] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\, d\omega . \tag{7}$$

Equation (7) can be directly obtained from the inverse Fourier transform $\gamma(h) = \mathcal{F}^{-1}\{S(w)\}(h)$ by setting $h = 0$.

Name	$X(t)$	$\gamma(h)$	$S(\omega)$
White noise	$dX(t) = \sigma_W dW(t)$	$\sigma_W^2 \delta(h)$	σ_W^2
OU-process	$dX(t) = -\lambda X(t)dt + \sigma_{OU} dW(t)$	$\frac{\sigma_{OU}^2}{2\lambda} e^{-\lambda\lvert h\rvert}$	$\frac{\sigma_{OU}^2}{\lambda^2+\omega^2}$
Cosine	$A\cos(\omega_0 t + \theta)$	$\frac{A^2}{2}\cos(\omega_0 h)$	$\frac{A^2}{2}\pi\left(\delta(\omega_0 - \omega) + \delta(\omega_0 + \omega)\right)$

Table 1: Examples of time series $X(t)$, their autocovariance $\gamma(h)$, and power spectral density $S(\omega)$ for uncorrelated white noise, an Ornstein-Uhlenbeck (OU) process, and a cosine function. $W(t)$ is the Wiener process. Mathematical proofs in Appendix A.

Table 1 gives examples of the PSD of uncorrelated white noise, an OU-process, and a cosine function. Figure 7 exemplifies a numerically simulated OU-process and its spectral representations obtained

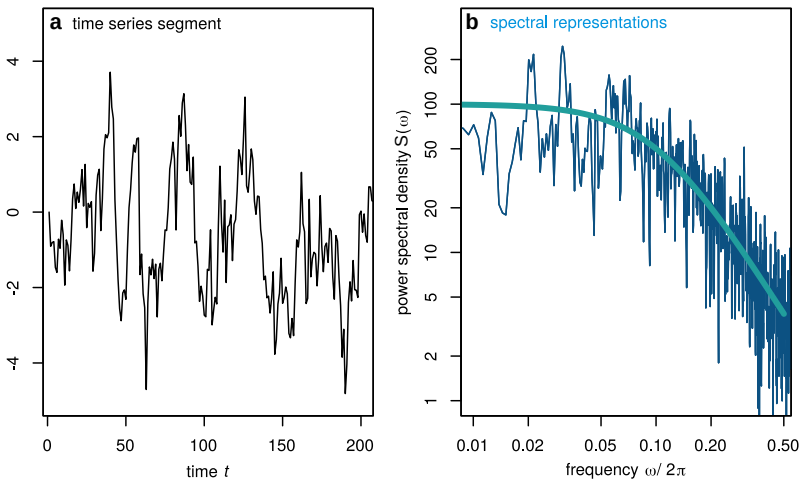


Figure 7: Example for the spectral representation of a time series, here numerically simulated for an OU-process with 10,000 timesteps, $\sigma = 1$, and $\lambda = 0.07$. **a** Time series segment of the full stochastic process. **b** Spectral estimates based on analytical (dark blue line) and numerical (cyan line) calculations.

from analytical (Tab. 1 and App. A) and numerical solutions. The numerical implementation of PSD calculations is based on Thomson's multitaper method (Percival & Walden, 1993) as described in **P1-P3**. In most parts of this thesis and **P1-P3**, the PSD is visualized over periods τ for illustrative purposes. The relations $S(\omega)\, d\omega = 2\pi S(f)\, df$ and $\tau = 1/f$ can be used to rewrite Equation (5) as a function of τ.

Linear time-invariant processes in time and frequency domain

In **P1**, spectral analysis is also used to analyze how external forcing drives fluctuations at specific frequencies. Linear response theory describes the relationship between the time and frequency responses of a time-variant linear system (Chatfield, 2003). It states that the response function $R(h)$ defines the output $Y(t)$ in response to the input $X(t)$ via

$$Y(t) = \int_{-\infty}^{\infty} R(h) X(t-h)\, dh\,.$$

The system is physically causal if $R(h) = 0$ for $h < 0$. The frequency response, or transfer, function results from the Fourier transform

$$\hat{H}(\omega) = \mathcal{F}\{R(h)\}(\omega)\,.$$

$\hat{H}(\omega)$ corresponds to a complex function $\hat{H}(\omega) = G(\omega)\mathrm{e}^{\mathrm{i}\phi(\omega)}$, where $G(\omega)$ and $\phi(\omega)$ are the real-valued gain and phase (Chatfield, 2003). Suppose that $X(t)$ and $Y(t)$ can be described as weakly-stationary processes, their PSDs relate via

$$S_Y(\omega) = |\hat{H}(\omega)|^2 S_X(\omega) = G^2(\omega) S_X(\omega). \tag{8}$$

P1 makes use of Equation (8) to compute the gain of the approximate linear response of global temperature to radiative forcing.

5.2.3 Scale invariance and long-range dependence

The intimate connectivity between numerous interacting components and their large number of degrees of freedom characterizes complex systems. As a result, complex systems feature comprehensive dynamics, including, for example, non-linear feedbacks. Many complex systems also show emergent phenomena such as self-organization that arise from comprehensive interactions within the system but are often independent of the details of individual components. Studying these emergent properties is critical for advancing our understanding of nature. Scale invariance, or scaling, is one of these properties.

Scaling bridges two sides of a parameter regime, for example, the occurrence of high and low-magnitude events or a system's fast and slow dynamics. As a universal behavior, scale invariance can link and classify the properties of different dynamical systems from areas such as biology, economics, quantum physics, and geology (Clauset et al., 2009; Mandelbrot & Van Ness, 1968). The scale-invariant function for a one-dimensional function $f : \mathbb{R} \to \mathbb{R}$ is a power law (proof in Appendix A)

$$f(x) = cx^b \,. \tag{9}$$

This relation states that rescaling x by a factor of x_0 leads to a change in $f(x)$ by a factor of x_0^b. The parameter b is called the scaling exponent.

Power-law scaling has been found in the distributions of the population of cities (Zipf, 1949), the magnitude of earthquakes (Gutenberg & Richter, 1944), and the sizes of wildfires (Newman, 2005), among others. These distributions exhibit so-called heavy tails, which imply higher probabilities for extreme magnitudes compared to exponentially decaying distributions, such as Gaussian distributions. Many time-varying processes also show power-law scaling. The main property of scale-invariant stochastic processes such as Brownian motion is that their statistical properties have no characteristic timescale (Mandelbrot & Van Ness, 1968). Accordingly, the statistical properties transfer from one scale to another. This is particularly useful when only one scale is within the possible range of observations.

Among others, the study of scale-invariant processes led to the concept of long-range dependence: In 1951, Hurst (Hurst, 1951) studied the Nile's water level fluctuations and related the change of the water level's variance to the time-span n of the observation[1]. He found that as $n \to \infty$, the variance increases following a power law n^H with Hurst exponent H. This so-called "Hurst effect" has been confirmed in various systems from physics and climate science to economics (Franzke et al., 2015). Formally, a stationary stochastic process $X(t)$ is said to possess long-range dependence, or memory, if the autocovariance function decays asymptotically

$$\gamma(h) \sim h^{2H-2} \tag{10}$$

with $0.5 < H \leq 1$. The Hurst exponent quantifies the temporal persistence of the process, that is, the strength of the correlation of successive values in the time series. In the frequency domain, Equation (10) becomes (Mandelbrot & Van Ness, 1968)

$$S(\omega) \sim \omega^{-2H+1} \quad \text{as} \quad \omega \to 0 \,. \tag{11}$$

For non-stationary processes, long-range memory corresponds to a diverging sum of the autocovariance

$$\int \gamma(h)\, dh \to \infty \,. \tag{12}$$

Surface air temperature has been suggested to show fluctuation behavior similar to the Hurst phenomenon (Fraedrich et al., 2004). Fractional Gaussian noise and fractional Brownian motion are two types of stochastic processes suited to describe temperatures (Nilsen et al., 2016). Their correlation properties define the process (Mandelbrot & Van Ness, 1968) and strongly relate to the Hurst phenomenon (10)-(12). **P1** gives a detailed description of these processes and uses them to study the persistence of temperature fluctuations.

[1]The specific quantity Hurst (1951) studied was the so-called "rescaled range", which is a statistical measure of the variance as a function of n.

5.3 Uncertainties in temperature variability

The quantification of temperature variability is pivotal to future projections (Katz & Brown, 1992; Schär et al., 2004). However, the understanding of temperature variability across scales still needs to be completed (Fig. 1), especially on interannual to millennial timescales. Uncertainties in local to global temperature fluctuations and their representation in climate model simulations could particularly impact the assessment of temperature-related risks, ECS estimates, and abrupt transitions. This points to a necessity of constraining temperature variability from climate simulations and paleoclimate reconstructions. Three specific problems complicating the characterization of interannual to millennial-scale variability have been identified:

1. *Data-model mismatches on the timescale-dependent temperature variance*

Agreement between paleoclimate reconstructions and model simulations on multidecadal global mean temperature variability has been suggested by estimates in the time (PAGES 2k Consortium, 2019) and frequency domain (Zhu et al., 2019). However, the scale breaks from Zhu et al. (2019) are different from those found by Lovejoy (2015) (Fig. 1 c). Moreover, comparisons of the variance from local temperature time series present a mismatch between model simulations and paleoclimate reconstructions (Fernández-Donado et al., 2013; Laepple & Huybers, 2014a; Parsons et al., 2017). Yet, a detailed comparison of the PSD and scaling behavior from reconstructions and simulations across spatiotemporal scales and mean climates is still missing.

2. *Insufficient knowledge about internal variability mechanisms and the role of external forcing*

Different mechanisms might explain the potential scale invariance of temperature fluctuations. One candidate is the linear integration of high-frequency forcing by slowly varying climate subsystems, such as the oceans and cryosphere, similar to Equation (3). In line with that, power-law behavior could resemble the one from an OU-process (Fig. 7). This idea has been successfully used to describe ocean surface temperatures by a linear model representing oceanic diffusion through box models (Fraedrich et al., 2004; Fredriksen & Rypdal, 2017). The overlap of short-range dependent processes might explain continuous scaling (Fredriksen & Rypdal, 2017). By contrast, long-range dependent processes such as fractional Brownian motion have also been suggested to describe temperature fluctuations (Franzke, 2012; Nilsen et al., 2016). However, their appearance in Earth dynamics is controversial since most geophysical equations of motion depend solely on the previous state but not the distant past (Franzke et al., 2020). Moreover, there are large uncertainties about the relative magnitude of internal and external variability (Frankcombe et al., 2015; Hawkins & Sutton, 2011; Hébert & Lovejoy, 2018), especially on decadal to centennial timescales. This requires quantifying contributions from externally forced and internally generated fluctuations to the spectrum.

3. *Unclear effects of changes in the mean climate on temperature variability*

Rehfeld et al. (2018) have shown that temperature variability depends on the mean climate state. As a result, global warming affects variability (Bathiany et al., 2018; Olonscheck et al., 2021). However, there is conflicting evidence on the mechanisms and spatiotemporal patterns of variability changes (Brown et al., 2017; Holmes et al., 2016; Huntingford et al., 2013; Rehfeld et al., 2020). Studies based on single strong volcanic eruptions (Berdahl & Robock, 2013; Muthers et al., 2014; Swingedouw et al., 2017; Zanchettin et al., 2016) and future projections (Hopcroft et al., 2018) suggest state-dependent effects in response to volcanic eruptions. It still needs to be determined whether this state dependency also holds for the response to weaker forcing. Sea ice could contribute to state-dependent effects as its extent follows the mean state and varies under volcanic forcing (Timmreck, 2012). However, the impact of sea ice dynamics on temperature variability across scales remains uncertain. There is a need to quantify the state-dependent changes of forced and unforced variability and the contribu-

tions from potentially state-dependent components such as sea ice.

Overall, there is a large uncertainty about the structure and causes of temperature variability. Timescale-, forcing-, and state-dependent variability changes are insufficiently constrained. Achieving improvement on the above elements of uncertainty could open new possibilities for reliable simulations of future variations, climate impacts, and associated risks. To this end, this work and **P1–P3** follow five specific research objects, summarized in the next section (Sec. 6). Altogether, this thesis presents and discusses new findings on the local-to-global structure, the mechanisms, and state-dependent effects of temperature variability to improve the understanding of temperature variations and assess their potential implications for future projections.

6 Research objectives

This thesis comprises the analysis of the timescale-, forcing-, and state-dependent structures and causes of temperature variability (Fig. 8) in three publications (**P1-P3**) and complementary investigations (**CI**), presented in this work. The underlying specific research objectives (**O1-O5**) and resulting key findings (*in italic*) are:

O1: To evaluate the agreement of reconstructed and simulated global temperature variability
→ *The agreement has been largely confirmed (**P1, P2, P3, CI**). This supports confidence in the simulated global temperature variability, especially if simulations include natural forcing and interactively coupled model components (**CI**).*

O2: To quantify the potential mismatch between reconstructed and simulated local temperature variability in terms of overall variance and scaling on interannual to millennial timescales
→ *Mismatches in the variance and scaling properties of reconstructed and simulated temperature on decadal to multicentennial timescales are significant (**P1, P2**). Using natural forcing (**P2**) and dynamic ice sheets (**CI**) in the simulation reduces discrepancies. If simulated temperatures miss variability, they could underestimate the recurrence of extremes (**CI**).*

O3: To isolate the timescale-dependent forced and unforced temperature variability
→ *A Bayesian energy balance framework can successfully separate forced and unforced global temperature variability (**P3**) by emulation from more complex models. The relative contribution of internal variability increases with model complexity and decreases with timescale (**P3**). The approach provides an alternative to computationally expensive methods for separating variability components (**P3**) and can support statistical estimates of climate-related risks (**CI**).*

O4: To test for state-dependent effects in forced and unforced temperature variability
→ *The global variability is predominately forcing-dependent while local variability is mainly state-dependent and decreases with warming (**P2, CI**). Sea ice dynamics is identified as a state-dependent mechanism that causes changes in local temperature variability (**P2**).*

O5: To identify key mechanisms contributing to the power spectral density of surface air temperature
→ *The global temperature spectrum of the last millennia results from the integrated linear response to mostly external forcing (**P1, P3**). A millennial-scale change in fluctuation behavior might be due to changes in meltwater forcing and stabilizing feedback in response to it (**CI**). Locally, sea ice and volcanic forcing influence temperature variability for decades (**P2**). Improved representations of feedbacks in climate models are necessary to reconcile simulated and reconstructed temperatures (**CI**).*

Common to these research objectives is the overarching question of the implications of temperature variability across scales. The combined findings of this work, discussed in the next section, suggest that *the presented findings are critical for the reliable representation of temperature variability in simulations of the past, present, and future, as well as the assessment of climate-related risks.*

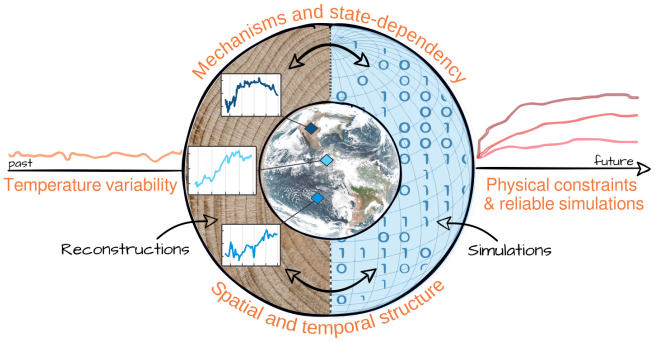


Figure 8: Visualization of the research strategy.

7 Results and discussion

This chapter comprises the summary and discussion of this work's findings on the structure and causes of surface air temperature variability across scales. Section 7.1 summarizes the main conclusions from **P1-P3**, which are discussed in detail in the respective manuscripts (App. F - App. H). Section 7.2 presents complementary analyses of an ensemble of comprehensive climate model simulations for the last 26 thousand years to expand the discussion of **P1-P3**. Section 7.3 discusses remaining limitations resulting from the underlying assumptions (Sec. 7.3.1) as well as uncertainties in simulations (Sec. 7.3.2) and reconstruction data (Sec. 7.3.3). Finally, implications for climate model improvements (Sec. 7.4.1), the projection and attribution of extremes (Sec. 7.4.2), as well as risk and impact assessments (Sec. 7.4.3) are discussed. The outlook (Sec. 7.4.4) provides some ideas for testing these implications in follow-up studies.

7.1 Summary of the results from the presented publications

P1: B. Ellerhoff and K. Rehfeld (2021)

The publication **P1** "Probing the timescale dependency of local and global variations in surface air temperature from climate simulations and reconstructions of the last millennia" (App. F) quantifies the agreement between temperature variability in model simulations and paleoclimate reconstructions of the last millennia. We provide a new estimate of the global and local mean PSD from periods of hours to millennia using paleoclimate reconstructions and simulation data from the Coupled Model Intercomparison Project phase 5 (CMIP5). The local paleoclimate records show higher multidecadal variability and scaling coefficients β of the local PSD $S \sim \tau^{\beta}$ than simulated temperatures. The global mean PSD is widely consistent across datasets, primarily due to the dominant linear response of global temperature to external forcing. However, local disagreement despite global agreement presents a dilemma. It appears that local variations in climate model simulations need to be enhanced without significantly affecting global properties.

P2: B. Ellerhoff, M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime, and K. Rehfeld (2022)

P2 "Contrasting state-dependent effects of natural forcing on global and local climate variability" (App. G) examines whether the climatic response to natural forcing depends on the mean climate state. We present an ensemble of forced and unforced Glacial/Interglacial simulations with the HadCM3 climate model and evaluate their variability on interannual to multicentennial scales. Global mean spectra are predominately determined by external forcing, in line with the results from **P1**. Natural forcing increases variability on all timescales. Local mean spectra are more characteristic of the mean state than of the applied forcing. The mismatch between reconstructed and simulated local temperatures is significant and increases with timescale. However, including natural forcing helps reduce this mismatch, even beyond decadal scales. Moreover, we quantify the contributions from sea ice dynamics to temperature fluctuations across scales. The variability of sea ice extent is significantly higher in forced compared to unforced simulations and in glacial compared to interglacial climates. By integrating the climate system's response to high-frequency natural forcing, sea ice dynamics plays a crucial role in amplifying local multidecadal variability.

P3: M. Schillinger, B. Ellerhoff, R. Scheichl, and K. Rehfeld (2022)

P3 "Separating internal and externally forced contributions to global temperature variability using a Bayesian stochastic energy balance framework" (App. H) emulates temperature fluctuations from simulations with climate models of intermediate to high complexity based on physical constraints. To this end, **P3** presents the "ClimBayes" software package (Schillinger et al., 2022b) which infers the free parameters of a stochastic two-box energy balance model (EBM) using a Bayesian approach.

The EBM reproduces the characteristics of simulated global temperature fluctuations even from comprehensive climate models. Furthermore, the ClimBayes framework allows for the separation of forced and unforced variability as a function of timescale. The relative contribution of internal dynamics increases with model complexity and decreases with timescale. Compared to observations, internal (external) variability in CMIP5 simulations is smaller (larger) on decadal and multidecadal scales.

Altogether, **P1, P2**, and **P3** largely confirm the agreement between the variability of simulated and reconstructed global temperatures for the last millennia. However, it is unclear whether the agreement can be confirmed for temperatures beyond the stable period of the Holocene and, if so, for which climate model simulations. Moreover, **P1** and **P2** point to pronounced differences in simulated and reconstructed local temperature variability. It remains to be clarified whether more advanced climate models can overcome these mismatches. In addition, it is necessary to test whether simulations with improved representations of feedback mechanisms can confirm the global forcing and local state dependency of variability (**P2**). Precise understanding of these dependencies could help reveal the extent to which global temperatures represent a linear response to external forcing (**P3**). These open points are addressed by complementary studies presented here.

7.2 Complementary studies with advanced model simulations

This section presents additional investigations to support the discussion of **P1-P3**. The complementary analyses use an ensemble of transient deglaciation simulations (Mikolajewicz, 2023; Mikolajewicz & Kapsch, 2022) for the last 26 thousand years with the Max Planck Institute for Meteorology Earth System Model (MPI-ESM) version 1.2 (Mauritsen et al., 2019) (App. B). The model experiments were performed by Uwe Mikolajewicz and Marie-Luise Kapsch (Kapsch et al., 2022; Mikolajewicz, 2023; Mikolajewicz & Kapsch, 2022). The runs provide state-of-the-art paleoclimate simulation with improved ice sheet representations compared to other PMIP3-type simulations (Kapsch et al., 2022). In particular, these runs allow differentiation of variability from volcanic forcing, ice sheet implementations, and meltwater fluxes.

Section 7.2.1 evaluates the robustness of our finding of forcing-dependent global and state-dependent local variability using the MPI-ESM simulation ensemble. Section 7.2.2 tests if advanced climate model simulations with volcanic forcing and interactively coupled ice sheets can reduce discrepancies in the multidecadal and multicentennial variability between reconstructions and simulations. Lastly, Section 7.2.3 provides a new composite of the global temperature PSD on timescales of years to hundreds of millennia by comprising paleoclimate data from state-of-the-art simulations and reconstruction.

7.2.1 Evaluation of global forcing and local state dependency

According to our analysis of a Glacial/Interglacial simulation ensemble with HadCM3, external forcing dominates global temperature variability (**P2**). This is confirmed by studies of the forced response in **P1** and **P3**. Conversely, local temperature variability depends primarily on the mean climate (**P2**). For validation of these findings, the analysis from **P2** is repeated with MPI-ESM simulations with optional volcanic forcing and an interactively coupled ice sheet, hereafter denoted by MPI-ESM (coupled) (App. B, Tab. 2). Ice sheet changes directly affect the river routing and meltwater fluxes in these runs through the interactive coupling. The volcanic forcing corresponds to a reconstruction for the last 130,000 years that compiles proxy records and stochastically generated synthetic eruptions (Toohey et al., 2023) (App. B). Following **P2**, “control” runs without volcanic forcing are examined

to investigate the forcing dependency of variance estimates across timescales. In contrast to the considered HadCM3 runs, the MPI-ESM simulations feature no solar forcing. Following **P1-P3**, the multitaper method with three windows (Percival & Walden, 1993; Yiou et al., 1996) is used to compute power spectra. Time series are linearly detrended to avoid biases arising from non-stationary trends. In addition, logarithmic Gaussian smoothing is applied to the PSDs for visual purposes.

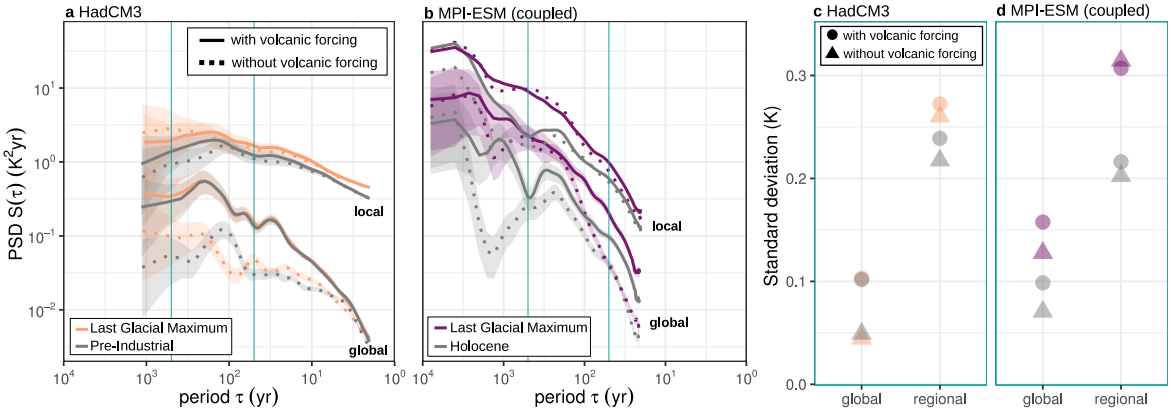


Figure 9: Local and global power spectral density for simulated temperature using HadCM3 (**a**, as in **P2**), and MPI-ESM (coupled) (**b**). Global spectra are calculated from global mean surface air temperature. Local refers to the area-weighted average of all local temperature spectra. Solid lines correspond to simulations with high-frequency natural forcing (MPI-ESM: volcanic, HadCM3: volcanic + solar forcing). Dotted lines show corresponding control runs without high-frequency natural forcing. The spectra were computed from simulated temperatures of climate states corresponding to a glacial and interglacial climate state. HadCM3 runs are of millennial length and correspond to the Last Glacial Maximum (LGM) and Pre-industrial state (**P2**). For MPI-ESM (coupled), decadally-resolved temperature time series were extracted from periods of 25 - 17 kiloyears ago (LGM) and the last eight kiloyears (Holocene). Lines show logarithmically smoothed (0.08 dB) mean spectra with shaded 95% confidence intervals. Confidence bands of HadCM3 are wider because they are obtained by sampling from a three-member ensemble for each configuration. The panels **c** and **d** summarize the forcing and state dependencies in terms of the standard deviation of temperature fluctuations over periods $50 \leq \tau \leq 500$ yrs, highlighted by green vertical lines in **a** and **b**.

Figure 9 compares the findings on the state-dependent variability from HadCM3 (**P2**) to those obtained with MPI-ESM (coupled). The results from MPI-ESM (coupled) confirm that state-dependent changes of internal variability dominate the simulated local temperature variability (Fig. 9 d). Local variability decreases with warming in the MPI-ESM simulations, as it does in the HadCM3 ensemble (Fig. 9 c). The effects of external forcing on global temperature variability also remain dominant in MPI-ESM. However, the MPI-ESM simulation shows enhanced variability at global scales during the LGM (Fig. 9 d), while there is no global state dependency in the HadCM3 simulations (Fig. 9 c). Using prescribed compared to coupled ice sheets does not alter the state-dependent effects significantly. This is shown by comparison against the same MPI-ESM model experiment with a prescribed ice sheet (App. D, Fig. 19).

Thus, the analysis of MPI-ESM simulations confirms the result from **P2** that local variability is dominated by internal state-dependent mechanisms. However, the complementary study also reveals the potential underestimation of state-dependent changes in the global temperature evolution simulated by HadCM3. Instead, more comprehensive climate model simulations can exhibit state-dependent effects in the global mean. These effects are likely due to a more sophisticated representation of variability mechanisms related to state-dependent cryospheric processes in MPI-ESM, addressed in Section 7.2.3.

7.2.2 Variability effects of volcanic forcing and interactive ice sheets

The findings from **P1** and **P2** reveal significant differences between simulated and reconstructed temperature variability on interannual to millennial scales. In particular, the decadal-to-centennial variability of the local mean PSD based on proxy records is one to two orders of magnitude higher than the one from state-of-the-art climate model simulations (**P1**). The spectral exponent on multi-decadal scales is smaller for simulation than for paleoclimate data. This disagreement is statistically significant, showing no agreement beyond chance for almost all model simulations (**P1**). In **P2**, we showed that it is possible to partially compensate for these mismatches by including natural forcing in simulations. Nevertheless, the mismatch in variance increases with timescale (**P2**).

Similar to the previous section, the improvement in model-data agreement for simulations with volcanic forcing is verified using the MPI-ESM (coupled) simulations. The local variance from MPI-ESM (coupled) runs with and without volcanic forcing are compared against paleoclimate data, following **P2**. The same proxy records from Rehfeld et al. (2018) and PAGES2k Consortium (2017) are analyzed as in **P2**. To be selected, records must be published and calibrated to temperature, contain more than 50 data points, cover at least three times the largest period of interest, and have a mean sampling frequency twice the highest frequency considered (**P2**). Records with gaps exceeding five times the required resolution are rejected.

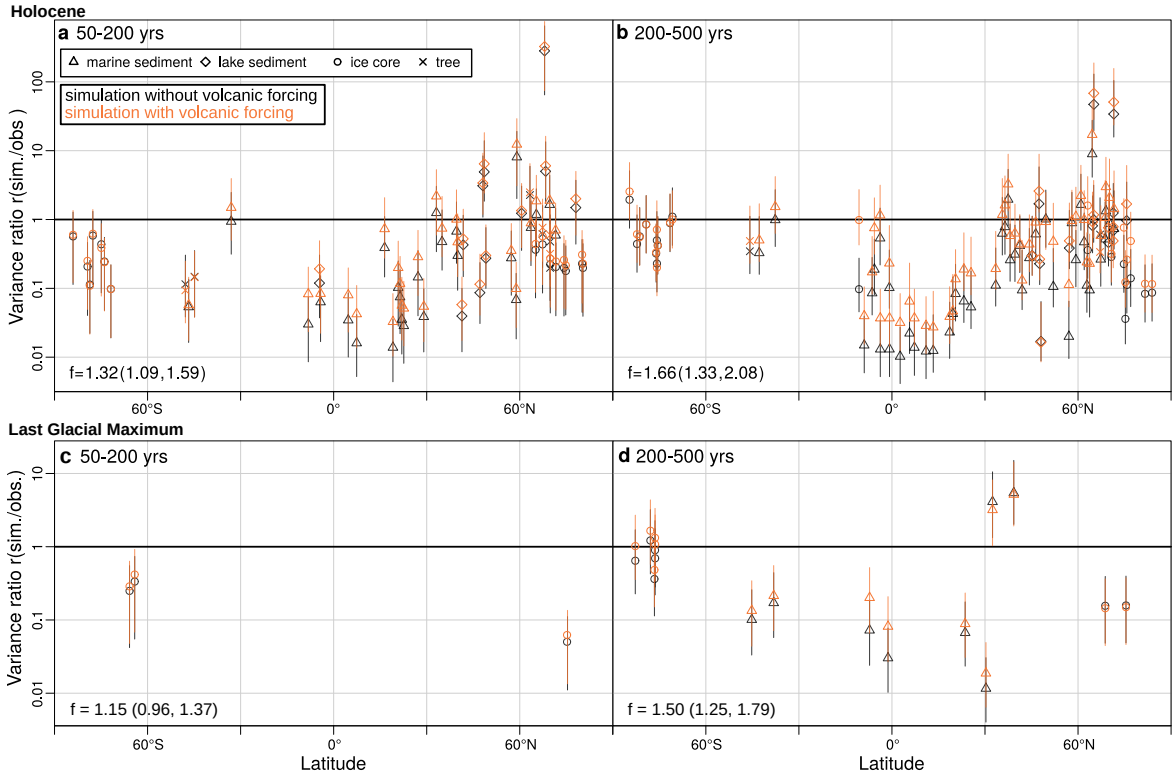


Figure 10: Ratios r(sim./obs.) of simulated to reconstructed temperature variance over latitude for control (black) and volcanically forced (orange) runs from the MPI-ESM (coupled) simulation for the Holocene (**a** and **b**, last eight kyrs, 78 data points) and LGM (**c** and **d**, 25-17 kyrs ago, 15 data points), as Fig. 4 in **P2**. The ratio of simulated to reconstructed temperature variance is computed on multidecadal (**a** and **c**) and multicentennial (**b** and **d**) timescales. Model data is bilinearly interpolated to the location of the observation. Symbols indicate the variance ratio and proxy archive. Vertical lines show 90% confidence intervals. The data points are slightly displaced along the x-axis to increase visibility. The local mean improvement f of the variance ratio is given in the lower left of each panel, with confidence intervals in parentheses (see Appendix A of **P2**).

Following **P2**, Figure 10 shows the ratio of simulated to reconstructed temperature variance on multidecadal ($50 \leq \tau \leq 200$ yrs) and multicentennial ($200 \leq \tau \leq 500$ yrs) timescales for the Last Glacial

Maximum (25-17 kyrs) and Holocene (last eight kyrs). The proxy records show a higher variance at both timescales and most locations compared to simulated temperatures. Including volcanic forcing in the MPI-ESM runs increases the variance, confirming effects from high-frequency natural forcing on the simulated variability (**P2**). Overall, this improvement is significant and quantified by the improvement factors (**P2**) $f = 1.32\,(1.09, 1.59)$ at multidecadal and $f = 1.66\,(1.33, 2.08)$ at multicentennial scales for the Holocene (Fig. 10 a and b). The factors are similar for LGM temperatures (Fig. 10 c and d) but less statistically robust due to fewer records. Simulations with prescribed instead of interactive ice sheets show similar improvements in both the Holocene and LGM (App. D, Fig. 20). In contrast to HadCM3, an increase in the mismatch between reconstructed and simulated variance with timescale is not discernible.

Mistmatch between simulated and reconstructed variance across model types

To further investigate this finding, Figure 11 shows the percentage agreement and the distribution of local mismatches as a function of the climate state and considered model. In addition to HadCM3 and MPI-ESM (coupled), an MPI-ESM simulation with volcanic forcing (Toohey et al., 2023), local meltwater, and prescribed ice sheets (Briggs et al., 2014; Kapsch et al., 2022; Tarasov et al., 2012) based on the GLAC-1D reconstruction is considered and denoted by MPI-ESM (GLAC-1D) (App. B). MPI-ESM (coupled) appears to provide the best abilities to simulate multicentennial variability among the considered model experiments (Fig. 11).

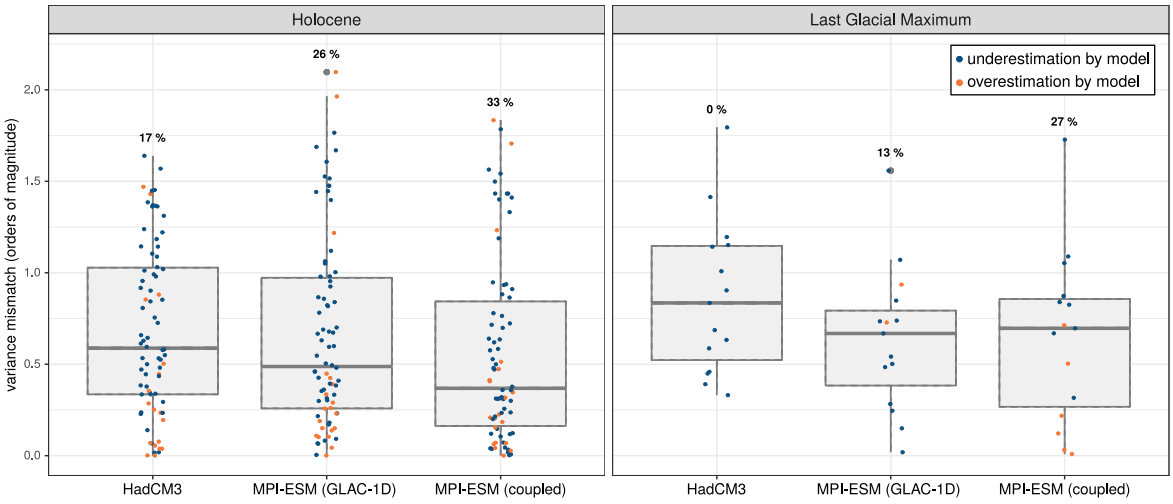


Figure 11: Comparison of simulated and reconstructed local temperature variance at multicentennial scales (200-500 yrs) using naturally forced climate model simulations of the Holocene (left panel) and LGM (right panel). Numbers give the percentage agreement, that is, the fraction of locations at which simulated and reconstructed variance agrees within uncertainties. Dots show the mismatch in orders of magnitude at single locations (see previous figure). Their colors indicate whether the simulated variance is smaller (blue) or larger (orange) than the one from proxy records. Within a simulation type, dots are slightly displaced along the x-axis to increase visibility. The boxes indicate the mean and quartiles (25% and 75% confidence levels). Vertical lines correspond to the 95% confidence interval of the ensemble. The MPI-ESM simulations are described in Appendix B, Tab. 2. HadCM3 corresponds to the solar and volcanically forced simulations from **P2**.

For Holocene temperatures, the median variance mismatch for MPI-ESM (coupled) is the lowest, and the percentage agreement is the highest (33%, corresponding to 26 out of 78 samples). It performs slightly better than MPI-ESM (GLAC-1D) (26% agreement). For HadCM3, only 17% of the samples agree. There are only minor differences between the MPI-ESM runs for LGM temperatures in the distribution of mismatches. The percentage agreement for MPI-ESM (coupled) is approximately twice as high as for MPI-ESM (GLAC-1D) (27% compared to 13%). Simulated LGM temperatures from HadCM3 show no agreement with the considered proxy records. Moreover, the HadCM3 experi-

ment uses a volcanic forcing reconstruction of the last millennium. Therefore, the comparability of the variability between the HadCM3 simulation and proxy records is limited, which could partially explain the strong disagreement for LGM temperatures. However, other factors due to the less comprehensive representation of cryosphere processes in HadCM3 are more likely to explain the inconsistency.

In summary, the analysis with the MPI-ESM simulation ensemble confirms, first, the disagreement between simulated and reconstructed variance on multidecadal and multicentennial timescales (**P1, P2**), and, second, the enhancement of the simulated variance by volcanic forcing (**P2**). In contrast to the 1000 years long HadCM3 simulation, the more comprehensive 26 kiloyears long runs with MPI-ESM show no increase in mismatches with timescales. This could be due to improved sampling of low-frequency variability or better representations of local long-term variability mechanisms. Improved agreement with MPI-ESM (coupled) compared to MPI-ESM (GLAC-1D) (Fig. 11) indicates that a greater time series length alone cannot explain the reduced mismatch. Instead, the representation of local feedbacks from interactively coupling the ice sheet could cause improved agreement in MPI-ESM (coupled). Differences in the representation of feedbacks relevant to the simulation of temperature variability in the MPI-ESM ensemble are discussed in Section 7.2.3.

The slightly enhanced agreement for local Holocene temperatures in MPI-ESM (coupled) is remarkable since the meltwater flux in the prescribed simulation can be expected to be close to zero over the last eight kiloyears (Kapsch et al., 2022). This might imply that meltwater-related fluctuations, which could occur in the coupled simulation in response to internal modes of the ice sheets, contribute to multicentennial variability in the Holocene. While this hypothesis requires further testing, the MPI-ESM (coupled) simulation opens new possibilities to validate temperature variability on timescales from decades to multiple millennia compared to PMIP3-type last millennium simulations, such as the one with HadCM3 (**P2**).

7.2.3 A new composite PSD of global temperature beyond the Common Era

Based on the above findings, Figure 12 a presents a composite estimate of the global temperature PSD that spans periods between years and hundreds of thousands of years. This PSD represents a new estimate compared to the previous findings from Huybers & Curry (2006), Lovejoy (2015), and Mitchell (1976) (Fig. 1), and **P1** for three reasons: First, it combines data from state-of-the-art instrumental observations, paleoclimate reconstructions, and climate model simulations (App. E, Tab. 3). Second, the reconstruction ensembles from PAGES2k (PAGES 2k Consortium, 2019) (1000 members per method, seven methods) and Temp12k (Kaufman et al., 2020) (500 members per method, five methods) help constrain the uncertainty range by allowing to sample the PSD from different reconstruction methods. Third, climate model simulations were selected based on their ability to represent relevant physical mechanisms contributing to variability across scales.

The high-frequency end of the spectrum, highlighted in Figure 12 c, is formed by an ensemble of 23 CMIP5 last millennium runs from seven models (PAGES 2k Consortium, 2019) (App. E, Tab. 3). The observed and reconstructed global temperatures from HadCRUT5 (Morice et al., 2021) and PAGES2k (PAGES 2k Consortium, 2019) are added. The low-frequency part of the spectrum (Fig. 12 b) consists of recent global temperature reconstructions from Snyder (2016) (Snyder GMST) and Kaufman et al. (2020) (Temp12k). In addition, simulations with the Earth system models of intermediate complexity Famous (Smith & Gregory, 2012) and CLIMBER-2 are shown. The CLIMBER-2 simulation provides a transient run of the past three million years with interactive Northern Hemisphere ice sheets and atmospheric CO_2 (Willeit et al., 2019). The (coupled) MPI-ESM run, which agrees best with paleoclimate data (Sec. 7.2.2), covers decadal to multimillennial timescales and, thus, allows for comparison

against all other datasets. The PSDs are calculated following **P1-P3**. The mean spectrum is computed by interpolating to the lowest resolution, binning into equally spaced log-frequency intervals, and taking the average with equal weights (Huybers & Curry, 2006).

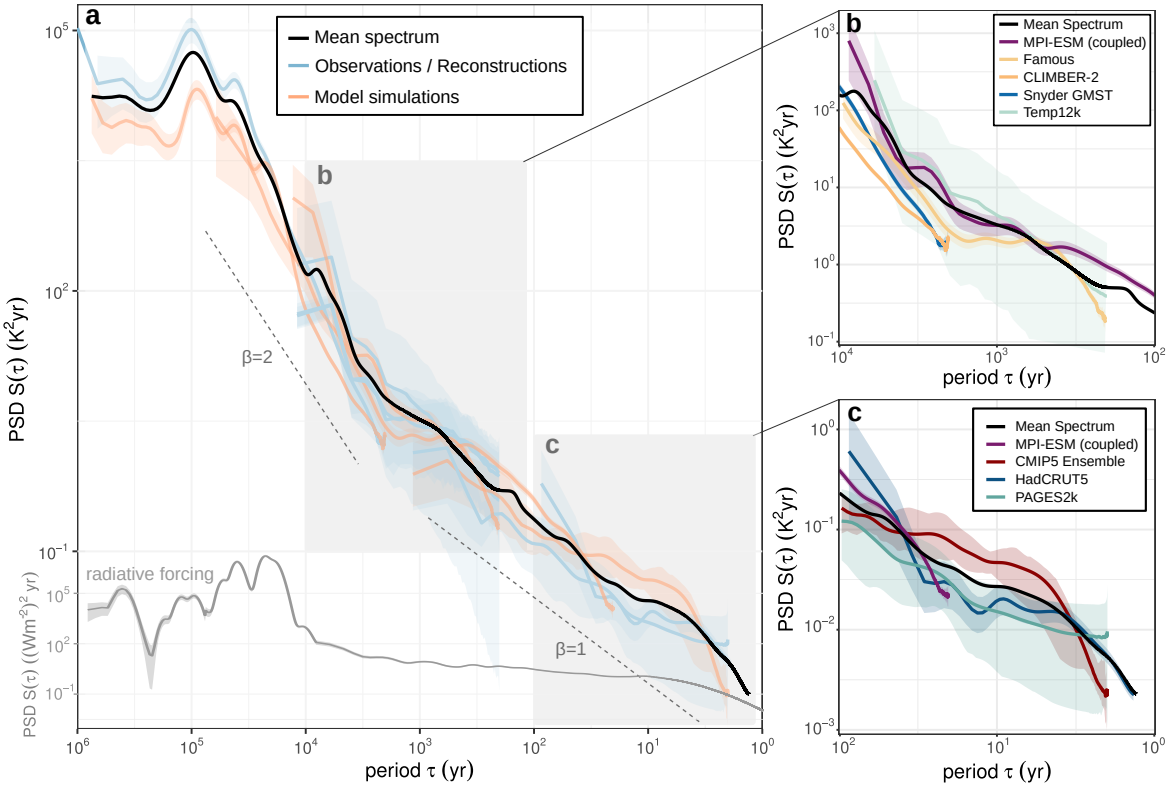


Figure 12: **a** New composite estimate of the PSD of global mean temperature from instrumental observations (HadCRUT5), reconstructions (PAGES2k, Temp12k, Snyder GMST), and climate model simulations (MPI-ESM (coupled), CLIMBER-2, Famous). In addition, 23 CMIP5 last millennium runs are combined as "CMIP5 ensemble" following the PAGES 2k Consortium (2019). References to all considered datasets are given in Appendix E, Tab. 3. Uncertainties are χ^2 distributed according to the degrees of freedom of the spectral estimate, except for the PAGES2k, Temp12k, and CMIP5 ensemble. In these cases, uncertainty bands correspond to the 95% confidence interval of the spectra sampled from the full ensemble (Temp12k: 5 reconstruction methods, 500 samples each; PAGES2k: 7 reconstruction methods, 1000 samples each; CMIP5: 23 members, 7 models). The uncertainty for the mean spectrum is not shown here for better visibility and because it is best represented by the envelope of all other PSDs. The light grey line at the bottom of panel **a** shows a composite estimate of the PSD from radiative forcing that considers orbital, volcanic, solar, and CO_2 forcing (App. D, Fig. 21). **b** The inset highlights the PSD from $100 \leq \tau \leq 10,000$ yrs at the centennial to multimillennial scale from all considered time series. **c** As **b**, but for $1 < \tau \leq 100$ yrs.

Figure 12 confirms a wide model-data agreement in global temperature variability. On interannual to decadal scales (Fig. 12 c), the CMIP5 last millennium simulation ensemble shows a higher PSD than the HadCRUT5 instrumental record and the PAGES2k paleoclimate reconstruction for the last two millennia. This is in line with model-data differences in the spectral gain (**P1**) and can be explained by an enhanced response to external forcing in CMIP5 simulations compared to observations (**P3**) (Fredriksen & Rypdal, 2016; Goosse et al., 2005; Scaife & Smith, 2018). Moreover, artificial amplification of the El Niño–Southern Oscillation (ENSO) in CMIP5 simulation using the atmospheric general circulation model ECHAM could contribute to this mismatch (**P1**) (Jungclaus, 2020). The approximate power-law scaling $\beta \approx 1$ ($10 < \tau < 1,000$ yrs) is in line with previous results (**P3**) (Huybers & Curry, 2006; Lovejoy, 2015; Nilsen et al., 2016; Rypdal et al., 2013). The linear response to forcing can explain this behavior (**P3**) (Fredriksen & Rypdal, 2016).

Millennial-scale change in global temperature fluctuation behavior

There are indications for a change in fluctuation behavior at millennial periods (Fig. 12 b). Above this period, the temperature scaling is more similar to $\beta \approx 2$ (for $1,000 < \tau < 100,000$ yrs). The change in scaling indicates two regimes with distinct behavior of temperature fluctuation. To illustrate this, the relative contribution from the two neighboring intervals $[\omega'/2, \omega']$ and $[\omega', 2\omega']$ within a scaling regime $[\omega_1, \omega_2]$ reads

$$\frac{\int_{\omega'/2}^{\omega'} S(\omega)\, d\omega}{\int_{\omega'}^{2\omega'} S(\omega)\, d\omega} = \frac{1 - 2^{\beta-1}}{2^{1-\beta} - 1} = 2^{\beta-1},$$

with $\omega_1 = \omega'/2$ and $\omega_2 = 2\omega'$. Hence, for $\beta = 1$, the contribution to the variance stays the same over timescales. For $\beta > 1$, variance is relatively stronger on long compared to short timescales. Huybers & Curry (2006) have reported higher scaling coefficients on millennial timescales, however, with a scale break at centennial scales. Conversely, Figure 12 shows no evidence for a scale break at the centennial scale and raises the question of why the transition occurs at $\tau \approx 1,000$ years.

To investigate the causes for this transition, Figure 13 highlights the PSD at centennial to multimillennial timescales from the Temp12k reconstruction (Kaufman et al., 2020) and MPI-ESM simulation ensemble. The simulated and reconstructed PSD agree well within the uncertainty. Restricting the analysis to simulated temperatures for the last twelve thousand years, which corresponds to the temporal overlap between the two datasets, confirms the agreement between the MPI-ESM simulations and Temp12k (Fig. 13 c) within uncertainties. However, estimates from the Temp12k reconstruction ensemble exhibit large uncertainty bands.

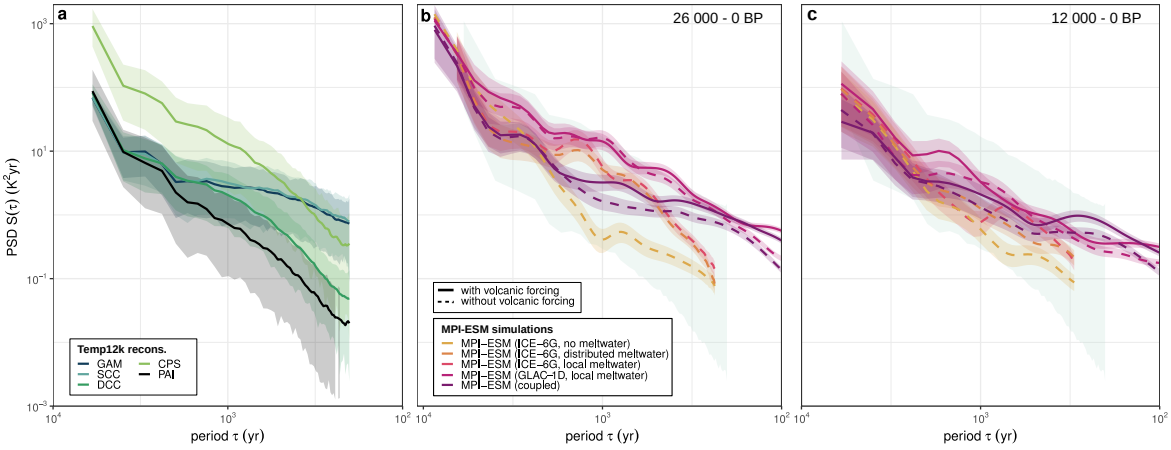


Figure 13: Global temperature PSD at the centennial to multimillennial scale ($100 \leq \tau \leq 10,000$ yrs) from (**a**) the Temp12k reconstruction (Kaufman et al., 2020) and (**b** and **c**) the MPI-ESM simulation ensemble (App. B, Tab. 2). **c** shows the PSD for the MPI-ESM runs considered in **b** but using only global temperatures since 12 000 kyrs before present (BP). In panel **b** and **c**, the 95% confidence interval of the spectra sampled from the full Temp12k reconstruction ensemble is highlighted in light blue. "Distributed meltwater" corresponds to an implementation that redistributes the local meltwater discharge homogeneously across all grid cells worldwide (App. B).

The Standard Calibrated Composite (SCC) and Generalized Additive Model (GAM) reconstruction show a change in scaling behavior from $\beta \approx 1$ to $\beta \approx 2$ at a period of approximately 1,000 years (Fig. 12 a). The Pairwise Comparison (PAI) and Composite Plus Scale (CPS) reconstructions do not show this transition. The reconstruction based on the Dynamic Calibrated Composite (DCC) method indicates a scale break, although the change in scaling behavior is not as pronounced as in SCC and GAM. The SCC, GAM, and DCC reconstructions are likely more robust than PAI and CPS. This is because PAI features an exceptionally low variability in all regions except for the northern latitudes (Kaufman et al., 2020). CPS stands out with its significant temperature changes compared to other

methods and proxy records, especially in the Early- and Mid-Holocene warmth (Kaufman et al., 2020). Thus, the reconstruction methods most likely to be trusted indicate a millennial-scale change in fluctuation behavior.

Regarding the MPI-ESM simulation ensemble, Figure 13 b compares the PSDs from the MPI-ESM (coupled) run against those from the full MPI-ESM simulations with prescribed ice sheets (GLAC-1D and ICE-6G) and different implementations of meltwater fluxes (App. B, Tab. 2). There are pronounced differences in the simulated PSD at centennial to millennial scales, depending on the implementation of ice sheet feedback and volcanic forcing in the simulations. These differences are smaller when considering only the last twelve thousand years of the simulation (Fig. 13 c). While volcanic forcing enhanced the PSD in all runs, differences between the spectra in Figure 13 b are primarily due to the considered meltwater and ice sheet implementations. Both simulations with local meltwater and prescribed ice sheets show the highest millennial-scale variability. The MPI-ESM (GLAC-1D) simulation without meltwater shows the lowest PSD. MPI-ESM (GLAC-1D) shows enhanced millennial-scale variability compared to MPI-ESM (ICE-6G). On the one hand, this could be due to the higher temporal resolution of the considered simulation output from MPI-ESM (GLAC-1D). On the other hand, Kapsch et al. (2022) argue that greater variability in meltwater release throughout the last deglaciation in MPI-ESM (GLAC-1D) could cause the increased temperature variability compared to MPI-ESM (ICE-6G). This is in line with reduced differences for the Holocene period (Fig. 13 c).

Improved representation of ice sheet feedbacks in coupled simulations

Together, the analysis of Temp12k reconstructions and MPI-ESM simulations suggests a millennial-scale change in scaling behavior. Here, it is hypothesized that the differences between the millennial-scale variability from MPI-ESM runs with a coupled or prescribed ice sheet (Fig. 13 b) are due to a different role of ice sheet feedbacks in the simulations. Ice sheet dynamics and freshwater forcing are essential feedback mechanisms operating at this timescale. Ice sheets predominately build up during glacial cold periods, recurring on orbital timescales with periods of approximately 41 to 100 kyrs. Freshwater discharge from ice sheets and icebergs can destabilize ocean circulation (Ziemen et al., 2019) and lead to sudden changes in local to global temperatures (Bakker et al., 2017; Klockmann et al., 2020), especially during deglaciations. The meltwater flux can be understood as a natural driver that forces temperature changes in the ocean, thereby enhancing the surface temperature variability on multimillennial timescales.

In MPI-ESM simulations with prescribed ice sheets, the meltwater flux is computed from the temporal derivative in ice thickness (Kapsch et al., 2022; Meccia & Mikolajewicz, 2018). By contrast, MPI-ESM (coupled) computes meltwater fluxes dynamically from ice sheet changes. Therefore, the interactively coupled version can account for feedbacks of the meltwater onto the ice sheet growth (Ziemen et al., 2019). In particular, the meltwater flux could cause a slowdown of the AMOC, leading to a cooling in the North Atlantic and, thus, reduced meltwater flux. Consequently, destabilizing meltwater fluxes and stabilizing negative feedback in response to it might compete at millennial scales. These mechanisms, however, can only unfold in interactively coupled simulations. In runs with prescribed ice sheets, the ocean's cooling can not feedback onto the meltwater flux, which keeps driving temperature variability. This could explain the increased millennial-scale variability in the prescribed compared to coupled runs as well as a better agreement of local temperatures from MPI-ESM (coupled) with reconstructions (Fig. 11).

The transition from $\beta \approx 2$ to $\beta \approx 1$ could be due to destabilizing meltwater fluxes becoming less relevant below millennial timescales, partially due to negative feedback from ocean cooling. This is in line with previous ideas that changes in fluctuation behavior might result from stabilizing versus

destabilizing slow processes in the climate system (Arnscheidt & Rothman, 2022). On the one hand, the change in fluctuation behavior at millennial scales could also be an artifact of the enhanced variability on orbital timescales according to Milankovitch cycles (Fig. 12 a). On the other hand, the MPI-ESM simulations, relying on the same orbital and CO_2 forcing, show pronounced differences in millennial-scale variability and scaling behavior depending on the implementation of meltwater fluxes and ice sheets in the simulation (Fig. 12 a). While the linear response to forcing can largely explain global temperature on timescales below centuries (**P1, P3**), internal mechanisms increasingly contribute to variability on millennial scales (Fig. 13 b). The substantial contributions to variability from ice sheet processes that persist at the global scale could also explain state-dependent changes in millennial-scale variability (Fig. 9 d). In line with Section 7.2.1 and Section 7.2.2, climate models with a realistic implementation of natural forcing and ice sheet feedbacks simulate temperature fluctuations that are more consistent with reconstructions. Therefore, the present findings indicate that the representation of feedback processes in response to meltwater fluxes could cause a profound change in fluctuation behavior on multicentennial to multimillennia scales and is, therefore, an essential component of reliable climate simulations on centennial and longer timescales. Testing the proposed mechanisms requires analyses of the ice sheet dynamics and meltwater forcing in the MPI-ESM simulation ensemble once more variables are available (Mikolajewicz, 2023; Mikolajewicz & Kapsch, 2022).

Nonetheless, the presented mean PSD and the indication of a change in scaling behavior on millennial timescales are subject to uncertainties (Fig. 12 a). Only datasets that are trusted because of their properties to reflect variability on the respective timescales are considered. Moreover, the mean spectrum is formed with equal weighting across all PSD. While this could introduce a sampling bias, small changes in the data selection and weights do not significantly alter the findings. Monte Carlo simulation of this sampling (as in **P1**) could yield realistic uncertainty bands for the mean global temperature and radiative forcing spectrum in follow-up studies. Improved statistical reconstructions based on proxy synthesis are required to provide better constraints on global millennial-scale temperature variability from paleoclimate evidence.

7.3 Limitations and uncertainties

In addition to the discussions in **P1-P3**, this section reviews remaining limitations and uncertainties. While the respective manuscripts (App. F - App. H) discuss limitations of the specific data and methods used, there are also uncertainties common to the studies presented. These include the assumption of weak stationarity of examined time series and the approximative linear response to external forcing, which are addressed in (Sec. 7.3.1). Moreover, uncertainties resulting from the considered paleoclimate reconstructions (Sec. 7.3.3) and simulation data (Sec. 7.3.2) are discussed as well as potential remedies.

7.3.1 Stationarity and linearity assumptions

As introduced in Section 5.2.2, the spectral analysis assumes weak stationarity of the underlying stochastic process (Chatfield, 2003). Linear detrending serves to remove trends from time series before computing spectra (Fredriksen & Rypdal, 2016; Laepple & Huybers, 2014b; Nilsen et al., 2016) (**P1-P3**). Supplementary analyses confirm that the recent global warming trend does not impact our main conclusions (**P2**, Fig. S6, S9, and S10, Fig. 22). However, time series beyond the stable period of the Holocene, considered in Section 7.2, could be subject to spectral biases since glacial cycles involve non-stationary transitions. The onset of the deglaciation affects the spectral estimate, but linear detrending does not significantly alter the presented findings of variability below millennial scales

(App. D, Fig. 22). This indicates that the change in PSD at millennial scales is due to physics-based changes in temperature fluctuations rather than spectral biases from the impact of non-stationary trends. The findings on state-dependent effects also imply that power spectral densities of time series covering glacial and interglacial periods might be more representative of the cold climate state, exhibiting more variability. Besides analyzing time series segments such as in Fig. 9, time-varying spectra might be best suited to disentangle power spectral density changes over time and prevent potential biases from trends.

To quantify the contribution from forced variability in **P1** and **P3**, it is assumed that the global temperature evolution can be described as a linear response to external drivers. However, this assumption neglects higher-order perturbations in response to forcing. Thus, potential impacts from radiative forcing on internal climate variations, for example, by modulation of the phase or amplitude of internal climate modes (Maher et al., 2015; Otterå et al., 2010) are not considered. Nevertheless, the finding that the stochastic two-box EBM successfully emulates the global temperature variability of CMIP5 simulations and AR5 EMICs well (**P3**) is consistent with previous findings on a linear relation between external forcing and global temperature for the last millennium (Fredriksen & Rypdal, 2017; Geoffroy et al., 2013; MacMynowski et al., 2011). One way to test the limitations of the linear assumption for other timescales and future climates could be the study of single-forcing experiments (Schurer et al., 2014) and large single-model ensembles (Fyfe et al., 2021) to separate contributions to temperature variability from different external, anthropogenic, and internal sources. Moreover, **P3** suggests several potential modifications of ClimBayes, such as extensions to other temporal resolutions and numerical optimizations to enable longer simulations. However, non-linear effects in response to radiative forcing or changes in variability with the mean climate could complicate variability analysis on centennial and longer timescales. The application of ClimBayes to deglaciation simulations and future projections with substantial trends, such as SSP5-8.5 experiments, remains to be tested.

The power spectral density estimator measures two-point correlations and is therefore limited in identifying non-linear processes. Inappropriate use of linear methods for analyzing highly non-linear processes such as DO-events can lead to false interpretations of the significance of peaks (Braun et al., 2010). This does not directly affect our analysis as we focus on the continuum of the climate spectrum instead of spectral peaks. However, since nonlinearities are inherent to the climate system, it will be necessary to develop additional tools to extend the analysis of non-linear effects on climate variability across scales. Moreover, power spectral analysis cannot be directly applied to infer causality (Chatfield, 2003). For example, the spectral gain cannot tell which process preceded the other. Cross-spectral and higher-order spectral analyses could be used to study how one process influences the other in a linear or non-linear setting (Liebrand & de Bakker, 2019). The cross-spectral density, extending the power spectrum to the case of two input time series, allows studying the temporal synchronicity and time offsets between these two (Pires & Hannachi, 2021). Yet, robust applications of these advanced spectral estimates are limited to time series that exhibit many similarly shaped cycles per considered time window, supporting high enough significance levels (Liebrand & de Bakker, 2019). Moreover, there is no common standard for defining significance levels against stochastic noise and normalizing these spectral estimates. Thus, their application for quantifying the strength and causality of climate interactions requires additional research.

7.3.2 Uncertainties in climate model simulations

Climate models have constantly been improved in their representation of climate processes (Knutti & Sedláček, 2013). Nevertheless, they represent simplifications of physical processes and cannot re-

produce the entire complexity of Earth's dynamic system. In simulations, relevant processes might be missing because of insufficient coupling between climate model components, such as the carbon cycle and ice sheets. Paleoclimate simulations are typically based on less comprehensive models or use lower model resolutions to simulate long periods within limited computational capacities (Braconnot et al., 2012). Unavoidable uncertainties arise from their fixed temporal and spatial resolution as well as their different level of detail in representing Earth system components. The necessary parametrization of sub-grid processes complicates the representation of important small-scale feedback processes, such as those from stratospheric or indirect aerosol effects in response to volcanic forcing, cloud feedbacks, or ocean mixing, needed to advance climate model simulations (Eyring et al., 2019). Parametrization schemes that simplify sub-grid processes can be prone to uncertainties (Lehner et al., 2020), for example, due to empirically estimated parameters.

In addition, paleoclimate simulations require forcing reconstructions, which are subject to uncertainties in paleoclimate analysis (Toohey & Sigl, 2017). The design of model experiments, which use different (Schmidt et al., 2012) or incomplete (O'Neill et al., 2016) forcing inputs, further contribute to the spread between simulations (Fyfe et al., 2021) and between simulated and observed climate variability (Fredriksen & Rypdal, 2016; Goosse et al., 2005; Scaife & Smith, 2018) (**P2**). Running ensemble simulations and considering multiple members in the statistical analysis can make statistical assessments more reliable (**P2**), assuming that the ensemble members sample the relevant uncertainties. Moreover, Bayesian inference based on Monte Carlo Markov chains can help infer model-specific parameters from a limited number of model runs (**P3**).

7.3.3 Uncertainties in proxy reconstructions

Paleoclimate reconstructions are also subject to uncertainties. This is because they are based on natural chemical, biological, geological, and physical processes (Cronin, 2010). Time series extracted from paleoclimate archives can exhibit irregular sampling, hiatuses, measurement uncertainties, varying resolutions, limited coverage, or a limited number of data points. Natural signal-altering processes such as bioturbation (Dolman et al., 2021; Mollenhauer et al., 2003) and diffusion (Johnsen et al., 2000; Münch et al., 2017) impact the records and can lead to archive- or proxy-specific limitations in representing variability. Examples are the tendency of reduced high-frequency variability in marine sediments and ice core records (Casado et al., 2020), as well as the reduced low-frequency variability in tree ring reconstructions (Lücke et al., 2019). Calibration and reconstruction techniques rely on statistical assumptions and vary between archive types and proxies. This can lead to additional biases and uncertainties (Birks & John, 2012; Büntgen et al., 2022). Global and hemispheric reconstructions are considered more robust since they build on multiple records (Christiansen & Ljungqvist, 2017). However, the sparse proxy coverage in some parts of the Earth can bias reconstructions (Mann et al., 2008), for example, towards more densely sampled regions.

The irregularity and sparse availability of proxy records also challenge their statistical analysis. In particular, spectral analysis of irregularly sampled proxy data is more prone to biases than that of regular time series. To reduce possible errors, proxy analyses in **P1**, **P2**, and this work build on the interpolation and uncertainty quantification for irregular temperature records of previous studies by Laepple & Huybers (2014b). Biases are minimized by selecting published proxy records from large proxy syntheses such as PAGES2k (PAGES2k Consortium, 2017). Selection criteria account for hiatus, target resolution, number of data points, and covered period and are chosen to allow for reliable estimates over the considered timescale (**P1** and **P2**). Comparison of statistical estimators (Baudouin et al., 2022), for example, for extracting scaling coefficients (**P1**), forms the basis for the development and application of robust statistical tools. Monte Carlo and pseudo-proxy approaches (**P1**)

address the problem that only a single observed realization of past climate exists. They can also help distinguish signal from noise in the presence of limited data. However, more reliable regional reconstructions are needed to exploit the full potential of climate model validation with paleoclimate reconstructions.

7.4 Implications and outlook

The presented results point to limitations in the simulated variability structure from climate model simulations. In particular, climate model simulations might show too few temperature variations over decades and centuries (**P1, P2**, Fig. 11). This section examines the potential for reconciling simulated and reconstructed local temperature variability (Sec. 7.4.1). Next, potential biases in temperature projections and attributions studies arising from our findings are discussed (Sec. 7.4.2). Numerical experiments with integrated assessment models (IAMs) show the importance of accounting for the full range of potential future variability by contrasting projections of climate-related economic damages with and without natural variability (Sec. 7.4.3). It remains to be clarified to what extent the presented paleoclimate findings transfer to future climate variability in an Earth system state subject to far-reaching anthropogenic changes (IPCC, 2022b). The section concludes by discussing potential future studies to extend this work's findings and implications (Sec. 7.4.4).

7.4.1 Reconciling simulated and reconstructed temperature variability

Different arguments can partly explain our findings on the local-to-global variability structure. Regarding global variability, simulated, observed, and reconstructed temperatures largely agree across timescales (**P1, P3**). The dependence of global temperature on external radiative forcing (**P1–P3**) is the main reason why less comprehensive models can reliably simulate its properties for the recent millennia. Our simple stochastic two-box EBM confirms a predominant linear dependence of surface air temperature on natural forcing by robustly emulating complex climate model simulations (**P3**). The agreement of forced climate model experiments with paleoclimate reconstructions demonstrates their ability to simulate the global temperature evolution on interannual to centennial scales.

However, simulations show a stronger contribution from forcing to global temperature variability at interannual to decadal scales than instrumental observations and paleoclimate reconstructions (**P1**, **P3**) (Chylek et al., 2020; Schurer et al., 2013). Likewise, the simulated internal variability is smaller over these periods (**P3**) (Yan et al., 2018). As the overall variability agrees, the enhanced global response to forcing might compensate for too low internal variability in simulations. The presented results also demonstrate that contributions from internal variability grow over periods from years to centuries when slow processes in the oceans, cryosphere, vegetation, and carbon cycle become increasingly relevant (**P3**). The simulation of long-term ice sheet feedbacks using an interactive ice sheet model can help reconcile the continuum of reconstructed and simulated variability on millennial scales from different data sources (Fig. 12).

Regarding temperatures at the local-to-regional level, many studies confirm disagreement on decadal to multicentennial scales (Cheung et al., 2017; Fredriksen & Rypdal, 2016; Hébert et al., 2022; Laepple & Huybers, 2014a,b; Parsons et al., 2017). Some uncertainties inherent to the model and proxy data can help explain this finding. Local variability could be enhanced in proxy records because they may include non-climatic "noise", for example, from site-specific features or the aliasing of the seasonal cycle (Laepple & Huybers, 2013). Conversely, potential biases, such as the bioturbation of sediments (Berger & Heath, 1968) and isotopic diffusion in ice (Johnsen et al., 2000), tend to deplete variability in proxy records. These effects can occur at specific, potentially distinct timescales.

In Laepple et al. (2023), we hypothesize that uncertainties in proxy reconstructions alone cannot explain the mismatch in variance and persistence at the local level. By reviewing the evidence on natural temperature variability and potential reasons for the data-model mismatch, we argue that regional climate variations persist for timescales longer than those currently simulated by climate models. In favor of this argument, statistical methods to correct potential biases from non-climatic processes have strongly improved the reliability of proxy reconstruction (Casado et al., 2020; Dolman & Laepple, 2018). In addition, comparisons to observations also show wide agreement (Hébert et al., 2022), as do reconstructions and simulations at interannual scales (**P1, P2**) (Bühler et al., 2021; Cheung et al., 2017; Hébert et al., 2022; Laepple & Huybers, 2014a,b). Therefore, neither an overestimation of local variability nor an underestimation of global variability by reconstructions can fully explain the systematic model-data discrepancies (Laepple et al., 2023). Instead, the mismatch might be due to model deficiencies. Climate model simulations might show too little local variability on multidecadal scales. Suppression of this variability could result from a high model diffusivity and, thus, fast energy dissipation over temporal scales in simulations. In line with this, comprehensive climate models are often too stable to simulate abrupt climate transitions (Valdes, 2011).

Several optimizations of climate models and simulation experiments can help improve the representation of variability. The enhanced model-data agreement between proxy and simulated temperature using MPI-ESM (coupled) points to an improved representation of climate variability in comprehensive climate models with interactive ice sheets. In line with this, models with sophisticated ice sheet feedbacks could potentially better simulate sudden climate shifts (Armstrong et al., 2022; Klockmann et al., 2020; Ziemen et al., 2019) and, thus, account for these contributions to variability. Missing representation of interactive ice sheets (Bakker et al., 2017) and dynamic vegetation (Braconnot et al., 2019; Hopcroft & Valdes, 2021) have been proposed to explain underestimated simulated variability, as well as too weak forcing at the local scale (Fredriksen & Rypdal, 2016; Goosse et al., 2005; Scaife & Smith, 2018). More interactive components and feedback processes could yield more independent modes of variability in climate models. Improved representation of sub-grid scale processes can also increase temperature variability in model simulations (Juricke et al., 2017).

When model experiments account for natural forcing, the simulated temperature variance agrees significantly better with palaeoclimate data, especially on decadal to multicentennial scales (**P2**, Fig. 9). Thus, including natural forcing provides a more accurate representation of climate variability. Consistent with previous arguments (Bethke et al., 2017), it is necessary to complement solely emissions-driven projections (O'Neill et al., 2016) by natural forcing. Representing stochastic and intermittent natural forcing from volcanic eruptions at the regional scale could also improve the representation of forced variability. It might correct for too strong responses at the global level (Fredriksen & Rypdal, 2016; Goosse et al., 2005; Scaife & Smith, 2018). If disagreement persists, regional projections will be affected by associated model biases. This raises the question of whether projections of the occurrence of extreme periods and the attribution of single events are sensitive to simulated long-term variability.

7.4.2 Projection and attribution of extremes under long-term variability

Climate variability influences projections of future climate (Blanusa et al., 2022), impact assessments (Calel et al., 2020; Schwarzwald & Lenssen, 2022), the occurrence of extreme events (Czymzik et al., 2016; Ionita et al., 2021), and attribution studies (Harrington et al., 2021). Open challenges remain in linking climate variability with the recurrence properties of extreme climate and weather events (Simolo & Corti, 2022). **P1** concluded that local temperature fluctuations at decadal and longer timescales tend to show statistically different behavior with increasing ($\beta > 1$) versus decreasing

($\beta < 1$) fluctuations on longer timescales in proxy records compared to simulations. As a result, climate model simulations might underestimate long-term correlations of local temperature at decadal to centennial scales compared to reconstructions.

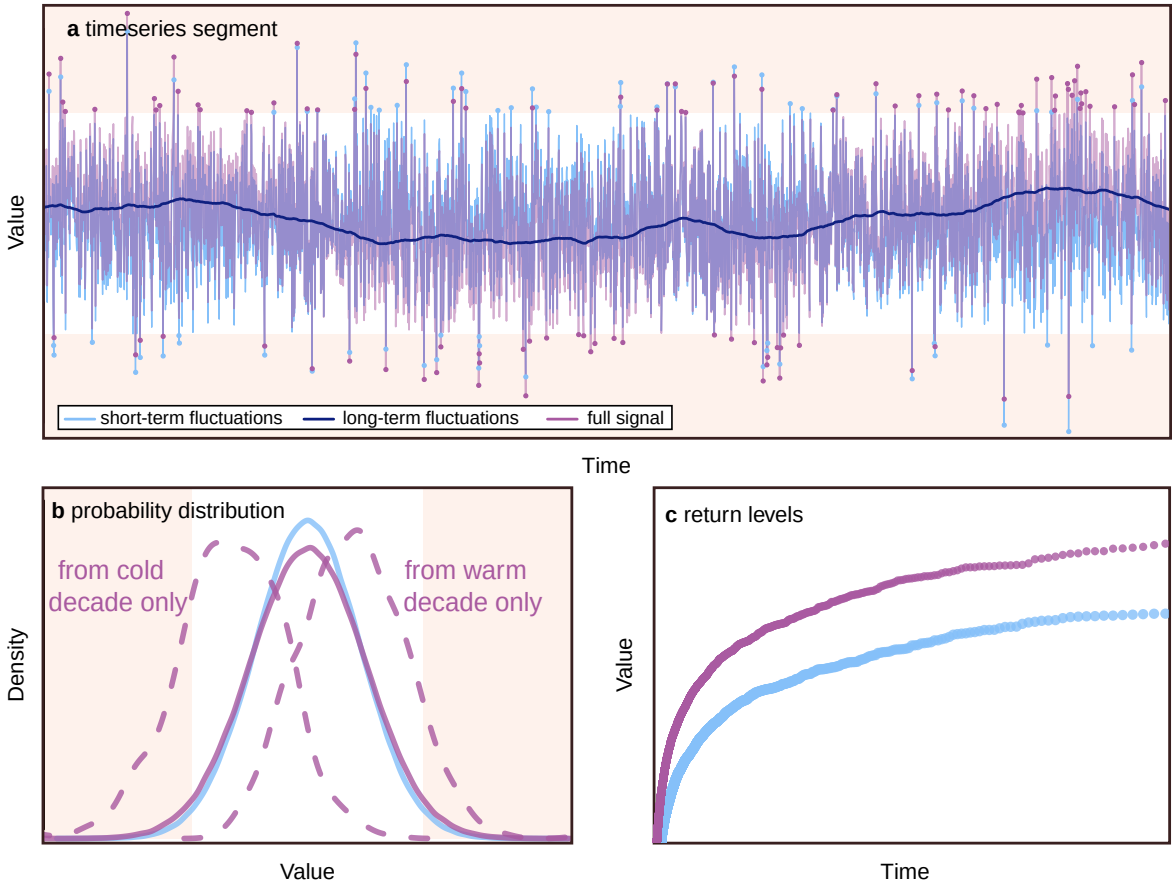


Figure 14: Conceptual representation of how long-term variability might influence the occurrence of extremes. **a** Light blue line shows a realization of a white noise process reflecting the short-term variations linked to weather phenomena. The dark blue line shows a realization of a self-similar stochastic process with $S(\tau) \sim \tau^{\beta}$ and $\beta \approx 1$. This long-term correlated time series is low-pass filtered (cutoff 1/10 yrs) to reflect decadal-to-centennial variability. The pink line is the sum of both processes and refers to a typical climate time series. Parameters reflect characteristic timescales of weeks (short-term) and decades (long-term), with a variance ratio of 5:1. The segment displays the numerical simulations of these time series corresponding to a time window of 50 years. Dots show extreme events that lie more than two standard deviations below or above the mean (white area). **b** Empirical distributions of the short-term time series and the full signal (solid lines). Dashed lines show distributions of the full signal estimated from an exceptionally cold and warm decade. **c** Empirical return levels from the full "short-term" and "short-term + long-term" time series. The empirical estimates are based on a numerically simulated time series of 10,000 years to ensure convergence of the empirically computed distributions. Figure adapted from Laepple et al. (2023).

Figure 14 visualizes the linkage between the statistical properties of time series fluctuations and the recurrence of extreme values. For the conceptual illustration of this link, numerical experiments with correlated and uncorrelated time series (Fig. 14 a) were performed. The short-term uncorrelated fluctuations correspond to a white noise process on a characteristic timescale of weeks. The long-term fluctuations correspond to a low-pass filtered self-similar stochastic process with $\beta \approx 1$, representing variability on decadal and longer timescales. Figure 14 b compares the probability distributions of the short-term fluctuations with that of the sum of short- and long-term fluctuations, reflecting a conceptual climate time series that features fast ("weather") and slow modulations, for example, from decadal variability. Figure 14 c displays the return levels of both time series. They were obtained from the Weibull formula, which relates the inverse rank i of the sorted time series

vector $X = (x_1, \ldots, x_N), x_i \geq x_{i+1}$ of length N to the return period $R(x_i)$ via

$$R(x_i) = 1/P(x \geq x_i) = \frac{N+1}{i}\,.$$

Long-term variability leads to a higher sensitivity of the estimated distribution on the observed time window (Fig. 14 b). Moreover, it increases the magnitude of extreme events and decreases their return time (Fig. 14 c). Underestimating long-term variability can cause statistical predictions of extremes that turn out too weak and, thus, miss losses and damages. Such an underestimation can, for example, result from relying on instrumental observations that are too short to constrain temperature variability beyond the decadal scale. This can bias distribution estimates of the underlying variable (Fig. 14 b). If strong and persistent local variability is partially missing in climate model simulations, future projections might underestimate the recurrence times of extremes (Fig. 14 c).

Moreover, a higher persistence increases the risk of extreme conditions, such as sustained high temperatures that increase the likelihood of extreme heat and droughts. Likewise, climate model simulations might underestimate the risk of compound heat and drought events and the risk of sustained agricultural losses and wildfires. Such combinations of extremes are essential for long-term planning, as they increase the vulnerability of ecosystems and populations (Zscheischler et al., 2020). Consequently, there is a need for impact assessment studies that consider climate variability (Calel et al., 2020; Schwarzwald & Lenssen, 2022). Uncertainties in policy-relevant projections due to incomplete knowledge of climate variability can also affect socio-economic planning (Deser et al., 2012).

As the simulation of variability is linked to the simulation of extremes, uncertainties in variability directly affect attribution studies. Attribution studies provide evidence for the anthropogenic influence on single extremes (Stott et al., 2016) and long-term changes (Hegerl et al., 1996). They decompose the climate response into anthropogenic and natural contributions based on climate model simulations. As such, they are susceptible to model biases in simulating variability (Harrington et al., 2021; van Oldenborgh et al., 2021). Extreme event attribution often assumes that decadal and longer internal variability plays only a minor role in the occurrence of extremes (Philip et al., 2020). If simulations fail to simulate the strength and persistence of local variability, tests for the significance of anthropogenic effects will tend to be biased positive (Laepple & Huybers, 2014b; Laepple et al., 2023).

7.4.3 Impacts of natural variability on damage and risk assessments

Besides links between climate variability and extreme events, variability also affects economic damages and losses (Calel et al., 2020; Schwarzwald & Lenssen, 2022). Here, the economic impact of global temperature variability is examined by combining the developed ClimBayes software package (**P3**) (Schillinger et al., 2022b) with an integrated assessment model (IAM). IAMs serve to represent the interaction between socio-economic and climate processes throughout the upcoming decades and centuries (Parson & Fisher-Vanden, 1997; van Beek et al., 2020). They are designed to explore scenarios of technical development, socio-economic transformation, climate-related damages, and their interaction with policy choices (Krey, 2014; Parson & Fisher-Vanden, 1997). Depending on the scope of application, IAMs vary in complexity, spatiotemporal resolution, and time horizon, among others (Baumstark et al., 2021; Krey, 2014). Examples include multiregional IAMs that address energy transformations in response to climate policies and economic development by modeling the energy-economic system (Baumstark et al., 2021). Other IAMs contrast the cost of global warming to that of climate mitigation and adaptation (Nordhaus, 2010).

Typically, IAMs rely on strong economic assumptions (Rotmans & van Asselt, 2001), such as simpli-

fied damage functions that extrapolate expected climate-related damages using empirical estimates (Weitzman, 2012). Moreover, IAM-based projections of losses and damages often focus on the mean values of climate parameters instead of their variability. Using a comparatively simple IAM (Nordhaus, 2017; Weitzman, 2012), Calel et al. (2020) demonstrated that projections of economic costs might be overly optimistic if future scenarios of global temperature change miss internal variability. The impacts of volcanic forcing as a substantial driver of interannual global temperature variability (**P3**) still need to be examined.

Here, the impacts from internal and volcanically forced variability on monetary costs from climate-related damages are estimated by combining the approach from Calel et al. (2020) with ClimBayes. Following **P3**, nine CMIP5 models[2] are emulated to generate an ensemble of global temperature time series with different variability representations, reflecting the *internal*, *volcanic*, and *internal + volcanic* variations. Next, the IAM computes the annual global economic damage based on global temperature anomalies ΔT, a damage function $D(\Delta T)$ (Fig. 15 a), population growth, and economic parameters such as per capita consumption and discount rate (App. C). The parameters are taken from Calel et al. (2020) and comply with economic and population models for 2020-2500 CE.

The emulated global temperature data consists of an ensemble of 10,000 time series with a length of 480 years for each of the nine CMIP5 models and three categories (*internal*, *volcanic*, and *internal + volcanic*). Example time series are shown in Appendix D, Fig. 18. The volcanic forcing is randomly sampled from distribution estimates of volcanic eruptions (Ammann & Naveau, 2010; Sprinz, 2023) (App. C). Other natural drivers, such as solar forcing, are neglected here since volcanism represents the dominant driver on interannual scales (**P1**). To isolate the damages from annual fluctuations, the mean temperature is held constant by fixing the free parameters T_0 and F_0 in ClimBayes. It is important to note that a constant mean temperature for 2020-2500 CE does not comply with realistic future scenarios. Here, potential impacts from global warming can only be approximated by comparing the damages for different mean temperatures.

Figure 15 b shows the damage distributions for global temperatures with *internal*, *volcanic* and *internal + volcanic* variability for a mean temperature anomaly of $\langle \Delta T(t) \rangle = 1.09$ K, corresponding to present day warming. The mean economic cost for all categories lies above a baseline scenario with constant $\Delta T(t) = 1.09$ K for all times t. This is because the non-linear damage function penalizes positive deviations ($\Delta T(t) > 1.09$ K) from the mean stronger than negative ones ($\Delta T(t) < 1.09$ K) in this scenario (Fig. 15 a). As a result, variability broadens the distribution of costs and increases its mean value (Fig. 15 b). The ensemble mean standard deviation $\sigma(\Delta T) = \sqrt{\langle (\Delta T(t) - \langle \Delta T(t) \rangle)^2 \rangle}$ is higher for *volcanic* ($\sigma(\Delta T) = 0.26$ K) compared to *internal* ($\sigma(\Delta T) = 0.12$ K). This explains increased mean economic costs and larger quartiles under volcanic forcing. Combining *internal + volcanic* further increases the variance, broadens the damage distributions, and increases the mean costs (Fig. 15 b).

The dependence of the overall extent of economic costs on the global mean temperature anomaly (Fig. 15 c and d) results directly from the damage function (Fig. 15 a). The costs are approximately one order of magnitude higher for $\langle \Delta T(t) \rangle = 3$ K compared to $\langle \Delta T(t) \rangle = 1.09$ K (Fig. 15 c), and one order of magnitude lower compared to $\langle \Delta T(t) \rangle = 5$ K (Fig. 15 d). The mean cumulative damage (127 Trillion USD) for $\langle \Delta T(t) \rangle = 3$ K corresponds to annual damages of approximately 3 % of today's gross world product ($\approx$ 80 Billion USD), in line with Calel et al. (2020) (App. C). Disregardless of the mean temperature, the distribution's quantiles span a range of costs that vary by approximately 10% due to *internal + volcanic* variability. Thus, missing volcanic forcing and internal variability in future

[2] All CMIP5 models used in **P3** are considered, except for MIROC, showing a substantial drift of the mean temperature, which could bias the analysis (**P3**). Table 1 of **P3** (App. H) specifies the simulations.

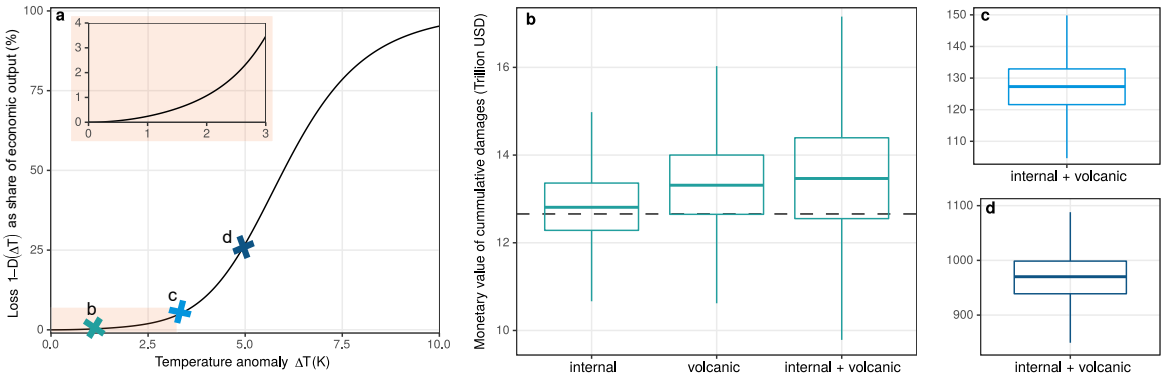


Figure 15: **a** Share of global economic output that will be lost according to the damage function $D(\Delta T(t)) = \left(1 + a\,\Delta T(t)^2 + b\,\Delta T(t)^6\right)^{-1}$ for $a = (1/20.46\,\mathrm{K})^2$ and $b = (1/6.081\,\mathrm{K})^6$ (Calel et al., 2020; Nordhaus, 2017; Weitzman, 2012) (App. C). The annual global temperature anomaly $\Delta T(t)$ refers to pre-industrial times (reference period 1850–1900 CE). Crosses indicate the mean anomaly 1.09 K, 3 K, and 5 K, for which idealized experiments computed the distribution of cumulative damages (**b, c,** and **d**). The inset highlights the loss for comparatively small temperature anomalies below 3K. **b** Distribution of cumulative damages in an idealized scenario with $\langle \Delta T(t) \rangle = 1.09\,\mathrm{K}$ considering *volcanic*, *internal* and *internal + volcanic* variability over the period 2020-2500 CE. The boxes indicate the quartiles, solid horizontal lines the median, and vertical lines the 5 % and 95 % quantiles. Each box represents the distribution estimates based on 90,000 samples from nine emulated CMIP5 last millennium simulations (10,000 per model). The horizontal dashed line corresponds to the cumulative damages obtained from a constant mean temperature without fluctuations. Panel **c** and **d** show the results for *internal+volcanic* variability and $\langle \Delta T(t) \rangle = 3\,\mathrm{K}$ and $\langle \Delta T(t) \rangle = 5\,\mathrm{K}$, respectively. Stephan Sprinz generated the global temperature time series using ClimBayes (Sprinz, 2023).

scenarios can cause an underestimation of the range of possible future damages.

Fundamental limitations of this conceptual IAM study arise from the damage function $D(\Delta T)$ (App. C), which does not account for uncertainties in the empirically estimated parameters. Moreover, the function is ill-defined for negative temperature anomalies. Substantial cooling after strong volcanic eruptions could bias the distribution estimates. Here, these effects occur in less than 0.7 % of all data points and do not significantly alter the ensemble estimates. Further, the damage does not capture secondary feedbacks of temperature variability change and non-linear non-temperature variable effects. The market-based cost-benefit or cost-efficiency optimization (Baumstark et al., 2021) of IAMs can often not reflect the potential for endogenous lifestyle changes. These changes, however, could significantly contribute to climate mitigation (van Vuuren et al., 2018). In addition, the ClimBayes package neglects potential changes in variability with warming (**P3**). Precise damage projections should be quantified with more complex IAMs and, ideally, simulation ensembles of naturally and anthropogenically forced projections based on more comprehensive models.

7.4.4 Remaining questions and potential for future studies

This dissertation comprises studies of various mechanisms of interannual to millennial temperature variability. In particular, the distinction between externally driven and intrinsic variability and the response to volcanic forcing on different timescales has been investigated. Moreover, the contributions of sea ice and meltwater flux feedbacks to variability have been discussed. However, as outlined in the introduction (Sec. 5.1 and Fig. 2), numerous other mechanisms can also influence climate variability. Thus, the presented analyses cover only a subset of relevant mechanisms of interannual to millennial variability.

This work does not explicitly study multidecadal modes such as the Atlantic Multidecadal Variability or the Pacific Decadal Oscillation, extensively covered in the literature (Fasullo et al., 2020; IPCC, 2021a). These and other modes can influence local and global temperature and are important mecha-

nisms for multidecadal variability (von der Heydt et al., 2021). However, variability is often considered as the response to external forcing plus internal modes, which may need to be revised because of the temperature fluctuations' scaling and persistence. The interaction of modes with the background continuum, particularly their contributions to scaling, remains open to future work.

Other essential mechanisms of variability, merely discussed in **P1–P3**, are short-term feedbacks on interannual timescales. They are especially relevant in response to forcing but also for the projection of extremes. Potentially relevant processes, such as stratospheric (Aubry et al., 2022) or cloud feedbacks (Hopcroft et al., 2018), require enhanced parameterization schemes, especially for paleoclimate simulations. If these can be improved, one critical question arises from the finding that local trends of short-term extremes, such as daily temperature maxima, differ in simulation and instrumental data (van Oldenborgh et al., 2022). On the one hand, a too-homogeneous structure of variability could explain why climate models are less capable of representing small-scale than large-scale changes. On the other hand, unresolved local feedbacks, potentially resulting from anthropogenic aerosols and land surface changes, could cause the observed differences. Contrasting the influences of long-term statistical changes and short-term local feedbacks will be necessary to improve the reliability of local projections.

The importance for simulated variability to couple model components interactively

Including land ice in climate simulations often relies on prescribing ice sheets based on reconstructions. Simulations with coupled ice sheets open new possibilities for improved representation of long-term variability, as the comparison of the MPI-ESM simulations shows. Similar results might be expected from including other interactive Earth system components in climate model simulations. In particular, paleoclimate simulations with ESMs typically prescribe CO_2 concentrations rather than simulating them within a dynamic carbon cycle (Brierley et al., 2020). Simulating all relevant climate feedback processes from the carbon cycle and vegetation dynamics with ESMs is challenging and computationally expensive. As a result, state-of-the-art PMIP4-CMIP6 simulations include only a few runs with an interactive carbon cycle. None of them represents a fully dynamic vegetation together with an interactive carbon cycle for a Mid-Holocene simulation (Brierley et al., 2020).

However, changes in the CO_2 concentrations from ocean biogeochemistry (Sigman & Boyle, 2000), enhanced weathering (Arnscheidt & Rothman, 2022), and vegetation changes (Braconnot et al., 2019) have been proposed to contribute to variability on millennial and longer scales. The implementation of an interactive carbon cycle in ESMs for last millennium simulations (Braconnot et al., 2012) and the first EMIC simulation of the past glacial cycles with interactive ice sheet and carbon cycle components (Willeit et al., 2019) could help assess slow carbon and vegetation feedbacks in the climate system. Moreover, dynamic vegetation can improve the representation of variability in simulations (Braconnot et al., 2019; Hopcroft & Valdes, 2021). Therefore, comprehensive climate model experiments that can dynamically simulate the interaction between climate, ice sheets, vegetation, and the carbon cycle over timescales longer than a thousand years will be necessary to complete the understanding of temperature variability.

Climate sensitivity, non-linear feedbacks, and future abrupt transitions

The equilibrium climate sensitivity (ECS) strongly depends on the strength of climate feedbacks, as outlined in Section 5.1.2. One major reason for the spread in ECS assessed by climate models is that they represent feedback mechanisms differently (Ceppi et al., 2017; Knutti & Rugenstein, 2015). As a result, the natural variability of feedback processes and their response to forcing could influence ECS estimates. The findings presented here can support ECS estimates by reducing model uncertainties and improving the understanding of internal variability. Future studies based on validation against paleoclimate reconstructions, such as the one presented for the MPI-ESM simulations with differ-

ent representations of ice sheet feedbacks, can help identify those models which provide a realistic temperature spectrum and which are most capable of simulating feedback processes. Therefore, a comparison of ECS estimates based on these models and scenario runs needs to follow.

Proper assessment of feedback mechanisms is not only critical for ECS estimates but also for assessing the risk of abrupt future changes. Both anthropogenic and natural forcings influence abrupt transitions and tipping points (Ashwin et al., 2012; Boers et al., 2022). Separating externally forced and internal variability could help understand whether abrupt transitions are due to external causes, internal mechanisms, or both (Ashwin et al., 2012). If climate models cannot simulate past transitions, they may also miss future abrupt changes due to similar mechanisms. Paleoclimate analyses can help determine whether climate model simulations capture the non-linear dynamics involved in abrupt transitions with sufficient accuracy (Valdes, 2011). Our results indicate that incorporating ice sheet feedbacks with interactively coupled ice sheet models may simulate temperature changes more reliably.

One possibility to project abrupt transitions is to determine critical warming values and associated forcing levels at which components of the Earth system might undergo a critical transition. However, estimates of these critical thresholds are still subject to significant uncertainties (Boers et al., 2022). Underestimating variability in climate model simulations can cause too optimistic predictions of the probability of crossing potentially critical thresholds in future projections. However, the link between long-term variability and critical regional or global climate thresholds in a future climate remains to be verified.

Transferring paleoclimate knowledge to future climates

This dissertation focuses on paleoclimate simulations and comparisons with existing paleoclimate data. Testing the relevance of the discussed state and forcing dependency for projections was beyond the scope of this thesis but should be readily possible using simulations of future climate scenarios. To this end, it will be necessary to compare simulated variability in past, present, and future climate states using climate models with a comprehensive representation of feedback processes. These are particularly needed to account for anthropogenically-modified mechanisms of variability. For example, air pollution could influence decadal modes of variability (Smith et al., 2016). In addition, land surface changes due to human activities strongly perturb the carbon cycle (Falkowski et al., 2000) and vegetation. These changes can impact future variability, for example, through albedo feedbacks. Moreover, the rapid melting of ice sheets, sea ice, and mountain glaciers could increase volcanic activity (Aubry et al., 2022; Swindles et al., 2018).

Nevertheless, the presented studies indicate strong implications for future climate states and their simulation: First, the lower mean local variability of surface air temperature in warmer climate states (**P2**) is likely to hold for future warming (Bathiany et al., 2018; Olonscheck et al., 2021; Rehfeld et al., 2018). While this pattern is robust for the high latitudes (**P2**), it remains to be verified for other regions. Second, the benefits of including high-frequency natural forcings in model experiments have been demonstrated for simulations of the past (**P2, P3**) and future climate (Bethke et al., 2017). It is crucial to ensure that climate models simulate natural forcing, internal variability, and their future changes sufficiently well to provide reliable projections of the impacts of local variability on society and ecosystems. Third, the representation of natural, long-term temperature variability in climate models influences projections of extremes in numerous ways. Representing variability in impact studies is critical to account for uncertainties in climate risk projections. Yet, the examined impact studies did not consider any adaptation behavior. Studies investigating the interaction between societal preparedness and the distribution of damages, for example, using coupled climate-social models (Moore et al., 2022), will be needed to further reduce uncertainties in impact assessments.

8 Conclusion

Due to anthropogenic influences on the climate, temperatures on Earth are rising faster than ever before. Mitigation and adaptation policies require reliable simulations of future climate variability. In particular, understanding and modeling temperature variability is crucial for projecting extreme climate and weather events like drought and heat. Model simulations are the primary method for assessing climate-related risks for humans and nature. However, simulated temperatures are subject to considerable uncertainties, especially on the societally relevant decadal and centennial timescales. Consequently, it is essential to comprehend the statistics of temperature fluctuations across spatiotemporal scales and the ability of climate models to represent them. This requires the comparison of simulated temperatures against paleoclimate reconstructions.

This dissertation aimed to improve the understanding and modeling of temperature variability across spatiotemporal scales and its potential impacts on projections and extremes. Three studies (**P1-P3**) and complementary analyses investigated temperature fluctuations across scales, focusing on their dependence on climate forcings and the mean climate state. With respect to the five research objectives **O1-O5** (Sec. 6), this work concludes (key findings in *italic*):

O1 To evaluate the agreement of reconstructed and simulated global temperature variability

Evaluating previously found agreement (PAGES 2k Consortium, 2019; Zhu et al., 2019) between reconstructed and simulated global temperature variability, the presented studies allow three conclusions: First, they confirm *wide agreement of observed, simulated, and reconstructed global temperature variability* throughout the last millennia (**P1, P2**). The dominant response of simulated temperatures to included external drivers explains this finding (**P1-P3**). However, this dependence also suggests that future projections will underestimate temperature variability and associated risks when ignoring natural forcing mechanisms, such as volcanism. Second, the strong response to volcanic eruptions yields an overestimation of the decadal variability by CMIP5 simulations (**P3**, Fig. 12 c). This raises the question for future research whether more localized in contrast to zonal implementations of volcanic forcing can correct this mismatch. Third, the broad agreement between climate models and observations cannot be generalized to timescales longer than centuries. This partially contradicts the arguments by Zhu et al. (2019) that climate models can correctly simulate the continuum of temperature variability. Instead, global temperature variability of the comprehensive climate models investigated here depends on their implementation of potentially non-linear feedback mechanisms from slow components of the dynamic climate system, such as ice sheets. In particular, the presented results demonstrate a strong effect of meltwater fluxes on global temperature variability (Fig. 13). Nevertheless, *our studies confirm the confidence in the simulation of global temperature, especially if climate model simulations represent the physically relevant building blocks of long-term variability in a realistic and interactively coupled manner.*

O2 To quantify the potential mismatch between reconstructed and simulated local temperature variability in terms of overall variance and scaling on interannual to millennial timescales

The main conclusion that can be drawn from comparing simulated and reconstructed local temperatures is that *mismatches on the examined decadal to multicentennial timescale are significant*. This includes both the scaling (**P1**) and the variance (**P2**). A too homogenous spatial structure of simulated variability might explain global agreement despite local disagreement of paleoclimate reconstructions and model simulations (Laepple et al., 2023). Incorporating *natural high-frequency forcing* (**P2**, Fig. 10) and *interactive ice sheets* (Fig. 11) in climate model experiments can *reduce the mismatch*. If climate model simulations do not sufficiently represent slow

feedback processes, the mismatch between simulated and reconstructed local temperatures likely increases with timescales. However, there is a need for more robust zonal or regional reconstructions and large ensembles of climate model experiments to constrain local variability. Lastly, reconciling existing discrepancies between simulated and reconstructed variability is crucial for projecting extremes. *If simulated temperatures miss temporal persistence or decadal variance*, as indicated by the present results, *they will underestimate the recurrence of extremes* (Fig. 14).

O3 To isolate the timescale-dependent forced and unforced temperature variability

In **P3**, we have developed the statistically robust and easy-to-use ClimBayes (Schillinger et al., 2022b) software package to isolate the timescale-dependent forced and unforced temperature statistics from climate model outputs. The main implication of the *successful emulation of global temperature from state-of-the-art climate model simulations* is that a few mechanisms, represented by a linear stochastic two-box EBM (**P3**), suffice to simulate global temperature variations over the last millennia. This finding can be considered a further validation of the predominant linear response to external forcing of simulated global mean temperatures (**P1, P3**) (Fredriksen & Rypdal, 2017; Geoffroy et al., 2013; MacMynowski et al., 2011). ClimBayes provides an *alternative to computationally expensive approaches for separating variability components* based on large ensembles of climate model simulations (Kay et al., 2015; Lehner et al., 2020). Therefore, ClimBayes can help identify uncertainties in variability estimates from internal stochastic dynamics, deterministic forcing, and additional proxy-, model-, or scenario-specific uncertainties. The developed approach also generates realistic global temperature time series very effectively. This supports more *reliable statistical estimates of potential climate-related risks and economic losses* (Fig. 15) compared to assessment studies that miss natural forcing (Calel et al., 2020; Schwarzwald & Lenssen, 2022). The many possibilities of combining integrated assessment models with ClimBayes warrants further investigation. In addition, future research could develop faster numerical implementations for ClimBayes that would allow emulating longer time series and, thus, validating limitations due to non-stationary trends and potential interactions between forced and internal variability.

O4 To test for state-dependent effects in forced and unforced temperature variability

Contrasting temperature variability in Glacial and Interglacial simulations with HadCM3 (**P2**) and MPI-ESM (Sec. 7.2.1), the present findings confirm *reduced local variability in warmer climates* (Bathiany et al., 2018; Berdahl & Robock, 2013; Bethke et al., 2017; Brown et al., 2017; Olonscheck et al., 2021; Rehfeld et al., 2018). Decreasing sea ice dynamics with warming is a potential mechanism for this change and causes the strongest reduction of variability in high latitudes (**P2**). In contrast to previously suggested state-dependent effects of volcanic forcing on the global and hemispheric climate based on ensemble simulations with large volcanic eruptions (Berdahl & Robock, 2013; Muthers et al., 2014; Swingedouw et al., 2017; Zanchettin et al., 2016), our study **P2** concludes that the *simulated global temperature response does not strongly depend on the mean climate*. This result is broadly consistent with the predominant linear relation between the global temperature response and volcanic forcing (see **O3**). While the local response to strong volcanic eruptions may depend on the climate state due to non-linear processes (Zanchettin et al., 2013), these effects are likely to vanish at the global scale and in more realistic forcing scenarios (**P2**). Since the present analyses reveal state-dependent differences according to the implementation of ice sheet feedbacks in climate models (Fig. 13), investigating the response to individual strong volcanic eruptions in these simulations might prove important to test the findings of **P2**. Overall, this work's results demonstrate a *substantial impact of state-*

dependent effects on local and a weak impact on global temperature variability. This implies high confidence in future projections of global mean temperature driven by greenhouse gas emissions. The dependence of internal variability on future scenarios and associated uncertainties in regional projections are issues for further research.

O5 To identify key mechanisms contributing to the power spectral density of surface air temperature

This overarching research goal draws on the insights into variability mechanisms from **O1-O4**. In summary, *external forcing dominates global temperature variability below centennial scales* (**P1-P3**). The *integration of uncorrelated, short-term fluctuations* ("weather forcing") (**P3**) *describes internal variability reliably on these timescales*, except for the impact of superposed pseudo-oscillatory climate modes, such as ENSO. With larger timescales, the contribution from internal variability arising from slow components of Earth dynamics, such as the cryosphere, becomes more important. At the period of approximately 1000 years, data from paleoclimate reconstructions and model simulations indicate a change in the behavior of global temperature fluctuations (Fig. 12). This finding can help reveal millennial-scale variability mechanisms. In particular, the *transition in fluctuation behavior might be due to changes in meltwater forcing and stabilizing feedbacks in response to it*. More research is needed to test this hypothesis. At the local level, temperature variability in simulation and reconstruction still deviates above decadal timescales. Thus, it is difficult to arrive at conclusions about mechanisms governing local variability. Nevertheless, the present results show that *sea ice and volcanic forcing influence temperature variability for decades* (**P2**). This is an interesting finding for understanding the continuum of temperature variability, as sea ice and volcanic forcing have much shorter characteristic timescales. However, the climate system's response to these processes integrates their dynamics to longer timescales. The representation of meltwater fluxes and coupled ice sheets can further improve the agreement of simulated temperatures with reconstructions on multicentennial scales (Fig. 11). Altogether, this strengthens the hypothesis that *more independent modes of variability and enhanced representations of feedbacks in climate model simulations are necessary to reconcile simulated and reconstructed temperature variability* (Laepple et al., 2023).

The contributions of this work to the above research objectives improved the understanding of reconstructed and simulated temperature variability, as well as the estimation of the future range of variability. The developed methods robustly estimated variability from proxy reconstructions and separated externally generated from internal variability in the Earth system. On this basis, they helped identify model experiments that simulate past climate processes best. The results shed light on the feedbacks from sea ice and meltwater flux at interannual to millennial timescales, which may be critical for estimating the ECS and the risk of abrupt changes. It can be concluded that trustworthy simulations of surface air temperature for decadal and longer-term projections require the inclusion of volcanic forcing, sea ice dynamics and interactive ice sheets. Future studies could continue to explore mechanisms of long-term variability by investigating the role of interactively coupling the carbon cycle and vegetation dynamics in simulations of temperature variability. Further efforts are needed to bridge the gaps between simulated and reconstructed temperature variability, providing the basis for reliable predictions of local impacts on society and ecosystems. In particular, a better understanding of temperature variability will benefit attribution studies, as well as projections of extremes, risks, and damages. The presented research on the structure and causes of temperature variability will support assessing future climate risks more accurately and developing optimal adaptation and mitigation strategies to temperature changes from natural and anthropogenic sources.

Appendix 1 (A-E)

A Mathematical proofs

Proof of Equation (6)

The Fourier transform and its inverse are here defined as

$$\hat{X}(\omega) := \mathcal{F}\{X(t)\}(\omega) = \int_{-\infty}^{\infty} X(t)e^{-i\omega t}dt \quad \text{and} \quad X(t) = \mathcal{F}^{-1}\{X(\omega)\}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{X}(\omega)e^{i\omega t}d\omega\,.$$

The complex conjugate of $X(\omega)$ is

$$\hat{X}^*(\omega) = \left(\int_{-\infty}^{\infty} X(t)\mathrm{e}^{-i\omega t}dt\right)^* = \hat{X}(-\omega)\,. \tag{13}$$

To derive mathematical equations for spectral analysis, $X(t)$ is treated as a real-valued ergodic process sampled over the time interval $t \in [-T/2, T/2]$ and zero otherwise. Therefore, the process is denoted by $X_T(t)$. Further, it is assumed that $X_T(t)$ has a zero mean $E[X_T(t)] = 0$. Then, the autocovariance function reads

$$\gamma(h) = E[X_T(t)X_T(t+h)] = \lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} X_T(t)X_T(t+h)\,dt\,.$$

Accordingly, the power spectral density becomes

$$\begin{aligned}
S(\omega) &= \int_{-\infty}^{\infty}\gamma(h)e^{-i\omega h}\,dh \\
&= \int_{-\infty}^{\infty}\left(\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} X_T(t)X_T(t+h)\,dt\right)e^{-i\omega h}dh \\
&\overset{\text{I}}{=} \int_{-\infty}^{\infty}\left(\lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty} X_T(t)X_T(t-h)\,dt\right)e^{-i\omega h}dh \\
&\overset{\text{II}}{=} \lim_{T\to\infty}\frac{1}{T}\mathcal{F}\{X_T(-t) * X_T(t)\}(\omega) \\
&\overset{\text{III}}{=} \lim_{T\to\infty}\frac{1}{T}|\hat{X}_T(\omega)|^2\,.
\end{aligned} \tag{14}$$

Step I follows from the symmetry $\gamma(h) = \gamma(-h)$. In step II, the integrand is rewritten as a convolution integral. Step III applies Equation (13) and the convolution theorem, which relates the convolution in time to a multiplication in the frequency domain. Equation (14) also shows why the compact support $t \in [-T/2, T/2]$ is needed. First, it assures that the Fourier transform exists because $X_T(t)$ is integrable over $t \in [-T/2, T/2]$. Second, it avoids convergence problems by preventing a diverging expectation value of $|X_T(t)|^2$.

Proof of expressions in Table 1

1. White noise process

The white noise process has the autocovariance function $\gamma(h) = \sigma_W^2\delta(h)$. As a result, the PSD

$$S(\omega) = \int_{-\infty}^{\infty}\sigma_W^2\delta(h)e^{-i\omega h}dh = \sigma_W^2$$

results directly from the property of the dirac delta function $\int_{-\infty}^{\infty}\delta(h)e^{-i\omega h}dh = 1\,.$

2. OU-process

The autocovariance function of an OU-process $dX(t) = -\lambda X(t)dt + \sigma_{OU}dW(t)$ reads

$$\gamma(h) = \frac{\sigma_{OU}^2}{2\lambda}e^{-\lambda|h|}\,.$$

The PSD results from

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} \frac{\sigma_{OU}^2}{2\lambda} e^{-\lambda|h|} e^{-i\omega h} dh \\ &= \frac{\sigma_{OU}^2}{2\lambda} \left(\int_{-\infty}^{0} e^{(\lambda - i\omega)h} dh + \int_{0}^{\infty} e^{-(\lambda + i\omega)h} dh \right) \\ &= \frac{\sigma_{OU}^2}{2\lambda} \left(\frac{1}{\lambda - i\omega} + \frac{1}{\lambda + i\omega} \right) \\ &= \frac{\sigma_{OU}^2}{2\lambda} \left(\frac{2\lambda}{\lambda^2 + \omega^2} \right) \\ &= \frac{\sigma_{OU}^2}{\lambda^2 + \omega^2} \,. \end{aligned}$$

3. Cosine function

For a sinusoidal function $A\cos(\omega_0 t + \theta)$ with random phase $\theta \in (0, 2\pi/\omega_0)$, the autocovariance reads

$$\begin{aligned} \gamma(h) &= \lim_{T\to\infty} \frac{A^2}{T} \int_{-T/2}^{T/2} \cos(\omega_0(t+h) + \theta) \cos(\omega_0 t + \theta) dt \\ &= \lim_{T\to\infty} \frac{A^2}{2T} \int_{-T/2}^{T/2} \cos(\omega_0(2t+h) + 2\theta) + \cos(\omega_0 h)\, dt \\ &= \frac{A^2}{2} \cos(\omega_0 h) \end{aligned}$$

For symmetry reasons, the integral over $\cos(\omega_0(2t+h) + 2\phi)$ vanishes. Using the Euler relation and properties of the Kronecker delta, the PSD becomes

$$\begin{aligned} S(\omega) &= \frac{A^2}{2} \int_{-\infty}^{\infty} \cos(\omega_0 h) e^{-i\omega h} dh \\ &= \frac{A^2}{4} \int_{-\infty}^{\infty} \left(e^{i(\omega_0 - \omega)h} + e^{-i(\omega_0 + \omega)h} \right) dh \\ &= \frac{A^2}{2} \pi \left(\delta(\omega_0 - \omega) + \delta(\omega_0 + \omega) \right) \,. \end{aligned}$$

Proof of Equation (9)

This proof shows that scale-invariant functions in one dimension, $f : \mathbb{R} \to \mathbb{R}$, are power-laws. First, scale invariance can be described by:

$$g(\alpha) f(x) = f(\alpha x), \quad \alpha \in \mathbb{R}^+, g : \mathbb{R}^+ \to \mathbb{R}^+ \tag{15}$$

Here, $g(\alpha)$ is an arbitrary function of the factor α. For simplicity, $g(\alpha)$ and α are assumed to be positive. Computing the derivative with respect to x and multiplying (15) by x yields

$$\frac{f'(x)}{f(x)} x = \frac{f'(\alpha x)}{f(\alpha x)} \alpha x \,.$$

This relation must hold for all values of $\alpha \in \mathbb{R}^+$. Thus, the ratio $\frac{f'(x)}{f(x)} x = b$ must be constant. Integration yields the form of f:

$$f(x) = c x^b$$

Hence, the scale-invariant function for a one-dimensional function is a power law. Independent of the remaining constant c, rescaling x by a factor of α leads to a change in $f(x)$ by a factor of $g(\alpha) = \alpha^b$. This intimately links the scale of x to the scale of $f(x)$. The parameter b is called the scaling exponent and can be estimated from the regression on the logarithm $\log(f(x)) = b \log(x) + \log(c)$.

B MPI-ESM model simulations

The complementary analysis in Section 7.2 studies variability in a seven-member ensemble (Kapsch et al., 2022; Mikolajewicz, 2023; Mikolajewicz & Kapsch, 2022) of climate model simulations with the Max Planck Institute for Meteorology Earth System Model (MPI-ESM) version 1.2 (Mauritsen et al., 2019) (Tab. 2 and Fig. 17). MPI-ESM version 1.2 includes the atmosphere general circulation model ECHAM6.3, the land surface model JSBACH3.2, the ocean biogeochemistry model HAMOCC6 and the dynamic ocean and sea ice component MPIOM1.6, as described in Figure 16 and Mauritsen et al. (2019). The atmosphere–land and ocean–marine biogeochemistry parts are coupled by the OASIS3-MCT component. The simulations were performed in coarse resolution. This includes ECHAM6.3 at T31 horizontal resolution ($\approx 3.75°$) and 31 vertical levels, coupled with daily frequency. JSBACH3.2 and MPIOM1.6 are run with a nominal resolution of $3°$ and 40 vertical levels.

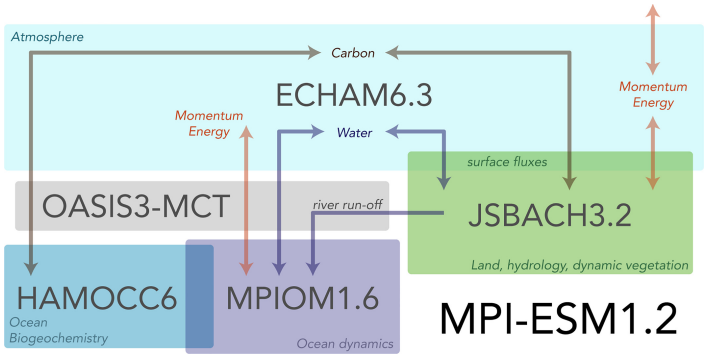


Figure 16: Schematic overview of the components of MPI-ESM version 1.2 and their interactions. The influence of external forcing and the ice-sheet component is not shown. Figure taken from Mauritsen et al. (2019).

The considered simulations represent transient model experiments that cover the period between the Last Glacial Maximum (26 kyrs ago) until present day (1950 CE), using prescribed orbital (Berger & Loutre, 1991) and CO_2 forcing (Köhler et al., 2017). Although not yet published, the simulations build on a series of MPI-ESM optimizations for transient deglaciation experiments (Erokhina, 2020; Kapsch et al., 2022; Ziemen et al., 2019). To investigate the contribution of cryosphere processes on temperature variability, some model experiments apply prescribed ice sheets, while others use an interactive coupling of an ice sheet model to the full model setup (Tab. 2).

Runs with prescribed ice sheets use the GLAC-1D (Briggs et al., 2014; Tarasov et al., 2012) or ICE-6G (Peltier et al., 2015) reconstruction with a native temporal resolution of 100 and 500 years, respectively. All forcing fields and the ice sheets are updated every ten years in the simulation and, for this, interpolated to decadal resolution (Kapsch et al., 2022). Ice sheet updates initiate changes in the topography, glacier mask, river pathways, ocean bathymetry, and land-sea mask (Kapsch et al., 2022; Meccia & Mikolajewicz, 2018; Riddick et al., 2018). Meltwater flux is computed from the temporal derivative in ice thickness and applied locally in the North Atlantic Ocean (Kapsch et al., 2022; Meccia & Mikolajewicz, 2018). The coupled ice sheet–climate model experiments use the modified Parallel Ice Sheet Model version 0.3 (Bueler & Brown, 2009; Winkelmann et al., 2011; Ziemen et al., 2019). Ice sheet changes directly affect the river routing and meltwater fluxes in these runs through the interactive coupling. All global mean temperature time series of the seven-member ensemble (Tab. 2) are shown in Figure 17.

Moreover, Section 7.2 quantifies the impact of volcanic forcing on temperature variability in the MPI-ESM simulation ensemble. Consequently, model experiments with and without volcanic forcing are contrasted for each ice sheet configuration (prescribed or coupled, Tab. 2). The volcanic forcing corresponds to a recent reconstruction for the last glacial cycle by Toohey et al. (2023) (Fig. 17 b). It was obtained by an ensemble reconstruction of volcanic stratospheric sulfur injection over the past 130,000 years, mainly using terrestrial and marine tephra records. To correct for the incompleteness

Run ID	Simulation type	Volcanic forcing	Ice sheet configuration	Resolution (yrs)	Meltwater forcing
M1	MPI-ESM (coupled)	-	interactively coupled ice sheet	10	interactively simulated
M2	MPI-ESM (coupled)	Toohey et al. (2023)	interactively coupled ice sheet	10	interactively simulated
M3	MPI-ESM (GLAC-1D)	-	prescribed, GLAC-1D (Briggs et al., 2014; Tarasov et al., 2012)	10	local meltwater
M4	MPI-ESM (GLAC-1D)	Toohey et al. (2023)	prescribed, GLAC-1D (Briggs et al., 2014; Tarasov et al., 2012)	10	local meltwater
M5	MPI-ESM (ICE-6G)	-	prescribed, ICE-6G (Peltier et al., 2015)	100	local meltwater
M6	MPI-ESM (ICE-6G)	-	prescribed, ICE-6G (Peltier et al., 2015)	100	globally distributed
M7	MPI-ESM (ICE-6G)	-	prescribed, ICE-6G (Peltier et al., 2015)	100	no meltwater

Table 2: Key specifications of considered MPI-ESM simulations performed by Uwe Mikolajewicz and Marie-Luise Kapsch (Kapsch et al., 2022; Mikolajewicz, 2023; Mikolajewicz & Kapsch, 2022). Local meltwater discharge is calculated from changes in ice sheet thickness. Globally distributed meltwater corresponds to an implementation that redistributes the local meltwater discharge homogeneously across all grid cells worldwide.

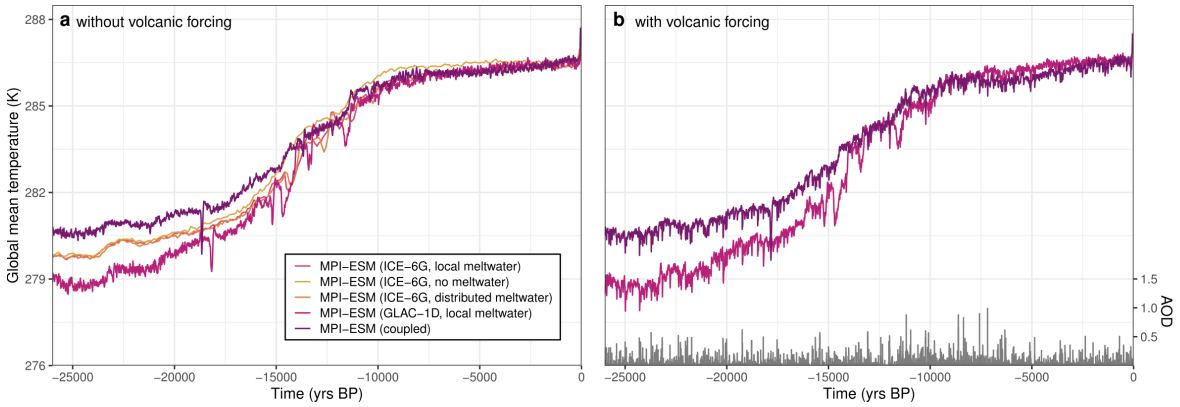


Figure 17: Global temperature evolution over the past 26 kyrs from the MPI-ESM simulation ensemble (Tab. 2) (Kapsch et al., 2022; Mikolajewicz, 2023; Mikolajewicz & Kapsch, 2022). Panel **a** displays runs without volcanic forcing. Panel **b** shows the volcanically-forced MPI-ESM (coupled) and MPI-ESM (GLAC-1D, local meltwater) runs as well as the considered aerosol optical depth (AOD) time series from Toohey et al. (2023).

of the compiled records, stochastically generated synthetic eruptions are added, assuming a constant background eruption frequency from the ice core Holocene record (Toohey et al., 2023). A more detailed description of all considered simulations is in preparation (Mikolajewicz, 2023).

C Economic damages from global temperature variability

The annual economic damage can be approximated from annual global mean temperature anomalies $\Delta T(t)$ compared to pre-industrial times (1850-1900 CE) following

$$D(\Delta T(t)) = \left(1 + a\,\Delta T(t)^2 + b\,\Delta T(t)^6\right)^{-1} ,$$

with empirically parameters $a = (1/20.46\,\mathrm{K})^2$ and $b = (1/6.081\,\mathrm{K})^6$ (Calel et al., 2020; Nordhaus, 2017; Weitzman, 2012). The economic loss $1 - D(\Delta T)$ is given as a share of global economic output (Fig. 15 a). The total monetary costs M from all economic losses over a time interval $0 < t < T$ is given by the difference between the expected potential and actual consumption (Calel et al., 2020)

$$M = \sum_{t=0}^{T} \left(\frac{(1+g)}{(1+r)}\right)^t c_0 P(t)(1 - D(\Delta T(t))) \,. \tag{16}$$

Free parameters correspond to the initial per capita consumption level c_0, global population $P(t)$, rate of consumption growth g, and discount rate r. Following Calel et al. (2020), the parameters are set to $c_0 = 10,666$ United States Dollars (USD), $r = 4.25\,\%$, and $g = 1.9\,\%$ (Calel et al., 2020) to represent realistic values for the year 2020 CE. The global population growth is modeled by $P(t+1) = P(t)^{1-\alpha} P_{\max}^{\alpha}$, assuming a population increases from $P(t=0) = 7.5$ billion in 2020 CE to an asymptotic value of $P_{\max} = \lim_{t\to\infty} P(t) = 11.5$ billion people at an annual rate of $\alpha = 2.68\,\%$ (Calel et al., 2020).

Sampling of internal and volcanically-forced temperature variability using ClimBayes

Samples of internal and volcanically-forced temperature variations $\Delta T(t)$ over the time period 2020-2500 CE were generated applying the ClimBayes packages (Schillinger et al., 2022b) and following **P3**. After fitting the stochastic two-box EBM to CMIP5 simulations, the EBM was used to run idealized experiments for 2020-2500 CE with and without volcanic forcing. Artificial volcanic forcing time series were obtained using the forcing generator by Ammann & Naveau (2010). The volcanic forcing reconstruction by Sigl et al. (2015) was used as input data. For each simulation, a new forcing sample was drawn from the distribution of eruption sizes (Sprinz, 2023). To convert aerosol optical depths (AOD) to radiative forcing, a factor of $-18\ \mathrm{Wm}^{-2}$ was used. Figure 18 displays exemplary temperature time series considering *internal*, *volcanic*, and *internal + volcanic* fluctuations.

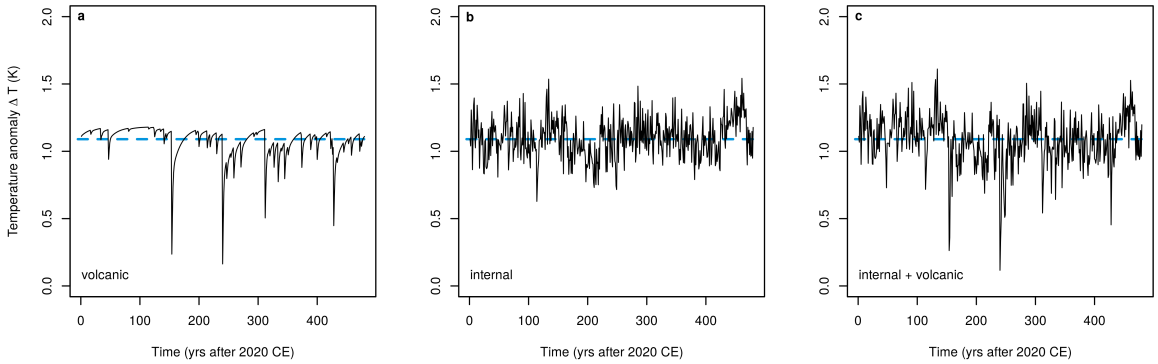


Figure 18: Example of emulated time series from the IPSL last millennium simulation (Dufresne et al., 2013; Hourdin et al., 2013) using the ClimBayes package (Schillinger et al., 2022b) and following **P3**. The emulation of global temperature anomalies can consider volcanically forced (**a**), internally generated (**b**), or both types (**c**) of fluctuations. Stephan Sprinz generated the data (Sprinz, 2023).

D Supplementary figures

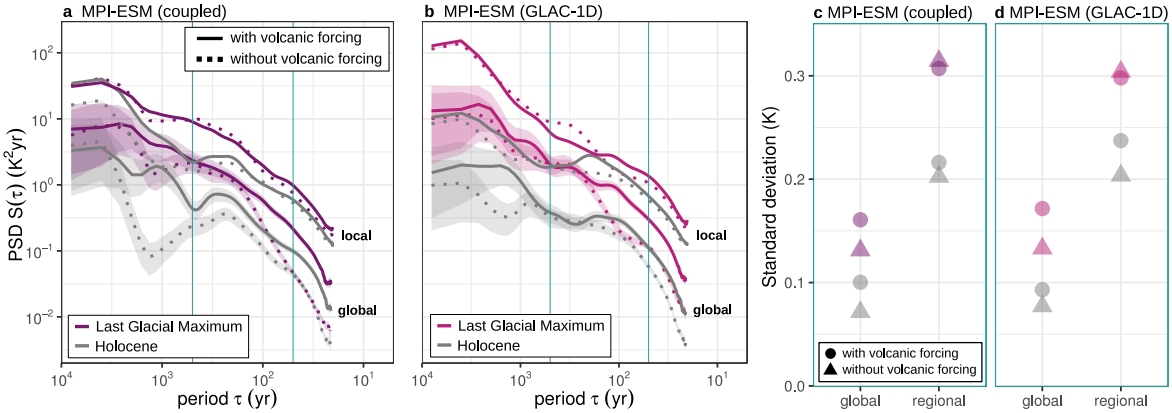


Figure 19: As Fig. 9, but comparing the forcing and state dependency of MPI-ESM (coupled) with interactively coupled ice sheets (**a** and **c**) and MPI-ESM (GLAC-1D) with prescribed ice sheets (**b** and **d**).

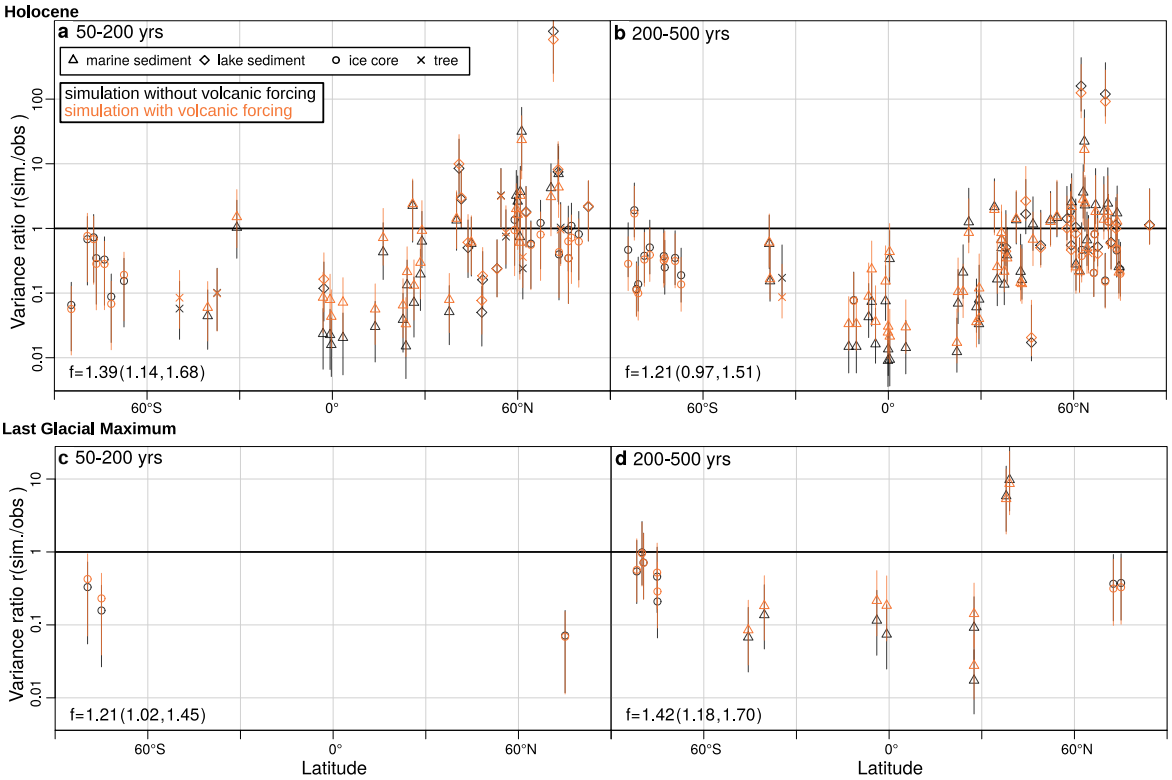


Figure 20: As Figure 10, but using the MPI-ESM (GLAC-1D) simulation with prescribed ice sheets.

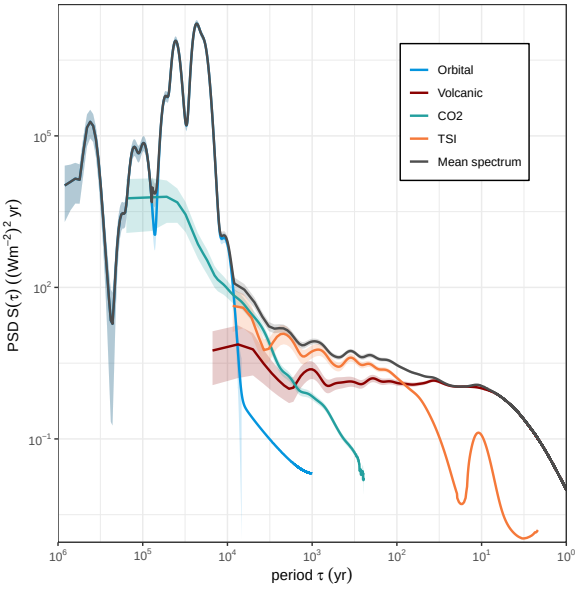


Figure 21: Composite spectral estimate of radiative forcing including orbital, volcanic, total solar irradiance (TSI) and carbon dioxide (CO2) forcing. The orbital forcing corresponds to radiative changes at 65°N (Berger, 1978), computed with the Palinsol Package (Crucifix, 2016) (as in **P1**). The solar forcing spectrum represents the mean of the spectral estimate from **P1** using the compiled data from Schmidt et al. (2012). Spectral estimates for solar forcing above centennial periods are obtained from surrogate time series. Surrogates are generated by applying the Amplitude Adjusted Fourier transform surrogate method (Theiler et al., 1992) based on TSI data from Steinhilber et al. (2009), as described by Wirths (2021). Changes in aerosol optical depth (AOD) due to volcanic eruptions are taken from Toohey et al. (2023). A scaling factor of $-18\ \mathrm{Wm}^{-2}$ was used to convert AOD values to effective radiative forcing (IPCC, 2021b). Atmospheric carbon dioxide concentrations correspond to the reconstruction by Köhler et al. (2017). The mean spectrum is formed as a sum of all individual spectra assuming independent and additive forcings (**P1**).

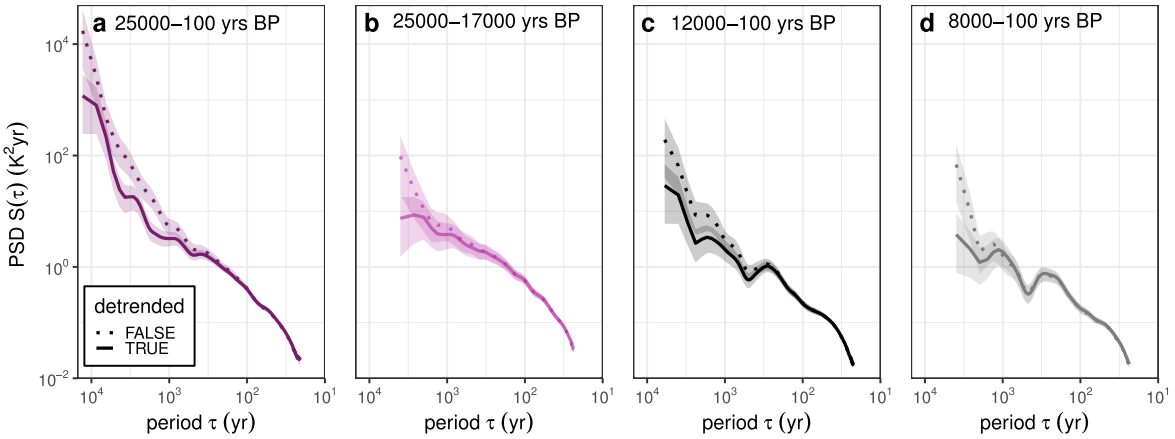


Figure 22: PSD of global temperature segments from the volcanically-forced MPI-ESM (coupled) simulation with and without linear detrending prior to spectral analysis. The time series represent the full transient run (**a**), the Last Glacial Maximum (**b**), the past twelve thousand years (**c**), and the past eight thousand years (**d**). The global warming trend (1850-1950 CE) was excluded from spectral analysis to isolate the effects from detrending the deglaciation. Detrending significantly impacts spectral analysis above millennial scales.

E Supplementary tables

Name	**Type**	**Time span**	**Resolution** (yrs)	**Reference**
Synder GMST	reconstruction	past 2 Million years	1000	Snyder (2016)
Temp12k	reconstruction	past 12 kiloyears	100	Kaufman et al. (2020)
PAGES2k	reconstruction	0-2000 CE	1	PAGES 2k Consortium (2019)
HadCRUT5	observation	1850-2019 CE	1	Morice et al. (2021)
CLIMBER-2	simulation	past 3 Million years	1000	Willeit et al. (2019)
Famous	simulation	past 120 kiloyears	100	Smith & Gregory (2012)
MPI-ESM (coupled)	simulation	past 26 kiloyears	10	Mikolajewicz (2023)
CMPI5 ensemble	simulation	850-2004 CE	1	BCC-CSM (Xiao-Ge et al., 2013), CCSM4 (Landrum et al., 2013), CESM-LME (Otto-Bliesner et al., 2016), CSIROMk3L-1-2 (Phipps et al., 2012), GISS-E2-R (Schmidt et al., 2006), HadCM3 (Schurer et al., 2013), MPI-ESM-P (Giorgetta et al., 2013; Jungclaus et al., 2013)

Table 3: Key specifications of datasets considered for the calculation of the global PSD presented in Figure 12. The resolution corresponds to the mean resolution of the considered time series after preprocessing. The references might also provide the datasets in higher resolutions. The PAGES2k and Temp12k present reconstruction ensembles of seven and five members, respectively. The CMIP5 ensemble corresponds to 23 simulations from seven models. Their simulated global temperature was aggregated annually by April–March as in PAGES 2k Consortium (2019). The MPI-ESM (coupled) simulation is described in Appendix B.

References

Ammann, C. M. & P. Naveau (2010). "A Statistical Volcanic Forcing Scenario Generator for Climate Simulations". *Journal of Geophysical Research* 115.D5. doi: 10.1029/2009JD012550.

Armstrong, E., K. Izumi & P. Valdes (2022). "Identifying the Mechanisms of DO-scale Oscillations in a GCM: A Salt Oscillator Triggered by the Laurentide Ice Sheet". *Climate Dynamics*. doi: 10.1007/s00382-022-06564-y.

Armstrong McKay, D. I., A. Staal, J. F. Abrams, R. Winkelmann, B. Sakschewski, S. Loriani, I. Fetzer, S. E. Cornell, J. Rockström & T. M. Lenton (2022). "Exceeding 1.5C Global Warming Could Trigger Multiple Climate Tipping Points". *Science* 377.6611. doi: 10.1126/science.abn7950.

Arnscheidt, C. W. & D. H. Rothman (2022). "Presence or Absence of Stabilizing Earth System Feedbacks on Different Time Scales". *Science Advances* 8.46. doi: 10.1126/sciadv.adc9241.

Ashwin, P., S. Wieczorek, R. Vitolo & P. Cox (2012). "Tipping Points in Open Systems: Bifurcation, Noise-Induced and Rate-Dependent Examples in the Climate System". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370.1962. doi: 10.1098/rsta.2011.0306.

Aubry, T. J., J. I. Farquharson, C. R. Rowell, S. F. L. Watt, V. Pinel, F. Beckett, J. Fasullo, P. O. Hopcroft, D. M. Pyle, A. Schmidt & J. S. Sykes (2022). "Impact of Climate Change on Volcanic Processes: Current Understanding and Future Challenges". *Bulletin of Volcanology* 84.6. doi: 10.1007/s00445-022-01562-8.

Bakker, P., P. U. Clark, N. R. Golledge, A. Schmittner & M. E. Weber (2017). "Centennial-Scale Holocene Climate Variations Amplified by Antarctic Ice Sheet Discharge". *Nature* 541.7635. doi: 10.1038/nature20582.

Bathiany, S., V. Dakos, M. Scheffer & T. M. Lenton (2018). "Climate Models Predict Increasing Temperature Variability in Poor Countries". *Science Advances* 4.5. doi: 10.1126/sciadv.aar5809.

Baudouin, J.-P., O. Bothe, M. Chevalier, B. Ellerhoff, M. Adam, P. Schoch, N. Weitzel & K. Rehfeld (2022). "PTBox, a Toolbox to Facilitate Palaeoclimate Model-Data Analyses". *EGU General Assembly 2022*. Vol. EGU22-3684. Vienna, Austria. doi: 10.5194/egusphere-egu22-3684.

Baumstark, L. et al. (2021). "REMIND2.1: Transformation and Innovation Dynamics of the Energy-Economic System within Climate and Sustainability Limits". *Geoscientific Model Development* 14.10. doi: 10.5194/gmd-14-6571-2021.

Benzi, R., G. Parisi, A. Sutera & A. Vulpiani (1983). "A Theory of Stochastic Resonance in Climatic Change". *SIAM Journal on Applied Mathematics* 43.3. doi: 10.1137/0143037.

Berdahl, M. & A. Robock (2013). "Northern Hemispheric Cryosphere Response to Volcanic Eruptions in the Paleoclimate Modeling Intercomparison Project 3 Last Millennium Simulations". *Journal of Geophysical Research Atmospheres* 118.22. doi: 10.1002/2013JD019914.

Berger, A. & M. Loutre (1991). "Insolation Values for the Climate of the Last 10 Million Years". *Quaternary Science Reviews* 10.4. doi: 10.1016/0277-3791(91)90033-Q.

Berger, A. (1978). "Long-Term Variations of Daily Insolation and Quaternary Climatic Changes". *Journal of the Atmospheric Sciences* 35.12. doi: 10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2.

Berger, W. H. & G. R. Heath (1968). "Vertical Mixing in Pelagic Sediments". 26.2. url: https://images.peabody.yale.edu/publications/jmr/jmr26-02-05.pdf.

Bethke, I., S. Outten, O. H. Otterå, E. Hawkins, S. Wagner, M. Sigl & P. Thorne (2017). "Potential Volcanic Impacts on Future Climate Variability". *Nature Climate Change* 7.11. doi: 10.1038/nclimate3394.

Birks, B. & H. John (2012). "Strengths and Weaknesses of Quantitative Climate Reconstructions Based on Late-Quaternary Biological Proxies". *Quaternary International* 279–280. doi: 10.1016/j.quaint.2012.07.228.

Blanusa, M. L., C. J. López-Zurita & S. Rasp (2022). "The Role of Internal Variability in Global Climate Projections of Extreme Events". *Arxiv Preprint*. doi: 10.48550/ARXIV.2208.08275.

Boers, N., M. Ghil & T. F. Stocker (2022). "Theoretical and Paleoclimatic Evidence for Abrupt Transitions in the Earth System". *Environmental Research Letters* 17.9. doi: 10.1088/1748-9326/ac8944.

Boucher, O. et al. (2020). "Presentation and Evaluation of the IPSL-CM6A-LR Climate Model". *Journal of Advances in Modeling Earth Systems* 12.7. doi: 10.1029/2019MS002010.

Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner & Y. Zhao (2012). "Evaluation of Climate Models Using Palaeoclimatic Data". *Nature Climate Change* 2.6. doi: 10.1038/nclimate1456.

Braconnot, P., D. Zhu, O. Marti & J. Servonnat (2019). "Strengths and Challenges for Transient Mid- to Late Holocene Simulations with Dynamical Vegetation". *Climate of the Past* 15.3. doi: 10.5194/cp-15-997-2019.

Braun, H., P. Ditlevsen, J. Kurths & M. Mudelsee (2010). "Limitations of Red Noise in Analysing Dansgaard-Oeschger Events". *Climate of the Past* 6.1. doi: 10.5194/cp-6-85-2010.

Brierley, C. M. et al. (2020). "Large-Scale Features and Evaluation of the PMIP4-CMIP6 midHolocene Simulations". *Climate of the Past* 16.5. doi: 10.5194/cp-16-1847-2020.

Briggs, R. D., D. Pollard & L. Tarasov (2014). "A Data-Constrained Large Ensemble Analysis of Antarctic Evolution since the Eemian". *Quaternary Science Reviews* 103. doi: 10.1016/j.quascirev.2014.09.003.

Brown, P. T., Y. Ming, W. Li & S. A. Hill (2017). "Change in the Magnitude and Mechanisms of Global Temperature Variability with Warming". *Nature Climate Change* 7.10. doi: 10.1038/nclimate3381.

Budyko, M. I. (1969). "The Effect of Solar Radiation Variations on the Climate of the Earth". *Tellus* 21.5. doi: 10.3402/tellusa.v21i5.10109.

Bueler, E. & J. Brown (2009). "Shallow Shelf Approximation as a "Sliding Law" in a Thermomechanically Coupled Ice Sheet Model". *Journal of Geophysical Research* 114.F3. doi: 10.1029/2008JF001179.

Bühler, J. C., C. Roesch, M. Kirschner, L. Sime, M. D. Holloway & K. Rehfeld (2021). "Comparison of the Oxygen Isotope Signatures in Speleothem Records and iHadCM3 Model Simulations for the Last Millennium". *Climate of the Past* 17.3. doi: 10.5194/cp-17-985-2021.

Büntgen, U. et al. (2022). "Recognising Bias in Common Era Temperature Reconstructions". *Dendrochronologia* 74. doi: 10.1016/j.dendro.2022.125982.

Calel, R., S. C. Chapman, D. A. Stainforth & N. W. Watkins (2020). "Temperature Variability Implies Greater Economic Damages from Climate Change". *Nature Communications* 11.1. doi: 10.1038/s41467-020-18797-8.

Casado, M., T. Münch & T. Laepple (2020). "Climatic Information Archived in Ice Cores: Impact of Intermittency and Diffusion on the Recorded Isotopic Signal in Antarctica". *Climate of the Past* 16.4. doi: 10.5194/cp-16-1581-2020.

Ceppi, P., F. Brient, M. D. Zelinka & D. L. Hartmann (2017). "Cloud Feedback Mechanisms and Their Representation in Global Climate Models". *WIREs Climate Change* 8.4. doi: 10.1002/wcc.465.

Chatfield, C. (2003). *The Analysis of Time Series*. Chapman and Hall/CRC. isbn: 978-0-203-49168-3. doi: 10.4324/9780203491683.

Cheung, A. H., M. E. Mann, B. A. Steinman, L. M. Frankcombe, M. H. England & S. K. Miller (2017). "Comparison of Low-Frequency Internal Climate Variability in CMIP5 Models and Observations". *Journal of Climate* 30.12. doi: 10.1175/JCLI-D-16-0712.1.

Christiansen, B. & F. C. Ljungqvist (2017). "Challenges and Perspectives for Large-Scale Temperature Reconstructions of the Past Two Millennia". *Reviews of Geophysics* 55.1. doi: 10.1002/2016RG000521.

Chylek, P., C. Folland, J. D. Klett & M. K. Dubey (2020). "CMIP5 Climate Models Overestimate Cooling by Volcanic Aerosols". *Geophysical Research Letters* 47.3. doi: 10.1029/2020GL087047.

Clark, P. U., D. Archer, D. Pollard, J. D. Blum, J. A. Rial, V. Brovkin, A. C. Mix, N. G. Pisias & M. Roy (2006). "The Middle Pleistocene Transition: Characteristics, Mechanisms, and Implications for Long-Term Changes in Atmospheric pCO2". *Quaternary Science Reviews* 25.23-24. doi: 10.1016/j.quascirev.2006.07.008.

Clauset, A., C. R. Shalizi & M. E. J. Newman (2009). "Power-Law Distributions in Empirical Data". *SIAM Review* 51.4. doi: 10.1137/070710111.

Comas-Bru, L., K. Atsawawaranunt, S. Harrison & S. W. G. Members (2020). *SISAL (Speleothem Isotopes Synthesis and AnaLysis Working Group) Database Version 2.0*. doi: 10.17864/1947.256.

Cronin, T. M. (2010). *Paleoclimates: Understanding Climate Change Past and Present*. New York: Columbia University Press. isbn: 978-0-231-14494-0 978-0-231-51636-5. url: https://cup.columbia.edu/book/paleoclimates/9780231144940.

Crucifix, M. (2012). "Oscillators and Relaxation Phenomena in Pleistocene Climate Theory". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370.1962. doi: 10.1098/rsta.2011.0315.

– (2016). *Palinsol Package*. url: https://CRAN.R-project.org/package=palinsol.

Crucifix, M. & J. Annan (2019). "Is the Concept of 'Tipping Point' Helpful for Describing and Communicating Possible Climate Futures?" *Contemporary Climate Change Debates*. Routledge. isbn: 978-0-429-44625-2. url: https://doi.org/10.4324/9780429446252.

Czymzik, M., R. Muscheler & A. Brauer (2016). "Solar Modulation of Flood Frequency in Central Europe during Spring and Summer on Interannual to Multi-Centennial Timescales". *Climate of the Past* 12.3. doi: 10.5194/cp-12-799-2016.

Dansgaard, W., S. J. Johnsen, H. B. Clausen, D. Dahl-Jensen, N. S. Gundestrup, C. U. Hammer, C. S. Hvidberg, J. P. Steffensen, A. E. Sveinbjörnsdottir, J. Jouzel & G. Bond (1993). "Evidence for General Instability of Past Climate from a 250-Kyr Ice-Core Record". *Nature* 364.6434. doi: 10.1038/364218a0.

Deser, C., A. Phillips, V. Bourdette & H. Teng (2012). "Uncertainty in Climate Change Projections: The Role of Internal Variability". *Climate Dynamics* 38.3-4. doi: 10.1007/s00382-010-0977-x.

Ditlevsen, P., T. Mitsui & M. Crucifix (2020). "Crossover and Peaks in the Pleistocene Climate Spectrum; Understanding from Simple Ice Age Models". *Climate Dynamics* 54.3-4. doi: 10.1007/s00382-019-05087-3.

Ditlevsen, P. D. & P. Ashwin (2018). "Complex Climate Response to Astronomical Forcing: The Middle-Pleistocene Transition in Glacial Cycles and Changes in Frequency Locking". *Frontiers in Physics* 6. doi: 10.3389/fphy.2018.00062.

Ditlevsen, P. D. & S. J. Johnsen (2010). "Tipping Points: Early Warning and Wishful Thinking". *Geophysical Research Letters* 37.19. doi: 10.1029/2010GL044486.

Dolman, A. M., J. Groeneveld, G. Mollenhauer, S. L. Ho & T. Laepple (2021). "Estimating Bioturbation From Replicated Small-Sample Radiocarbon Ages". *Paleoceanography and Paleoclimatology* 36.7. doi: 10.1029/2020PA004142.

Dolman, A. M. & T. Laepple (2018). "Sedproxy: A Forward Model for Sediment-Archived Climate Proxies". *Climate of the Past* 14.12. doi: 10.5194/cp-14-1851-2018.

Dortmans, B., W. F. Langford & A. R. Willms (2019). "An Energy Balance Model for Paleoclimate Transitions". *Climate of the Past* 15.2. doi: 10.5194/cp-15-493-2019.

Dufresne, J.-L. et al. (2013). "Climate Change Projections Using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5". *Climate Dynamics* 40.9-10. doi: 10.1007/s00382-012-1636-1.

Ellerhoff, B., M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime & K. Rehfeld (2022). "Contrasting State-Dependent Effects of Natural Forcing on Global and Local Climate Variability". *Geophysical Research Letters* 49.10. doi: 10.1029/2022GL098335.

Ellerhoff, B. & K. Rehfeld (2021). "Probing the Timescale Dependency of Local and Global Variations in Surface Air Temperature from Climate Simulations and Reconstructions of the Last Millennia". *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 104.6. doi: 10.1103/PhysRevE.104.064136.

Erokhina, O. (2020). "A New Eulerian Iceberg Module for Climate Studies: Formulation and Application to the Investigation of the Sensitivity of the AMOC to Iceberg Calving". *PhD Thesis, Universität Hamburg, Hamburg*. doi: 10.17617/2.3270743.

Evans, M., S. Tolwinski-Ward, D. Thompson & K. Anchukaitis (2013). "Applications of Proxy System Modeling in High Resolution Paleoclimatology". *Quaternary Science Reviews* 76. doi: 10.1016/j.quascirev.2013.05.024.

Eyring, V. et al. (2019). "Taking Climate Model Evaluation to the next Level". *Nature Climate Change* 9.2. doi: 10.1038/s41558-018-0355-y.

Falkowski, P. et al. (2000). "The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System". *Science* 290.5490. doi: 10.1126/science.290.5490.291.

Fasullo, J. T., A. S. Phillips & C. Deser (2020). "Evaluation of Leading Modes of Climate Variability in the CMIP Archives". *Journal of Climate* 33.13. doi: 10.1175/JCLI-D-19-1024.1.

Fergus, G. (2014). *Wikimedia Commons: Global Average Temperature Estimates for the Last 540 My (CC BY-SA 3.0 License)*. url: https://commons.wikimedia.org/wiki/File:All_palaeotemps.svg.

Fernández-Donado, L. et al. (2013). "Large-Scale Temperature Response to External Forcing in Simulations and Reconstructions of the Last Millennium". *Climate of the Past* 9.1. doi: 10.5194/cp-9-393-2013.

Fischer, E. M. & R. Knutti (2015). "Anthropogenic Contribution to Global Occurrence of Heavy-Precipitation and High-Temperature Extremes". *Nature Climate Change* 5.6. doi: 10.1038/nclimate2617.

Forster, P. M. (2016). "Inference of Climate Sensitivity from Analysis of Earth's Energy Budget". *Annual Review of Earth and Planetary Sciences* 44.1. doi: 10.1146/annurev-earth-060614-105156.

Fraedrich, K., U. Luksch & R. Blender (2004). "1 / f Model for Long-Time Memory of the Ocean Surface Temperature". *Physical Review E* 70.3. doi: 10.1103/PhysRevE.70.037301.

Frankcombe, L. M., M. H. England, M. E. Mann & B. A. Steinman (2015). "Separating Internal Variability from the Externally Forced Climate Response". *Journal of Climate* 28.20. doi: 10.1175/JCLI-D-15-0069.1.

Franzke, C. (2012). "Nonlinear Trends, Long-Range Dependence, and Climate Noise Properties of Surface Temperature". *Journal of Climate* 25.12. doi: 10.1175/JCLI-D-11-00293.1.

Franzke, C. L. E., S. M. Osprey, P. Davini & N. W. Watkins (2015). "A Dynamical Systems Explanation of the Hurst Effect and Atmospheric Low-Frequency Variability". *Scientific Reports* 5.1. doi: 10.1038/srep09068.

Franzke, C. L. E. et al. (2020). "The Structure of Climate Variability Across Scales". *Reviews of Geophysics* 58.2. doi: 10.1029/2019RG000657.

Fredriksen, H.-B. & K. Rypdal (2016). "Spectral Characteristics of Instrumental and Climate Model Surface Temperatures". *Journal of Climate* 29.4. doi: 10.1175/JCLI-D-15-0457.1.

Fredriksen, H.-B. & M. Rypdal (2017). "Long-Range Persistence in Global Surface Temperatures Explained by Linear Multibox Energy Balance Models". *Journal of Climate* 30.18. doi: 10.1175/JCLI-D-16-0877.1.

Fyfe, J. C., V. V. Kharin, B. D. Santer, J. N. S. Cole & N. P. Gillett (2021). "Significant Impact of Forcing Uncertainty in a Large Ensemble of Climate Model Simulations". *Proceedings of the National Academy of Sciences* 118.23. doi: 10.1073/pnas.2016549118.

Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon & S. Tytéca (2013). "Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter

Calibration Using CMIP5 AOGCM Experiments". *Journal of Climate* 26.6. doi: 10.1175/JCLI-D-12-00195.1.

Ghil, M. (1976). "Climate Stability for a Sellers-Type Model". *Journal of the Atmospheric Sciences* 33.1. doi: 10.1175/1520-0469(1976)033<0003:CSFAST>2.0.CO;2.

– (1984). "Climate Sensitivity, Energy Balance Models, and Oscillatory Climate Models". *Journal of Geophysical Research* 89.D1. doi: 10.1029/JD089iD01p01280.

Ghil, M. & V. Lucarini (2020). "The Physics of Climate Variability and Climate Change". *Reviews of Modern Physics* 92.3. doi: 10.1103/RevModPhys.92.035002.

Giorgetta, M. A. et al. (2013). "Climate and Carbon Cycle Changes from 1850 to 2100 in MPI-ESM Simulations for the Coupled Model Intercomparison Project Phase 5: Climate Changes in MPI-ESM". *Journal of Advances in Modeling Earth Systems* 5.3. doi: 10.1002/jame.20038.

Goosse, H., H. Renssen, A. Timmermann & R. S. Bradley (2005). "Internal and Forced Climate Variability during the Last Millennium: A Model-Data Comparison Using Ensemble Simulations". *Quaternary Science Reviews* 24.12-13. doi: 10.1016/j.quascirev.2004.12.009.

Gutenberg, B. & C. F. Richter (1944). "Frequency of Earthquakes in California*". *Bulletin of the Seismological Society of America* 34.4. doi: 10.1785/BSSA0340040185.

Hansen, J., M. Sato, G. Russell & P. Kharecha (2013). "Climate Sensitivity, Sea Level and Atmospheric Carbon Dioxide". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.2001. doi: 10.1098/rsta.2012.0294.

Harrington, L. J., C.-F. Schleussner & F. E. L. Otto (2021). "Quantifying Uncertainty in Aggregated Climate Change Risk Assessments". *Nature Communications* 12.1. doi: 10.1038/s41467-021-27491-2.

Hasselmann, K. (1976). "Stochastic Climate Models Part I. Theory". *Tellus* 28.6. doi: 10.3402/tellusa.v28i6.11316.

Hawkins, E. & R. Sutton (2011). "The Potential to Narrow Uncertainty in Projections of Regional Precipitation Change". *Climate Dynamics* 37.1-2. doi: 10.1007/s00382-010-0810-6.

Hébert, R., U. Herzschuh & T. Laepple (2022). "Millennial-Scale Climate Variability over Land Overprinted by Ocean Temperature Fluctuations". *Nature Geoscience* 15.11. doi: 10.1038/s41561-022-01056-4.

Hébert, R. & S. Lovejoy (2018). "Regional Climate Sensitivity- and Historical-Based Projections to 2100". *Geophysical Research Letters* 45.9. doi: 10.1002/2017GL076649.

Hébert, R., K. Rehfeld & T. Laepple (2021). "Comparing Estimation Techniques for Temporal Scaling in Palaeoclimate Time Series". *Nonlinear Processes in Geophysics* 28.3. doi: 10.5194/npg-28-311-2021.

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch & P. D. Jones (1996). "Detecting Greenhouse-Gas-Induced Climate Change with an Optimal Fingerprint Method". *Journal of Climate* 9.10. doi: 10.1175/1520-0442(1996)009<2281:DGGICC>2.0.CO;2.

Heinrich, H. (1988). "Origin and Consequences of Cyclic Ice Rafting in the Northeast Atlantic Ocean During the Past 130,000 Years". *Quaternary Research* 29.2. doi: 10.1016/0033-5894(88)90057-9.

Henry, L. G., J. F. McManus, W. B. Curry, N. L. Roberts, A. M. Piotrowski & L. D. Keigwin (2016). "North Atlantic Ocean Circulation and Abrupt Climate Change during the Last Glaciation". *Science* 353.6298. doi: 10.1126/science.aaf5529.

Holmes, C. R., T. Woollings, E. Hawkins & H. de Vries (2016). "Robust Future Changes in Temperature Variability under Greenhouse Gas Forcing and the Relationship with Thermal Advection". *Journal of Climate* 29.6. doi: 10.1175/JCLI-D-14-00735.1.

Hopcroft, P. O., J. Kandlbauer, P. J. Valdes & R. S. J. Sparks (2018). "Reduced Cooling Following Future Volcanic Eruptions". *Climate Dynamics* 51.4. doi: 10.1007/s00382-017-3964-7.

Hopcroft, P. O. & P. J. Valdes (2021). "Paleoclimate-Conditioning Reveals a North Africa Land–Atmosphere Tipping Point". *Proceedings of the National Academy of Sciences* 118.45. doi: 10.1073/pnas.2108783118.

Hourdin, F. et al. (2013). “Impact of the LMDZ Atmospheric Grid Configuration on the Climate and Sensitivity of the IPSL-CM5A Coupled Model”. *Climate Dynamics* 40.9-10. doi: 10.1007/s00382-012-1411-3.

Huber, M., U. Beyerle & R. Knutti (2014). “Estimating Climate Sensitivity and Future Temperature in the Presence of Natural Climate Variability”. *Geophysical Research Letters* 41.6. doi: 10.1002/2013GL058532.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton & P. M. Cox (2013). “No Increase in Global Temperature Variability despite Changing Regional Patterns”. *Nature*. doi: 10.1038/nature12310.

Hurst, H. E. (1951). “Long-Term Storage Capacity of Reservoirs”. *Transactions of the American Society of Civil Engineers* 116.1. doi: 10.1061/TACEAT.0006518.

Huybers, P. & W. Curry (2006). “Links between Annual, Milankovitch and Continuum Temperature Variability”. *Nature* 441.7091. doi: 10.1038/nature04745.

Ionita, M., M. Dima, V. Nagavciuc, P. Scholz & G. Lohmann (2021). “Past Megadroughts in Central Europe Were Longer, More Severe and Less Warm than Modern Droughts”. *Communications Earth & Environment* 2.1. doi: 10.1038/s43247-021-00130-w.

IPCC (2021a). “Annex IV: Modes of Variability”. Climate Change 2021: The Physical Science Basis. Cambridge, UK, and New York, NY, USA: Cambridge University Press. url: https://www.ipcc.ch/report/ar6/wg1/.

– (2021b). “Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change”. *Climate Change 2021: The Physical Science Basis*. Cambridge, UK, and New York, NY, USA: Cambridge University Press. url: https://www.ipcc.ch/report/ar6/wg1/.

– (2022a). “Annex II: Glossary”. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Climate Change 2022: Impacts, Adaptation and Vulnerability. Cambridge, UK, and New York, NY, USA: Cambridge University Press. url: https://www.ipcc.ch/report/ar6/wg2/.

– (2022b). “Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change”. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. Cambridge, UK, and New York, NY, USA: Cambridge University Press. url: https://www.ipcc.ch/report/ar6/wg2/.

Johnsen, S. J., H. B. Clausen, K. M. Cuffey, G. Hoffmann, J. Schwander & T. Creyts (2000). “Diffusion of Stable Isotopes in Polar Firn and Ice: The Isotope Effect in Firn Diffusion”. *Physics of Ice Core Records*. Hokkaido University Press. doi: 10.7916/D8KW5D4X.

Jouzel, J. et al. (2007). “Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years”. *Science* 317.5839. doi: 10.1126/science.1141038.

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz & J. S. Storch (2013). “Characteristics of the Ocean Simulations in the Max Planck Institute Ocean Model (MPIOM) the Ocean Component of the MPI-Earth System Model”. *Journal of Advances in Modeling Earth Systems* 5.2. doi: 10.1002/jame.20023.

Jungclaus, J. H. (2020). *Personal Communication*.

Juricke, S., T. N. Palmer & L. Zanna (2017). “Stochastic Subgrid-Scale Ocean Mixing: Impacts on Low-Frequency Variability”. *Journal of Climate* 30.13. doi: 10.1175/JCLI-D-16-0539.1.

Kapsch, M.-L., U. Mikolajewicz, F. Ziemen & C. Schannwell (2022). “Ocean Response in Transient Simulations of the Last Deglaciation Dominated by Underlying Ice-Sheet Reconstruction and Method of Meltwater Distribution”. *Geophysical Research Letters* 49.3. doi: 10.1029/2021GL096767.

Katz, R. W. & B. G. Brown (1992). “Extreme Events in a Changing Climate: Variability Is More Important than Averages”. *Climatic Change* 21.3. doi: 10.1007/BF00139728.

Kaufman, D., N. McKay, C. Routson, M. Erb, C. Dätwyler, P. S. Sommer, O. Heiri & B. Davis (2020). “Holocene Global Mean Surface Temperature, a Multi-Method Reconstruction Approach”. *Scientific Data* 7.1. doi: 10.1038/s41597-020-0530-7.

Kay, J. E. et al. (2015). "The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability". *Bulletin of the American Meteorological Society* 96.8. doi: 10.1175/BAMS-D-13-00255.1.

Khintchine, A. (1934). "Korrelationstheorie der stationären stochastischen Prozesse". *Mathematische Annalen* 109.1. doi: 10.1007/BF01449156.

Klockmann, M., U. Mikolajewicz, H. Kleppin & J. Marotzke (2020). "Coupling of the Subpolar Gyre and the Overturning Circulation During Abrupt Glacial Climate Transitions". *Geophysical Research Letters* 47.21. doi: 10.1029/2020GL090361.

Knutti, R. & M. A. A. Rugenstein (2015). "Feedbacks, Climate Sensitivity and the Limits of Linear Models". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373.2054. doi: 10.1098/rsta.2015.0146.

Knutti, R. & J. Sedláček (2013). "Robustness and Uncertainties in the New CMIP5 Climate Model Projections". *Nature Climate Change* 3.4. doi: 10.1038/nclimate1716.

Köhler, P., C. Nehrbass-Ahles, J. Schmitt, T. F. Stocker & H. Fischer (2017). "A 156 Kyr Smoothed History of the Atmospheric Greenhouse Gases CO_2, CH_4, and N_2O and Their Radiative Forcing". *Earth System Science Data* 9.1. doi: 10.5194/essd-9-363-2017.

Krey, V. (2014). "Global Energy-Climate Scenarios and Models: A Review: Global Energy-Climate Scenarios and Models". *Wiley Interdisciplinary Reviews: Energy and Environment* 3.4. doi: 10.1002/wene.98.

Laepple, T. & P. Huybers (2013). "Reconciling Discrepancies between Uk37 and Mg/Ca Reconstructions of Holocene Marine Temperature Variability". *Earth and Planetary Science Letters* 375. doi: 10.1016/j.epsl.2013.06.006.

– (2014a). "Global and Regional Variability in Marine Surface Temperatures". *Geophysical Research Letters* 41.7. doi: 10.1002/2014GL059345.

– (2014b). "Ocean Surface Temperature Variability: Large Model–Data Differences at Decadal and Longer Periods". *Proceedings of the National Academy of Sciences* 111.47. doi: 10.1073/pnas.1412077111.

Laepple, T., E. Ziegler, N. Weitzel, R. Hébert, B. Ellerhoff, P. Schoch, B. Martrat, O. Bothe, E. Moreno-Chamarro, M. Chevalier, A. Herbert & K. Rehfeld (2023). "The Enigma of Multidecadal to Centennial Temperature Variability in Climate Simulations and Reconstructions". *submitted*.

Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom & H. Teng (2013). "Last Millennium Climate and Its Variability in CCSM4". *Journal of Climate* 26.4. doi: 10.1175/JCLI-D-11-00326.1.

Lehner, F., C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti & E. Hawkins (2020). "Partitioning Climate Projection Uncertainty with Multiple Large Ensembles and CMIP5/6". *Earth System Dynamics* 11.2. doi: 10.5194/esd-11-491-2020.

Lenton, T. M., H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf & H. J. Schellnhuber (2008). "Tipping Elements in the Earth's Climate System". *Proceedings of the National Academy of Sciences* 105.6. doi: 10.1073/pnas.0705414105.

Lenton, T. M., J. Rockström, O. Gaffney, S. Rahmstorf, K. Richardson, W. Steffen & H. J. Schellnhuber (2019). "Climate Tipping Points — Too Risky to Bet Against". *Nature* 575.7784. doi: 10.1038/d41586-019-03595-0.

Lenton, T. M. & H. J. Schellnhuber (2007). "Tipping the Scales". *Nature Climate Change* 1.712. doi: 10.1038/climate.2007.65.

Liebrand, D. & A. T. M. de Bakker (2019). "Bispectra of Climate Cycles Show How Ice Ages Are Fuelled". *Climate of the Past* 15.6. doi: 10.5194/cp-15-1959-2019.

Lisiecki, L. E. & M. E. Raymo (2005). "A Pliocene-Pleistocene Stack of 57 Globally Distributed Benthic δ^{18} O Records". *Paleoceanography* 20.1. doi: 10.1029/2004PA001071.

Lovejoy, S. & D. Schertzer (2012). "Stochastic and Scaling Climate Sensitivities: Solar, Volcanic and Orbital Forcings". *Geophysical Research Letters* 39.11. doi: 10.1029/2012GL051871.

Lovejoy, S. (2015). "A Voyage through Scales, a Missing Quadrillion and Why the Climate Is Not What You Expect". *Climate Dynamics* 44.11-12. doi: 10.1007/s00382-014-2324-0.

Lücke, L. J., G. C. Hegerl, A. P. Schurer & R. Wilson (2019). "Effects of Memory Biases on Variability of Temperature Reconstructions". *Journal of Climate* 32.24. doi: 10.1175/JCLI-D-19-0184.1.

MacMynowski, D. G., H.-J. Shin & K. Caldeira (2011). "The Frequency Response of Temperature and Precipitation in a Climate Model". *Geophysical Research Letters* 38.16. doi: 10.1029/2011GL048623.

Maher, N., S. McGregor, M. H. England & A. S. Gupta (2015). "Effects of Volcanism on Tropical Variability: Effects of Volcanism". *Geophysical Research Letters* 42.14. doi: 10.1002/2015GL064751.

Manabe, S. & R. T. Wetherald (1967). "Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity". *Journal of the Atmospheric Sciences* 24.3. doi: 10.1175/1520-0469(1967)024<0241:TEOTAW>2.0.CO;2.

Mandelbrot, B. B. & J. W. Van Ness (1968). "Fractional Brownian Motions, Fractional Noises and Applications". *SIAM Review* 10.4. doi: 10.1137/1010093.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford & F. Ni (2008). "Proxy-Based Reconstructions of Hemispheric and Global Surface Temperature Variations over the Past Two Millennia". *Proceedings of the National Academy of Sciences* 105.36. doi: 10.1073/pnas.0805721105.

Marcott, S. A., J. D. Shakun, P. U. Clark & A. C. Mix (2013). "A Reconstruction of Regional and Global Temperature for the Past 11,300 Years". *Science* 339.6124. doi: 10.1126/science.1228026.

Mauritsen, T. et al. (2019). "Developments in the MPI-M Earth System Model Version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2". *Journal of Advances in Modeling Earth Systems* 11.4. doi: 10.1029/2018MS001400.

McInerney, F. A. & S. L. Wing (2011). "The Paleocene-Eocene Thermal Maximum: A Perturbation of Carbon Cycle, Climate, and Biosphere with Implications for the Future". *Annual Review of Earth and Planetary Sciences* 39.1. doi: 10.1146/annurev-earth-040610-133431.

Meccia, V. L. & U. Mikolajewicz (2018). "Interactive Ocean Bathymetry and Coastlines for Simulating the Last Deglaciation with the Max Planck Institute Earth System Model (MPI-ESM-v1.2)". *Geoscientific Model Development* 11.11. doi: 10.5194/gmd-11-4677-2018.

Meinshausen, M., S. C. B. Raper & T. M. L. Wigley (2011). "Emulating Coupled Atmosphere-Ocean and Carbon Cycle Models with a Simpler Model, MAGICC6 – Part 1: Model Description and Calibration". *Atmospheric Chemistry and Physics* 11.4. doi: 10.5194/acp-11-1417-2011.

Mengis, N., D. P. Keller, A. H. MacDougall, M. Eby, N. Wright, K. J. Meissner, A. Oschlies, A. Schmittner, A. J. MacIsaac, H. D. Matthews & K. Zickfeld (2020). "Evaluation of the University of Victoria Earth System Climate Model Version 2.10 (UVic ESCM 2.10)". *Geoscientific Model Development* 13.9. doi: 10.5194/gmd-13-4183-2020.

Mikolajewicz, U. (2023). *In Preparation*.

Mikolajewicz, U. & M.-L. Kapsch (2022). *Personal Communication*.

Milankovitch, M. (1941). "Kanon Der Erdbestrahlung Und Seine Anwendung Auf Das Eiszeitenproblem". *Royal Serbian Academy Special Publication* 133.

Mitchell, J. M. (1976). "An Overview of Climatic Variability and Its Causal Mechanisms". *Quaternary Research* 6.4. doi: 10.1016/0033-5894(76)90021-1.

Mollenhauer, G., T. Eglinton, N. Ohkouchi, R. Schneider, P. Müller, P. Grootes & J. Rullkötter (2003). "Asynchronous Alkenone and Foraminifera Records from the Benguela Upwelling System". *Geochimica et Cosmochimica Acta* 67.12. doi: 10.1016/S0016-7037(03)00168-6.

Moore, F. C., K. Lacasse, K. J. Mach, Y. A. Shin, L. J. Gross & B. Beckage (2022). "Determinants of Emissions Pathways in the Coupled Climate–Social System". *Nature* 603.7899. doi: 10.1038/s41586-022-04423-8.

Morice, C. P., J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. H. Dunn, T. J. Osborn, P. D. Jones & I. R. Simpson (2021). "An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set". *Journal of Geophysical Research: Atmospheres* 126.3. doi: 10.1029/2019JD032361.

Münch, T., S. Kipfstuhl, J. Freitag, H. Meyer & T. Laepple (2017). "Constraints on Post-Depositional Isotope Modifications in East Antarctic Firn from Analysing Temporal Changes of Isotope Profiles". *The Cryosphere* 11.5. doi: 10.5194/tc-11-2175-2017.

Muthers, S., J. G. Anet, C. C. Raible, S. Brönnimann, E. Rozanov, F. Arfeuille, T. Peter, A. I. Shapiro, J. Beer, F. Steinhilber, Y. Brugnara & W. Schmutz (2014). "Northern Hemispheric Winter Warming Pattern after Tropical Volcanic Eruptions: Sensitivity to the Ozone Climatology". *Journal of Geophysical Research* 119.3. doi: 10.1002/2013JD020138.

Myrvoll-Nilsen, E., S. H. Sørbye, H.-B. Fredriksen, H. Rue & M. Rypdal (2020). "Statistical Estimation of Global Surface Temperature Response to Forcing under the Assumption of Temporal Scaling". *Earth System Dynamics* 11.2. doi: 10.5194/esd-11-329-2020.

Newman, M. (2005). "Power Laws, Pareto Distributions and Zipf's Law". *Contemporary Physics* 46.5. doi: 10.1080/00107510500052444.

Nilsen, T., K. Rypdal & H.-B. Fredriksen (2016). "Are There Multiple Scaling Regimes in Holocene Temperature Records?" *Earth System Dynamics* 7.2. doi: 10.5194/esd-7-419-2016.

Nordhaus, W. D. (2010). "Economic Aspects of Global Warming in a Post-Copenhagen Environment". *Proceedings of the National Academy of Sciences* 107.26. doi: 10.1073/pnas.1005985107.

– (2017). "Revisiting the Social Cost of Carbon". *Proceedings of the National Academy of Sciences* 114.7. doi: 10.1073/pnas.1609244114.

North, G. R., J. G. Mengel & D. A. Short (1983). "Simple Energy Balance Model Resolving the Seasons and the Continents: Application to the Astronomical Theory of the Ice Ages". *Journal of Geophysical Research* 88.C11. doi: 10.1029/JC088iC11p06576.

North, G. R., L. Howard, D. Pollard & B. Wielicki (1979). "Variational Formulation of Budyko-Sellers Climate Models". *Journal of the Atmospheric Sciences* 36.2. doi: 10.1175/1520-0469(1979)036<0255:VFOBSC>2.0.CO;2.

North Greenland Ice Core Project members (2004). "High-Resolution Record of Northern Hemisphere Climate Extending into the Last Interglacial Period". *Nature* 431.7005. doi: 10.1038/nature02805.

O'Neill, B. C. et al. (2016). "The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6". *Geoscientific Model Development* 9.9. doi: 10.5194/gmd-9-3461-2016.

Olonscheck, D., A. P. Schurer, L. Lücke & G. C. Hegerl (2021). "Large-Scale Emergence of Regional Changes in Year-to-Year Temperature Variability by the End of the 21st Century". *Nature Communications* 12.1. doi: 10.1038/s41467-021-27515-x.

Olson, R., R. Sriver, W. Chang, M. Haran, N. M. Urban & K. Keller (2013). "What Is the Effect of Unresolved Internal Climate Variability on Climate Sensitivity Estimates?" *Journal of Geophysical Research: Atmospheres* 118.10. doi: 10.1002/jgrd.50390.

Otterå, O. H., M. Bentsen, H. Drange & L. Suo (2010). "External Forcing as a Metronome for Atlantic Multidecadal Variability". *Nature Geoscience* 3.10. doi: 10.1038/ngeo955.

Otto-Bliesner, B. L., E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai & G. Strand (2016). "Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model". *Bulletin of the American Meteorological Society* 97.5. doi: 10.1175/BAMS-D-14-00233.1.

PAGES 2k Consortium (2019). "Consistent Multidecadal Variability in Global Temperature Reconstructions and Simulations over the Common Era". *Nature Geoscience* 12.8. doi: 10.1038/s41561-019-0400-0.

PAGES2k Consortium (2017). “A Global Multiproxy Database for Temperature Reconstructions of the Common Era”. *Scientific Data* 4.1. doi: 10.1038/sdata.2017.88.

PALAEOSENS Project Members (2012). “Making Sense of Palaeoclimate Sensitivity”. *Nature* 491.7426. doi: 10.1038/nature11574.

Parker, D. E., T. P. Legg & C. K. Folland (1992). “A New Daily Central England Temperature Series, 1772–1991”. *International Journal of Climatology* 12.4. doi: 10.1002/joc.3370120402.

Parson, E. A. & a. K. Fisher-Vanden (1997). “Integrated Assessment Models of Global Climate Change”. *Annual Review of Energy and the Environment* 22.1. doi: 10.1146/annurev.energy.22.1.589.

Parsons, L. A., G. R. Loope, J. T. Overpeck, T. R. Ault, R. Stouffer & J. E. Cole (2017). “Temperature and Precipitation Variance in CMIP5 Simulations and Paleoclimate Records of the Last Millennium”. *Journal of Climate* 30.22. doi: 10.1175/JCLI-D-16-0863.1.

Peixoto, J. P. & A. H. Oort (1992). *Physics of Climate*. New York: American Institute of Physics. isbn: 978-0-88318-711-1 978-0-88318-712-8.

Peltier, W. R., D. F. Argus & R. Drummond (2015). “Space Geodesy Constrains Ice Age Terminal Deglaciation: The Global ICE-6G_C (VM5a) Model: Global Glacial Isostatic Adjustment”. *Journal of Geophysical Research: Solid Earth* 120.1. doi: 10.1002/2014JB011176.

Percival, D. B. & A. T. Walden (1993). *Spectral Analysis for Physical Applications*. First. Cambridge University Press. isbn: 978-0-521-35532-2 978-0-521-43541-3 978-0-511-62276-2. doi: 10.1017/CBO9780511622762.

Philip, S., S. Kew, G. J. van Oldenborgh, F. Otto, R. Vautard, K. van der Wiel, A. King, F. Lott, J. Arrighi, R. Singh & M. van Aalst (2020). “A Protocol for Probabilistic Extreme Event Attribution Analyses”. *Advances in Statistical Climatology, Meteorology and Oceanography* 6.2. doi: 10.5194/ascmo-6-177-2020.

Phipps, S. J., L. D. Rotstayn, H. B. Gordon, J. L. Roberts, A. C. Hirst & W. F. Budd (2012). “The CSIRO Mk3L Climate System Model Version 1.0 – Part 2: Response to External Forcings”. *Geoscientific Model Development* 5.3. doi: 10.5194/gmd-5-649-2012.

Pires, C. A. L. & A. Hannachi (2021). “Bispectral Analysis of Nonlinear Interaction, Predictability and Stochastic Modelling with Application to ENSO”. *Tellus A: Dynamic Meteorology and Oceanography* 73.1. doi: 10.1080/16000870.2020.1866393.

Rehfeld, K., R. Hébert, J. Lora, M. Lofverstrom & C. Brierley (2020). “Variability of Surface Climate in Simulations of Past and Future”. *Earth System Dynamics*. doi: 10.5194/esd-2019-92.

Rehfeld, K., T. Münch, S. L. Ho & T. Laepple (2018). “Global Patterns of Declining Temperature Variability from the Last Glacial Maximum to the Holocene”. *Nature* 554.7692. doi: 10.1038/nature25454.

Reschke, M., K. Rehfeld & T. Laepple (2019). “Empirical Estimate of the Signal Content of Holocene Temperature Proxy Records”. *Climate of the Past* 15.2. doi: 10.5194/cp-15-521-2019.

Riddick, T., V. Brovkin, S. Hagemann & U. Mikolajewicz (2018). “Dynamic Hydrological Discharge Modelling for Coupled Climate Model Simulations of the Last Glacial Cycle: The MPI-DynamicHD Model Version 3.0”. *Geoscientific Model Development* 11.10. doi: 10.5194/gmd-11-4291-2018.

Rohling, E. J., G. Marino, G. L. Foster, P. A. Goodwin, A. S. von der Heydt & P. Köhler (2018). “Comparing Climate Sensitivity, Past and Present”. *Annual Review of Marine Science* 10.1. doi: 10.1146/annurev-marine-121916-063242.

Rotmans, J. & M. B. van Asselt (2001). “Uncertainty in Integrated Assessment Modelling: A Labyrinthic Path”. *Integrated Assessment* 2.2. doi: 10.1023/A:1011588816469.

Rypdal, K., L. Østvand & M. Rypdal (2013). “Long-Range Memory in Earth’s Surface Temperature on Time Scales from Months to Centuries”. *Journal of Geophysical Research: Atmospheres* 118.13. doi: 10.1002/jgrd.50399.

Rypdal, K. & M. Rypdal (2016). “Comment on ”Scaling Regimes and Linear/Nonlinear Responses of Last Millennium Climate to Volcanic and Solar Forcing” by S. Lovejoy and C. Varotsos (2016)”. *Earth System Dynamics* 7.3. doi: 10.5194/esd-7-597-2016.

Rypdal, M. & K. Rypdal (2014). "Long-Memory Effects in Linear Response Models of Earth's Temperature and Implications for Future Global Warming". *Journal of Climate* 27.14. doi: 10.1175/JCLI-D-13-00296.1.

Scaife, A. A. & D. Smith (2018). "A Signal-to-Noise Paradox in Climate Science". *npj Climate and Atmospheric Science* 1.1. doi: 10.1038/s41612-018-0038-4.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger & C. Appenzeller (2004). "The Role of Increasing Temperature Variability in European Summer Heatwaves". *Nature* 427.6972. doi: 10.1038/nature02300.

Schillinger, M., B. Ellerhoff, R. Scheichl & K. Rehfeld (2022a). "Separating Internal and Externally Forced Contributions to Global Temperature Variability Using a Bayesian Stochastic Energy Balance Framework". *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32.11. doi: 10.1063/5.0106123.

– (2022b). *The ClimBayes Package in* R. Zenodo. doi: 10.5281/ZENODO.7317984.

Schmidt, G. A. et al. (2012). "Climate Forcing Reconstructions for Use in PMIP Simulations of the Last Millennium (v1.1)". *Geoscientific Model Development* 5.1. doi: 10.5194/gmd-5-185-2012.

Schmidt, G. A. et al. (2006). "Present-Day Atmospheric Simulations Using GISS ModelE: Comparison to In Situ, Satellite, and Reanalysis Data". *Journal of Climate* 19.2. doi: 10.1175/JCLI3612.1.

Schurer, A. P., G. C. Hegerl, M. E. Mann, S. F. B. Tett & S. J. Phipps (2013). "Separating Forced from Chaotic Climate Variability over the Past Millennium". *Journal of Climate* 26.18. doi: 10.1175/JCLI-D-12-00826.1.

Schurer, A. P., S. F. B. Tett & G. C. Hegerl (2014). "Small Influence of Solar Variability on Climate over the Past Millennium". *Nature Geoscience* 7.2. doi: 10.1038/ngeo2040.

Schwarzwald, K. & N. Lenssen (2022). "The Importance of Internal Climate Variability in Climate Impact Projections". *Proceedings of the National Academy of Sciences* 119.42. doi: 10.1073/pnas.2208095119.

Sellers, W. D. (1969). "A Global Climatic Model Based on the Energy Balance of the Earth-Atmosphere System". *Journal of Applied Meteorology* 8.3. doi: 10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2.

Shackleton, N. J. (2000). "The 100,000-Year Ice-Age Cycle Identified and Found to Lag Temperature, Carbon Dioxide, and Orbital Eccentricity". *Science* 289.5486. doi: 10.1126/science.289.5486.1897.

Shakun, J. D. & A. E. Carlson (2010). "A Global Perspective on Last Glacial Maximum to Holocene Climate Change". *Quaternary Science Reviews* 29.15-16. doi: 10.1016/j.quascirev.2010.03.016.

Sherwood, S. C. et al. (2020). "An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence". *Reviews of Geophysics* 58.4. doi: 10.1029/2019RG000678.

Sigl, M. et al. (2015). "Timing and Climate Forcing of Volcanic Eruptions for the Past 2,500 Years". *Nature* 523.7562. doi: 10.1038/nature14565.

Sigman, D. M. & E. A. Boyle (2000). "Glacial/Interglacial Variations in Atmospheric Carbon Dioxide". *Nature* 407.6806. doi: 10.1038/35038000.

Simolo, C. & S. Corti (2022). "Quantifying the Role of Variability in Future Intensification of Heat Extremes". *Nature Communications* 13.1. doi: 10.1038/s41467-022-35571-0.

Smith, D. M., B. B. B. Booth, N. J. Dunstone, R. Eade, L. Hermanson, G. S. Jones, A. A. Scaife, K. L. Sheen & V. Thompson (2016). "Role of Volcanic and Anthropogenic Aerosols in the Recent Global Surface Warming Slowdown". *Nature Climate Change* 6.10. doi: 10.1038/nclimate3058.

Smith, R. S. & J. Gregory (2012). "The Last Glacial Cycle: Transient Simulations with an AOGCM". *Climate Dynamics* 38.7-8. doi: 10.1007/s00382-011-1283-y.

Snyder, C. W. (2016). "Evolution of Global Temperature over the Past Two Million Years". *Nature* 538.7624. doi: 10.1038/nature19798.

Sprinz, S. (2023). *In preparation, Bachelor's thesis, Heidelberg University*.

Steinhilber, F., J. Beer & C. Fröhlich (2009). "Total Solar Irradiance during the Holocene". *Geophysical Research Letters* 36.19. doi: 10.1029/2009GL040142.

Stott, P. A., N. Christidis, F. E. L. Otto, Y. Sun, J.-P. Vanderlinden, G. J. van Oldenborgh, R. Vautard, H. von Storch, P. Walton, P. Yiou & F. W. Zwiers (2016). "Attribution of Extreme Weather and Climate-related Events". *WIREs Climate Change* 7.1. doi: 10.1002/wcc.380.

Swindles, G. T., E. J. Watson, I. P. Savov, I. T. Lawson, A. Schmidt, A. Hooper, C. L. Cooper, C. B. Connor, M. Gloor & J. L. Carrivick (2018). "Climatic Control on Icelandic Volcanic Activity during the Mid-Holocene". *Geology* 46.1. doi: 10.1130/G39633.1.

Swingedouw, D., J. Mignot, P. Ortega, M. Khodri, M. Menegoz, C. Cassou & V. Hanquiez (2017). "Impact of Explosive Volcanic Eruptions on the Main Climate Variability Modes". *Global and Planetary Change* 150. doi: 10.1016/j.gloplacha.2017.01.006.

Tarasov, L., A. S. Dyke, R. M. Neal & W. Peltier (2012). "A Data-Calibrated Distribution of Deglacial Chronologies for the North American Ice Complex from Glaciological Modeling". *Earth and Planetary Science Letters* 315–316. doi: 10.1016/j.epsl.2011.09.010.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian & J. Doyne Farmer (1992). "Testing for Nonlinearity in Time Series: The Method of Surrogate Data". *Physica D: Nonlinear Phenomena* 58.1-4. doi: 10.1016/0167-2789(92)90102-S.

Timmreck, C. (2012). "Modeling the Climatic Effects of Large Explosive Volcanic Eruptions". *Wiley Interdisciplinary Reviews: Climate Change* 3.6. doi: 10.1002/wcc.192.

Toohey, M., J. Belo, M. Jegen-Kulcsar, S. Kutterolf & K. Rehfeld (2023). "A Proxy-Based Semi-Stochastic Ensemble Reconstruction of Volcanic Climate Forcing for the Last Glacial Cycle (130,000 – 50 BP)". *XXI INQUA Congress 2023*.

Toohey, M. & M. Sigl (2017). "Volcanic Stratospheric Sulfur Injections and Aerosol Optical Depth from 500 BCE to 1900 CE". *Earth System Science Data* 9.2. doi: 10.5194/essd-9-809-2017.

Valdes, P. (2011). "Built for Stability". *Nature Geoscience* 4.7. doi: 10.1038/ngeo1200.

van Beek, L., M. Hajer, P. Pelzer, D. van Vuuren & C. Cassen (2020). "Anticipating Futures through Models: The Rise of Integrated Assessment Modelling in the Climate Science-Policy Interface since 1970". *Global Environmental Change* 65. doi: 10.1016/j.gloenvcha.2020.102191.

van Oldenborgh, G. J., K. van der Wiel, S. Kew, S. Philip, F. Otto, R. Vautard, A. King, F. Lott, J. Arrighi, R. Singh & M. van Aalst (2021). "Pathways and Pitfalls in Extreme Event Attribution". *Climatic Change* 166.1-2. doi: 10.1007/s10584-021-03071-7.

van Oldenborgh, G. J., M. F. Wehner, R. Vautard, F. E. L. Otto, S. I. Seneviratne, P. A. Stott, G. C. Hegerl, S. Y. Philip & S. F. Kew (2022). "Attributing and Projecting Heatwaves Is Hard: We Can Do Better". *Earth's Future* 10.6. doi: 10.1029/2021EF002271.

van Vuuren, D. P., E. Stehfest, D. E. H. J. Gernaat, M. van den Berg, D. L. Bijl, H. S. de Boer, V. Daioglou, J. C. Doelman, O. Y. Edelenbosch, M. Harmsen, A. F. Hof & M. A. E. van Sluisveld (2018). "Alternative Pathways to the 1.5 °C Target Reduce the Need for Negative Emission Technologies". *Nature Climate Change* 8.5. doi: 10.1038/s41558-018-0119-8.

von der Heydt, A. S., P. Ashwin, C. D. Camp, M. Crucifix, H. A. Dijkstra, P. Ditlevsen & T. M. Lenton (2021). "Quantification and Interpretation of the Climate Variability Record". *Global and Planetary Change* 197. doi: 10.1016/j.gloplacha.2020.103399.

Weitzman, M. L. (2012). "GHG Targets as Insurance Against Catastrophic Climate Damages". *Journal of Public Economic Theory* 14.2. doi: 10.1111/j.1467-9779.2011.01539.x.

Wiener, N. (1930). "Generalized Harmonic Analysis". *Acta Mathematica* 55.0. doi: 10.1007/BF02546511.

Willeit, M., A. Ganopolski, R. Calov & V. Brovkin (2019). "Mid-Pleistocene Transition in Glacial Cycles Explained by Declining CO_2 and Regolith Removal". *Science Advances* 5.4. doi: 10.1126/sciadv.aav7337.

Wing, S. L., G. J. Harrington, F. A. Smith, J. I. Bloch, D. M. Boyer & K. H. Freeman (2005). "Transient Floral Change and Rapid Global Warming at the Paleocene-Eocene Boundary". *Science* 310.5750. doi: 10.1126/science.1116913.

Winkelmann, R., M. A. Martin, M. Haseloff, T. Albrecht, E. Bueler, C. Khroulev & A. Levermann (2011). "The Potsdam Parallel Ice Sheet Model (PISM-PIK) – Part 1: Model Description". *The Cryosphere* 5.3. doi: 10.5194/tc-5-715-2011.

Wirths, C. (2021). "Forcing-Induced Climate Response from the LGM to Present Day in an Energy Balance Model". *Master's thesis, Heidelberg University*.

Wunderling, N., J. F. Donges, J. Kurths & R. Winkelmann (2021). "Interacting Tipping Elements Increase Risk of Climate Domino Effects under Global Warming". *Earth System Dynamics* 12.2. doi: 10.5194/esd-12-601-2021.

Xiao-Ge, X., W. Tong-Wen & Z. Jie (2013). "Introduction of CMIP5 Experiments Carried out with the Climate System Models of Beijing Climate Center". *Advances in Climate Change Research* 4.1. doi: 10.3724/SP.J.1248.2013.041.

Yan, X., R. Zhang & T. R. Knutson (2018). "Underestimated AMOC Variability and Implications for AMV and Predictability in CMIP Models". *Geophysical Research Letters* 45.9. doi: 10.1029/2018GL077378.

Yiou, P., E. Baert & M. F. Loutre (1996). "Spectral Analysis of Climate Data". *Surveys in Geophysics* 17.6. doi: 10.1007/BF01931784.

Zachos, J. C., G. R. Dickens & R. E. Zeebe (2008). "An Early Cenozoic Perspective on Greenhouse Warming and Carbon-Cycle Dynamics". *Nature* 451.7176. doi: 10.1038/nature06588.

Zanchettin, D., O. Bothe, H. F. Graf, S. J. Lorenz, J. Luterbacher, C. Timmreck & J. H. Jungclaus (2013). "Background Conditions Influence the Decadal Climate Response to Strong Volcanic Eruptions". *Journal of Geophysical Research Atmospheres* 118.10. doi: 10.1002/jgrd.50229.

Zanchettin, D. et al. (2016). "The Model Intercomparison Project on the Climatic Response to Volcanic Forcing (VolMIP): Experimental Design and Forcing Input Data for CMIP6". *Geoscientific Model Development* 9.8. doi: 10.5194/gmd-9-2701-2016.

Zhu, F., J. Emile-Geay, N. P. McKay, G. J. Hakim, D. Khider, T. R. Ault, E. J. Steig, S. Dee & J. W. Kirchner (2019). "Climate Models Can Correctly Simulate the Continuum of Global-Average Temperature Variability". *Proceedings of the National Academy of Sciences* 116.18. doi: 10.1073/pnas.1809959116.

Ziegler, E. (2022). *Personal Communication*.

Ziegler, E. & K. Rehfeld (2021). "TransEBM v. 1.0: Description, Tuning, and Validation of a Transient Model of the Earth's Energy Balance in Two Dimensions". *Geoscientific Model Development* 14.5. doi: 10.5194/gmd-14-2843-2021.

Ziemen, F. A., M.-L. Kapsch, M. Klockmann & U. Mikolajewicz (2019). "Heinrich Events Show Two-Stage Climate Response in Transient Glacial Simulations". *Climate of the Past* 15.1. doi: 10.5194/cp-15-153-2019.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, Massachusetts: Addison-Wesley.

Zscheischler, J. et al. (2020). "A Typology of Compound Weather and Climate Events". *Nature Reviews Earth & Environment* 1.7. doi: 10.1038/s43017-020-0060-z.

Appendix 2 (F-G)

F Publication 1

B. Ellerhoff and K. Rehfeld. "Probing the Timescale Dependency of Local and Global Variations in Surface Air Temperature from Climate Simulations and Reconstructions of the Last Millennia". *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 104.6 (2021).
doi: 10.1103/PhysRevE.104.064136.

Probing the timescale dependency of local and global variations in surface air temperature from climate simulations and reconstructions of the last millennia

Beatrice Ellerhoff [1,*] and Kira Rehfeld [1,2,†]

[1]*Institute of Environmental Physics, Ruprecht-Karls-Universität Heidelberg, INF 229, 69120 Heidelberg, Germany*

[2]*Geo- und Umweltforschungszentrum (GUZ), Universität Tübingen, Schnarrenbergstr. 94-96, 72076 Tübingen, Germany*

(Received 2 February 2021; revised 15 June 2021; accepted 9 December 2021; published 27 December 2021)

Earth's climate can be understood as a dynamical system that changes due to external forcing and internal couplings. Essential climate variables, such as surface air temperature, describe this dynamics. Our current interglacial, the Holocene (11 700 yr ago to today), has been characterized by small variations in global mean temperature prior to anthropogenic warming. However, the mechanisms and spatiotemporal patterns of fluctuations around this mean, called temperature variability, are poorly understood despite their socioeconomic relevance for climate change mitigation and adaptation. Here we examine discrepancies between temperature variability from model simulations and paleoclimate reconstructions by categorizing the scaling behavior of local and global surface air temperature on the timescale of years to centuries. To this end, we contrast power spectral densities (PSD) and their power-law scaling using simulated and observation-based temperature series of the last 6000 yr. We further introduce the spectral gain to disentangle the externally forced and internally generated variability as a function of timescale. It is based on our estimate of the joint PSD of radiative forcing, which exhibits a scale break around the period of 7 yr. We find that local temperature series from paleoclimate reconstructions show a different scaling behavior than simulated ones, with a tendency towards stronger persistence (i.e., correlation between successive values within a time series) on periods of 10 to 200 yr. Conversely, the PSD and spectral gain of global mean temperature are consistent across data sets. Our results point to the limitation of climate models to fully represent local temperature statistics over decades to centuries. By highlighting the key characteristics of temperature variability, we pave a way to better constrain possible changes in temperature variability with global warming and assess future climate risks.

DOI: 10.1103/PhysRevE.104.064136

I. INTRODUCTION

The variability of surface air temperature is present on all spatial and temporal scales, from synoptic and seasonal changes to long-term variations on periods of years to multimillennia. On the one hand, it arises from internal processes, such as the El Niño-Southern Oscillation (ENSO) [1]. On the other hand, the temperature varies due to external forcing, such as the greenhouse effect [2,3]. Understanding the internally generated and externally forced variability has been suggested to be at least as necessary for evaluating climate risks for society and ecosystems as projecting the global mean temperature [4]. Available instrumental observations are limited to a small time span, leading to challenges in quantifying temperature variability. Paleoclimate reconstructions extend the characterization of temperature variability and can be compared to global circulation models (GCMs) [5–7]. However, discrepancies between model and paleoclimate data remain to be resolved, especially on the local level and on periods between years and centuries [8–12].

Characterizing local temperature variability is crucial for predicting extremes [6], not only to minimize short-term damage but also to design long-term strategies, including urban planning and food cultivation [13]. Variability of global temperature on periods above years is relevant to the understanding of long-term changes [14] as well as climate sensitivity [15]. Assessing the temporal correlation structure of temperature series by means of scaling behavior and persistence is particularly important for distinguishing externally forced trends from natural changes [16]. It could affect the confidence in future projections and attribution studies [17,18]. Therefore, one of the main topics to be investigated here is the characteristics of local and global temperature variability on periods of years to centuries from model simulations and observation-based data of the last millennia.

To determine how the variability of a temperature series is distributed with timescales τ, we make use of the power spectral density (PSD) $S(\tau)$, known as spectrum. It can be obtained from the Fourier transform of the autocorrelation function (see Appendix A) [19,20]. The spectrum was shown to often follow a power law

$$S(\tau) \sim \tau^{\beta}, \tag{1}$$

with spectral exponent β and period τ [21,23–27], especially on decadal-to-centennial scales [22,28,29]. We refer to this behavior (1) as temporal scaling since the temperature signal has no preferred timescale and is statistically similar across periods τ. The exact determination of the start and end points of a scaling interval is not part of this study.

*beatrice.ellerhoff@iup.uni-heidelberg.de

†Now at Geo- und Umweltforschungszentrum (GUZ), Universität Tübingen, Schnarrenbergstr. 94-96, 72076 Tübingen, Germany.

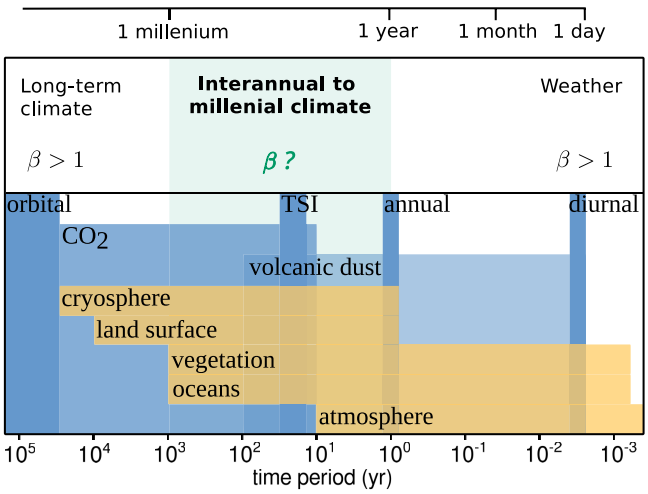


FIG. 1. Characteristic timescales relevant to surface air temperature variability of climatic drivers (dark blue) and climate subsystems (yellow) [32–34]. The weather and long-term climate is characterized by $\beta > 1$ for local and global mean temperature. On interannual to millennial timescales the statistical properties of temperature fluctuations remain to be determined, especially at the local scale. The TSI bar highlights the dominant variations in recent total solar irradiance observations.

Long-range memory stochastic processes are suitable to describe temperature signals with temporal scaling [25,26,30]. Among those, fractional Gaussian noise (fGn) is a stationary process and exhibits a spectral exponent $\beta \in (-1, 1)$ on sufficiently long periods (see Appendix B). Fractional Brownian motion (fBm) is a nonstationary process that shows $\beta \in (1, 3)$. The scaling exponent β relates to the decay of the autocovariance function (Appendix B) and indicates how strongly the values within a time series are correlated (or anticorrelated). It is therefore regarded as a measure of the strength of temporal persistence (or antipersistence) [26,31].

Particular scaling behavior with $\beta \approx 2$ [23,25] is typical for the weather regime (hours to weeks) and can be explained by atmospheric turbulence [35,36]. In the long-term climate, regional and global mean temperatures show similar behavior ($\beta > 1$) [24,25,29] due to the presence of nonlinear processes, such as the temperature-albedo feedback [37]. On timescales between years to millennia, the temperature is constantly influenced by the interaction of all climate subsystems and by volcanic, solar, as well as CO_2 forcing (Fig. 1). Estimates of the spatially dependent scaling behavior of local temperature on these timescales differ [22,25]. On the global scale, many studies find $\beta \approx 1$ [28,29]. However, Lovejoy *et al.* has identified a change from the so-called macroweather regime ($\beta \approx 0.8$ on periods of 10 days to 40 yr) to the climate regime ($\beta \approx 1.8$ on periods from 40 yr to 80 000 yr) [25].

In this manner, previous works find ambiguity in the interpretation of local and global temperature scaling, and it remains to be determined whether simulations and reconstructions qualitatively agree in scaling behavior $\beta < 1$ or $\beta > 1$. The so-called "$1/f$ noise" ($\beta = 1$) corresponds to a process with power spectral density proportional to the period. For $\beta > 1$, the relative contribution

$$\frac{\int_{f'/2}^{f'} S(f)\,df}{\int_{f'}^{2f'} S(f)\,df} = \frac{1-2^{\beta-1}}{2^{1-\beta}-1} = 2^{\beta-1} \tag{2}$$

to the variance is larger from slow timescales compared to faster ones for all frequency intervals $f'/2$ to $2f'$ within a scaling interval [38]. With increasing $\beta > 1$, the fBM is said to exhibit "nonlinear pseudotrends" [30] (Appendix B). Thus, for understanding climate variability and for modeling purposes, the systematic estimate of the scaling exponent β allows to assess the behavior of fluctuation levels across timescales [26]. Moreover, the differentiation between forced and unforced changes poses a challenge to understanding temperature variability [39,40]. Beyond the analysis of Haar fluctuations of a few forcing reconstructions [41–43], spectral analysis of climatic drivers and their frequency-dependent linkage to the temperature response remains incomplete.

We investigate the timescale dependency of local and global surface air temperature variability by analyzing power spectral densities from a few hours to a thousand years, thereby extending and improving on earlier work [24,25,29,44]. We use model simulations and observation-based data, which we introduce in Sec. II. To estimate the PSD and determine its power-law scaling on periods of 10 to 200 yr, we use state-of-the-art methods described in Sec. III. This allows us to contrast regional and global spectra (Sec. IV A), spatial patterns (Sec. IV B), and the agreement of simulated and observation-based estimates (Sec. III C). Along with that, we discuss the joint PSD from various radiative forcings, which allows us to calculate the spectral gain and study the externally forced variability in Sec. III D. Based on our reconstruction of the PSD of surface air temperature for the last millennia, we evaluate the consistency of spectral characteristics across the data sets considered. In Sec. V we elaborate on the stronger persistence of temperature on local than global level as well as remaining discrepancies. Finally, we discuss how our findings could help improve climate model simulations and understand Earth's climate dynamics.

II. DATA

We investigate the timescale-dependent distribution of surface air temperature variability using model simulations, observation-based data, and radiative forcing reconstructions. The model simulations include ten transient runs from GCM experiments [45]. The observation-based data consists of reanalysis data, instrumental measurements, and the paleoclimate reconstructions from the Past Global Changes 2k (PAGES2k) network [46]. We use 12 reconstructions of climatic drivers, including solar, volcanic, orbital, and CO_2 forcing. All temperature and radiative forcing signals are specific to the Mid- and Late-Holocene (the last 6000 yr), with a focus on the Common Era (0 to 2000 CE). The supplemental tables S1–S3 [47] summarize their key specifications.

A. Model simulations

Each of the ten GCM runs considered features a transient, albeit different forcing and a comparable spatiotemporal resolution. The CESM-LME 1 [48] and MPI-M LM [49] ex-

periments serve as representative runs of the last millennium. We analyze them at two temporal resolutions (one month, six hours) to capture both the high- and low-frequency variability within our available computing capacities (see Fig. S7 [47]). CESM 1 past 2k [50] is included as a slightly newer run for the Common Era. To cover the Mid-Holocene, we use simulations from the IPSL [51] (denoted IPSL-p6k) and ECHAM5/MPI-OM [52] (denoted ECH5/MPIOM-p6k) of the last 6000 yr. From the TraCE-21k [53] simulation, we also consider only the last 6000 yr to retain comparability and to avoid potential biases due to significant shifts in the mean state of climate. The Mid-Holocene runs were temporally averaged to a bimonthly resolution to reduce computational costs. To test for the influence of human-induced climate change on our results, we include the HadCM3 LM1 simulation [12], covering the period from 850 to 1850 CE. Furthermore, we compare our results to the pre-industrial (PI) control runs from CESM-LME 1 and MPI-M LM, as well as the TraCE-21k-ORB run, which is solely forced by orbital changes.

B. Observation-based data

In addition to the simulations, we analyze the monthly resolved HadCRUT4 (Hadley Centre/Climatic Research Unit Temperature) instrumental records, ranging from 1850 to 2019 [54]. However, most of the grid-box time series are not available as continuous measurements as required for spectral analysis. Therefore, we retain only those 104 grid boxes with coverage greater than 150 yr after interpolating gaps of up to two months. While the Northern Hemisphere is comparatively well covered up to 72.5°N, only nine grid boxes remain for the Southern Hemisphere. Therefore, this selection comes at the expense of spatial resolution but offers a higher spectral resolution on longer timescales. To further explore the potential effect of these spatiotemporal constraints, we include the ERA5 (European Centre for Medium-Range Weather Forecasts Reanalysis 5th generation) temperature reanalysis for the years 1979 to 2019 [55]. Along with CESM-LME 1 and MPI-M LM, we analyze the ERA5 data at both six-hourly and monthly resolution for the same reasons as mentioned earlier.

In addition to direct temperature observations and reanalysis, we analyze paleoclimate data. Paleoclimate records hold preserved biological, chemical, and physical tracers ("proxies") of past climate. The number of temperature records from paleoclimate data with subcentennial resolution is limited. Recent progress has been made by improved calibration and pseudoproxy methods within the PAGES2k network [56]. Therefore, we base our analysis on their newest global multiproxy database for temperature reconstructions of the Common Era [46]. It gathers 692 records from trees, ice, sediment, corals, speleothems, and documentary evidence with a resolution between weeks and centuries. The records are spread over 648 locations, including all continental regions and major ocean basins.

For investigating the variability of global mean surface temperature, we use the seven spatially weighted statistical reconstructions for the last 2000 yr provided by PAGES2k [46]. To estimate the mean of local spectra, we choose records from the PAGES2k database according to their resolution ($\leqslant$80 yr), their number of data points ($\geqslant$20), their coverage ($\geqslant$20 yr), as well as their maximum hiatuses ($\leqslant$160 yr). To reliably deduce the scaling of the PSD from individual records, we select the records according to our scales of interest (Table I), similar to [26,57]. Ice core records were excluded from our analysis since they require additional consideration of signal-to-noise ratios at the subcentennial timescales [58,59].

TABLE I. Requirements on irregularly sampled time series $x(t)$ for analyzing power-law scaling on timescales $\tau \in [\tau_1, \tau_2]$. We apply this scheme for $\tau_1 = 10$ and $\tau_2 = 200$ yr in Secs. IV B and IV C.

Parameter	Value
Number of data points (N)	$\geqslant 50$
Mean temporal resolution $[\langle t_{i+1} - t_i \rangle]$	$\leqslant \tau_1$
Coverage ($t_N - t_1$)	$\geqslant 3\tau_2$
Length of hiatuses $[\max(t_{i+1} - t_i)]$	$\leqslant 5\tau_1$

C. Radiative forcing

External forcing contributes significantly to temperature variability and is an essential part of reliable climate projections [40,60,61]. We study its spectral properties using forcing reconstructions, widely implemented in GCM experiments and coordinated within the Palaeoclimate Model Intercomparison Project (PMIP3/PMIP4) [62,63]. This includes five solar [64–68], one CO_2 [63], and two volcanic [60,69] forcing reconstructions as well as Berger's numerical solution for orbital forcing [70]. Furthermore, we calculate diurnal insolation changes from the hour angle of the sun [71]. We also use a more recently published volcanic [72] and high-resolution solar forcing [73] reconstruction as well as CO_2 measurements [74]. We neglect land-use forcing [75] which is much lower in amplitude and variability than the other forcings considered here.

All forcing reconstructions are rescaled to radiative forcing equivalents, which express their respective change in the Earth's radiation balance in Watts per square meter (Wm^{-2}). We apply the widely used formula $5.35\,\ln([\mathrm{CO_2}]/278\mathrm{ppm})\,\mathrm{Wm}^{-2}$ to rescale CO_2 concentrations $[\mathrm{CO_2}]$, given in parts per million (ppm) [76]. The stratospheric aerosol optical depth (AOD) from volcanic eruptions is rescaled by $(-20)^{-1}\,\mathrm{Wm}^{-2}/\mathrm{AOD}$ [77]; however, the optimal conversion factor is still a matter of debate [78]. Additional uncertainties arise from the wide spread of reconstructions for volcanic and solar forcing. To account for this and the choice of conversion factor, we simulate the joint PSD of radiative forcing by a Monte Carlo approach described in Appendix E. Here "joint" indicates that the PSD of radiative forcing is calculated by linear summation of the mean PSD from different types of climatic drivers, rescaled to their radiative forcing equivalents.

III. METHODS

Spectral analysis is the primary tool used here for studying the timescale-dependent variability and scaling of temperature series. To minimize uncertainties in the spectral analysis of proxy records, we use state-of-the-art approaches for irregularly sampled time series [79]. Statistical estimators further

test for the agreement between simulations and paleoclimate data. We apply linear response theory to derive the spectral gain and investigate the forced temperature response.

A. Spectral analysis

Power spectral analysis requires the assumption that the underlying time series can be described as a weakly stationary, stochastic process with time-independent mean and autocovariance [80]. We therefore linearly detrend all time series as it is standard for temperature analysis [9,26,81,82]. The agreement of the PSD from disjoint time intervals in Fig. S13 [47] provides evidence that stationarity is sufficiently fulfilled. We use the multitaper method with three windows [83,84] and chi-square distributed uncertainties to compute the PSD. The two lowest frequencies were omitted to reduce biases of the multitaper method [24]. For visual purposes, we apply a logarithmic Gaussian smoothing filter of constant width (0.005 decibels) [85]. Mean spectra were calculated by interpolation to the lowest resolution, binning into equally spaced log-frequency intervals, and taking the average with equal weights [24]. This requires the statistical independence of the averaged values [43]. The spectral exponent β is calculated by linear regression to the logarithm of (1) on periods between $\tau_1 = 10$ and $\tau_2 = 200$ yr after binning the PSD into equally spaced log-frequency intervals to more uniformly weight the estimate and avoid low-frequency biases [24,26,29,86]. In the case of seven proxy records with an insufficient resolution, the scaling is estimated on their corresponding spectral resolution, but always at least between 20 and 200 yr (Fig. S4 [47]). The uncertainty of the spectral exponent, $\Delta\beta$, is given by the standard error of the linear regression model $\Delta\beta_{lm}$, except for irregularly temperature series.

B. Uncertainties for irregular temperature series

Spectral analysis of proxy records, which are typically not sampled in regular time steps, is more prone to errors than that of regular time series. We aim to minimize biases by accounting for the number of data points, temporal resolution, total coverage, and hiatuses' length when selecting the records (Table I). We find that the mean temporal resolution of a proxy record approximates well the optimal interpolation time step. Nevertheless, the interpolation introduces uncertainties which are not captured by $\Delta\beta_{lm}$. Similar to Laepple *et al.* [79], we quantify this additional uncertainty $\Delta\beta_{\text{int}}$ in four steps: (1) For each record with spectral exponent β, we simulate $N = 100$ surrogate time series with annual resolution and a power-law scaling $\beta_n \approx \beta$ and $n \in [1, N]$. (2) We form the surrogate's block average over the proxy record's irregular time steps and obtain N surrogate time series at record resolution. (3) We interpolate the surrogate time series, calculate the multitaper spectrum, and extract the scaling exponent $\beta_{n,lm}$ from linear regression in the same way as for the proxy record (Fig. S8 [47]). (4) We calculate the mean deviation $\Delta\beta_{\text{int}} = \frac{1}{N}\sum_{n=1}^{N} |\beta_{n,lm} - \beta_n|$ of the ensemble. The uncertainty of the individual fits $\Delta\beta_{n,lm}$ is negligible compared to the mean deviation $\Delta\beta_{\text{int}}$. We obtain the uncertainty of the record's spectral exponent from both, the uncertainty of the initial fit $\Delta\beta_{lm}$ and due to interpolation $\Delta\beta_{\text{int}}$ via quadratic summation: $\Delta\beta = \sqrt{(\Delta\beta_{lm})^2 + (\Delta\beta_{\text{int}})^2}$.

C. Statistical analysis of spectral exponents

We quantify the agreement of simulated and reconstructed β-values using percent agreement, categorical agreement, and Kappa statistics. Beforehand, we extract the simulated temperature at the proxy record location by bilinear interpolation of neighboring grid boxes to achieve the best possible comparability between record and simulation. Percent agreement p_0 gives the percentage of locations at which the confidence range $\beta \pm \Delta\beta$ from simulation and reconstruction overlap. The agreement by category, here referred to as categorical agreement p_c, is calculated with the help of $\nu = 0.32$, the mean uncertainty of β from all proxy records considered. We then assign the three categories *low* ($\beta < 1 - \nu$), *high* ($1 + \nu \leqslant \beta$), and *intermediate* ($1 - \nu \leqslant \beta < 1 + \nu$) to the spectral exponent β. The *intermediate* regime prevents incorrect assignment. To verify the reliability of categorical agreement, we calculate the kappa statistics

$$\kappa = (p_c - p_e)/(1 - p_e) \tag{3}$$

with expected percent agreement p_e by category [87]. The latter can be obtained from $p_e = \frac{1}{N^2}\sum_{c=1}^{3} n_{c,m} n_{c,p}$ where c is the category, N the number of locations and n the number of times that models (m) and proxy records (p) have predicted category c. The κ-coefficient quantifies the reliability from no agreement beyond chance ($\kappa = 0$) to full agreement ($\kappa = 1$). Negative κ indicates agreement that is beyond change, for example, due to systematic biases.

D. Spectral gain

We investigate how climatic drivers influence the global mean temperature at period τ by calculating the spectral gain

$$G^2(\tau) = \frac{S_T(\tau)}{S_F(\tau)}. \tag{4}$$

Here $S_T(\tau)$ is the PSD of the global mean temperature and $S_F(\tau)$ the PSD of radiative forcing (see also Appendix C). The gain requires the assumption that the global mean temperature can be well approximated as a linear function of the forcing [27,88,89] and that different types of radiative forcing add linearly [90–93]. To this end, we focus on timescales between years and centuries when additivity is a valid assumption and nonlinearities in the global mean temperature are sufficiently small [42,43]. The main practical problem that confronts us is that the gain might be subject to a sampling bias due to our data sets choice. Therefore, we perform a Monte Carlo simulation of the PSD of radiative forcing and the global mean temperature, as well as the spectral gain as described in Appendix E.

IV. RESULTS AND DISCUSSION

A. Global mean and mean of local spectra

In order to study the timescale dependency of global mean temperature, we present its power spectral density in Fig. 2(b). It shows the characteristic background continuum, spectral

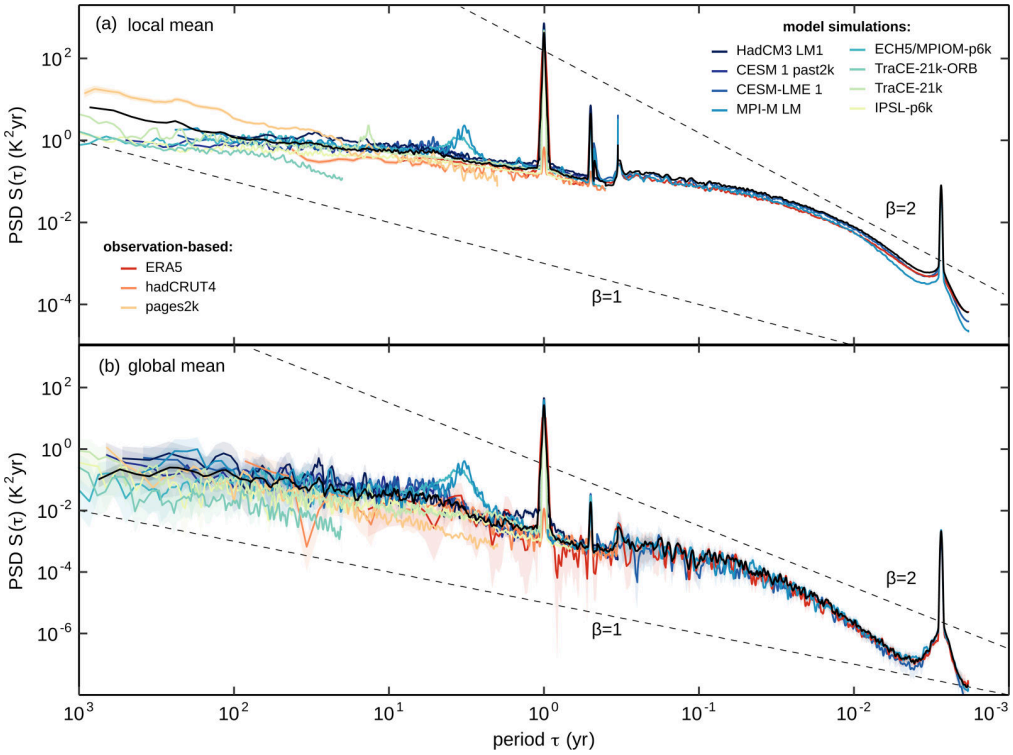


FIG. 2. (a) Mean power spectral densities (PSD) of local temperature from model simulations and observation-based data on periods from hours to 1000 yr for the Holocene. (b) PSD of global mean temperature. The dashed lines with slope β and arbitrary y-intercept in the log-log graph indicate the scaling behavior for visual comparison. The ensemble means (black solid lines) were formed using equal weights across the model group M_0 (see Table S1 [47]).

peaks, and higher harmonics associated with the diurnal and annual cycle. Overall, the PSDs tend to agree between the data sets, albeit with some differences on the interannual scale and when compared to the Trace21k ORB run. The Trace21k-ORB run is solely forced by orbital changes and therefore shows less variability than the ensemble mean. The broad spectral peak on interannual periods reveals an artificially amplified ENSO in the shared MPI-M LM and ECHAM5/MPI-OM ocean component [94]. For a better visibility, PI control runs are separately shown in the supplementary Fig. S6 [47]. Overall, the PSD largely agrees among different data sets, especially towards shorter timescales.

We find a power-law scaling of $\beta \approx 1$ on timescales longer than 10 yr in line with previous results [25,26,28]. The PSD decreases more strongly towards shorter periods, which is characteristic of the weather regime [25,36]. Similar to Nilsen *et al.* [26], we find no evidence for significant changes in scaling behavior around the centennial scale. One limitation of previous work that found scale breaks is that the spectra were estimated across nonstationary shifts in climate, such as the deglaciation [29], and with a change in proxies and archives [24].

We present the area-weighted mean spectra of the local (grid box) temperature in Fig. 2(a). Compared to the global mean in Fig. 2(b), the power increases and the spectral slope decreases, in line with [81]. The spectra agree on periods below 10 yr, except for the artificially amplified ENSO signal mentioned earlier. Moreover, we find a narrow peak at 13 yr, associated with an unrealistic variability in the northern North Atlantic of the TraCE-21k run, similar to [95,96]. Remarkably, the decadal-to-centennial variability of the reconstructed temperature is increased by one to two orders of magnitude compared to the simulations. The spectral exponent is smaller for models ($\beta < 1$) compared to paleoclimate data ($\beta \approx 1$).

This finding verifies that models show less regional temperature variability and that the mismatch increases towards longer timescales. The results are robust to sampling from the PAGES2k database and the influence of anthropogenic climate change (Fig. S10 [47]). One shortcoming of forming the area-weighted mean PSD is that the uncertainty quantification requires the assumption of independent spatial degrees of freedom of the temperature field. Due to the presence of spatial correlations, an estimate of the effective spatial degrees of freedom and their dependence on the underlying timescale would be needed to resolve this limitation [97].

B. Spatial patterns of persistence

To further investigate the mismatch on local scaling properties, we compare the spatial dependence of temperature persistence from simulations and paleoclimate data in Fig. 3. The simulations largely exhibit small-magnitude scaling exponents ($-1 < \beta < 1$), whereas proxy records were found to also show $\beta > 1$. In this manner, the magnitude of local temperature fluctuations from model simulations often shows no dependence on the decadal-to-centennial timescale. However, approximately half of the proxy records show a variance that grows on increasingly long periods (see also Fig. S11 [47]).

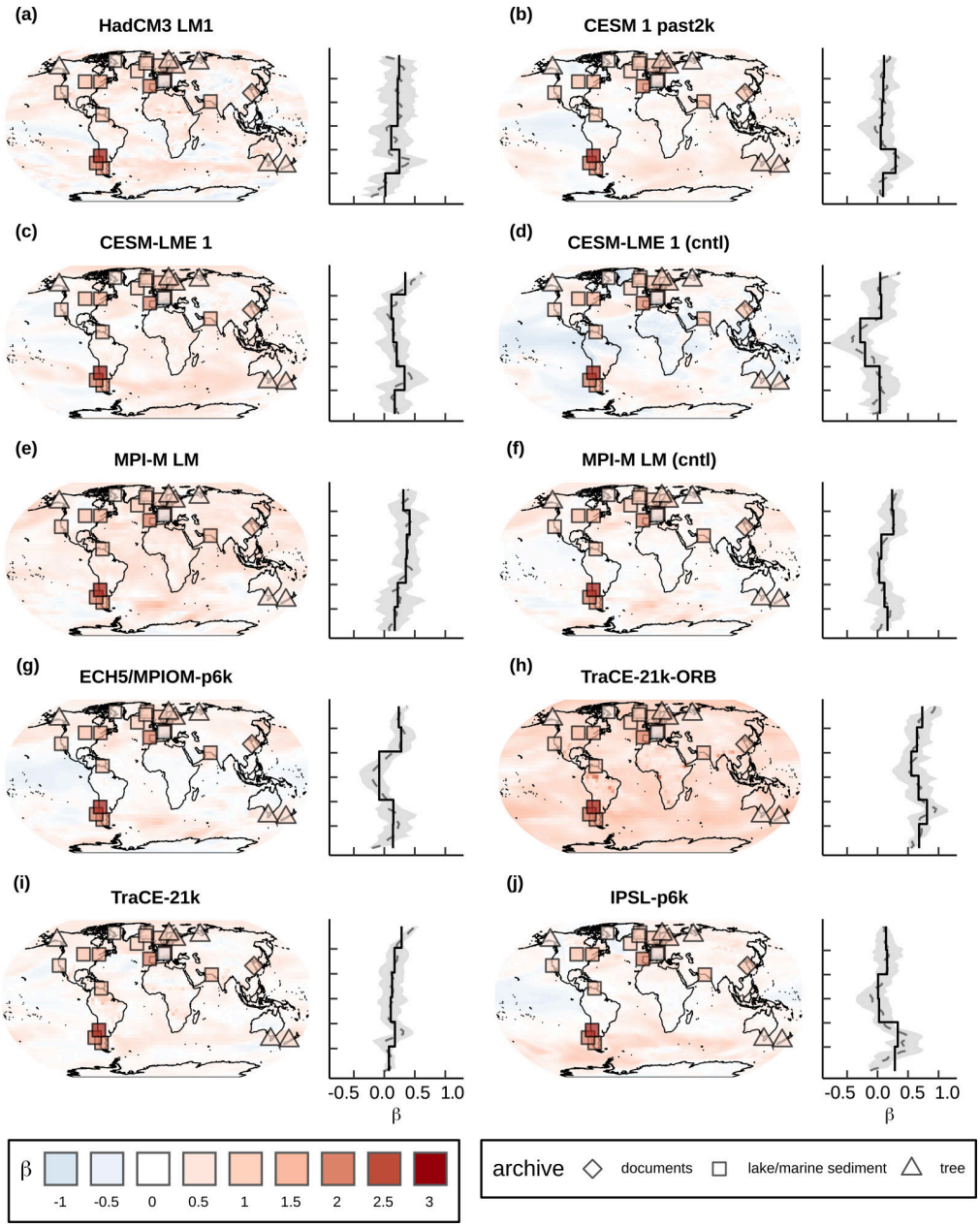


FIG. 3. Local temperature persistence on timescales from $\tau_1 = 10$ to $\tau_2 = 200$ yr across multiple climate simulations and selected proxy records from the PAGES2k database. Colors from blue to red indicate the scaling behavior ranging from $\beta = -1$ to $\beta = 3$. Symbols indicate the scaling of proxy records from different natural archives. The background of each panel shows the β-values fitted to the PSD of the local grid box temperature from simulations. Zonal mean values (dashed curves) are given next to the map, with means (solid curves) over latitude intervals (with breaks at -60, -30, 0, 30, and 60° N) and gray shaded confidence intervals. The spatial coverage of proxy records is not sufficient for robust mean estimates, which is why only simulation data are shown here.

From both simulations and paleoclimate data, we can strengthen the argument by Fredriksen *et al.* [81] that there is no latitudinal dependence of β (Fig. 3), in contrast to previous studies, suggesting a possible linkage to the strength of the seasonal cycle [24]. Inspecting the simulations' β-values (background of Fig. 3), we find a small land-sea contrast. Strongest scaling occurs in the Southern Oceans in line with previous findings [81]. Ocean-sea ice interactions with characteristic timescales of the order of centuries and a generally increased internal variability over the oceans might explain these results.

We find generally lower values for the slope β in the ENSO and Indo-Pacific region. This could be attributed to the fact that (quasi-)oscillatory signals, such as active modes of internal variability, are reflected in the PSD as broad peaks and hence cannot be described by a scaling law. On the other hand, this finding is stronger in PI control runs compared to fully forced runs [Figs. 3(c)–3(f)]. Thus, residual effects of the re-

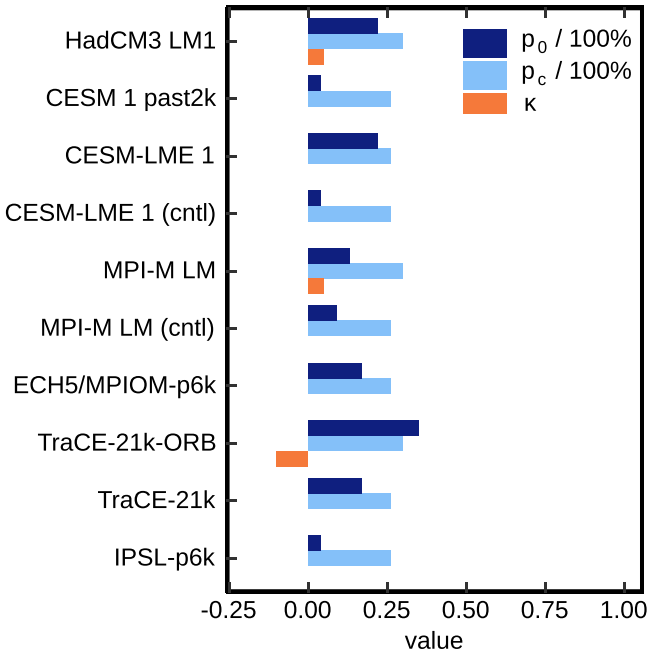


FIG. 4. Percentage agreement p_0, categorical agreement p_c and interrater reliability κ of local temperature persistence from simulations and paleoclimate data. The measures were calculated from a set of bilinearly interpolated simulation records and the proxy record at 23 different locations. Missing orange bars indicate no agreement beyond chance and, therefore, zero interrater reliability ($\kappa = 0$).

cent global warming trend might play an additional role [98]. A systematic bias becomes clear from the spatially almost uniform β-values of Trace21k-ORB [Fig. 3(h)]. In line with Fig. 2, we explain this by the lack of forcing mechanisms on interannual to multidecadal timescales in the aforementioned simulation.

Marine and lake sediments, as well as the archived documents, follow the general trend of increased β-values compared to simulations. Tree ring records agree well with most simulations in North America and Siberia, but not necessarily at the coast of Australia and northern Europe. Discrepancies such as those in southern South America could reflect the proxies' strength in representing local conditions, for example, topography. However, noise sources in the climate signal recording and preservation, such as bioturbation, can influence proxy records. Further separating the signal content from noise sources in paleoclimate reconstructions can help refine our findings [99,100].

C. Statistical agreement of temperature persistence

We further investigate the question of temperature scaling by a statistical analysis of β-values from simulations and paleoclimate reconstructions. It is based on the detailed uncertainty quantification outlined in Sec. III B. Our results show that reconstructions and simulations agree in less than 30% of locations within the scope of uncertainties (Fig. 4). To single out the scaling behavior of temperature signals, we study the agreement by category. We find approximately 25% of agreement within the categories $\beta < 1 - \nu$ (*low*) and $\beta > 1 + \nu$ (*high*). Although widely accepted [101], categorical and percentage agreement suffer from the limitation to ignore any agreement by chance. Therefore, we investigate the κ-statistics (orange bar in Fig. 4) and verify that there is no agreement beyond chance ($\kappa = 0$) for almost all models. Only MPI-M LM and HadCM3 LM1 show any, if poor agreement ($\kappa \approx 0.1$), whereas Trace21k-ORB shows even lower agreement than expected by chance ($\kappa < 0$) due to its systematic bias.

The disagreement could be attributed to both paleoclimate data and simulations. A systematic bias could arise, for example, through the recent, nonstationary global warming trend. Therefore, we repeat our analysis with all time series cut at 1850. In particular, anthropogenic warming slightly increases long-term temperature variability and thus scaling behavior, but not significantly (Figs. S6, S9, and S10 [47]). Further uncertainties could arise from our choice of statistical estimator for the scaling exponent β. Maximum likelihood estimation (MLE) should generally be preferred over linear regression (LR) because of its mathematical soundness and skillfulness [102]. We find that MLE is indeed more accurate for regular time series with $\beta > 0$ (Fig. S14 [47]). However, LR allows for estimation of $\beta < 1$, unlike MLE which assumes $\beta > 1$ [102]. In addition, for the characteristics of our empirical data, the differences between the two methods are not significant for $\beta > 0$ (Fig. S15 [47]). Therefore, linear regression represents the preferred estimator for our analysis. Regardless of the chosen estimator, we observe a slight tendency towards increased scaling exponents for irregularly sampled data (Fig. S15 [47]), similar to Lucke *et al.* [100]. Our uncertainty quantification carefully accounts for these potential errors due to irregular sampling and interpolation by simulating their influence using surrogates (Fig. S8 [47]).

We do not expect other systematic biases for the paleoclimate data since we base our results on multiple archives and proxies, and no systematic spatial pattern is discernible (Fig. S11 [47]). In particular, the cross-correlations between the 23 proxy data sets are weakly positive (0.02 on average with 95% quantiles of -0.17 to 0.21). The assumption of spatial independence necessary for robust statistical analysis (Fig. S16 [47]) therefore appears fully satisfied. The models' resolutions are another possible element of uncertainty that impacts variability over a wide range of timescales [103–105]. We here facilitate intermodel comparison by using state-of-the-art GCMs with comparable spatial and temporal resolutions, but computational costs precluded higher resolutions. The latter might be necessary to improve the representation of decadal variability and response to external forcing. In particular, the increased scaling exponents ($\beta > 1$) from paleoclimate data could indicate that nonlinear processes from an interactive carbon cycle and dynamical ice sheets might not be sufficiently represented in models.

D. The forced temperature response

Climatic drivers are not constant in time and thus affect the surface air temperature on multiple timescales. To investigate the forced temperature response, we present spectra for the main climatic drivers in Fig. 5. The PSD of orbital forcing consists of the diurnal and annual cycle as well as a back-

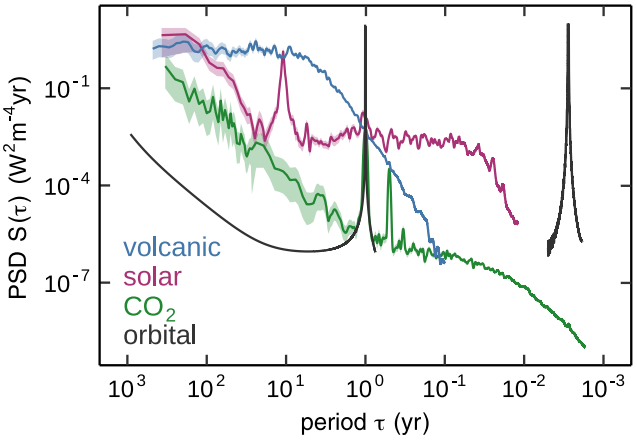


FIG. 5. Power spectral densities from radiative forcings. Details on the reconstructions considered here are summarized in Table S2 and Fig. S5 [47].

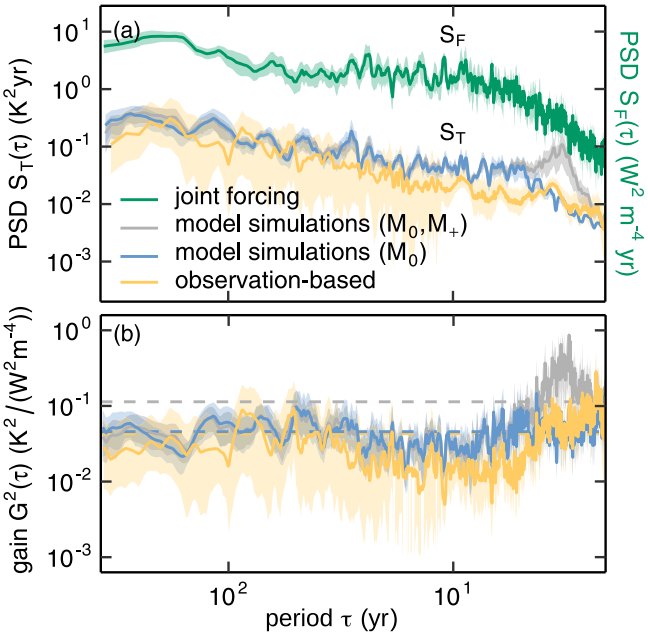


FIG. 6. Monte Carlo simulation of PSD (a) and spectral gain (b) using temperature and forcing reconstructions as well as model simulations. Shaded confidence intervals lie between the 5% and 95% quantiles. We consider only models from the groups M_0 and M_+ (Table S1 [47]) to exclude model artifacts and to represent the historical temperature response in the best possible way. Notably, M_+ contains those simulations with amplified ENSO [94]. (b) Dashed lines indicate the mean variance ratio $\langle S_T \rangle / \langle S_F \rangle$.

ground continuum on longer timescales. Higher harmonics on monthly timescales were omitted. We calculate the mean volcanic, solar, and CO_2 spectra using an equally weighted average of spectra from multiple data sets (Fig. S5 [47]). The CO_2 forcing follows the orbital forcing. The PSD of solar forcing again contains more power and has a pronounced peak around the 11 yr solar cycle. Multiple theories and paleoclimate reconstructions suggest the increased variability on centennial to millennial periods due to the long-term behavior of solar activity [106].

Volcanic forcing dominates interannual to centennial scales and undergoes a scale break around the period of 7 yr, estimated using the goodness of fit [102]. Above decadal scales, it follows a white noise spectrum with constant variance. However, the intermittency of volcanic eruptions might have led to biases in the spectral characteristics [42]. We verify our results using an analytical approach described in Appendix D. Remarkably, the derived PSD of an ideal, intermittent time series with Poisson distributed return times explains our findings. We further demonstrate the scale break by a Monte Carlo simulation of the joint PSD of radiative forcing in Fig. 6(a). This finding raises the question of how the spectrum with a scale break translates into the continuous spectrum in Fig. 2(b).

We address this question by calculating the spectral gain (4) on periods between years and centuries in Fig. 6(b). Here observation-based data include HadCRUT4, ERA5, and PAGES2k again. To account for the model artifacts explained above, we calculate the gain from the model simulation group M_0 and together with group M_+ (see Table S1 [47]). We find that the spectral gain is similar from observation-based data and the model simulation group M_0, which is the one without artificially amplified ENSO. This suggests that both follow a similar distribution of timescale-dependent variability, as already indicated by Fig. 2(b). Large parts of the gain show constant behavior, which is most pronounced in M_0. In a simplified way, the gain might be approximated by an ideal linear amplifier or damper of the forcing with comparable internal variability on all timescales. However, we also find a dip around decadal scales, which is strongest in the gain from measurements. Inspecting Fig. 6(a), this can be explained by forming the ratio between a spectrum with a scale break ($\beta > 1 \rightarrow \beta \approx 0$) and one with moderate scaling ($\beta \approx 1$).

From this standpoint, internal variability slightly grows on periods from years to centuries when slow processes in the oceans, vegetation, land surface, and cryosphere become increasingly active (Fig. 1). While the model simulations follow this general pattern, they may not represent its amplitude correctly, for example, due to the lack of feedback mechanisms. In addition, a too high model diffusivity could cause the suppression of low-frequency variability in model simulations due to a faster energy dissipation over temporal scales [8]. The PAGES2k multiproxy reconstruction, stemming from palaeoclimate data, possibly underestimates internal variability on interannual scales. However, the mean variance ratios in Fig. 5(b) of the model estimates agree with those from observations in the global mean. This leaves us with a conundrum: the global mean temperature based on model simulations and observations is mostly consistent in its variability, scaling, and response to forcing. Notwithstanding, locally, the models show a much lower variance on longer timescales and different scaling behavior than reconstructions. Thus, it appears that the statistics of local fluctuations need to be optimized in models but without significantly altering global properties. To this end, the study of unforced ("spontaneous") oscillations [107] and abrupt transitions [108,109] in the climate system is one promising approach to improve the representation of local variation. Furthermore, higher-resolved ocean and at-

mosphere models with additional mechanisms such as ice sheet dynamics and an interactive carbon cycle might increase long-range dependence and persistence of local temperature in the future.

V. CONCLUSION

In summary, we have investigated the question of temperature variability on the timescale of years to centuries. To this end, we have presented power spectral densities for both local and global surface air temperature from simulation and observation-based data of the last millennia. On this basis, we concluded that locally there is a stronger scaling and increased variance in reconstructions as compared to simulations. Using statistical analysis, we found that local temperature series extracted from simulations and paleoclimate reconstructions show different scaling behavior, with proxy records hinting at a stronger persistence. Furthermore, we have largely extended the spectral analysis of climatic drivers by estimating the joint PSD from CO_2, solar, volcanic, and orbital forcing using Monte Carlo simulation. Hereby, we discovered a scale break at the period of approximately 7 yr. Moreover, we have presented the spectral gain, describing the timescale-dependent forced temperature response. We found that it is mostly consistent across data sets and indicates an increasing internal variability on timescales of decades to centuries.

Our analysis of the spectral gain was limited to global average values and those timescales where linearity can be reasonably assumed [42,43,110]. Nonlinearities are inherent to the climate system, for example, due to the temperature-albedo feedback. Thus, it will be necessary to examine their possible effects on multiple spatiotemporal scales to further extend this work. Studying nonlinearities could also shine new light on the mechanisms of scaling in Earth's climate, which are not yet fully understood and might be linked to nonlinearities as well [6]. Furthermore, we have focused on the current interglacial, the Holocene. This is because climate variability has been demonstrated to depend on the mean climate state [82]. Furthermore, major shifts in climate could potentially violate the basic assumption of weak stationarity for spectral analysis. Thus, the conclusions laid out here cannot be readily applied to other climate states, such as glacial periods, which is an issue for future studies. Clearly, understanding the dependence of temperature variability on global warming demands additional work.

Ideally, our findings should be replicated by employing models with increased internal variability on longer timescales and paleoclimate data that provides improved spatiotemporal resolution. In particular, investigating the relationship between spatial and temporal disagreement is a key task for future analyses. Optimized analysis of noise sources and spectral analysis of (pseudo-)proxy records could help to expand the data basis of proxy records with decadal resolution [59,111,112]. Regarding climate models, an improved representation of processes that increase Earth's long-term memory, such as an interactive carbon cycle and dynamical ice sheets, might strengthen the long-range dependence and persistence of surface air temperature. A better understanding of unforced low-frequency oscillations as well as abrupt changes will be necessary to improve the representation of local fluctuations and could further help to understand nonlinear feedback and possible bifurcations in the climate system. Future studies could also continue to explore how internally generated and externally forced variability compares on different spatial scales. Research on the interrelation between internal and forced changes, as well as local, regional, and global variability, might prove important and could be conducted using single-forcing experiments from ensembles of model simulations.

Managing climate risks requires a detailed understanding of temperature variability. Locally and on timescales between years and centuries, there is an urgency to address discrepancies to make further progress in climate modeling. In this study, we have singled out the key characteristics of temperature variability and showed that the timescale dependency of local temperature variations from observation-based data and model simulations differs. Our results have demonstrated that the scaling behavior and spectral gain are easy-to-use yet effective and promising tools for investigating variability in Earth's dynamic climate.

Code to reproduce all figures is available at [113].

ACKNOWLEDGMENTS

This manuscript is based upon data provided by the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP and PMIP. We thank the research groups listed in Tables S1 and S2 for producing and making available their data from model outputs, measurements, paleoclimate, and forcing reconstructions. This study benefited from discussions within the CVAS working group, a working group of the Past Global Changes (PAGES) project. We thank T. Gasenzer, T. Kunz, and N. Weitzel for discussions and J. Bühler, M. Casado, M. Schillinger, and E. Ziegler for helpful comments on the manuscript. We are grateful to Aimé Fournier and one anonymous referee for their constructive and valuable review. This research has been funded by the Heidelberg Graduate School for Physics, by the PalMod project (subProject no. 01LP1926C), and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project No. 395588486.

APPENDIX A: RELATION BETWEEN POWER SPECTRAL DENSITY AND VARIANCE

The power spectral density of a weakly stationary, stochastic process is given by the Fourier transform of the autocorrelation $S(f) = \mathcal{F}\{R(h)\}$ with frequency f and lag $h = t_2 - t_1$ between two points in time [19,20]. For zero lag and zero mean, the integral of the PSD corresponds to the variance of the signal [80]. Instead of frequency, we use the period $\tau = 1/f$ to express the PSD and spectral gain. The integration of expression (1) is divergent for $\beta < 1$ and $f \to \infty$ which requires a high-frequency cutoff, such as described by Lovejoy *et al.* [38]. In case of temperature time series considered here, this is naturally defined by the temporal resolution, setting the maximum frequency.

APPENDIX B: AUTOCOVARIANCE OF LONG-RANGE MEMORY PROCESSES

Fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) are fully described by their correlation properties [30,114], summarized below. The autocovariance function of fBm $B(t)$ reads

$$\gamma(t',t) = \langle B(t')B(t)\rangle = \frac{V_\beta}{2}(|t|^{\beta-1} + |t'|^{\beta-1} - |t'-t|^{\beta-1}) \propto 1 + \left|\frac{t'}{t}\right|^{\beta-1} - \left|1-\frac{t'}{t}\right|^{\beta-1} \tag{B1}$$

for $1 < \beta < 3$. V_β is a positive constant factor related to $\langle (B(t') - B(t))^2\rangle = V_\beta\,|t'-t|^{\beta-1}$. By definition, fGn is the series of stationary increments $B(t') - B(t)$ and shows spectral exponent $-1 < \beta' = \beta - 2 < 1$ for $f \ll 1/\pi\Delta t$ with $\Delta t = t' - t$. Its autocovariance

$$\gamma(h) = \langle [B(t+1+h) - B(t+h)][B(t+1) - B(t)]\rangle = \frac{V_\beta}{2}|h-1|^{\beta'+1} - 2|h|^{\beta'+1} + |h+1|^{\beta'+1} \tag{B2}$$

depends only on the lag $h \in \mathbb{Z}$, where we set $\Delta t = 1$ without loss of generality. The fGn has a power spectrum of the form [115]

$$S(f) \propto \frac{\sin^2(\pi\Delta t f)}{|2\pi\Delta t f|^{\beta'+2}}, \tag{B3}$$

with the slowly varying factor

$$\sin^2(\pi\Delta t f) \xrightarrow[f/f_{\max}\to 0]{} (\pi\Delta t)^2 f^2, \qquad f_{\max} = 1/\pi\Delta t. \tag{B4}$$

Considering positive frequencies $f > 0$, the spectrum (B3) can be approximated by the power law $S(f) \sim 1/f^{\beta'}$ if $f \ll f_{\max}$. For $f \gtrsim f_{\max}$, however, the fGn has a similar spectral shape to fBm [114]. We account for this by considering sufficiently long periods. To give an example, $10^{0.58}\ \mathrm{yr}^{-1} \lesssim f_{\max} \lesssim 10^{2.7}\ \mathrm{yr}^{-1}$ corresponds to 6 h $\lesssim \Delta t \lesssim$ 1 mo.

For all $|t'/t| \gg 1$, the covariances (B1) keep growing for $\beta > 2$ (persistence) and stay bounded for $\beta < 2$ (antipersistence). As a result, Eq. (B1) involves "nonlinear pseudo-trends" [30] for $B(t')$ conditioned on $B(t)$, which diverge for $\beta > 2$ and converge for $\beta < 2$. According to Eq. (B2), fGn is persistent for $\beta' > 0$ and antipersistent for $\beta' < 0$. Ordinary Brownian motion corresponds to $\beta = 2$ and white noise to $\beta' = 0$. The sequence of partial sums of the autocovariance function diverges for fGn with $\beta' > 0$ and fBm with $\beta > 2$. The process is nonsummable and said to possess long-range memory.

APPENDIX C: SPECTRAL GAIN FOR LINEAR SYSTEMS

In a time-invariant linear system, the output

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)\,du \tag{C1}$$

is given by the input time series $x(t)$ and the impulse response function $h(u)$ [80]. The Fourier transform $H(f) = \mathcal{F}\{h(u)\} = G(f)e^{i\phi(f)}$ gives the frequency response function, also called the transfer function. $G(f)$ and $\phi(f)$ are the gain and phase, respectively. The integral (C1) corresponds to a product in frequency space $\mathcal{F}\{y(t)\} = H(f)\mathcal{F}\{x(t)\}$. This relates the PSD of the output $S_y(f)$ to the one of the input $S_x(f)$ via

$$S_y(f) = |H(f)|^2 S_x(f) = G^2(f)S_x(f). \tag{C2}$$

APPENDIX D: ANALYTICAL SOLUTION TO THE PSD OF INTERMITTENT VOLCANIC FORCING

We investigate the power spectral density of intermittent volcanic forcing by approximating the eruption time series in a simplified way as a stochastic signal $X(t) = \delta(t - t_i)$. This function is zero at all times except t_i, when an event of unique amplitude occurs. We denote $T_i = t_i - t_{i-1}$ the time intervals between two events. We use the fact that the PSD cannot be calculated only from the covariance, but also from the Laplace transform $S(X,f) = 2\lim_{\epsilon\to 0}\langle|\mathcal{L}(X(t), \frac{\epsilon}{2} - 2\pi\mathrm{i}f)|^2\rangle$ [116]. Based on this approach, the power spectral density

$$S(f) = \mu_T \frac{1-|\rho(f)|^2}{|1-\rho(f)|^2}, \quad f > 0 \tag{D1}$$

becomes a function of the Fourier transform of the probability density function $\rho(f) = \mathcal{F}\{\rho(T)\}$ and the inverse mean interval between two events $\mu_T = \langle T\rangle^{-1}$ [116,117]. An exponentially decaying probability distribution $\rho(T) = \mu_T \exp(-\mu_T T)\Theta(T)$ for volcanic forcing is suggested [118], and we have checked this for the data sets considered. The Fourier transform reads $\rho(f) = \mu_T(\mu_T + 2\pi\mathrm{i}f)^{-1}$ such that $1 - |\rho(f)|^2 = |1-\rho(f)|^2$. As a consequence, the PSD (D1) takes a constant value. We can observe this white noise behavior in Figs. 5 and 6(a) on timescales longer than a few years, which is on the order of characteristic return times for eruptions. Below these timescales, the variability considerably drops. This analytical result provides an independent verification of the PSD for volcanic forcing and its scale break.

APPENDIX E: MONTE CARLO SAMPLING OF THE SPECTRAL GAIN

We simulate the spectral gain (4), as well as the PSD of global mean temperature and the joint PSD of radiative forcing using a Monte Carlo approach with $N = 1000$ realizations to account for sampling biases. The PSD of global mean temperature is sampled for three groups: the observation-based data, the model simulations from group M_0, and those from M_0 together with M_+ (Table S1 [47]). Here only models from the groups M_0 and M_+ are considered to exclude model artifacts and to represent the historical temperature response in the best possible way.

We sample the simulation-based PSD from the average PSD of the simulations using uniformly distributed random weights. To obtain the observation-based PSD, we use the global mean temperature from HadCRUT4, ERA5, and a 7000-member reconstruction ensemble provided by PAGES2k [46]. This ensemble allows us to sample the PSD by randomly selecting one ensemble member and form the mean of its spectrum with that of the ERA5 and HadCRUT4 temperature. The joint PSD of radiative forcing is calculated from all forcing reconstructions considered in this work ex-

cept the Fröhlich *et al.* solar forcing, which has too low temporal resolution above interannual scales (Table S3 and Fig. S5 [47]). We assume the PSD of CO_2 and orbital forcing as fixed since its spectral power is comparatively low on multidecadal scales. We sample the PSD of solar forcing by using uniformly distributed weights when forming the average PSD of all solar reconstructions. Similarly, the PSD of volcanic forcing is obtained. In addition, we randomly vary the conversion factor between $(-18)^{-1}$ and $(-25)^{-1}$ $\mathrm{Wm}^{-2}/\mathrm{AOD}$ [78]. The joint PSD of radiative forcing is calculated by linear summation of the PSD from CO_2, orbital, solar, and volcanic forcing.

Using this sampling scheme, our Monte Carlo produces two outcomes: First, we compute the PSD of global mean temperature and the joint PSD of radiative forcing by simulating an ensemble of N realizations for both forcing and response. Second, we sample the spectral gain directly from the quotient (4) in each of the N realizations. In both cases, the average of the generated N-member ensemble and its 5% and 95% quantiles constitute the result of our Monte Carlo simulation.

[1] J. Bjerknes, A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature, Tellus **18**, 820 (1966).

[2] S. Arrhenius, XXXI. On the influence of carbonic acid in the air upon the temperature of the ground, London Edinburgh Dublin Philos. Mag. J. Sci. **41**, 237 (1896).

[3] J. B. J. Fourier, Remarques générales sur les températures du globe terrestre et des espaces planétaires, Ann. Chim. Phys. **27**, 136 (1824).

[4] R. W. Katz and B. G. Brown, Extreme events in a changing climate: Variability is more important than averages, Clim. Chan. **21**, 289 (1992).

[5] M. Ghil and V. Lucarini, The physics of climate variability and climate change, Rev. Mod. Phys. **92**, 035002 (2020).

[6] C. L. E. Franzke, S. Barbosa, R. Blender, H.-B. Fredriksen, T. Laepple, F. Lambert, T. Nilsen, K. Rypdal, M. Rypdal, M. G. Scotto *et al.*, The structure of climate variability across scales, Rev. Geophys. **58**, e2019RG000657 (2020).

[7] J. E. Tierney *et al.*, Past climates inform our future, Science **370**, eaay3701 (2020).

[8] T. Laepple and P. Huybers, Global and regional variability in marine surface temperatures, Geophys. Res. Lett. **41**, 2528 (2014).

[9] T. Laepple and P. Huybers, Ocean surface temperature variability: Large model-data differences at decadal and longer periods, Proc. Nat. Acad. Sci. U. S. A. **111**, 16682 (2014).

[10] L. A. Parsons, G. R. Loope, J. T. Overpeck, T. R. Ault, R. Stouffer, and J. E. Cole, Temperature and precipitation variance in CMIP5 simulations and paleoclimate records of the last millennium, J. Clim. **30**, 8885 (2017).

[11] F. C. Ljungqvist, Q. Zhang, G. Brattström, P. J. Krusic, A. Seim, Q. Li, Q. Zhang, and A. Moberg, Centennial-scale temperature change in last millennium simulations and proxy-based reconstructions, J. Clim. **32**, 2441 (2019).

[12] J. C. Bühler, C. Roesch, M. Kirschner, L. Sime, M. D. Holloway, and K. Rehfeld, Comparison of the oxygen isotope signatures in speleothem records and iHadCM3 model simulations for the last millennium, Clim. Past **17**, 985 (2021).

[13] W. B. Anderson, R. Seager, W. Baethgen, M. Cane, and L. You, Synchronous crop failures and climate-forced production variability, Sci. Adv. **5**, eaaw1976 (2019).

[14] M. Crucifix, A. D. Vernal, and C. Franzke, Centennial to millennial climate variability, Past Global Changes Mag. **25**, 131 (2017).

[15] M. Rypdal, H. B. Fredriksen, E. Myrvoll-Nilsen, K. Rypdal, and S. H. Sørbye, Emergent scale invariance and climate sensitivity, Climate **6**, 93 (2018).

[16] C. L. E. Franzke and T. J. O'Kane (eds.), *Nonlinear and Stochastic Climate Dynamics* (Cambridge University Press, Cambridge, 2017).

[17] T. P. Barnett *et al.*, Detection and attribution of recent climate change: A status report, Bull. Am. Meteorol. Soc. **80**, 2631 (1999).

[18] N. Bindoff *et al.*, Detection and attribution of climate change: From global to regional, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Intergovernmental Panel on Climate Change (Cambridge University Press, Cambridge, 2013), pp. 867–952.

[19] N. Wiener, Generalized harmonic analysis, Acta Math. **55**, 117 (1930).

[20] A. Khintchine, Korrelationstheorie der stationären stochastischen Prozesse, Math. Ann. **109**, 604 (1934).

[21] C. Wunsch, The spectral description of climate change including the 100 ky energy, Clim. Dyn. **20**, 353 (2003).

[22] K. Fraedrich, U. Luksch, and R. Blender, $1/f$ model for long-time memory of the ocean surface temperature, Phys. Rev. E **70**, 037301 (2004).

[23] C. Franzke, Nonlinear trends, long-range dependence, and climate noise properties of surface temperature, J. Clim. **25**, 4172 (2012).

[24] P. Huybers and W. Curry, Links between annual, Milankovitch and continuum temperature variability, Nature (London) **441**, 329 (2006).

[25] S. Lovejoy, A voyage through scales, a missing quadrillion and why the climate is not what you expect, Clim. Dyn. **44**, 3187 (2015).

[26] T. Nilsen, K. Rypdal, and H. B. Fredriksen, Are there multiple scaling regimes in Holocene temperature records? Syst. Dyn. Rev. **7**, 419 (2016).

[27] H. B. Fredriksen and M. Rypdal, Long-range persistence in global surface temperatures explained by linear multibox energy balance models, J. Clim. **30**, 7157 (2017).

[28] K. Rypdal, L. Østvand, and M. Rypdal, Long-range memory in Earth's surface temperature on time scales from months to centuries, J. Geophys. Res.: Atmosph. **118**, 7046 (2013).

[29] F. Zhu, J. Emile-Geay, N. P. McKay, G. J. Hakim, D. Khider, T. R. Ault, E. J. Steig, S. Dee, and J. W. Kirchner, Climate models can correctly simulate the continuum of global-

average temperature variability, Proc. Nat. Acad. Sci. U. S. A. **116**, 8728 (2019).
[30] B. B. Mandelbrot and J. W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. **10**, 422 (1968).
[31] B. D. Malamud and D. L. Turcotte, Self-affine time series: Measures of weak and strong persistence, J. Stat. Plan. Infer. **80**, 173 (1999).
[32] J. P. Peixto and A. H. Oort, Physics of climate, Rev. Mod. Phys. **56**, 365 (1984).
[33] E. J. Rohling *et al.*, Making sense of palaeoclimate sensitivity, Nature (London) **491**, 683 (2012).
[34] E. J. Rohling, G. Marino, G. L. Foster, P. A. Goodwin, A. S. von der Heydt, and P. Köhler, Comparing climate sensitivity, past and present, Ann. Rev. Marine Sci. **10**, 261 (2018).
[35] F. Zhang, Y. Qiang Sun, L. Magnusson, R. Buizza, S. J. Lin, J. H. Chen, and K. Emanuel, What is the predictability limit of midlatitude weather? J. Atmos. Sci. **76**, 1077 (2019).
[36] J. D. Pelletier, Natural variability of atmospheric temperatures and geomagnetic intensity over a wide range of time scales, Proc. Nat. Acad. Sci. U. S. A. **99**, 2546 (2002).
[37] J. Lohmann and P. D. Ditlevsen, Random and externally controlled occurrences of Dansgaard–Oeschger events, Clim. Past. **14**, 609 (2018).
[38] S. Lovejoy and F. Lambert, Spiky fluctuations and scaling in high-resolution EPICA ice core dust fluxes, Clim. Past **15**, 1999 (2019).
[39] K. Marvel, G. A. Schmidt, R. L. Miller, and L. S. Nazarenko, Implications for climate sensitivity from the response to individual forcings, Nat. Clim. Change **6**, 386 (2016).
[40] A. P. Schurer, G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps, Separating forced from chaotic climate variability over the past millennium, J. Clim. **26**, 6954 (2013).
[41] S. Lovejoy and D. Schertzer, Stochastic and scaling climate sensitivities: Solar, volcanic and orbital forcings, Geophys. Res. Lett. **39**, L11702 (2012).
[42] S. Lovejoy and C. Varotsos, Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings, Syst. Dyn. Rev. **7**, 133 (2016).
[43] K. Rypdal and M. Rypdal, Comment on "Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcing" by S. Lovejoy and C. Varotsos (2016), Syst. Dyn. Rev. **7**, 597 (2016).
[44] J. Mitchell, An overview of climatic variability and its causal mechanisms, Quat. Res. **6**, 481 (1976).
[45] A. Henderson-Sellers and K. McGuffie, *A Climate Modelling Primer* (Wiley, Chichester, 1987).
[46] PAGES 2k Consortium, Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nat. Geosci. **12**, 643 (2019).
[47] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.104.064136 for further information on data sets and benchmarks.
[48] B. L. Otto-Bliesner, E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai, and G. Strand, Climate variability and change since 850 CE an ensemble approach with the Community Earth System Model, Bull. Am. Meteorol. Soc. **97**, 735 (2016).
[49] J. H. Jungclaus *et al.*, Climate and carbon-cycle variability over the last millennium, Clim. Past **6**, 723 (2010).
[50] Y. Zhong, A. Jahn, G. H. Miller, and A. Geirsdottir, Asymmetric cooling of the atlantic and pacific arctic during the past two millennia: A dual observation-modeling study, Geophys. Res. Lett. **45**, 12,497 (2018).
[51] P. Braconnot, D. Zhu, O. Marti, and J. Servonnat, Strengths and challenges for transient Mid- to Late Holocene simulations with dynamical vegetation, Clim. Past **15**, 997 (2019).
[52] N. Fischer and J. H. Jungclaus, Evolution of the seasonal temperature cycle in a transient Holocene simulation: Orbital forcing and sea-ice, Clim. Past **7**, 1139 (2011).
[53] Z. Liu, Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming, Science **325**, 310 (2009).
[54] C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones, Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, J. Geophys. Res.: Atmos. **117**, D08101 (2012).
[55] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara *et al.*, The ERA5 global reanalysis, Q. J. R. Meteorol. Soc. **146**, 1999 (2020).
[56] PAGES2k Consortium., A global multiproxy database for temperature reconstructions of the Common Era, Sci. Data **4**, 170088 (2017).
[57] J. W. Kantelhardt, Fractal and multifractal time series, in *Mathematics of Complexity and Dynamical Systems*, edited by R. A. Meyers (Springer, New York, 2011), pp. 463–487.
[58] T. Laepple, T. Münch, M. Casado, M. Hoerhold, A. Landais, and S. Kipfstuhl, On the similarity and apparent cycles of isotopic variations in East Antarctic snow pits, Cryosphere **12**, 169 (2018).
[59] M. Casado, T. Münch, and T. Laepple, Climatic information archived in ice cores: Impact of intermittency and diffusion on the recorded isotopic signal in Antarctica, Clim. Past **16**, 1581 (2020).
[60] T. J. Crowley, Causes of climate change over the past 1000 years, Science **289**, 270 (2000).
[61] G. C. Hegerl, T. J. Crowley, M. Allen, W. T. Hyde, H. N. Pollack, J. Smerdon, and E. Zorita, Detection of human influence on a new, validated 1500-year temperature reconstruction, J. Clim. **20**, 650 (2007).
[62] P. Braconnot, S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao, Evaluation of climate models using palaeoclimatic data, Nat. Clim. Change **2**, 417 (2012).
[63] G. A. Schmidt *et al.*, Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), Geosci. Model Dev. **5**, 185 (2012).
[64] G. Delaygue and E. Bard, An Antarctic view of Beryllium-10 and solar activity for the past millennium, Clim. Dyn. **36**, 2201 (2011).
[65] F. Steinhilber, J. Beer, and C. Fröhlich, Total solar irradiance during the Holocene, Geophys. Res. Lett. **36**, L19704 (2009).
[66] Y. Wang, J. L. Lean, and N. R. Sheeley, Jr., Modeling the Sun's magnetic field and irradiance since 1713, Astrophys. J. **625**, 522 (2005).
[67] R. Muscheler, F. Joos, J. Beer, S. A. Müller, M. Vonmoos, and I. Snowball, Solar activity during the last 1000 yr inferred from radionuclide records, Quat. Sci. Rev. **26**, 82 (2007).

[68] L. E. Vieira and S. K. Solanki, Evolution of the solar magnetic flux on time scales of years to millenia, Astron. Astrophys. **509**, A100 (2010).

[69] C. Gao, A. Robock, and C. Ammann, Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models, J. Geophys. Res.: Atmos. **113**, D23111 (2008).

[70] A. L. Berger, Long-term variations of daily insolation and Quaternary climatic changes, J. Atmos. Sci. **35**, 2362 (1978).

[71] M. Crucifix, Palinsol–Package (R) (2016), https://CRAN.R-project.org/package=palinsol.

[72] M. Toohey and M. Sigl, Reconstructed volcanic stratospheric sulfur injections and aerosol optical depth, 500 BCE to 1900 CE, version 2, World Data Center for Climate (WDCC) at DKRZ, https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=eVolv2k_v2 (2017).

[73] C. Fröhlich, Solar irradiance variability since 1978: Revision of the PMOD composite during solar cycle 21, Space Sci. Rev. **125**, 53 (2006).

[74] C. D. Keeling, R. B. Bacastow, A. E. Bainbridge, C. A. Ekdahl Jr., P. R. Guenther, L. S. Waterman, and John F. S. Chin, Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii, Tellus **28**, 538 (1976).

[75] J. Pongratz, C. Reick, T. Raddatz, and M. Claussen, A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochem. Cycles **22**, GB3018 (2008).

[76] G. Myhre, E. J. Highwood, K. P. Shine, and F. Stordal, New estimates of radiative forcing due to well mixed greenhouse gases, Geophys. Res. Lett. **25**, 2715 (1998).

[77] G. A. Schmidt *et al.*, Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geosci. Model Dev. **4**, 33 (2011).

[78] G. Myhre *et al.*, Anthropogenic and natural radiative forcing supplementary naterial, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Intergovernmental Panel on Climate Change (Cambridge University Press, Cambridge, 2014), pp. 1–30.

[79] T. Laepple and P. Huybers, Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability, Earth Planet. Sci. Lett. **375**, 418 (2013).

[80] C. Chatfield and H. Xing, *The Analysis of Time Series: An Introduction with R*, 7th ed., Chapman & Hall/CRC Texts in Statistical Science (CRC Press, Boca Raton, FL, 2019).

[81] H. B. Fredriksen and K. Rypdal, Spectral characteristics of instrumental and climate model surface temperatures, J. Clim. **29**, 1253 (2016).

[82] K. Rehfeld, T. Münch, S. L. Ho, and T. Laepple, Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene, Nature (London) **554**, 356 (2018).

[83] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge University Press, Cambridge, 1993).

[84] P. Yiou, E. Baert, and M. F. Loutre, Spectral analysis of climate data, Surv. Geophys. **17**, 619 (1996).

[85] J. W. Kirchner, Aliasing in $1/f^a$ noise spectra: Origins, consequences, and remedies, Phys. Rev. E **71**, 066110 (2005).

[86] L. Østvand, T. Nilsen, K. Rypdal, D. Divine, and M. Rypdal, Long-range memory in internal and forced dynamics of millennium-long climate model simulations, Syst. Dyn. Rev. **5**, 295 (2014).

[87] J. L. Fleiss and J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. **33**, 613 (1973).

[88] O. Geoffroy, D. Saint-Martin, D. J. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments, J. Clim. **26**, 1841 (2013).

[89] D. G. MacMynowski, H.-J. Shin, and K. Caldeira, The frequency response of temperature and precipitation in a climate model, Geophys. Res. Lett. **38**, L16711 (2011).

[90] A. Kirkevåg, T. Iversen, J. E. Kristjánsson, Ø. Seland, and J. B. Debernard, On the additivity of climate response to anthropogenic aerosols and CO_2, and the enhancement of future global warming by carbonaceous aerosols, Tellus, Ser. A: Dyn. Meteorol. Oceanogr. **60**, 513 (2008).

[91] H. Shiogama, D. A. Stone, T. Nagashima, T. Nozawa, and S. Emori, On the linear additivity of climate forcing-response relationships at global and continental scales, Int. J. Climatol. **33**, 2542 (2013).

[92] G. A. Meehl, W. M. Washington, C. M. Ammann, J. M. Arblaster, T. M. Wigley, and C. Tebaldi, Combinations of natural and anthropogenic forcings in twentieth-century climate, J. Clim. **17**, 3721 (2004).

[93] V. Ramaswamy and C. T. Chen, Linear additivity of climate response for combined albedo and greenhouse perturbations, Geophys. Res. Lett. **24**, 567 (1997).

[94] J. Jungclaus (private communication) (2020).

[95] G. Danabasoglu, On multidecadal variability of the Atlantic meridional overturning circulation in the community climate system model version 3, J. Clim. **21**, 5524 (2008).

[96] T. Kunz and T. Laepple, Frequency-dependent estimation of effective spatial degrees of freedom, J. Clim. **34**, 7373 (2021).

[97] T. Kunz and T. Laepple, Time-scale dependent estimation of spatial degrees of freedom, in *EGU General Assembly Conference Abstracts* (EGU General Assembly, Vienna, Austria, 2018), p. 17995.

[98] S.-W. Yeh, J.-S. Kug, B. Dewitte, M.-H. Kwon, B. P. Kirtman, and F.-F. Jin, El Niño in a changing climate, Nature (London) **461**, 511 (2009).

[99] M. Reschke, K. Rehfeld, and T. Laepple, Empirical estimate of the signal content of Holocene temperature proxy records, Clim. Past **15**, 521 (2019).

[100] L. J. Lücke, G. C. Hegerl, A. P. Schurer, and R. Wilson, Effects of memory biases on variability of temperature reconstructions, J. Clim. **32**, 8713 (2019).

[101] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*, 3rd ed., Wiley Series in Probability and Statistics (Wiley, New York, 2013).

[102] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, SIAM Rev. **51**, 661 (2009).

[103] B. P. Kirtman *et al.*, Impact of ocean model resolution on CCSM climate simulations, Clim. Dyn. **39**, 1303 (2012).

[104] J. M. Klavans, A. Poppick, S. Sun, and E. J. Moyer, The influence of model resolution on temperature variability, Clim. Dyn. **48**, 3035 (2017).

[105] D. L. Hodson and R. T. Sutton, The impact of resolution on the adjustment and decadal variability of the Atlantic meridional overturning circulation in a coupled climate model, Clim. Dyn. **39**, 3057 (2012).

[106] L. J. Gray *et al.*, Solar influences on climate, Rev. Geophys. **48**, (2010).

[107] E. Hertwig, F. Lunkeit, and K. Fraedrich, Low-frequency climate variability of an aquaplanet, Theor. Appl. Climatol. **121**, 459 (2015).

[108] W. R. Peltier and G. Vettoretti, Dansgaard-Oeschger oscillations predicted in a comprehensive model of glacial climate: A "kicked" salt oscillator in the Atlantic, Geophys. Res. Lett. **41**, 7306 (2014).

[109] M. Klockmann, U. Mikolajewicz, H. Kleppin, and J. Marotzke, Coupling of the subpolar gyre and the overturning circulation during abrupt glacial climate transitions, Geophys. Res. Lett. **47**, e2020GL090361 (2020).

[110] L. Fernández-Donado *et al.*, Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, Clim. Past **9**, 393 (2013).

[111] T. Kunz, A. M. Dolman, and T. Laepple, A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records—Part 1: Theoretical concept, Clim. Past **16**, 1469 (2020).

[112] A. M. Dolman, T. Kunz, J. Groeneveld, and T. Laepple, A spectral approach to estimating the timescale-dependent uncertainty of paleoclimate records–Part 2: Application and interpretation, Clim. Past. **17**, 825 (2021).

[113] https://github.com/paleovar/TimescaleDependency.

[114] A. Fournier (private communication) (2021).

[115] S. Mallat, Wavelet zoom, in *A Wavelet Tour of Signal Processing*, 3rd ed. (Academic Press, Boston, 2009), pp. 205–261.

[116] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1967).

[117] B. Lindner, Superposition of many independent spike trains is generally not a Poisson process, Phys. Rev. E **73**, 022901 (2006).

[118] P. Papale, Global time-size distribution of volcanic eruptions on Earth, Sci. Rep. **8**, 6838 (2018).

Supplementary Material for
Probing the timescale dependency of local and global surface air temperature variations from climate simulations and reconstructions of the last millennia

Beatrice Ellerhoff and Kira Rehfeld

LIST OF TABLES

LIST OF FIGURES

TABLE S1. Key specifications of model simulations and observation-based data used to estimate temperature variability. We give the main references (Doc.) of the simulation runs, observation-based data sets, and their forcing (Forc.). The groups M_+ and M_0 were assigned to better distinguish spectral properties when performing a Monte Carlo simulation of the spectral gain in the main manuscript. The temporal resolution Δt is given in months (m), hours (hr), years (yrs), and days. The spatial resolutions of atmosphere and ocean are denoted by the subscripts $^\circ\text{deg}_{at.}$ and $^\circ\text{deg}_{oc.}$, respectively.

Name	Doc.	Model	Δt	$^\circ\text{deg}_{at.}$	$^\circ\text{deg}_{oc.}$	Forc.	Time (CE)
Model simulations							
CESM 1 past2k[a]	[6]	CESM1	1 m	2	1	[7]	1-2005
CESM-LME 1[a]	[8]	CESM1	1 m, 6 hr	2	1	[9]	850-2006
CESM-LME 1 cntl	[8]	CESM1	1 m	2	1	[9]	850-2006
MPI-M LM[b]	[10]	MPI-ESM	1 m, 6 hr	3.75	GR30[c]	[10]	800-2005
MPI-M LM cntl	[10]	MPI-ESM	1 m	3.75	GR30[c]	[10]	800-2005
HadCM3 LM1[a]	[11]	iHadCM3	1 m	2.5 x 3.75	1.25	[12]	800-1850
IPSL-p6k[a]	[13]	IPSL-CM5A	2 m	2.5 x 1.27	2	[13]	4000 BCE - 2000
TraCE-21k	[14]	CCSM3	2 m	≈ 3.75	3.6 x v[d]	[15, 16]	4,000 BCE - 1990
TraCE-21k-ORB	[14]	CCSM3	10 yrs	≈ 3.75	3.6 x v[d]	[15, 16]	4,000 BCE - 1990
ECH5/MPIOM-p6k[b]	[17]	ECHAM5/MPI-OM	1 m	3.75	2	[15]	4000 BCE - 2000
Observation-based							
ERA5	[18]		1 m, 6 hr	2	2		1979-2019
HadCRUT4	[19]		7 days	5	5		1850-2019
PAGES2k	[3]		1 yr			[3]	0-2000

[a] assigned to group M_0

[b] assigned to M_+

[c] curvilinear grid with nominal resolution of 3.0°

[d] the latitudinal resolution is variable (v), with finer resolution near the equator ($\approx 0.90°$)

TABLE S2. Key specifications of climatic drivers used to estimate the power spectral density of radiative forcing as well as the gain function of the forced temperature response.

Name	Δt	Time (CE)
Volcanic forcing		
Crowley et al.[a] [21]	10 days	500BCE-1900
Gao et al.[a] [22]	1 yr	850-2000
Toohey et al. [23]		850-2000
Total solar irradiance		
Delaygue et al.[ab] [24]	1 yr	850-1850
Muscheler et al.[ab] [25]	1 yr	850-1850
Steinhilber et al.[a] [26]	1 yr	850-1850
Vieira et al.[a] [27, 28]	1 yr	850-1850
Wang et al.[ab] [29]	1 yr	1610-2009
Fröhlich et al. [30]	36 hr	1978-2017
CO_2		
Schmidt et al.[a] [20]	1 yr	850-2000
Keeling et al. [31]	1 hr	1970-2016
Insolation at 65°N		
Berger[ac] [15]	1 yr, 3 hr	0-2000

[a] from the PMIP simulations of the Last Millennium [20]

[b] when multiple versions were provided by PMIP, the "with-background"-version was considered here

[c] computed with the **R**-package "Palinsol" [32]

TABLE S3. Key specification of proxy records used to estimate local temperature variability. The first six columns ("ID"-"Proxy") are taken from the PAGES2k database [3]. The ID is additionally marked with a start ($\star$) in case the proxy record was used for calculating the agreement in scaling behavior between models and data. The cutoff identifies whether the proxy record was used for analyzing the potential impact of the recent global warming trend (Figure S6 and Figure S10). It indicates whether the necessary criteria for time series selection were fulfilled before 1850 (PI) and / or over the full historical (hist) period. Similarly, the selection column gives criteria for calculating the mean of local spectra and their comparison (Figure S6). The table spans multiple pages.

ID	Name	*Lat*	*Lon*	*Elev_masl*	Archive	Proxy	cutoff	selection
2	Africa-LakeTanganyi.Tierney.2010	−6	28.5	28	lake sediment	TEX86	PI/hist	loose
3	Africa-Malawi.Powers.2011	−10	34.3	34	lake sediment	TEX86	PI/hist	loose
4	Africa-P178-15.Tierney.2015	12	44.3	44	marine sediment	TEX86	PI/hist	strong/loose
39	Arc-BrayaSo.DAndrea.2011	67	−50.7	−51	lake sediment	alkenone	PI/hist	loose
40	Arc-Clegg2010	61.4	−143.6	−144	lake sediment	midge	PI/hist	loose
49*	Arc-GulfofAlaska.Wiles.2014	61	−146.6	−147	tree	TRW	PI/hist	strong/loose
50	Arc-HalletLake.McKay.2008	61.5	−146.2	−146	lake sediment	BSi	PI/hist	strong/loose
53	Arc-Iceland.Bergthorsson.1969	64.8	−18.4	−18	documents	historic	PI/hist	loose
56	Arc-Kongressvatn.DAndrea.2012	78	13.9	14	lake sediment	alkenone	PI/hist	loose
57	Arc-Lake4.Rolland.2009	65.1	−83.8	−84	lake sediment	chironomid	PI/hist	loose
59	Arc-LakeE.DAndrea.2011	67	−50.7	−51	lake sediment	alkenone	PI/hist	loose
60	Arc-LakePieni-Ka.Luoto.2010	64.3	30.1	30	lake sediment	chironomid	PI/hist	loose
65	Arc-Luoto2009	60.3	25.4	25	lake sediment	midge	PI/hist	strong/loose
66	Arc-MackenzieDelta.Porter.2013	68.6	−133.9	−134	tree	*missing*	hist	strong/loose
69	Arc-MD992275.Jiang.2005	66.5	−17.7	−18	marine sediment	diatom	PI/hist	strong/loose
81	Arc-SoperLakeBaf.Hughen.2000	62.9	−69.9	−70	lake sediment	varve thickness	PI/hist	strong/loose
82*	Arc-StoreggaSlid.Sejrup.2011	63.8	5.3	5	marine sediment	foram d18O	PI/hist	strong/loose
83*	Arc-Thomas2008	69.9	−68.8	−69	lake sediment	varve thickness	PI/hist	strong/loose
85*	Arc-Tornetrask.Melvin.2013	68.3	19.6	20	tree	TRW	PI/hist	strong/loose
88*	Arc-Yamalia.Briffa.2013	66.8	68	68	tree	TRW	PI/hist	strong/loose
122	Asia-CentralChina.Wang.1998	29	113	113	documents	historic	PI/hist	strong/loose
130	Asia-Chu.2012.Sihailongwanlake	42.2	126.4	126	lake sediment	alkenone	PI/hist	strong/loose
151	Asia-EastChina.Wang.1990	30	117.5	118	documents	historic	PI/hist	strong/loose
152*	Asia-EastChinareg.Wang.1998	34	120	120	documents	historic	PI/hist	strong/loose
157	Asia-FujianandTai.Wang.1998	24	121	121	documents	historic	PI/hist	strong/loose
167	Asia-Guangdong.Zheng.1982	23.2	113.2	113	documents	historic	PI/hist	strong/loose
168	Asia-Guangdongand.Zhang.1980	23.5	112.5	112	documents	historic	PI/hist	strong/loose
182	Asia-HunanJiangsu.Zhang.1980	28	116.5	116	documents	historic	PI/hist	strong/loose
205	Asia-KunashirIsla.Demezhko.2009	44	145.7	146	hybrid	hybrid	PI/hist	strong/loose
212	Asia-Lowerreaches.Zhang.1980	32.1	118.8	119	documents	historic	PI/hist	strong/loose
220	Asia-Middlereache.Zhang.1980	30.5	114.5	114	documents	historic	PI/hist	strong/loose
273*	Asia-SO9039KGSO1.Munz.2015	24.8	65.9	66	marine sediment	foraminifera	PI/hist	strong/loose
275	Asia-SourthandMid.Demezhko.2007	55	59.5	60	borehole	borehole	PI/hist	loose

ID	Name	*Lat*	*Lon*	*Elev_masl*	Archive	Proxy	cutoff	selection
276	Asia-SouthChina.Wang.1998	23	114	114	documents	historic	PI/hist	strong/loose
331	Asia-ZhejiangandF.Zhang.1980	25	118	118	documents	historic	PI/hist	strong/loose
342*	Aus-MtRead.Cook.2006	−41.8	145.5	146	tree	TRW	PI/hist	strong/loose
343*	Aus-Oroko.Cook.2002	−43.2	170.3	170	tree	TRW	PI/hist	strong/loose
347	Eur-CentralandEa.Pla.2005	42.5	0.8	1	lake sediment	chrysophyte	PI/hist	loose
349	Eur-CentralEu.Dobrovolny.2010	49	13	13	documents	Documentary	PI/hist	strong/loose
350*	Eur-CoastofPortu.Abrantes.2011	41.1	−8.9	−9	marine sediment	alkenone	PI/hist	strong/loose
352*	Eur-FinnishLakes.Helama.2014	62	28.3	28	tree	MXD	PI/hist	strong/loose
354	Eur-LakeSilvapla.Larocque-Tobler.2010	46.5	9.8	10	lake sediment	chironomid	PI/hist	strong/loose
355*	Eur-LakeSilvapla.Trachsel.2010	46.5	9.8	10	lake sediment	reflectance	PI/hist	strong/loose
357*	Eur-NorthIceland.Ran.2011	66.5	−17.7	−18	marine sediment	diatom	PI/hist	strong/loose
359*	Eur-Seebergsee.Larocque-Tobler.2012	46.1	7.5	8	lake sediment	midge	PI/hist	strong/loose
364	Eur-Stockholm.Leijonhufvud.2010	59.3	18.1	18	documents	historic	PI/hist	strong/loose
365	Eur-Tallinn.Tarand.2001	59.4	24.8	25	documents	historic	PI/hist	strong/loose
391	NAm-BasinP	44.5	−70.1	−70	lake sediment	pollen	PI/hist	loose
450	NAm-ClearP	33.8	−79	−79	lake sediment	pollen	PI/hist	loose
457	NAm-ConroyL	46.3	−67.9	−68	lake sediment	pollen	PI/hist	loose
459	NAm-DarkL	45.3	−91.5	−92	lake sediment	pollen	PI/hist	loose
463	NAm-HellKt	46.2	−89.7	−90	lake sediment	pollen	PI/hist	loose
466*	NAm-LakeMina	45.9	−95.5	−96	lake sediment	pollen	PI/hist	strong/loose
467	NAm-LClouds	48	−91	−91	lake sediment	pollen	PI/hist	strong/loose
468	NAm-LittlePineL	45.3	−91.5	−92	lake sediment	pollen	PI/hist	loose
469*	NAm-LNoir	45.8	−75.1	−75	lake sediment	pollen	PI/hist	strong/loose
485	NAm-RubyL	45.3	−91.5	−92	lake sediment	pollen	PI/hist	loose
507	O2kLR-AlboranSea-TTR17-1384B.Nieto-Moreno.2012	36	−4.7	−5	marine sediment	alkenone	PI/hist	loose
508	O2kLR-AlboranSea-TTR17-1436B.Nieto-Moreno.2012	36.2	−4.3	−4	marine sediment	alkenone	PI/hist	loose
510	O2kLR-ArabianSea.Doose-Rolinski.2001	24.8	65.9	66	marine sediment	alkenone	PI/hist	loose
511	O2kLR-CapeGhir.Kim.2007	30.9	−10.3	−10	marine sediment	alkenone	PI/hist	loose
512	O2kLR-CapeGhir.McGregor.2007	30.8	−10.1	−10	marine sediment	alkenone	PI/hist	loose
513	O2kLR-CapeHatteras.Clroux.2012	35	−75.2	−75	marine sediment	foram Mg/Ca	PI/hist	strong/loose
514*	O2kLR-CariacoBasin.Black.2007	10.8	−64.8	−65	marine sediment	foram Mg/Ca	PI/hist	strong/loose
517	O2kLR-ChileanMargin.Lamy.2002	−41	−74.5	−74	marine sediment	alkenone	PI/hist	loose
518	O2kLR-DryTortugas.Lund.2006	24.6	−83.6	−84	marine sediment	foram Mg/Ca	PI/hist	loose

ID	Name	*Lat*	*Lon*	*Elev_masl*	Archive	Proxy	cutoff	selection
519	O2kLR-DryTortugasA.Lund.2006	24.3	−83.3	−83	marine sediment	foram Mg/Ca	PI/hist	loose
520	O2kLR-EasternTropicalNorthAtlantic.Kuhnert.2011	16.8	−16.7	−17	marine sediment	foram Mg/Ca	PI/hist	loose
521	O2kLR-EmeraldBasin.Keigwin.2003	45.9	−62.8	−63	marine sediment	alkenone	PI/hist	loose
523	O2kLR-FeniDriftRichter.2009	55.5	−13.9	−14	marine sediment	foram Mg/Ca	PI/hist	loose
524	O2kLR-FiskBasin.Richey.2009	27.6	−93.9	−94	marine sediment	foram Mg/Ca	PI/hist	strong/loose
525	O2kLR-GarrisonBasin.Richey.2009	26.7	−93.9	−94	marine sediment	foram Mg/Ca	hist	loose
526	O2kLR-GreatBahamaBank.Lund.2006	24.6	−79.3	−79	marine sediment	foram Mg/Ca	PI/hist	loose
527	O2kLR-GreatBahamabAtlantic0326aLund2006125MC	24.8	−79.3	−79	marine sediment	foram Mg/Ca	PI/hist	loose
528	O2kLR-GreatBarrier.Hendy.2002	−18.3	146.6	147	coral	Coral Sr/Ca	PI/hist	strong/loose
529	O2kLR-GulfofGuinea.Weldeab.2007	2.5	9.4	9	marine sediment	foram Mg/Ca	PI/hist	loose
530	O2kLR-JacafFjord.Sepulveda.2009	−44.3	−73	−73	marine sediment	alkenone	PI/hist	loose
532	O2kLR-KuroshioCurrent.Isono.2009	36	141.8	142	marine sediment	alkenone	PI/hist	loose
535	O2kLR-MakassarStrait-MD98-2177.Newton.2011	1.4	119.1	119	marine sediment	foram Mg/Ca	PI/hist	strong/loose
536	O2kLR-MakassarStrait.Linsley.2010	−7.4	115.2	115	marine sediment	foram Mg/Ca	PI/hist	loose
538	O2kLR-MakassarStrait.Oppo.2009	−3.5	119.2	119	marine sediment	foram Mg/Ca	PI/hist	loose
539	O2kLR-MD952011.Calvo.2002	67	7.6	8	marine sediment	alkenone	PI/hist	loose
540	O2kLR-Minorca.Moreno.2012	40.5	4	4	marine sediment	alkenone	PI/hist	loose
541*	O2kLR-NorthIceland.Sicre.2011	66.5	−17.4	−17	marine sediment	alkenone	PI/hist	strong/loose
542	O2kLR-NWPacific.Harada.2004	46.3	152.5	152	marine sediment	alkenone	PI/hist	loose
544	O2kLR-OkinawaTrough.Wu.2012	24.8	122.5	122	marine sediment	TEX86	PI/hist	loose
546	O2kLR-Philippines-MD98-2181.Stott.2007	6.3	125.8	126	marine sediment	foram Mg/Ca	PI/hist	loose
548	O2kLR-PigmyBasin.Richey.2009	27.2	−91.4	−91	marine sediment	foram Mg/Ca	PI/hist	strong/loose
550*	O2kLR-SantaBarbara.Zhao.2000	34.2	−120	−120	marine sediment	alkenone	PI/hist	strong/loose
551	O2kLR-SouthAtlantic.Leduc.2010	−29.1	16.7	17	marine sediment	alkenone	PI/hist	loose
552	O2kLR-SouthChinaSea.Zhao.2006	8.7	109.9	110	marine sediment	alkenone	PI/hist	loose
553	O2kLR-SouthernChile.Mohtadi.2007	−44.1	−75.2	−75	marine sediment	alkenone	PI/hist	loose
554*	O2kLR-SouthIceland.Sicre.2011	57.5	−27.9	−28	marine sediment	alkenone	PI/hist	strong/loose
555	O2kLR-SubTropicalEasternAtlantic.deMenocal.2000	20.8	−18.6	−19	marine sediment	planktonic foraminifera	PI/hist	loose
561	O2kLR-WesternSvalbard.Spielhagen.2011	78.9	6.8	7	marine sediment	planktonic foraminifera	PI/hist	loose
562	O2kLR-WestSpitzbergen.Bonnet.2010	79	5.9	6	marine sediment	dynocist MAT	PI/hist	loose
563	O2kLRMakassarStrait-MD98-2160.Newton.2011	−5.2	117.5	118	marine sediment	foram Mg/Ca	PI/hist	loose
654*	SAm-LagunaAculeo.vonGunten.2009	−33.9	−70.9	−71	lake sediment	reflectance	PI/hist	strong/loose
655*	SAm-LagunaChepical.deJong.2013	−32.3	−70.5	−70	lake sediment	reflectance	PI/hist	strong/loose
656*	SAm-LagunaEscondida.Elbert.2013	−45.5	−71.8	−72	lake sediment	BSi	PI/hist	strong/loose

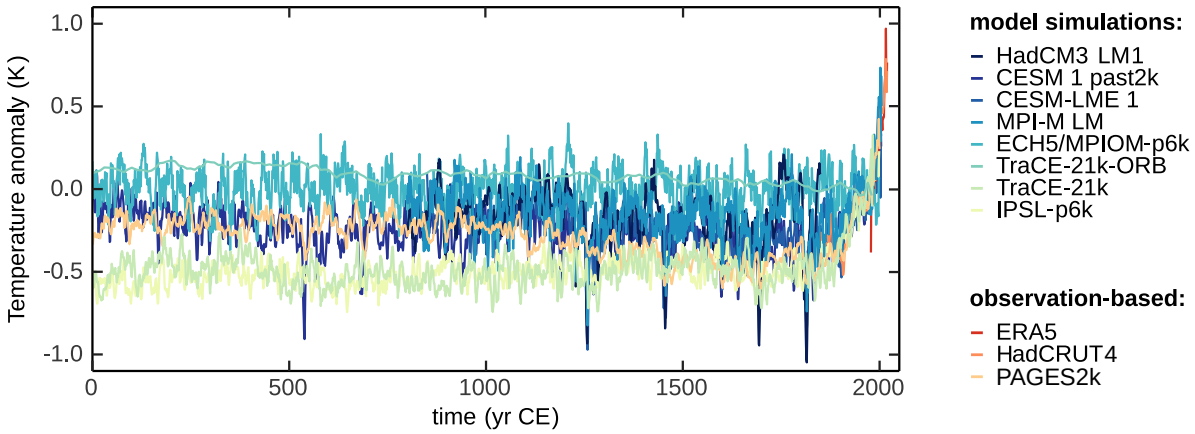


FIG. S1. Evolution of global mean temperature over the Common Era from model simulations and observation-based data as in Table S1. Anomalies are given with respect to the reference period 1961-1990 (HadCM3 LM1:1800-1850) and as a running average of five years.

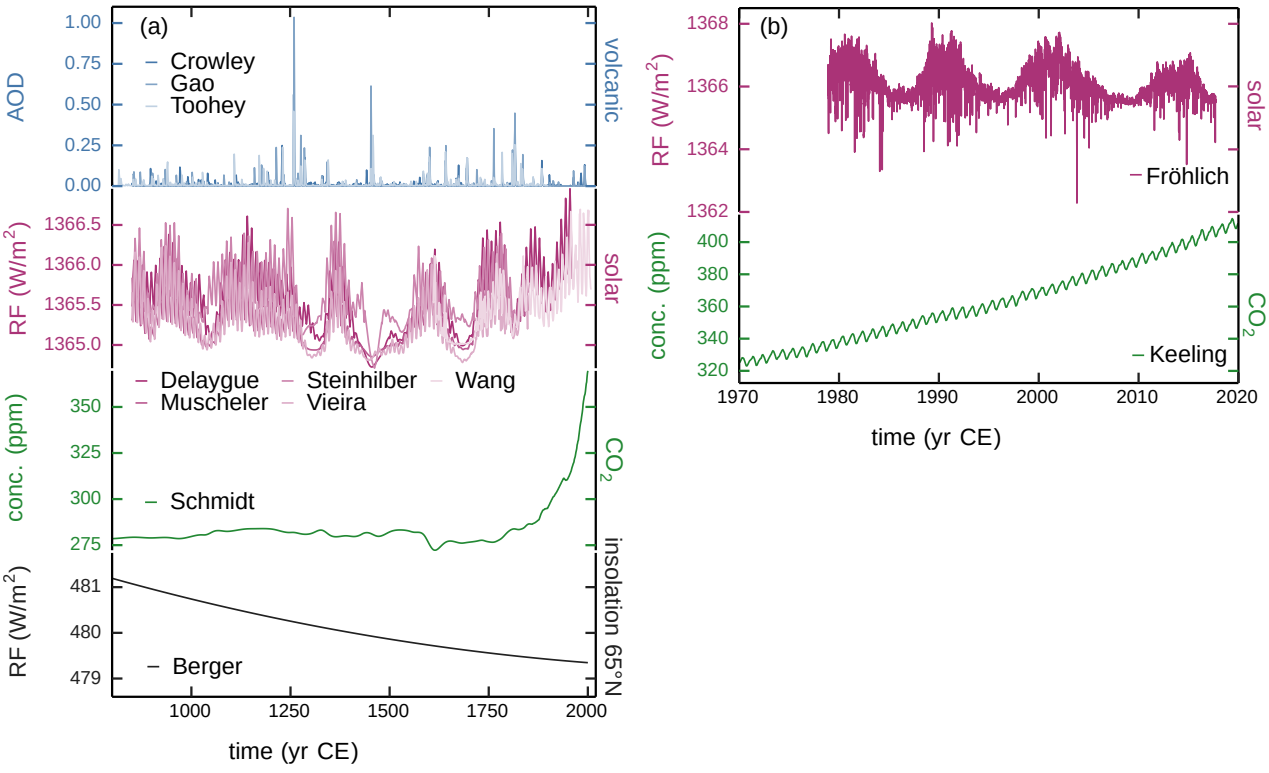


FIG. S2. (a) Reconstruction of climate drivers over the Common Era used to estimate the PSD of radiative forcing (RF). Labels indicate the data reference as given in Table S2. (b) Additional observational data for solar and CO_2 forcing used to obtain high-frequency spectral estimates. Highly resolved insolation changes due to the diurnal and annual cycle are not shown here.

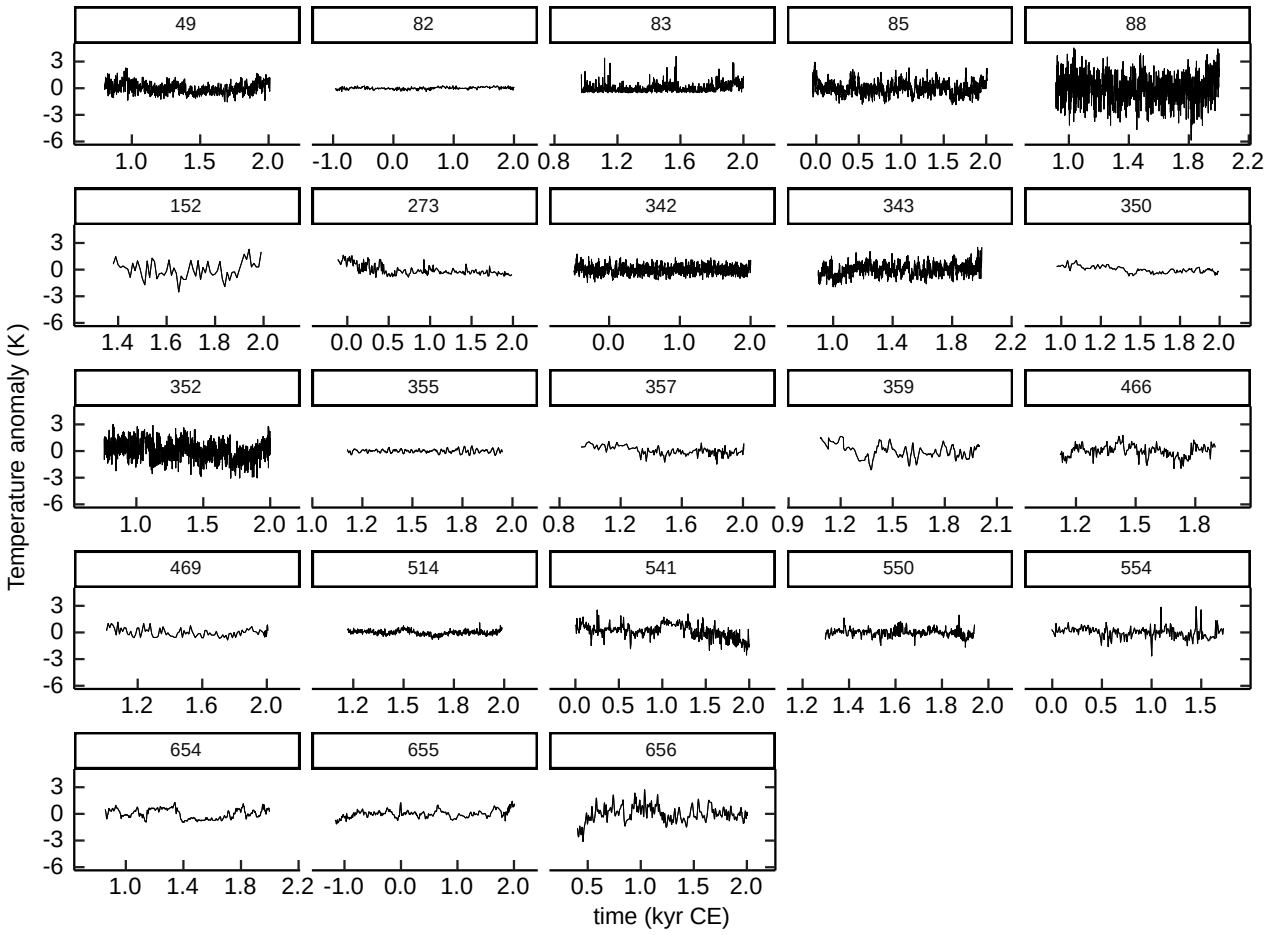


FIG. S3. Temperature anomalies from proxy records used to estimate local temperature variability. The labels give the ID from the PAGES2k database as indicated by a star ($\star$) in Table S3.

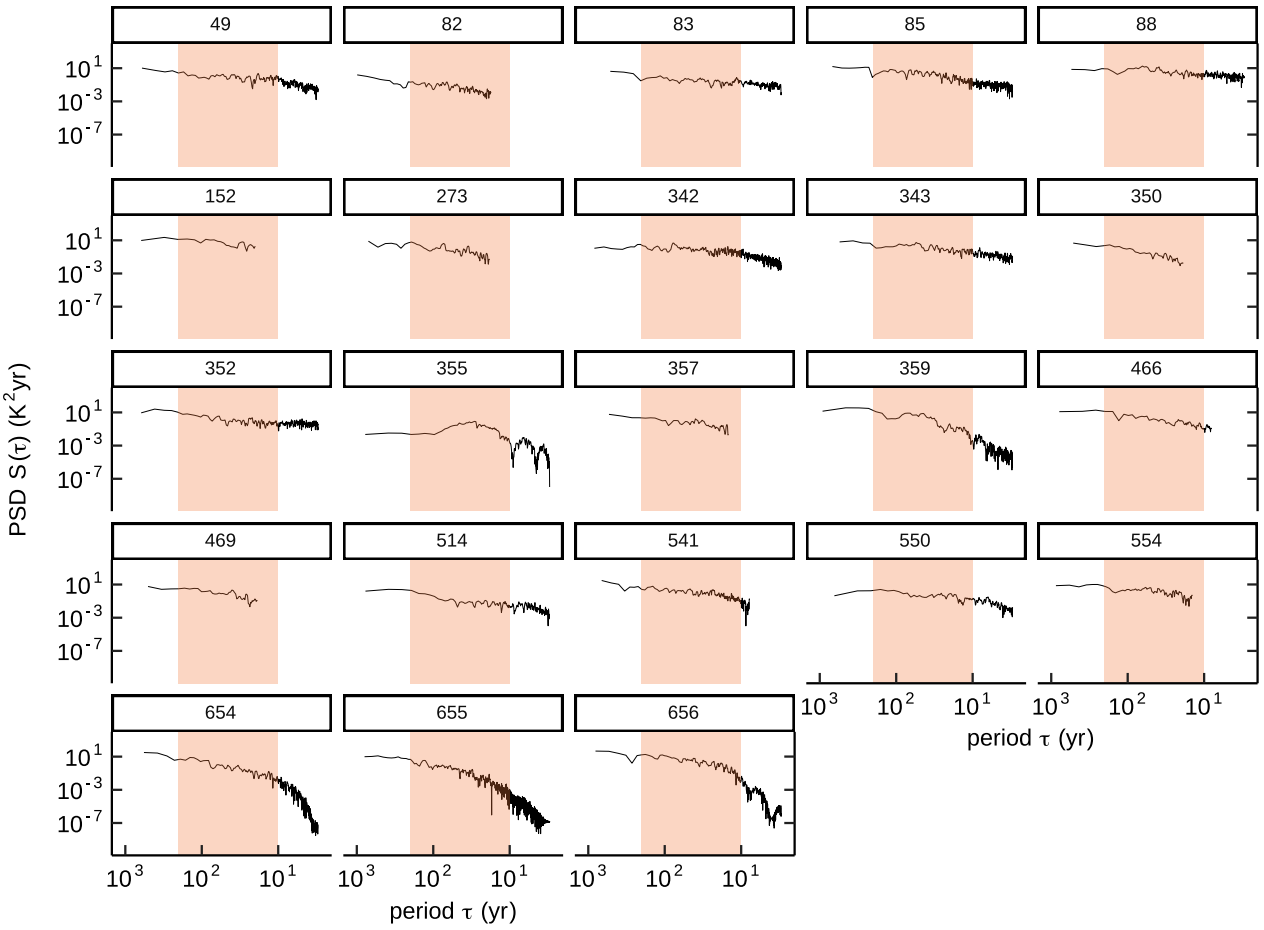


FIG. S4. As Figure S3, but showing the PSD of the proxy records. The shaded area indicates the periods between 10 and 200 years, used to estimate the scaling relationship. Scaling exponents for spectra that do not cover the full period were estimated on their corresponding timescales, but at least between 20 and 200 years.

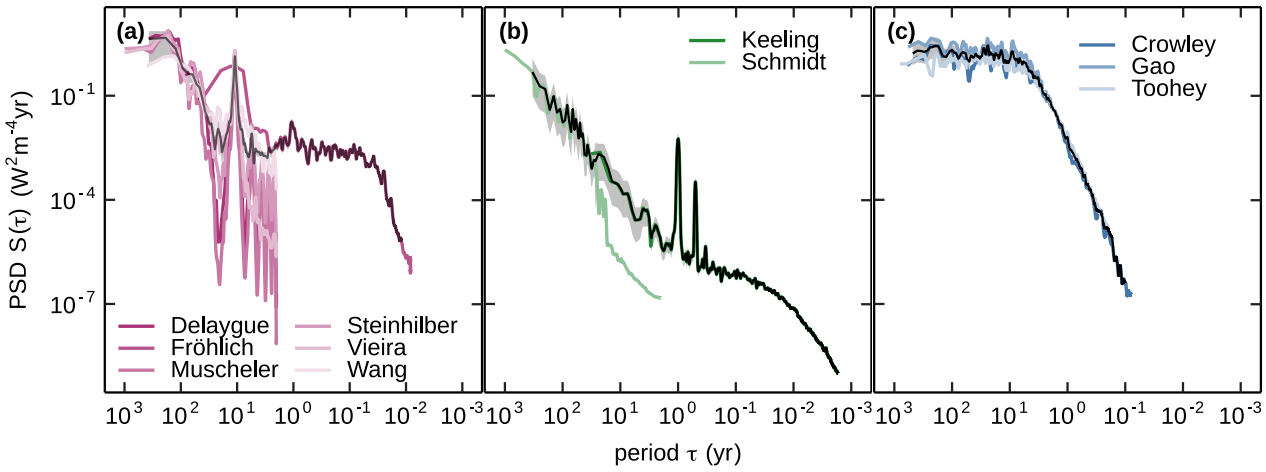


FIG. S5. Power spectral densities of solar (a), CO_2 (b), and volcanic (c) from multiple data sets and their mean spectra (black). Low- and high-frequency tails, likely prone to spectral biases, were omitted in the average.

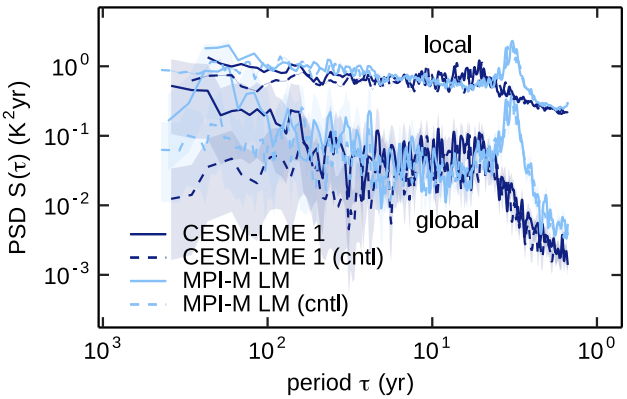


FIG. S6. PSD for global mean and mean of local spectra of surface air temperature from last millennium runs and their PI control (cntl).

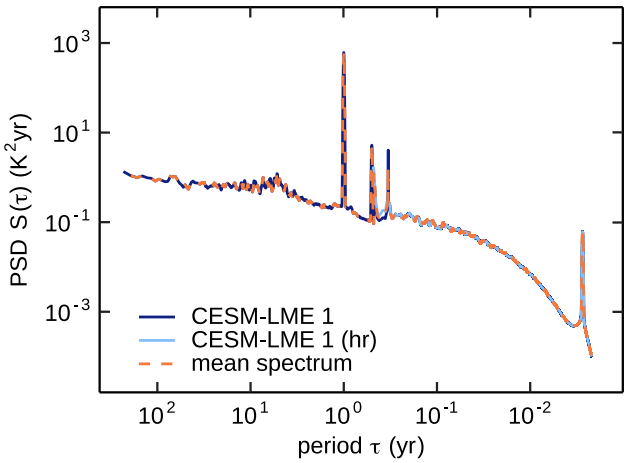


FIG. S7. Joint PSD (dashed orange line) from the mean spectrum of a highly (hr) and lower resolved time series on the example of surface air temperature from the CESM-LME 1 run. To improve computational efficiency, we formed this joint PSD for ERA5, MPI-M LM, and CESM-LME 1. The high-frequency component was estimated from data at hourly resolved temperature series from 1981 to 1990 CE since this is a part of the climatic period that is shared among the three data sets.

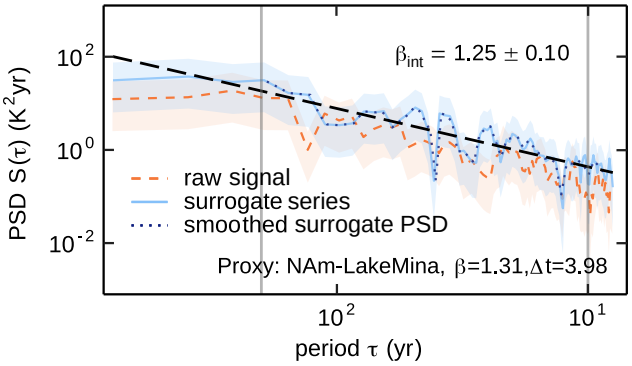


FIG. S8. Step-wise estimation of uncertainty on β-scaling relation for irregularly sampled proxy records on the example of the PSD from the "NAm-LakeMina"-record. When interpolated to its mean temporal resolution Δt, the raw signal has a scaling relation $\beta = 1.31$ (orange dashed line). The uncertainties were calculated from 100 surrogates that are random time series with approximately the same power-law scaling β (black dotted line). The PSD is altered after forming the block-average of the surrogate time series and interpolation to Δt (blue solid line). Linear regression on this spectrum (black dashed line) gives the scaling coefficient β_{int} of the surrogate series.

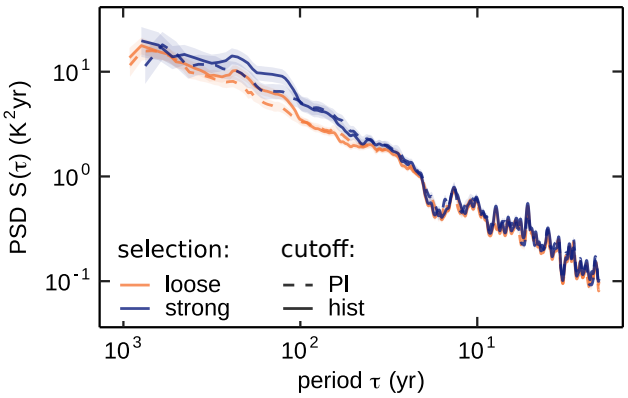


FIG. S9. Global mean of local PSD from proxy records subject to sampling from the PAGES2k data base and global warming. Dashed and solid lines refer to the mean spectra of local proxy records cut at 1850 CE (PI) and 2000 CE (hist) respectively. The plot compares strong (blue) and loose (orange) criteria. Loose criteria are described in the data section of the main manuscript and were used to calculate the mean of local spectra from reconstructions. For the stronger criteria we require the mean temporal resolution ($\langle t_{i+1} - t_i \rangle$) $\leq$ 20years, a coverage ($t_N - t_1$) $\geq$ 30years and the number of data points $N \geq 30$. Hiatus are tolerated up to max($t_{i+1} - t_i$) $\geq$ 40 years.

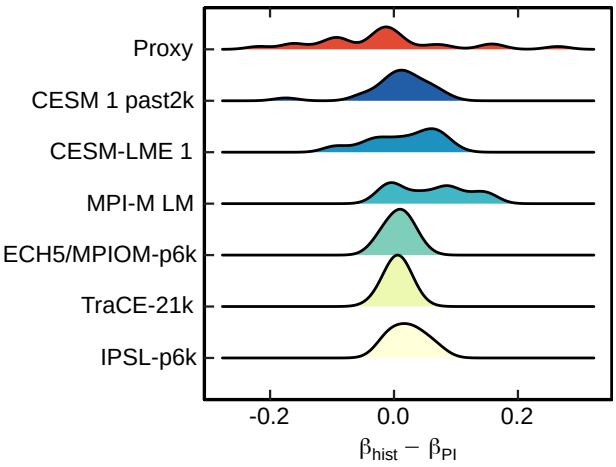


FIG. S10. Normalized density plots of the deviations of the spectral exponent β extracted from time series up to 2020 CE (β_{hist}) and time series up to 1850 CE (β_{PI}). The densities were computed from the local power-law scaling used to estimate the statistical agreement between models and data (Table S3).

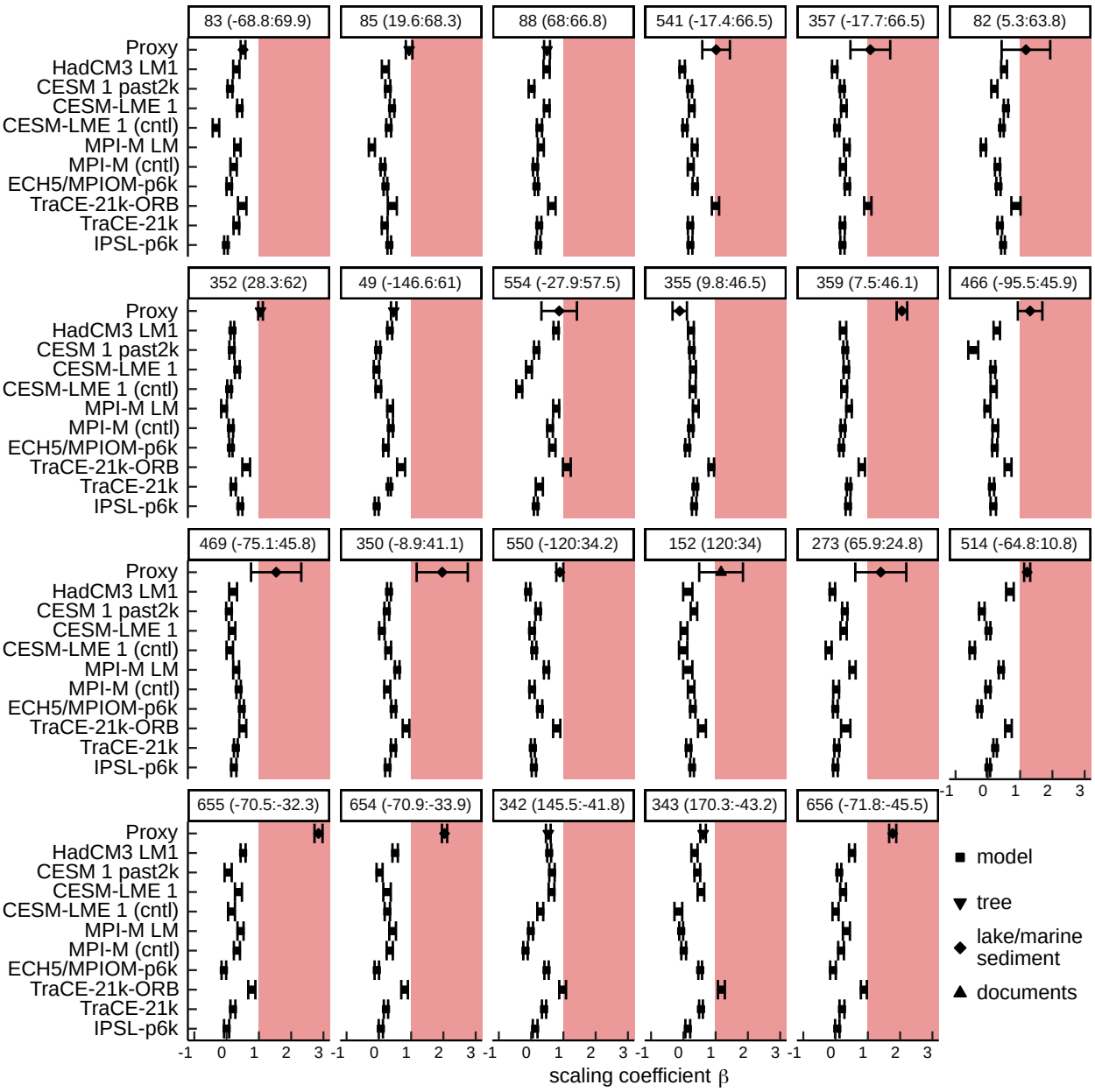


FIG. S11. Local scaling coefficients and its confidence intervals from simulations and paleoclimate data. Each panel denotes the ID within the PAGES2k database and its coordinates (longitude °E, latitude °N). Symbols indicate the archive of the proxy record or whether the scaling exponent belongs to a model simulation or not. Red shading refers to scaling exponents $\beta > 1$, whereas the white background indicates $\beta < 1$.

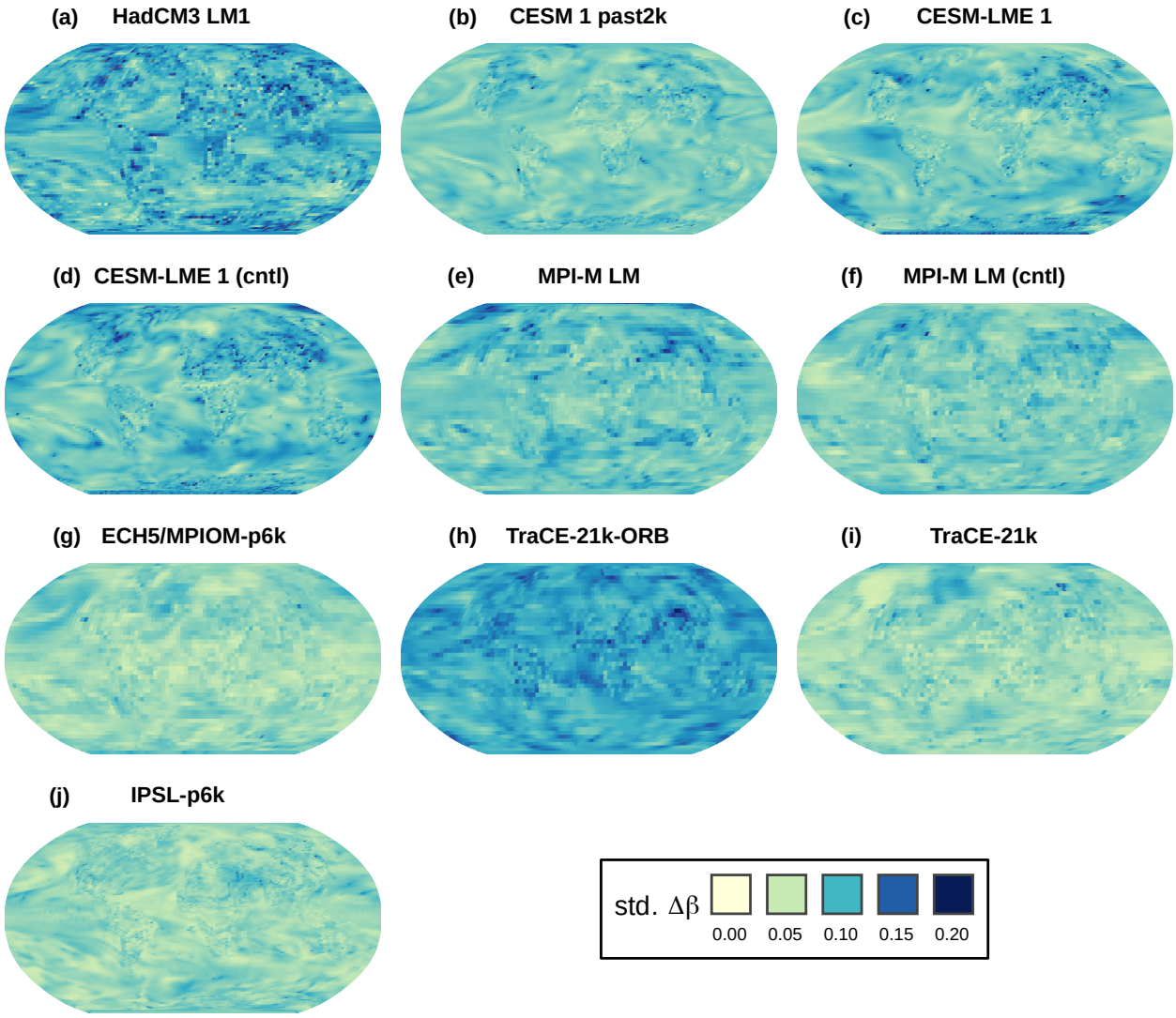


FIG. S12. Standard deviation of the least-square regression of local scaling coefficient β for all model simulations. The goodness of fit is generally lower for models with a comparatively low temporal resolution (Trace21k-ORB) or coverage (HadCM3 LM1). Furthermore, there is a slightly increased deviation in areas of active modes of variability, such as the ENSO region ((b), (c), (e), (j)). This could be due to the fact, that spectral peaks describe these (quasi-)periodic signals better than power-law scaling.

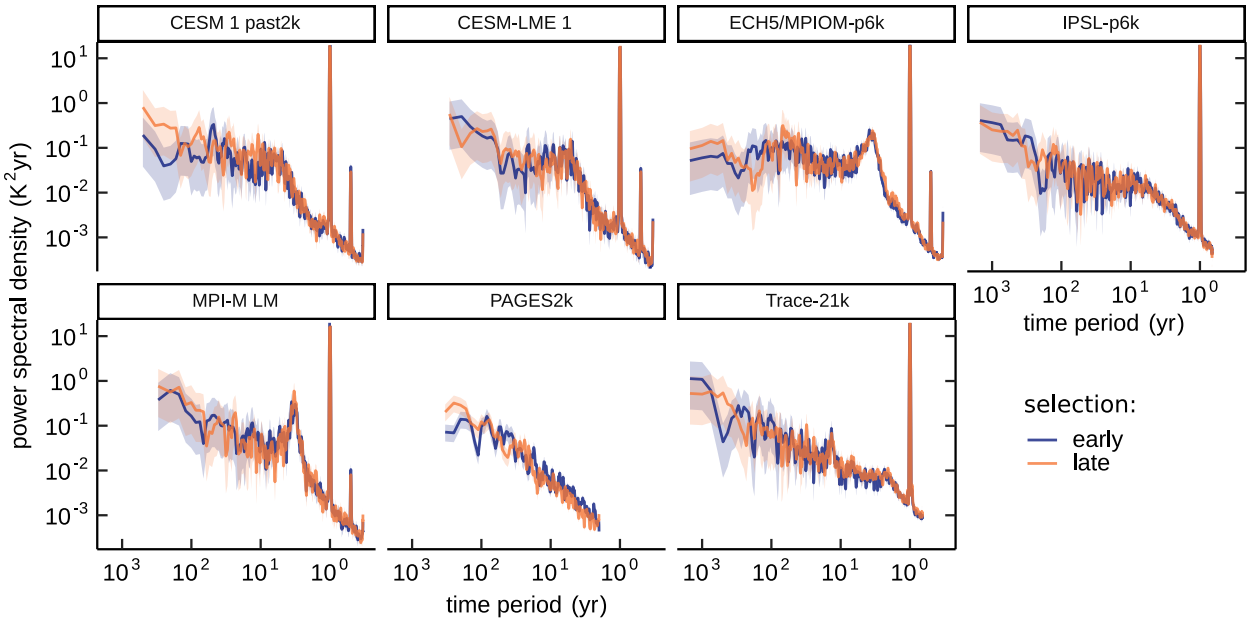


FIG. S13. PSD of global mean surface temperature from model simulations and PAGES2k reconstruction. To test the robustness of our spectral analysis against potentially non-trending, non-stationary signals in the data sets, each temperature signal was split into two disjoint time intervals of equal length, representing the "early" and the "late" part of the signal. The latter contains the anthropogenic global warming period, starting approximately 1850 CE. Trace-21k-ORB (orbitally forced only) as well as HadCM3 LM1 (800-1850 CE) were excluded from the discussion since they do not represent the recent global warming.

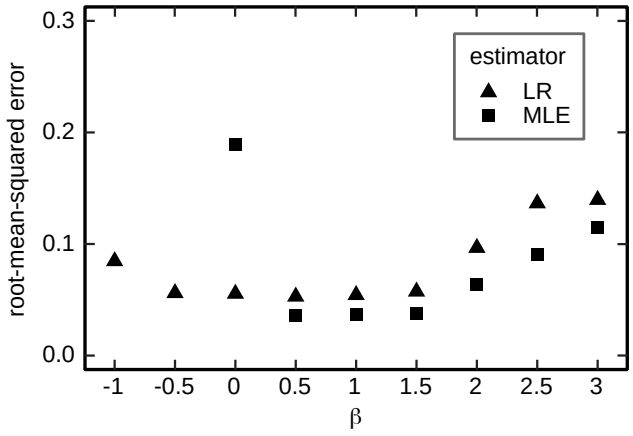


FIG. S14. Root-mean-squared error of linear regression (LR) and maximum likelihood estimation (MLE) [1, 2] computed from 200 surrogate time series with 6000 data points and power-law scaling β. The grey dashed line marks the mean confidence $\Delta\beta_{\mathrm{LR}}$ of the linear regression. Outliers (> 0.3) of the MLE for negative scaling exponents are not shown here. The implemented (standard) MLE is generally not appropriate to estimate $\beta \leq 0$, which explains deviations in this range [2].

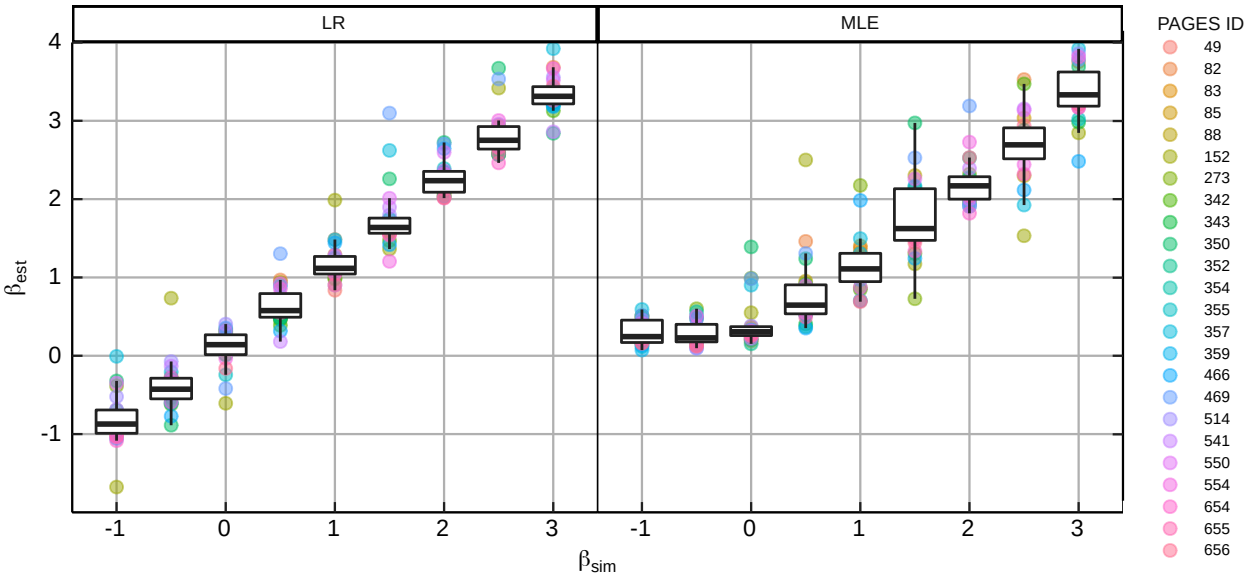


FIG. S15. Distribution of estimated beta (β_{est}) from irregularly sampled pseudo proxy records using LR and MLE. For each possible β_{sim}, a set of 200 surrogate time series was block-averaged to the resolution of each proxy record before computing its power spectral density. The horizontal thick lines within the boxes correspond to the median. The ends of the box denote the upper (75%) and lower (25%) quartiles. The vertical line extends from the upper ending of the box to the largest value no further than 1.5 times the interquartile range, and vice versa for the lower bound. Points further outside are potential outliers, of which the IDs 152, 273, 350, and 469 can be explained by their comparatively low mean resolution of 9-10 years.

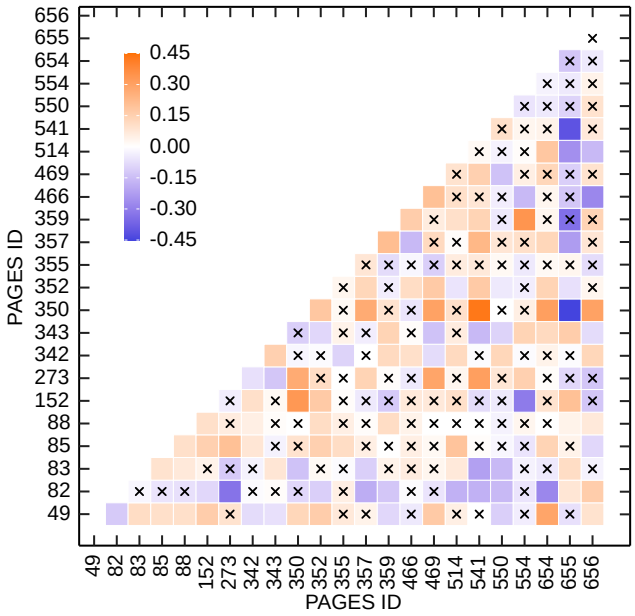


FIG. S16. Cross-correlations among the 23 proxy records, denoted by the ID of the PAGES2k database [3]. Insignificant correlations are marked by a cross. The cross-correlations for the irregularly sampled data were computed using the *nest*-package [4, 5] in R.

[1] C. Gillespie, poweRlaw - Package (R) (2020).
[2] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, SIAM Review **51**, 661–703 (2009).
[3] PAGES 2k Consortium., Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nature Geoscience **12**, 643 (2019).
[4] K. Rehfeld, N. Marwan, J. Heitzig, and J. Kurths, Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics 10.5194/npg-18-389-2011 (2011).
[5] K. Rehfeld and J. Kurths, Similarity estimators for irregular and age-uncertain time series, Climate of the Past , 107 (2014).
[6] Y. Zhong, A. Jahn, G. H. Miller, and A. Geirsdottir, Asymmetric Cooling of the Atlantic and Pacific Arctic During the Past Two Millennia: A Dual Observation-Modeling Study, Geophysical Research Letters **45**, 12,412 (2018).
[7] J. H. Jungclaus et al., The PMIP4 contribution to CMIP6 – Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 simulations, Geoscientific Model Development **10**, 4005 (2017).
[8] B. L. Otto-Bliesner, E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai, and G. Strand, Climate variability and change since 850 ce an ensemble approach with the community earth system model, Bulletin of the American Meteorological Society 10.1175/BAMS-D-14-00233.1 (2016).
[9] G. A. Schmidt et al., Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geoscientific Model Development **4**, 33 (2011).
[10] J. H. Jungclaus et al., Climate and carbon-cycle variability over the last millennium, Climate of the Past 10.5194/cp-6-723-2010 (2010).
[11] J. C. Bühler, C. Roesch, M. Kirschner, L. Sime, M. D. Holloway, and K. Rehfeld, Comparison of the oxygen isotope signatures in speleothem records and iHadCM3 model simulations for the last millennium, Climate of the Past Discussions **2020**, 1 (2020).
[12] A. P. Schurer, S. F. B. Tett, and G. C. Hegerl, Small influence of solar variability on climate over the past millennium, Nature Geoscience **7**, 104 (2014).
[13] P. Braconnot, D. Zhu, O. Marti, and J. Servonnat, Strengths and challenges for transient Mid- to Late Holocene simulations with dynamical vegetation, Climate of the Past **15**, 997 (2019).
[14] Z. Liu, Transient simulation of last deglaciation with a new mechanism for Bølling–Allerød warming, Science **325**, 10.1126/science.1171041 (2009).
[15] A. L. Berger, Long-term variations of daily insolation and Quaternary climatic changes., Journal of Atmospheric Sciences 10.1175/1520-0469(1978)035¡2362:ltvodi¿2.0.co;2 (1978).
[16] F. Joos and R. Spahni, Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years, Proceedings of the National Academy of Sciences **105**, 1425 LP (2008).
[17] N. Fischer and J. H. Jungclaus, Evolution of the seasonal temperature cycle in a transient Holocene simulation: Orbital forcing and sea-ice, Climate of the Past 10.5194/cp-7-1139-2011 (2011).
[18] H. Hersbach, B. Bell, P. Berrisford, and S. Hirahara, The ERA5 global reanalysis, Quarterly

Journal of the Royal Meteorological Society , 1999 (2020).
[19] C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones, Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research Atmospheres **117**, 10.1029/2011JD017187 (2012).
[20] G. A. Schmidt et al., Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), Geoscientific Model Development **5**, 185 (2012).
[21] T. J. Crowley and M. B. Unterman, Technical details concerning development of a 1200 yr proxy index for global volcanism, Earth System Science Data **5**, 187 (2013).
[22] C. Gao, A. Robock, and C. Ammann, Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models, Journal of Geophysical Research: Atmospheres **113**, 10.1029/2008JD010239 (2008).
[23] M. Toohey and M. Sigl, Reconstructed volcanic stratospheric sulfur injections and aerosol optical depth, 500 BCE to 1900 CE, version 2, World Data Center for Climate (WDCC) at DKRZ 10.1594/WDCC/eVolv2k_v2 (2017).
[24] G. Delaygue and E. Bard, An Antarctic view of Beryllium-10 and solar activity for the past millennium, Climate Dynamics **36**, 2201 (2011).
[25] R. Muscheler, F. Joos, J. Beer, S. A. Müller, M. Vonmoos, and I. Snowball, Solar activity during the last 1000yr inferred from radionuclide records, Quaternary Science Reviews **26**, 82 (2007).
[26] F. Steinhilber, J. Beer, and C. Fröhlich, Total solar irradiance during the Holocene, Geophysical Research Letters **36**, 10.1029/2009GL040142 (2009).
[27] L. E. Vieira and S. K. Solanki, Evolution of the solar magnetic flux on time scales of years to millenia, Astronomy and Astrophysics 10.1051/0004-6361/200913276 (2010).
[28] N. A. Krivova, L. Balmaceda, and S. K. Solanki, Reconstruction of solar total irradiance since 1700 from the surface magnetic flux, Astronomy and Astrophysics 10.1051/0004-6361:20066725 (2007).
[29] Y. Wang, J. L. Lean, and N. R. Sheeley, Jr., Modeling the Sun's Magnetic Field and Irradiance since 1713, The Astrophysical Journal 10.1086/429689 (2005).
[30] C. Fröhlich, Solar irradiance variability since 1978: Revision of the PMOD composite during solar cycle 21, Space Science Reviews **125**, 53 (2006).
[31] C. D. Keeling, R. B. Bacastow, and A. E. Bainbridge, Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii, TELLUS 10.3402/tellusa.v28i6.11322 (1976).
[32] M. Crucifix, Palinsol - Package (R) (2016).

G Publication 2

B. Ellerhoff, M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime, and K. Rehfeld. "Contrasting State-Dependent Effects of Natural Forcing on Global and Local Climate Variability". *Geophysical Research Letters* 49.10 (2022). doi: 10.1029/2022GL098335.

Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL098335

Key Points:

- We present Glacial/Interglacial climate simulations and quantify effects of time-varying volcanic and solar forcing on climate variability
- The mean global and local response to these forcings is similar in Glacial and Interglacial climate, suggesting low state dependency
- In both climate states, modeled temperature variance agrees better with palaeoclimate data when volcanic and solar forcing is included

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Rehfeld,
kira.rehfeld@uni-tuebingen.de

Citation:

Ellerhoff, B., Kirschner, M. J., Ziegler, E., Holloway, M. D., Sime, L., & Rehfeld, K. (2022). Contrasting state-dependent effects of natural forcing on global and local climate variability. *Geophysical Research Letters*, *49*, e2022GL098335. https://doi.org/10.1029/2022GL098335

Received 16 FEB 2022
Accepted 19 APR 2022

Contrasting State-Dependent Effects of Natural Forcing on Global and Local Climate Variability

Beatrice Ellerhoff[1,2], Moritz J. Kirschner[2], Elisa Ziegler[1,2], Max D. Holloway[3], Louise Sime[4], and Kira Rehfeld[1,2]

[1]Department of Geosciences and Department of Physics, Tübingen University, Tübingen, Germany, [2]Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany, [3]Scottish Association for Marine Science, Oban, UK, [4]British Antarctic Survey, Cambridge, UK

Abstract Natural forcing from solar and volcanic activity contributes significantly to climate variability. The post-eruption cooling of strong volcanic eruptions was hypothesized to have led to millennial-scale variability during Glacials. Cooling induced by volcanic eruption is potentially weaker in the warmer climate. The underlying question is whether the climatic response to natural forcing is state-dependent. Here, we quantify the response to natural forcing under Last Glacial and Pre-Industrial conditions in an ensemble of climate model simulations. We evaluate internal and forced variability on annual to multicentennial scales. The global temperature response reveals no state dependency. Small local differences result mainly from state-dependent sea ice changes. Variability in forced simulations matches paleoclimate reconstructions significantly better than in unforced scenarios. Considering natural forcing is therefore important for model-data comparison and future projections.

Plain Language Summary Climate variability describes the spatial and temporal variations of Earth's climate. Understanding these variations is important for estimating the occurrence of extreme climate events such as droughts. Yet, it is unclear whether climate variability depends on the mean surface temperature of the Earth or not. Here, we investigate the effects of natural forcing from volcanic eruptions and solar activity changes on climate variability. We compare simulations of a past (cold) and present (warm) climate with and without volcanism and solar changes. We find that overall, the climate system responds similarly to natural forcing in the cold and warm state. Small local differences mainly occur where ice can form. To evaluate the simulated variability, we use data from paleoclimate archives, including trees, ice-cores, and marine sediments. Climate variability from forced simulations agrees better with the temperature variability obtained from data. Natural forcing is therefore critical for reliable simulation of variability in past and future climates.

1. Introduction

Climate variability, that is variations in the statistics of climate parameters, characterizes Earth's dynamical system and is the primary influence on extreme events (Katz & Brown, 1992). Variability arises from unforced processes, internal to the climate system, and from forced processes, caused by external natural and anthropogenic drivers. Natural drivers include volcanic and solar forcing, contributing significantly to climate variability (Crowley & Unterman, 2013b). Due to anthropogenic activities, the recent trend of global mean surface temperature (GMST) and other variables has clearly emerged beyond the range of natural variability (Bindoff et al., 2013; Hasselmann, 1997; Marcott et al., 2013; Sippel et al., 2020).

Global warming also affects climate variability (Bathiany et al., 2018; Olonscheck et al., 2021). The underlying mechanisms remain poorly understood. There is conflicting and incomplete evidence on the spatio-temporal patterns of change (Brown et al., 2017; Holmes et al., 2016; Huntingford et al., 2013; Pendergrass et al., 2017; Rehfeld et al., 2020). This is a major source of uncertainty for regional climate projections. To accurately simulate climate variability models must resolve internal variability, its response to natural forcing across scales, and the mean climate states (Rehfeld et al., 2018).

Large explosive volcanic eruptions are suggested to have driven millennial-scale climate variations during glacial periods (Baldini et al., 2015). The largest eruption was hypothesized to have caused a human population bottleneck (Ambrose, 1998). The extent and impact of this event remains unclear (Svensson et al., 2013; Timmreck et al., 2010). Strong tropical volcanic eruptions have also been shown to influence daily temperature

and precipitation extremes (Wang et al., 2021). These eruptions could induce a somewhat weaker response in warmer climates (Hopcroft et al., 2018), but volcanism will continue to play an important role in future variability (Bethke et al., 2017). These studies, however, do not examine the dependency of forced variability on the mean climate because they rely on future projections or the responses to large eruptions.

The paleoclimate record is crucial to assess whether a colder planet is more sensitive to natural forcing than a warmer one. Yet, temperature variability shows a mismatch between paleoclimate simulations and proxy data on the decadal-to-multicentennial scale (Ellerhoff & Rehfeld, 2021; Laepple & Huybers, 2014a). Despite major efforts, such as the Paleoclimate Modeling Intercomparison Project (PMIP; Braconnot et al., 2012; Kageyama et al., 2018), this discrepancy remains unresolved. While PMIP experiments successfully demonstrated the influence of natural and anthropogenic forcing on temperature variability over the last millennium (Otto-Bliesner et al., 2016), similar studies for Glacial states are missing. Transient paleoclimate simulations for the Last Glacial Maximum (LGM) have mostly been performed without high-frequency solar and volcanic forcing (Liu et al., 2009; Smith & Gregory, 2012). This lack of time-dependent forcing could potentially explain discrepancies between reconstructed and simulated variability. Additional uncertainty remains about the mechanisms of local, long-term variability (Franzke et al., 2020; Fredriksen & Rypdal, 2017; Huybers & Curry, 2006).

Separating internal and external variability has improved the understanding of climate dynamics and processes (Frankcombe et al., 2015; Haustein et al., 2019; Mann et al., 2022; Schurer et al., 2013). The approach should allow to identify drivers of local, decadal-to-multicentennial variability in cold and warm climates. This requires the comparison of unforced and forced climate simulations under Glacial and Interglacial conditions, and their validation against paleoclimate data over a wide range of timescales. Studying contributions to surface climate variability of system components that bridge internal and external factors is also necessary. Sea ice, for example, follows in extent the mean state. Natural forcing could, however, also drive multidecadal variability via modulation of the Atlantic Meridional Overturning Circulation (AMOC; Halloran et al., 2020). This highlights contributions to variability from climate components and mechanisms that bridge intrinsic and external factors.

Here, we contrast unforced and naturally forced simulations under LGM and PI conditions in an ensemble using the Hadley Centre Coupled Model Version 3.4 (HadCM3 (Gordon et al., 2000; Pope et al., 2000; Reichler & Kim, 2008; Stott et al., 2000);). We examine the mean local response of the surface climate to volcanism in the two climate states (Section 3.1). Spectral analysis (Section 3.2) further quantifies the state- and timescale-dependent effects of natural forcing on local, zonal, and global scales. It confirms a robust response to natural forcing across climate states, but a mean decline in local temperature variability with warming. To aid interpretation of the spectra, we investigate sea ice dynamics as it appears a main contributor to local, long-term variability. We validate simulated variances using proxy data (Section 3.3) to confirm that the addition of natural forcing significantly reduces the model-data mismatch on multidecadal and longer timescales. Thus, the inclusion of natural forcing provides a more accurate representation of climate variability, needed for climate simulations.

2. Data and Methods

2.1. Model Setup

Our ensemble consists of 12 simulations using LGM or PI boundary conditions (Table S1, Figure S1 in Supporting Information S1). We performed them using HadCM3, a three-dimensional, coupled atmosphere-ocean general circulation model (AOGCM) that is widely used for paleoclimate study (Armstrong et al., 2022; Bühler et al., 2021; Collins et al., 2001; Flato et al., 2014; Reichler & Kim, 2008; Tindall et al., 2009; Valdes et al., 2017). Computational efficiency allows for long-term integrations, with a simulated climate comparable to other AOGCMs and observations (Gordon et al., 2000; Jackson & Vellinga, 2013).

The simulations are monthly resolved and of millennial length. The boundary conditions (orography, orbital parameters, greenhouse gas concentrations) defining the mean state are constant over these runs. All runs start from the same LGM/PI spin-up simulation at consecutive years. Six runs are akin to control simulations without transient forcing. The remaining runs feature time-varying solar and volcanic forcing. We mark these with a star (*). Thus, three runs exist for each mean state and each forcing scenario (Table S2 in Supporting Information S1). Unless otherwise specified, our results represent average values of these sub-ensembles. Temperature, precipitation, sea level pressure, and wind fields are shown in Figure S2 in Supporting Information S1. The Last Glacial GMST is decidedly colder (9.5 ± 1.4)°C and the global mean precipitation rate (GMPR) is

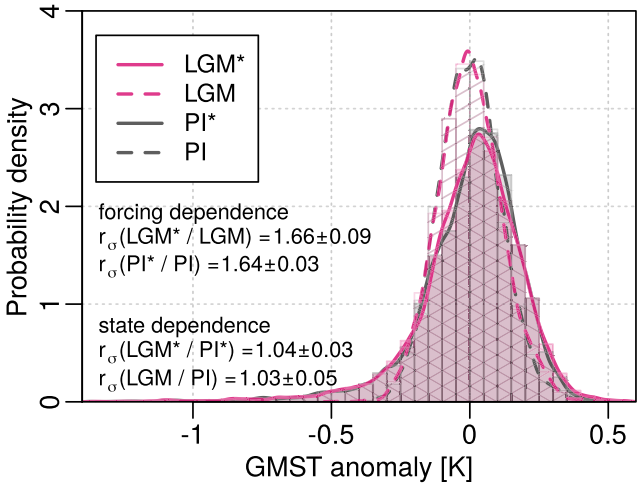


Figure 1. Probability density (unitless) of simulated yearly global mean surface temperature anomalies from all Pre-Industrial and Last Glacial Maximum runs. Forced scenarios are marked with a (*). The ratio of the distributions' standard deviations is given by r_σ.

lower (935 ± 20) mm yr^{-1}, with a steeper equator-to-pole temperature gradient than the Pre-Industrial with (15.1 ± 1.3) °C and (1,048 ± 21) mm yr^{-1}, respectively.

We use transient volcanic and solar forcing (Figure S1 in Supporting Information S1), following the PMIP3 protocol for the last millennium (850–1850 CE; Schmidt et al., 2012) and updated every 10 days. We apply the same forcing in both states, as no reconstructions of solar and volcanic forcing for the LGM exist to date. This also makes comparing forced variability between the states easier. Total solar irradiance (from Steinhilber et al. (2009) and Wang et al. (2005)) is time dependent, with a superposed 11-year cycle (Schmidt et al., 2012). Volcanic forcing (Crowley & Unterman, 2013) is supplied as Aerosol Optical Depth (AOD) at four equal-area latitude bands (90-30°S, 30°S-0, 0–30°N, 30–90°N). It describes the attenuation of incoming radiation by volcanic aerosols at a wavelength of 0.55 µm and is converted into an aerosol mass loading factor (Schmidt et al., 2012).

Figure 1 shows the distribution of simulated GMST anomalies for the Last Glacial and Pre-Industrial. Forced scenarios (LGM*, PI*) exhibit larger fluctuations. In both states the GMST standard deviation increases by a factor of approximately 1.65 compared to unforced runs. There is no strong difference in the GMST distribution attributable to the mean climate.

2.2. Observations and Paleoclimate Reconstructions

We use observations and paleoclimate reconstructions to validate the variance from model simulation on interannual to multicentennial scales (2–5, 5–50, 50–200, and 200–500 years). We consider proxy records from Rehfeld et al. (2018) and the PAGES2k-Consortium (2017), and observations from the Met Office Hadley Centre's sea surface temperature data set (HadISST downloaded 11/2019; Rayner et al. (2003)). We focus on sea surface temperature observations as much of our proxy data is of marine origin. We select records that (a) are published and calibrated to temperature, (b) contain more than 50 data points, (c) cover at least three times the largest period of interest, and (d) have a mean sampling frequency of twice the highest frequency considered (Ellerhoff & Rehfeld, 2021). We exclude records with gaps larger than five times the required resolution. Ice core records are not considered on timescales below 50 years, where signal-to-noise ratios are low (Casado et al., 2020; Laepple et al., 2018). Our ensemble consists of 41 observations and 115 proxy records from six archive types (Data Set S1, Figures S9 and S10 in Supporting Information S1).

2.3. Effect Analysis

We analyze the global and local state-dependent effects of natural forcing in time and spectral domain. Following Swingedouw et al. (2017), we quantify local effects of moderate to large-magnitude volcanic eruptions using the mean standardized anomaly (MSA). The MSA represents the average value of the standardized anomalies across ensemble members. It is computed for 12-month means surrounding periods with high aerosol imprint (AOD > 0.13, corresponding to approx. −2.6 W/m^2 (Forster et al., 2021)) as follows

$$\mathrm{MSA} = \frac{1}{j}\sum_{j}\frac{\frac{1}{12}\sum_{i\in T_j} X_i - \mu}{\sigma}, \tag{1}$$

with mean $\mu = \mathrm{E}[X]$ and standard deviation $\sigma = \sqrt{\mathrm{E}\left[(X-\mu)^2\right]}$ of each individual gridbox time series X. The index i specifies the 12 months of the time series X corresponding to the set of periods T_j for run j of each climate state. The normalization to the local variability σ allows detecting forced variations caused by volcanic eruptions. We test for statistical significance by bootstrapping using 400 block samples of X with a fixed length of 48 months.

We quantify the timescale-dependent variance of surface air temperature using the power spectral density (PSD, denoted spectrum). We use the multitaper method (Percival & Walden, 1993) with three windows and chi-square

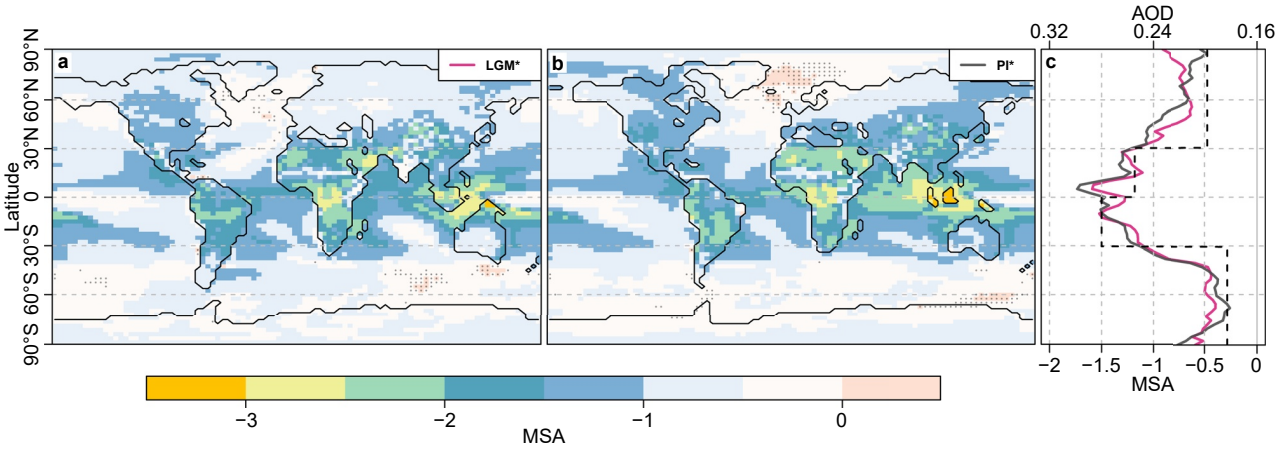


Figure 2. (a) Mean standardized anomalies (MSA) of surface air temperature in the Last Glacial Maximum* (b) and the Pre-Industrial* state after volcanic eruptions. Dots indicate insignificant anomalies within the 99% quantile range of local variability. Gray shaded crosses show land ice. Hatched areas indicate areas with >50% yearly sea ice coverage. (c) Zonally averaged MSA and Aerosol Optical Depth (AOD) (black dashed). MSA and AOD are unitless quantities.

distributed uncertainties. The required assumption of weak stationarity (Davies & Chatfield, 1990) is reasonably fulfilled, given that we linearly detrend all time series (Fredriksen & Rypdal, 2016; Laepple & Huybers, 2014b; Nilsen et al., 2016). Spectra are smoothed logarithmically using a Gaussian kernel. Following Huybers and Curry (2006), we compute mean spectra after interpolation to the lowest resolution and binning into equally spaced log-frequency intervals.

We use variance ratios, as in Laepple and Huybers (2014b), Rehfeld et al. (2018) and Ellerhoff and Rehfeld (2021), to compare model simulations and observational data. First, observation and proxy data are interpolated onto an equidistant time axis of the same mean resolution as the raw signal. We compute the spectrum and obtain the variance by integration over the considered timescale (2–5, 5–20, 50–200, 200–500 years). Finally, we calculate the variance ratio by dividing the simulated by the reconstructed variance. Confidence intervals (CI) are obtained from a F-distribution, based on the degrees of freedom of the variance estimates. The "lgm3" and "pi2" (Table S2 in Supporting Information S1) runs are excluded for the longest timescale (200–500 years) as they are shorter than 1,000 years. Changes in variance ratios between forced and unforced runs are quantified by the area-weighted mean of the improvement factor (Appendix A).

3. Results

3.1. Mean Response to Volcanic Forcing

Volcanic eruptions cause mean temperature decline at almost every location (Figure 2) as expected (Robock, 2000). The mean response, quantified by MSA, is weaker over the oceans than over land. The response is stronger between 30°N and 30°S than in high-latitude regions, following the mean AOD imprint (Figure 2c). The strongest cooling (up to three standard deviations) occurs over the Southeast Asian Archipelago (Figure 2b). These patterns are largely robust against changes in the mean climate. This also applies to precipitation, sea level pressure, and 500mbar wind speed (Figure S3 in Supporting Information S1).

The zonally averaged MSA (Figure 2 (c)) reveals small differences between the states during LGM* and PI* around the equator, 60°S, 50°N, and toward the North Pole. We identify differences in Southeast Asia, the Antarctic Ocean, over the Northern Hemisphere (NH) ice sheets, and the Barents Sea (Figures 2a and 2b). In Southeast Asia, the enhanced PI* surface climate response could be linked to the high AOD imprint from strong tropical volcanic eruptions (Fasullo et al., 2017), such as the 1,257 Samalas eruption. Changes in the land-sea mask in the region could alter the local coupling between ocean and atmosphere. In the LGM*, cooling in response to eruptions is enhanced at the Antarctic sea ice edge and in the Barents Sea. Both regions feature a higher amount of sea ice cover during the LGM. The variations in MSA extend toward the Arctic Ocean and Northern North Atlantic. Differences between the states could therefore be related to the potential for sea ice formation, likely amplifying the local response to volcanic eruptions (Timmreck, 2012). Remaining small differences are found in

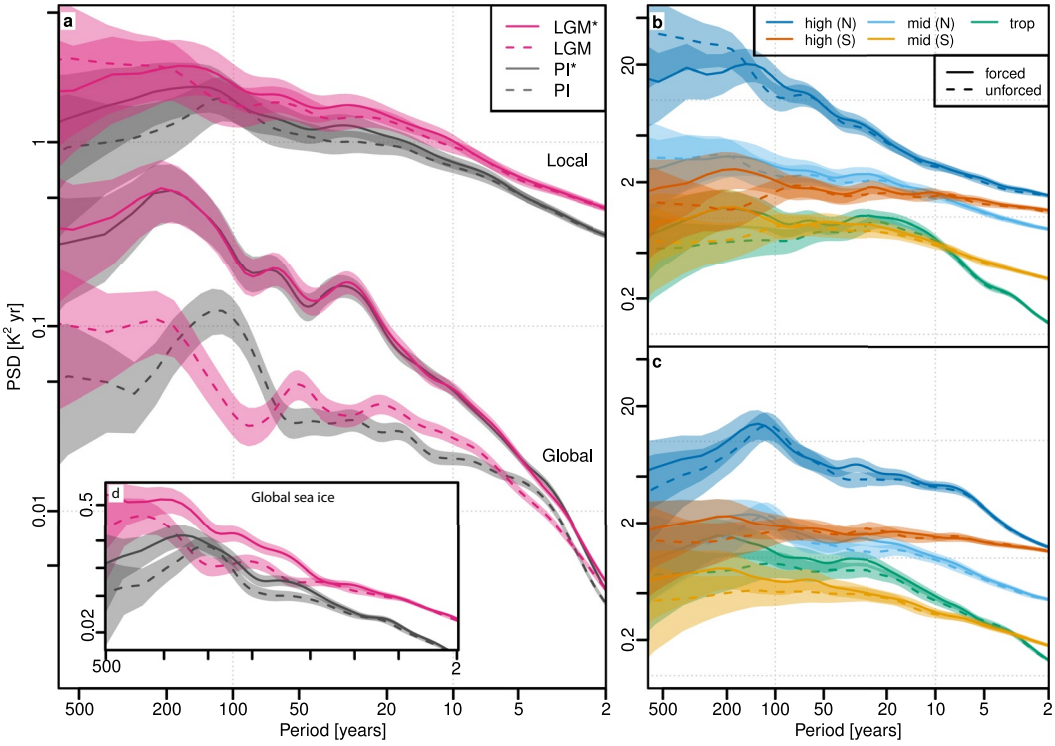


Figure 3. (a) Local and global power spectral density for simulated temperature using Hadley Centre Coupled Model Version 3.4. Global spectra are computed from global mean surface temperature. Local refers to the area-weighted average of all local spectra. (b and c) Area-weighted average of local spectra by climate zone, given by the tropics (−23.5 to 23.5°N), mid (23.5–66.5°), and high latitudes (>66.5°) for Last Glacial Maximum and Pre-Industrial. (d) Global spectra (units in $\%^2$ yr) of global mean sea ice concentration, defined as percentage of the globe covered in sea ice. Lines show logarithmically smoothed (0.08 dB) mean spectra with shaded 95% CIs.

regions with state-dependent changes of NH ice sheets, with a tendency toward enhanced cooling over NH land masses in the Pre-Industrial.

3.2. Spectral Response at the Global and Local Scale

Examining power spectra for the global and local scale highlights the timescale-dependent impact of natural forcing. Global mean spectra of simulated temperatures (Figure 3a) are predominately determined by natural forcing. Including the forcing increases the power, and thus variance, on all timescales. At multidecadal scales, the forced GMST shows approximately five times more variance than unforced runs. State-dependent effects of the forced response are not discernible in these spectra. There are no pronounced spectral peaks. Enhanced centennial-scale variability in the Pre-Industrial could be attributed to a more variable AMOC (Figure S8 in Supporting Information S1).

Local mean spectra (Figure 3a) are characteristic for the mean state and less affected by natural forcing. They point to a greater temperature variance during the LGM. Differences between the states are the strongest on interannual scales, where LGM$^{(*)}$ variance is higher by a factor of approximately two compared to PI$^{(*)}$. Zonal mean spectra (Figures 3b and 3c) reveal that the decrease in variability with warming is greatest at mid-, and especially high-latitudes, supporting a potential link to sea ice dynamics. Tropical variability widely agrees across states. Differences between forced and unforced local and zonal mean spectra are within uncertainties, but most pronounced for high-latitude, long-term variability. The PSD of global mean sea ice concentration is smaller under Pre-Industrial than Glacial conditions (Figure 3d). Above decadal scales, sea ice variability is significantly higher in forced compared to unforced scenarios.

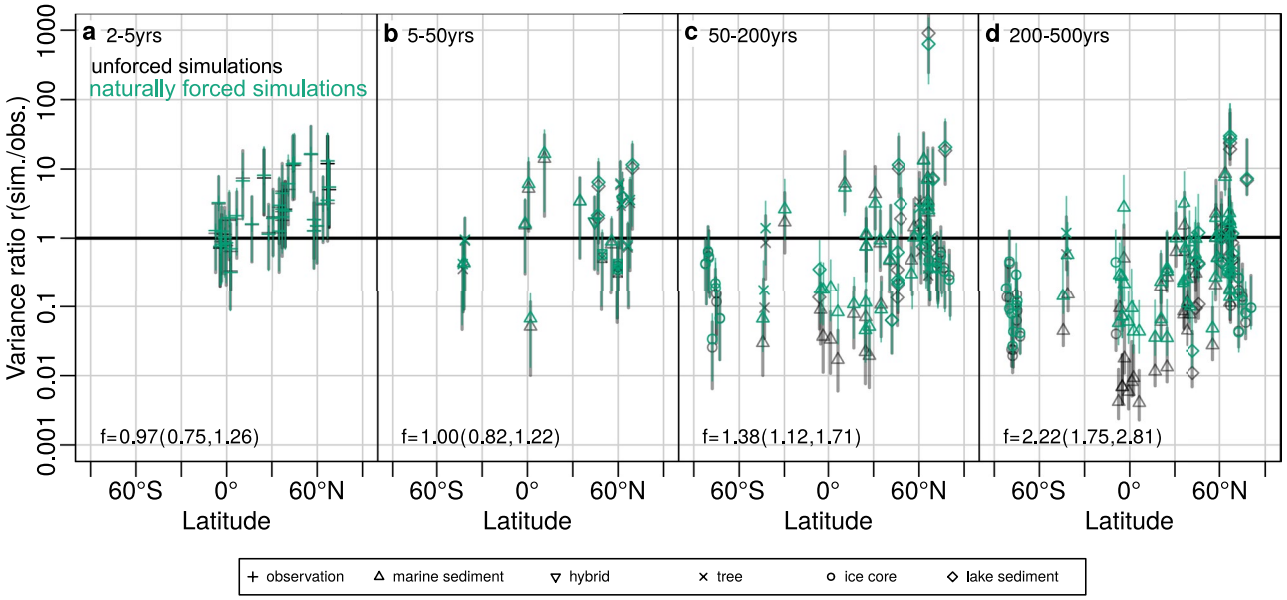


Figure 4. Ratio *r*(sim./obs.) of simulated to observed variance over latitude for unforced (black) and naturally forced (green) Hadley Centre Coupled Model Version 3.4 simulations for the Pre-Industrial. Model data is bilinearly interpolated to the location of the observation. We show the ratio of simulated PI temperature to observations for periods of 2–5 years (a), and to proxies spanning the last 8,000 years on interannual to decadal (b), multidecadal (c), and multicentennial (d) timescales. Symbols indicate the variance ratio and vertical lines their 90% CI. The local mean improvement *f* of the variance ratio is given in the lower left of each panel, with CIs in parentheses (Appendix A).

3.3. Comparison of Observed and Modeled Variability

We validate the simulated variability against observational and paleoclimate data and revisit the local, long-term variability mismatch (Ellerhoff & Rehfeld, 2021; Laepple & Huybers, 2014a; Rehfeld et al., 2018). Figure 4 shows the model-data mismatch as variance ratios. Proxy variance is increasingly larger on longer timescales compared to simulations. There is no major difference in the variance ratios between unforced and naturally forced runs on short timescales (2–5 and 5–50 years; Figures 3a and 3b). This can be explained by internal processes dominating simulated local variability at these scales. The PI simulation slightly overestimates interannual variability in the mid and high latitudes compared to sea surface temperature observations.

Beyond periods of 50 years (Figures 3c–3d), simulated local variance is consistently smaller than proxy-based reconstructions. Including natural forcing in simulations decreases the mismatch for the majority of proxy sample sites. On periods of 50–200 years, the ratio bias is decreased by a factor (local mean improvement, Appendix A) of f = 1.38 (1.12, 1.71, 90% CI). The local mean improvement increases toward multicentennial scales, reducing the discrepancy. On periods of 200–500 years, the mismatch is reduced by a factor of 2.22 (1.75, 2.81). This is not sufficient to achieve consistency between modeled and proxy variance, but the mismatch is significantly smaller.

4. Discussion

We confirm that including natural forcing promotes temperature variability in model simulations across a range of timescales. In contrast to some experiments in the literature, we find that the modeled response of GMST does not strongly depend on the mean climate (Figures 1 and 3). Locally, weak state-dependent effects occur (Figures 2 and 3). Considering natural forcing reduces the model-data mismatch on local temperature variability, in particular on multidecadal and multicentennial scales (Figure 4).

4.1. Forced and Unforced Variability

Previous studies suggested state-dependent effects of volcanic forcing on global and hemispheric climate (Berdahl & Robock, 2013; Muthers et al., 2015; Swingedouw et al., 2017; Zanchettin et al., 2016). These results were obtained using ensembles of large volcanic eruptions. State dependency in these has been primarily linked

to nonlinear processes and initial conditions (Zanchettin et al., 2013). We argue that the response to individual volcanic eruptions may well depend on the climate state. However, globally averaged effects from changes in response mechanisms are small considering realistic forcing scenarios, in line with a linear relation between GMST and external forcing (Fredriksen & Rypdal, 2017; Geoffroy et al., 2013; MacMynowski et al., 2011). In our ensemble, the GMST response to an eruption of the size of Samalas 1,257 shows no difference between LGM* and PI* (Figure S7a in Supporting Information S1). Global precipitation and sea ice concentration is only slightly enhanced in the LGM* (Figures S7b and S7c in Supporting Information S1). Interannual GMST variability and equilibrium climate sensitivity (ECS) can also be linked in a linear feedback framework (Cox, 2001). Assuming the above framework, and that HadCM3 simulates all relevant feedbacks, the similar forced response could indicate that ECS in LGM and PI are not strongly different.

4.2. Internal Variability Across States

The question of state-dependent variability has long motivated studies of past (Ditlevsen et al., 1996; Rehfeld et al., 2018; Shao & Ditlevsen, 2016) and future (Huntingford et al., 2013; Olonscheck et al., 2021; Rehfeld et al., 2020) climate. Our results reveal a decrease in mean local variability with warming (Figure 3). Decreasing sea ice dynamics and a smaller meridional temperature gradient are suggested as major causes. In line with other studies (Bathiany et al., 2018; Berdahl & Robock, 2013; Bethke et al., 2017; Brown et al., 2017; Olonscheck et al., 2021; Rehfeld et al., 2018), we find a clear zonal pattern, with greater reduction of variability in the mid and high latitudes (Figures 3b and 3c). This is corroborated by the small discrepancy between short-term variability from observations and simulations in the mid and high latitudes (Figure 4a). The sea surface temperature observations contain the recent global warming trend and sea ice retreat, our PI(*) simulations do not. This could contribute to the decrease in local, high-latitude variability. The mean climate also changes AMOC variability in HadCM3 (Figure S8 in Supporting Information S1). It is smaller in the LGM on multidecadal and multicentennial scales (Jackson & Vellinga, 2013). Under LGM conditions, the AMOC strength and correlation length is also increased by natural forcing (Figure S8 in Supporting Information S1). Potential mechanisms of the intensification are debated (Iwi et al., 2012; Mignot et al., 2011), and they could contribute to state-dependent enhancement of long-term regional variability through natural forcing.

4.3. Mechanisms Leading to Long-Term Variability

Across our experiments, sea ice variability and regions with varying sea ice extent, primarily the Southern Oceans and Barents Sea, are most affected by natural forcing. This is further supported by mean standardized anomalies of precipitation, sea level pressure, and wind speed over the North American ice sheet, the North Atlantic Ocean, Antarctica, and the Southern Oceans (Figure S3 in Supporting Information S1). Comparing simulations with the two-dimensional energy balance model (TransEBM; Ziegler and Rehfeld (2021); Figure S5 in Supporting Information S1) adds support to the role of sea ice in forced temperature variability. TransEBM is a fairly linear model with no atmospheric and oceanic dynamics. We use it to differentiate the contribution from deterministic forcing and sea ice to the variance. In TransEBM experiments we prescribe sea ice changes from HadCM3. Forming the ratio of the local mean TransEBM and HadCM3 variability (Figure S6 in Supporting Information S1) supports the strong sea ice contribution to interannual variability (sea also Figure 3a). The contribution remains significant on decadal and longer timescales, promoting sea ice variations as a key mechanism of local, long-term variability.

Our results provide crucial insights into the discrepancy between modeled and reconstructed local, long-term variability (Ellerhoff & Rehfeld, 2021; Laepple & Huybers, 2014a). Internal variability dominates the local temperature variance on annual to decadal scales (Goosse et al., 2005), but contributions from natural forcing are detectable beyond decadal timescales (Figure 3), increasing variance on longer timescales. This is supported by increased scaling coefficients (Figure S4 in Supporting Information S1), and, hence, increased persistence of forced temperatures on periods of 50–500 years, similar to (Vyushin et al., 2004). Including natural forcing in simulations improved model-data agreement of local variability on multidecadal and multicentennial scales (Figure 4). This is perhaps surprising given that the forcing has no centennial scale variability (Ellerhoff & Rehfeld, 2021). There is no change in agreement from interannual to decadal timescales, implying that the gain from forcing on local temperatures is small on these short timescales. Hence not only the integrated response to strong (Timmreck, 2012) but also to weak natural forcing contributes to long-term variability. Time-varying forcing appears thus beneficial for reliable simulations of global mean (Figure 3) and local, long-term variability.

Consistent with previous arguments (Bethke et al., 2017), our results challenge the common usage of external forcing that shows no time-varying changes (O'Neill et al., 2016).

4.4. Limitations and Potential

We used the same volcanic forcing reconstruction to drive simulations in LGM and PI, in an idealized setup. The volcanic record shows that large eruptions occurred throughout the Glacial (Brown et al., 2014). We do not account for the possibly lower rate than during the last millennium, which could also be due to undersampling. Furthermore, we may miss feedbacks in HadCM3 that are relevant for local climate variability. This could explain the underestimation of local variability compared to proxy data (Figure 4a). Sea ice dynamics, stratospheric and cloud-related feedbacks are key nonlinear mechanisms that can alter the response to volcanic forcing in a warmer climate (Aubry et al., 2021; Fasullo et al., 2017; Hopcroft et al., 2018). Projections of tropical eruptions with a newer model show enhanced (dampened) radiative forcing from strong (moderate) eruptions (Aubry et al., 2021). Cloud-related feedback, likely underestimated in HadCM3, is generally weaker than feedback from sea ice, but may be enhanced with warming (Hopcroft et al., 2018).

Future work could examine the response in simulations with models that show a higher ECS (Gettelman et al., 2019; Tatebe et al., 2019; Voldoire et al., 2019; Wu et al., 2019) and better sea-ice (Guarino et al., 2020). Insufficient sea ice and vegetation cover changes may significantly alter the response in extreme warming scenarios. Future studies could test long-term impacts of volcanism and local state dependency with more advanced climate models, including better representation of radiative-chemical feedback, aerosol indirect, stratospheric and sea ice processes. Moreover, future research could apply probabilistic eruption projections (Bethke et al., 2017) in larger ensembles (Zanchettin et al., 2016) to study forcing scenarios with localized eruptions at high and low latitude. This will aid understanding of long-term Earth-system sensitivity and the state-dependent response of multidecadal modes to natural forcing (Swingedouw et al., 2017).

5. Conclusion

Presenting the first millennial-length, naturally forced simulation for the LGM, we investigated state-dependent effects of volcanic and solar forcing on global and local climate variability. The modeled global temperature response shows no dependency on the mean climate. Weak local differences resulted primarily from sea ice dynamics, providing a key mechanism of long-term variability. Including natural forcing in climate model simulations improves the agreement between modeled and observed variability and, thus, calls into question the lack of time-dependent volcanic forcing in projections and model-data comparison. The robust temperature response suggests that findings on the ability of models to simulate past variability can help constrain forced variability across spatial and temporal scales.

Appendix A: Variance Ratio Improvement

We quantify the change in variance ratio r from unforced and naturally forced simulations to proxy records using the logarithmic measure $l(x) = |\log_{10}(x)|$. Let $r_i^{(*)} = \mathrm{var}\left(S_i^{(*)}\right)/\mathrm{var}\left(S_i'\right)$ be the variance ratio from the simulated $S_i^{(*)}$ and proxy spectrum S_i' at site i, with (*) denoting the climate state. The distance $l_i = l\left(r_i^*\right) - l(r_i)$ gives the change of the variance ratio bias between the forced and unforced simulation. For a set of N sites, we quantify the mean change $\Delta l = \frac{1}{N}\sum_i^N l_i w_i$ with local area weights w_i derived from the HadCM3 grid $\left(\sum_i^N w_i = 1\right)$. We convert the logarithmic distance to the variance ratio improvement $f = 10^{\Delta l}$ and estimate confidence intervals using the area-weighted mean of the error propagation

$$\delta l_i = \sqrt{\left(\frac{\delta r_i^*}{r_i^* \ln(10)}\right)^2 + \left(\frac{\delta r_i}{r_i \ln(10)}\right)^2}. \tag{A1}$$

We ensure a conservative coverage of the CIs by using the upper limit on $\delta r_i^{(*)}$ from F-distributed uncertainties of the variance ratio estimate.

Data Availability Statement

The presented model simulations are available at Zenodo via https://doi.org/10.5281/zenodo.6074747 with CC-BY-SA 4.0 license. They were carried out using version three of the Hadley Center Coupled Model, HadCM3, as described in Valdes et al. (2017) and Tindall et al. (2009). Paleoclimate and observation datasets for this research are included in Rehfeld et al. (2018), PAGES2k-Consortium (2017) and Rayner et al. (2003). Supplemental analysis used the two-dimensional TransEBM model described by Ziegler and Rehfeld (2021) which is based on Zhuang et al. (2017). Code and data to reproduce all figures is available at https://github.com/paleovar/StateDependency, and also available at the following Zenodo: https://doi.org/10.5281/zenodo.6474769.

Acknowledgments
This research used the Archer UK National Supercomputing Services. It benefited from discussions within the CVAS working group, a working group of the Past Global Changes (PAGES) project. We thank M. Casado, T. Laepple and A. Schurer for discussion, and C. Wirths for setting up volcanic forcing over latitude intervals in TransEBM. Funding is acknowledged by the PalMod project (www.palmod.de, subproject no. 01LP1926C), the Deutsche Forschungsgemeinschaft (German Research Foundation) – project no. 395588486 and no. 316076679, and by the Heinrich-Böll-Stiftung. Open access funding enabled and organized by Projekt DEAL.

References

Ambrose, S. H. (1998). Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. *Journal of Human Evolution*, *34*(6), 623–651. https://doi.org/10.1006/jhev.1998.0219

Armstrong, E., Izumi, K., & Valdes, P. (2022). *No title. Climate dynamics*. https://doi.org/10.21203/rs.3.rs-715149/v1

Aubry, T. J., Staunton-Sykes, J., Marshall, L. R., Haywood, J., Abraham, N. L., & Schmidt, A. (2021). Climate change modulates the stratospheric volcanic sulfate aerosol lifecycle and radiative forcing from tropical eruptions. *Nature Communications*, *12*(1), 4708. https://doi.org/10.1038/s41467-021-24943-7

Baldini, J. U., Brown, R. J., & McElwaine, J. N. (2015). Was millennial scale climate change during the Last Glacial triggered by explosive volcanism? *Scientific Reports*, *5*, 1–9. https://doi.org/10.1038/srep17442

Bathiany, S., Dakos, V., Scheffer, M., & Lenton, T. M. (2018). Climate models predict increasing temperature variability in poor countries. *Science Advances*, *4*(5), eaar5809. https://doi.org/10.1126/sciadv.aar5809

Berdahl, M., & Robock, A. (2013). Northern hemispheric cryosphere response to volcanic eruptions in the paleoclimate modeling intercomparison project 3 last millennium simulations. *Journal of Geophysical Research: Atmospheres*, *118*, 12359–12370. https://doi.org/10.1002/2013JD019914

Bethke, I., Outten, S., Otterå, O. H., Hawkins, E., Wagner, S., Sigl, M., & Thorne, P. (2017). Potential volcanic impacts on future climate variability. *Nature Climate Change*, *7*(11), 799–805. https://doi.org/10.1038/nclimate3394

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., et al. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 867–952). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.022

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., et al. (2012). Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, *2*(6), 417–424. https://doi.org/10.1038/nclimate1456

Brown, P. T., Ming, Y., Li, W., & Hill, S. A. (2017). Change in the magnitude and mechanisms of global temperature variability with warming. *Nature Climate Change*, *7*(10), 743–748. https://doi.org/10.1038/nclimate3381

Brown, S. K., Crosweller, H. S., Sparks, R. S. J., Cottrell, E., Deligne, N. I., Guerrero, N. O., et al. (2014). Characterisation of the quaternary eruption record: Analysis of the large magnitude explosive volcanic eruptions (LaMEVE) database. *Journal of Applied Volcanology*, *3*(1), 5. https://doi.org/10.1186/2191-5040-3-5

Bühler, J. C., Roesch, C., Kirschner, M., Sime, L., Holloway, M., & Rehfeld, K. (2021). Comparison of the oxygen isotope signatures in speleothem records and ihadcm3 model simulations for the last millennium. *Climate of the Past*, 1–30. https://doi.org/10.5194/cp-2020-121

Casado, M., Münch, T., & Laepple, T. (2020). Climatic information archived in ice cores: Impact of intermittency and diffusion on the recorded isotopic signal in Antarctica. *Climate of the Past*, *16*(4), 1581–1598. https://doi.org/10.5194/cp-16-1581-2020

Collins, M., Tett, S., & Cooper, C. (2001). The internal climate variability of HadCM3, a version of the hadley centre coupled model without flux adjustments. *Climate Dynamics*, *17*(1), 61–81. https://doi.org/10.1007/s003820000094

Cox, P. M., Huntingford, C., & Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, *553*(7688), 319–322. https://doi.org/10.1038/nature25450

Crowley, T. J. (2000). Causes of climate change over the past 1000 years. *Science*, *289*(5477), 270–277. https://doi.org/10.1126/science.289.5477.270

Crowley, T. J., & Unterman, M. B. (2013). Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth System Science Data*, *5*(1), 187–197. https://doi.org/10.5194/essd-5-187-2013

Davies, N., & Chatfield, C. (1990). The analysis of time series: An introduction. *The Mathematical Gazette*, *74*(468), 194–195. https://doi.org/10.2307/3619403

Ditlevsen, P. D., Svensmark, H., & Johnsen, S. (1996). Contrasting atmospheric and climate dynamics of the last-glacial and Holocene periods. *Nature*, *379*(6568), 810–812. https://doi.org/10.1038/379810a0

Ellerhoff, B., & Rehfeld, K. (2021). Probing the timescale dependency of local and global variations in surface air temperature from climate simulations and reconstructions of the last millennia. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *104*(6), 064136. https://doi.org/10.1103/PhysRevE.104.064136

Fasullo, J. T., Tomas, R., Stevenson, S., Otto-Bliesner, B., Brady, E., & Wahl, E. (2017). The amplifying influence of increased ocean stratification on a future year without a summer. *Nature Communications*, *8*(1), 1236. https://doi.org/10.1038/s41467-017-01302-z

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., et al. (2014). *Evaluation of climate models*. In T. Stocker et al. (Eds.), (Vol. 9781107057999, pp. 741–866). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.020

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J. L., Frame, D., & Zhang, H. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021: The physical science basis. contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Frankcombe, L. M., England, M. H., Mann, M. E., & Steinman, B. A. (2015). Separating internal variability from the externally forced climate response. *Journal of Climate*, *28*(20), 8184–8202. https://doi.org/10.1175/JCLI-D-15-0069.1

Franzke, C. L., Barbosa, S., Blender, R., Fredriksen, H., Laepple, T., Lambert, F., et al. (2020). The structure of climate variability across scales. *Reviews of Geophysics*, *58*(2). https://doi.org/10.1029/2019RG000657

Fredriksen, H. B., & Rypdal, K. (2016). Spectral characteristics of instrumental and climate model surface temperatures. *Journal of Climate*, *29*(4), 1253–1268. https://doi.org/10.1175/JCLI-D-15-0457.1

Fredriksen, H. B., & Rypdal, M. (2017). Long-range persistence in global surface temperatures explained by linear multibox energy balance models. *Journal of Climate*, *30*(18), 7157–7168. https://doi.org/10.1175/JCLI-D-16-0877.1

Geoffroy, O., Saint-martin, D., Olivié, D. J., Voldoire, A., Bellon, G., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *Journal of Climate*, *26*(6), 1841–1857. https://doi.org/10.1175/JCLI-D-12-00195.1

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A., Danabasoglu, G., et al. (2019). othersHigh climate sensitivity in the community Earth system model version 2 (cesm2). *Geophysical Research Letters*, *46*(14), 8329–8337. https://doi.org/10.1029/2019gl083978

Goosse, H., Renssen, H., Timmermann, A., & Bradley, R. S. (2005). Internal and forced climate variability during the last millennium: A model-data comparison using ensemble simulations. *Quaternary Science Reviews*, *24*(12–13), 1345–1360. https://doi.org/10.1016/j.quascirev.2004.12.009

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., et al. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the hadley centre coupled model without flux adjustments. *Climate Dynamics*, *716*(2–3), 147–168. https://doi.org/10.1007/s003820050010

Guarino, M.-V., Sime, L. C., Schröeder, D., Malmierca-Vallet, I., Rosenblum, E., Ringer, M., et al. (2020). Sea-ice-free arctic during the last interglacial supports fast future loss. *Nature Climate Change*, *10*(10), 928–932. https://doi.org/10.1038/s41558-020-0865-2

Halloran, P. R., Hall, I. R., Menary, M., Reynolds, D. J., Scourse, J. D., Screen, J. A., et al. (2020). Natural drivers of multidecadal Arctic sea ice variability over the last millennium. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-57472-2

Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics*, *13*(9), 601–611. https://doi.org/10.1007/s003820050185

Haustein, K., Otto, F. E., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., et al. (2019). A limited role for unforced internal variability in twentieth-century warming. *Journal of Climate*, *32*(16), 4893–4917. https://doi.org/10.1175/JCLI-D-18-0555.1

Holmes, C. R., Woollings, T., Hawkins, E., & de Vries, H. (2016). Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *Journal of Climate*, *29*(6), 2221–2236. https://doi.org/10.1175/JCLI-D-14-00735.1

Hopcroft, P. O., Kandlbauer, J., Valdes, P. J., & Sparks, R. S. J. (2018). Reduced cooling following future volcanic eruptions. *Climate Dynamics*, *51*(4), 1449–1463. https://doi.org/10.1007/s00382-017-3964-7

Huntingford, C., Jones, P. D., Livina, V. N., Lenton, T. M., & Cox, P. M. (2013). No increase in global temperature variability despite changing regional patterns. *Nature*, *500*(7462), 327–330. https://doi.org/10.1038/nature12310

Iwi, A. M., Hermanson, L., Haines, K., & Sutton, R. T. (2012). Mechanisms linking volcanic aerosols to the Atlantic meridional overturning circulation. *Journal of Climate*, *25*(8), 3039–3051. https://doi.org/10.1175/2011JCLI4067.1

Jackson, L., & Vellinga, M. (2013). Multidecadal to centennial variability of the AMOC: HadCM3 and a perturbed physics ensemble. *Journal of Climate*, *26*(7), 2390–2407. https://doi.org/10.1175/jcli-d-11-00601.1

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., et al. (2018). The PMIP4 contribution to CMIP6–Part 1: Overview and over-arching analysis plan. *Geoscientific Model Development*, *11*(3), 1033–1057. https://doi.org/10.5194/gmd-11-1033-2018

Katz, R. W., & Brown, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, *21*(3), 289–302. https://doi.org/10.1007/bf00139728

Laepple, T., & Huybers, P. (2014a). Global and regional variability in marine surface temperatures. *Geophysical Research Letters*, *41*(7), 2528–2534. https://doi.org/10.1002/2014GL059345

Laepple, T., & Huybers, P. (2014b). Ocean surface temperature variability: Large model–data differences at decadal and longer periods. *Proceedings of the National Academy of Sciences*, *111*(47), 16682–16687. https://doi.org/10.1073/pnas.1412077111

Laepple, T., Münch, T., Casado, M., Hoerhold, M., Landais, A., & Kipfstuhl, S. (2018). On the similarity and apparent cycles of isotopic variations in East Antarctic snow pits. *The Cryosphere*, *12*(1), 169–187. https://doi.org/10.5194/tc-12-169-2018

Liu, Z., Otto-Bliesner, B., He, F., Brady, E., Tomas, R., Clark, P., et al. (2009). Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. *Science*, *325*(5938), 310–314. https://doi.org/10.1126/science.1171041

MacMynowski, D. G., Shin, H.-J., & Caldeira, K. (2011). The frequency response of temperature and precipitation in a climate model. *Geophysical Research Letters*, *38*. https://doi.org/10.1029/2011GL048623

Mann, M. E., Steinman, B. A., Brouillette, D. J., Fernandez, A., & Miller, S. K. (2022). On the estimation of internal climate variability during the preindustrial past millennium. *Geophysical Research Letters*, e2021GL096596. https://doi.org/10.1029/2021GL096596

Marcott, S. A., Shakun, J. D., Clark, P. U., & Mix, A. C. (2013). A reconstruction of regional and global temperature for the past 11,300 years. *Science*, *339*(6124), 1198–1201. https://doi.org/10.1126/science.1228026

Mignot, J., Khodri, M., Frankignoul, C., & Servonnat, J. (2011). Volcanic impact on the Atlantic ocean over the last millennium. *Climate of the Past*, *7*(4), 1439–1455. https://doi.org/10.5194/cp-7-1439-2011

Muthers, S., Arfeuille, F., Raible, C. C., & Rozanov, E. (2015). The impacts of volcanic aerosol on stratospheric ozone and the northern hemisphere polar vortex: Separating radiative-dynamical changes from direct effects due to enhanced aerosol heterogeneous chemistry. *Atmospheric Chemistry and Physics*, *15*(20), 11461–11476. https://doi.org/10.5194/acp-15-11461-2015

Nilsen, T., Rypdal, K., & Fredriksen, H. B. (2016). Are there multiple scaling regimes in Holocene temperature records? *Earth System Dynamics*, *7*(2), 419–439. https://doi.org/10.5194/esd-7-419-2016

Olonscheck, D., Schurer, A. P., Lücke, L., & Hegerl, G. C. (2021). Large-scale emergence of regional changes in year-to-year temperature variability by the end of the 21st century. *Nature Communications*, *12*(1), 7237. https://doi.org/10.1038/s41467-021-27515-x

O'Neill, B. C., Tebaldi, C., Vuuren, D. P. V., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., et al. (2016). Climate variability and change since 850 CE: An ensemble approach with the community Earth system model. *Bulletin of the American Meteorological Society*, *97*(5), 735–754. https://doi.org/10.1175/bams-d-14-00233.1

PAGES2k-Consortium. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*, 1–33. https://doi.org/10.1038/sdata.2017.88

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., & Sanderson, B. M. (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*, *7*(1), 17966. https://doi.org/10.1038/s41598-017-17966-y

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*. Cambridge University Press. https://doi.org/10.1017/CBO9780511622762

Pope, V., Gallani, M., Rowntree, P., & Stratton, R. (2000). The impact of new physical parametrizations in the Hadley Centre Climate model: HadAM3. *Climate Dynamics*, *16*(2–3), 123–146. https://doi.org/10.1007/s003820050009

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Rehfeld, K., Hébert, R., Lora, J., Lofverstrom, M., & Brierley, C. (2020). Variability of surface climate in simulations of past and future. *Earth System Dynamics*, 1–30. https://doi.org/10.5194/esd-2019-92

Rehfeld, K., Münch, T., Ho, S. L., & Laepple, T. (2018). Global patterns of declining temperature variability from the last glacial maximum to the Holocene. *Nature*, *554*(7692), 356–359. https://doi.org/10.1038/nature25454

Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3), 303–312. https://doi.org/10.1175/BAMS-89-3-303

Robock, A. (2000). Volcanic eruptions and climate. *Reviews of Geophysics*, *38*(2), 191–219. https://doi.org/10.1029/1998RG000054

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., et al. (2012). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1). *Geoscientific Model Development*, *5*(1), 185–191. https://doi.org/10.5194/gmd-5-185-2012

Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B., & Phipps, S. J. (2013). Separating forced from chaotic climate variability over the past millennium. *Journal of Climate*, *26*(18), 6954–6973. https://doi.org/10.1175/JCLI-D-12-00826.1

Shao, Z. G., & Ditlevsen, P. D. (2016). Contrasting scaling properties of interglacial and glacial climates. *Nature Communications*, *7*(1), 10951. https://doi.org/10.1038/ncomms10951

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020). Climate change now detectable from any single day of weather at global scale. *Nature Climate Change*, *10*(1), 35–41. https://doi.org/10.1038/s41558-019-0666-7

Smith, R. S., & Gregory, J. (2012). The last glacial cycle: Transient simulations with an AOGCM. *Climate Dynamics*, *38*(7–8), 1545–1559. https://doi.org/10.1007/s00382-011-1283-y

Steinhilber, F., Beer, J., & Fröhlich, C. (2009). Total solar irradiance during the Holocene. *Geophysical Research Letters*, *36*(19), 1–5. https://doi.org/10.1029/2009GL040142

Stott, P. A., Tett, S. F., Jones, G. S., Allen, M., Mitchell, J., & Jenkins, G. (2000). External control of 20th century temperature by natural and anthropogenic forcings. *Science*, *290*(5499), 2133–2137. https://doi.org/10.1126/science.290.5499.2133

Svensson, A., Bigler, M., Blunier, T., Clausen, H. B., Dahl-Jensen, D., Fischer, H., et al. (2013). Direct linking of Greenland and Antarctic ice cores at the Toba eruption (74 ka BP). *Climate of the Past*, *9*(2), 749–766. https://doi.org/10.5194/cp-9-749-2013

Swingedouw, D., Mignot, J., Ortega, P., Khodri, M., Menegoz, M., Cassou, C., & Hanquiez, V. (2017). Impact of explosive volcanic eruptions on the main climate variability modes. *Global and Planetary Change*, *150*, 24–45. https://doi.org/10.1016/j.gloplacha.2017.01.006

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Timmreck, C. (2012). Modeling the climatic effects of large explosive volcanic eruptions. *Wiley Interdisciplinary Reviews: Climate Change*, *3*(6), 545–564. https://doi.org/10.1002/wcc.192

Timmreck, C., Graf, H.-F., Lorenz, S. J., Niemeier, U., Zanchettin, D., Matei, D., et al. (2010). Aerosol size confines climate response to volcanic super-eruptions. *Geophysical Research Letters*, *37*(24). https://doi.org/10.1029/2010GL045464

Tindall, J. C., Valdes, P. J., & Sime, L. C. (2009). Stable water isotopes in HadCM3: Isotopic signature of El Niño-southern oscillation and the tropical amount effect. *Journal of Geophysical Research*, *114*(D4), 1–12. https://doi.org/10.1029/2008JD010825

Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., et al. (2017). The bridge HadCM3 family of climate models: Hadcm3@bristol v1.0. *Geoscientific Model Development*, *10*(10), 3715–3743. https://doi.org/10.5194/gmd-10-3715-2017

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 deck experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177–2213. https://doi.org/10.1029/2019ms001683

Vyushin, D., Zhidkov, I., Havlin, S., Bunde, A., & Brenner, S. (2004). Volcanic forcing improves atmosphere-ocean coupled general circulation model scaling performance. *Geophysical Research Letters*, *31*(10). https://doi.org/10.1029/2004gl019499

Wang, T., Miao, J., Wang, H., & Sun, J. (2021). Influence of strong tropical volcanic eruptions on daily temperature and precipitation extremes across the globe. *Journal of Meteorological Research*, *35*(3), 428–443. https://doi.org/10.1007/s13351-021-0160-9

Wang, Y., Lean, J. L., & Sheeley, N. R. Jr. (2005). Modeling the Sun's magnetic field and irradiance since 1713. *The Astrophysical Journal*, *625*(1), 522–538. https://doi.org/10.1086/429689

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., et al. (2019). The Beijing Climate Center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*(4), 1573–1600. https://doi.org/10.5194/gmd-12-1573-2019

Zanchettin, D., Bothe, O., Graf, H. F., Lorenz, S. J., Luterbacher, J., Timmreck, C., & Jungclaus, J. H. (2013). Background conditions influence the decadal climate response to strong volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, *118*(10), 4090–4106. https://doi.org/10.1002/jgrd.50229

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., et al. (2016). The model intercomparison project on the climatic response to volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6. *Geoscientific Model Development*, *9*(8), 2701–2719. https://doi.org/10.5194/gmd-9-2701-2016

Zhuang, K., North, G. R., & Stevens, M. J. (2017). A NetCDF version of the two-dimensional energy balance model based on the full multigrid algorithm. *Software*, *6*, 198–202. https://doi.org/10.1016/j.softx.2017.07.003

Ziegler, E., & Rehfeld, K. (2021). Transebm V. 1.0: Description, tuning, and validation of a transient model of the Earth's energy balance in two dimensions. *Geoscientific Model Development*, *14*(5), 2843–2866. https://doi.org/10.5194/gmd-14-2843-2021

References From the Supporting Information

Berger, A. L. (1978). Long-term variations of daily insolation and Quaternary climatic changes. *Journal of the Atmospheric Sciences*, *35*(12), 2361–2367. https://doi.org/10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2

Brad Adams, J., Mann, M. E., & Ammann, C. M. (2003). Proxy evidence for an El Niño-like response to volcanic forcing. *Nature*, *426*(6964), 274–278. https://doi.org/10.1038/nature02101

Cox, P. M. (2001). *Description of the "TRIFFID" dynamic global vegetation model*. Hadley Centre Technical Note.

Crowley, T. J., & Unterman, M. B. (2013). Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth System Science Data*, *5*(1), 187–197. https://doi.org/10.5194/essd-5-187-2013

Danabasoglu, G., Yeager, S. G., Kwon, Y.-O., Tribbia, J. J., Phillips, A. S., & Hurrell, J. W. (2012). Variability of the Atlantic meridional overturning circulation in CCSM4. *Journal of Climate*, *25*(15), 5153–5172. https://doi.org/10.1175/jcli-d-11-00463.1

Fredriksen, H. B., & Rypdal, M. (2017). Long-range persistence in global surface temperatures explained by linear multibox energy balance models. *Journal of Climate*, *30*(18), 7157–7168. https://doi.org/10.1175/JCLI-D-16-0877.1

Huybers, P., & Curry, W. (2006). Links between annual, Milankovitch and continuum temperature variability. *Nature*, *441*(7091), 329–332. https://doi.org/10.1038/nature04745

Lovejoy, S., & Varotsos, C. (2016). Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings. *Earth System Dynamics*, *7*(1), 133–150. https://doi.org/10.5194/esd-7-133-2016

PAGES2k-Consortium. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*, 1–33. https://doi.org/10.1038/sdata.2017.88

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Rehfeld, K., Münch, T., Ho, S. L., & Laepple, T. (2018). Global patterns of declining temperature variability from the last glacial maximum to the Holocene. *Nature*, *554*(7692), 356–359. https://doi.org/10.1038/nature25454

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., et al. (2012). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1). *Geoscientific Model Development*, *5*(1), 185–191. https://doi.org/10.5194/gmd-5-185-2012

Sear, C. B., Kelly, P. M., Jones, P. D., & Goodess, C. M. (1987). Global surface-temperature responses to major volcanic eruptions. *Nature*, *330*(6146), 365–367. https://doi.org/10.1038/330365a0

Singarayer, J. S., & Valdes, P. J. (2010). High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quaternary Science Reviews*, *29*(1–2), 43–55. https://doi.org/10.1016/j.quascirev.2009.10.011

Steinhilber, F., Beer, J., & Fröhlich, C. (2009). Total solar irradiance during the Holocene. *Geophysical Research Letters*, *36*(19), 1–5. https://doi.org/10.1029/2009GL040142

Wang, Y., Lean, J. L., & Sheeley, N. R. Jr. (2005). Modeling the Sun's magnetic field and irradiance since 1713. *The Astrophysical Journal*, *625*(1), 522–538. https://doi.org/10.1086/429689

Zhuang, K., North, G. R., & Stevens, M. J. (2017). A NetCDF version of the two-dimensional energy balance model based on the full multigrid algorithm. *Software*, *6*, 198–202. https://doi.org/10.1016/j.softx.2017.07.003

Ziegler, E., & Rehfeld, K. (2021). Transebm V. 1.0: Description, tuning, and validation of a transient model of the Earth's energy balance in two dimensions. *Geoscientific Model Development*, *14*(5), 2843–2866. https://doi.org/10.5194/gmd-14-2843-2021

Supporting Information for "Contrasting state-dependent effects of natural forcing on global and local climate variability"

Beatrice Ellerhoff[1,2], Moritz J. Kirschner[2], Elisa Ziegler[1,2], Max D. Holloway[3], Louise Sime[4], Kira Rehfeld[1,2]

[1]Department of Geosciences and Department of Physics, Tübingen University, Tübingen, Germany

[2]Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany

[3]Scottish Association for Marine Science, Oban, United Kingdom

[4]British Antarctic Survey, Cambridge, United Kingdom

Contents of this file

1. Figures S1 to S10
2. Tables S1 and S2

Additional Supporting Information (Files uploaded separately)

1. Captions for Datasets S1

Introduction This supporting material provides additional information on boundary conditions, surface climate, and spectral properties of the HadCM3 simulation. We show the power spectra of all simulated and reconstructed time series, used for variance ratio

estimates. The separately uploaded dataset S1 contains detailed information about the considered paleoclimate records from Rehfeld, Münch, Ho, and Laepple (2018), Rayner et al. (2003), and the PAGES2k-Consortium (2017). We provide a supporting analysis on the contribution of sea ice dynamics to variability using the two-dimensional TransEBM model (Ziegler & Rehfeld, 2021).

Data Set S1. Key specification of proxy records used to estimate local temperature variance ratios. The records were collected from Rehfeld et al. (2018), Rayner et al. (2003) and the PAGES2k-Consortium (2017). The first six columns denote the reconstruction name, assigned ID, location (Latitude, Longitude, Elevation), archive type, and proxy used. The last column denotes the climate state ("LGM" or "PI") for which the proxy reconstruction was considered. Surface temperature observations were taken from the location closest to the proxy location and specified by "HadISST@...".

References

Berger, A. L. (1978). Long-term variations of daily insolation and Quaternary climatic changes. *Journal of Atmospheric Sciences*, *35*(12), 2361–2367. doi: 10.1175/1520-0469(1978)035⟨2362:ltvodi⟩2.0.co;2

Brad Adams, J., Mann, M. E., & Ammann, C. M. (2003). Proxy evidence for an El Niño-like response to volcanic forcing. *Nature*, *426*(6964), 274–278. doi: 10.1038/nature02101

Cox, P. M. (2001). Description of the" triffid" dynamic global vegetation model. *Hadley Centre Technical Note*.

Crowley, T. J., & Unterman, M. B. (2013). Technical details concerning development of

a 1200 yr proxy index for global volcanism. *Earth System Science Data*, *5*, 187-197. doi: 10.5194/essd-5-187-2013

Danabasoglu, G., Yeager, S. G., Kwon, Y.-O., Tribbia, J. J., Phillips, A. S., & Hurrell, J. W. (2012). Variability of the atlantic meridional overturning circulation in ccsm4. *Journal of climate*, *25*(15), 5153–5172.

Fredriksen, H. B., & Rypdal, M. (2017). Long-range persistence in global surface temperatures explained by linear multibox energy balance models. *Journal of Climate*, *30*, 7157-7168. doi: 10.1175/JCLI-D-16-0877.1

Huybers, P., & Curry, W. (2006). Links between annual, milankovitch and continuum temperature variability. *Nature*, *441*, 329-332. doi: 10.1038/nature04745

Lovejoy, S., & Varotsos, C. (2016). Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings. *Earth System Dynamics*, *7*, 133-150. doi: 10.5194/esd-7-133-2016

PAGES2k-Consortium. (2017). A global multiproxy database for temperature reconstructions of the common era. *Scientific Data*, *4*, 1-33. doi: DOI:10.1038/sdata.2017.88

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(D14). doi: https://doi.org/10.1029/2002JD002670

Rehfeld, K., Münch, T., Ho, S. L., & Laepple, T. (2018). Global patterns of declining temperature variability from the last glacial maximum to the holocene. *Nature*, *554*,

356-359. doi: 10.1038/nature25454

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., ... Vieira, L. E. A. (2012). Climate forcing reconstructions for use in pmip simulations of the last millennium (v1.1). *Geoscientific Model Development*, *5*, 185-191. doi: 10.5194/gmd-5-185-2012

Sear, C. B., Kelly, P. M., Jones, P. D., & Goodess, C. M. (1987). Global surface-temperature responses to major volcanic eruptions. *Nature*, *330*(6146), 365–367. Retrieved from `https://doi.org/10.1038/330365a0` doi: 10.1038/330365a0

Singarayer, J. S., & Valdes, P. J. (2010). High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quaternary Science Reviews*, *29*, 43-55. doi: 10.1016/j.quascirev.2009.10.011

Steinhilber, F., Beer, J., & Fröhlich, C. (2009). Total solar irradiance during the holocene. *Geophysical Research Letters*, *36*, 1-5. doi: 10.1029/2009GL040142

Wang, Y., Lean, J. L., & Jr., N. R. S. (2005). Modeling the sun's magnetic field and irradiance since 1713. *The Astrophysical Journal*, *625*, 522-538. doi: 10.1086/429689

Zhuang, K., North, G. R., & Stevens, M. J. (2017). A netcdf version of the two-dimensional energy balance model based on the full multigrid algorithm. *SoftwareX*, *6*, 198-202. doi: 10.1016/j.softx.2017.07.003

Ziegler, E., & Rehfeld, K. (2021). Transebm v. 1.0: Description, tuning, and validation of a transient model of the earth's energy balance in two dimensions. *Geoscientific Model Development*, *14*, 2843-2866. doi: 10.5194/gmd-14-2843-2021

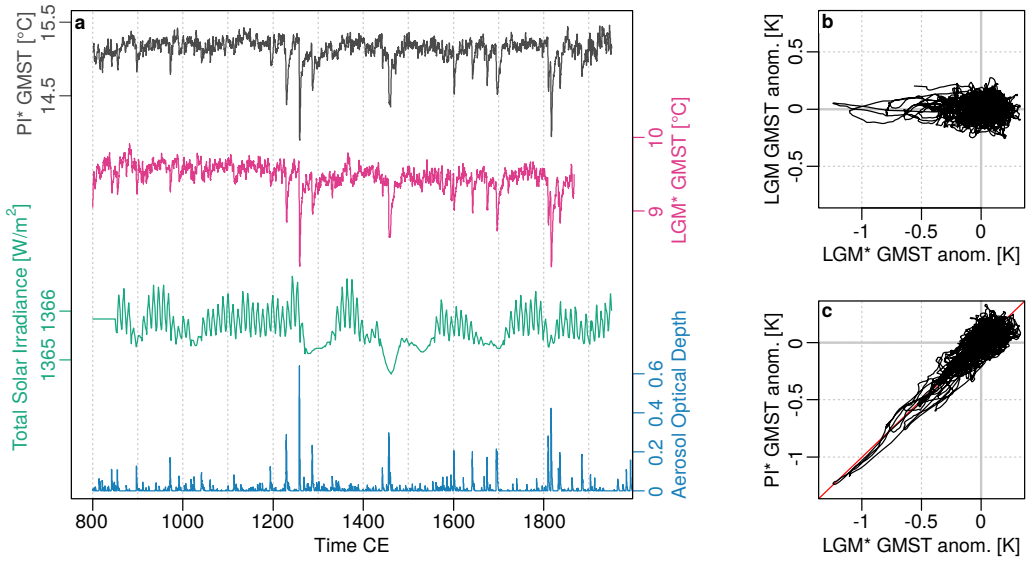


Figure S1. **a** Simulated PI* and LGM* global mean surface temperature (GMST) averaged over all runs in a state, total solar irradiance (Steinhilber et al., 2009), and aerosol optical depth (Crowley & Unterman, 2013). The solar forcing was kept constant the first 50 years due to missing reconstructions. **b** LGM over LGM* GMST anomalies from HadCM3 after linear detrending and subtracting the mean of the full time series. **c** As **b**, with PI* over LGM* GMST anomalies.

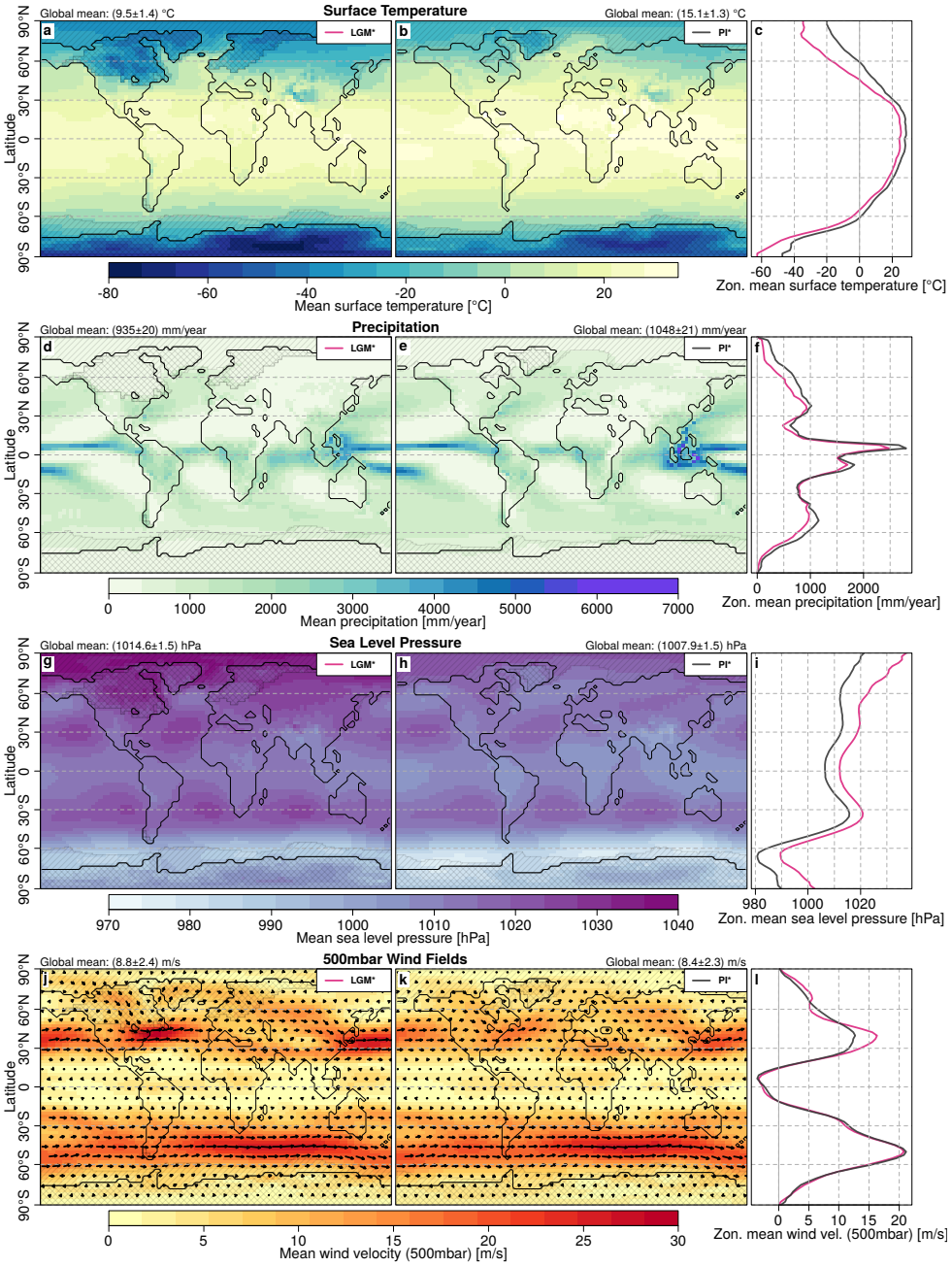


Figure S2. Surface temperature (**a-c**), precipitation rate (**d-f**), sea level pressure (**g-i**), and wind fields at 500mbar (**j-l**) as simulated by HadCM3 for the LGM* and PI*. Means over latitude intervals are displayed in the right-hand panels. Global mean values and their standard deviation are given above the maps.

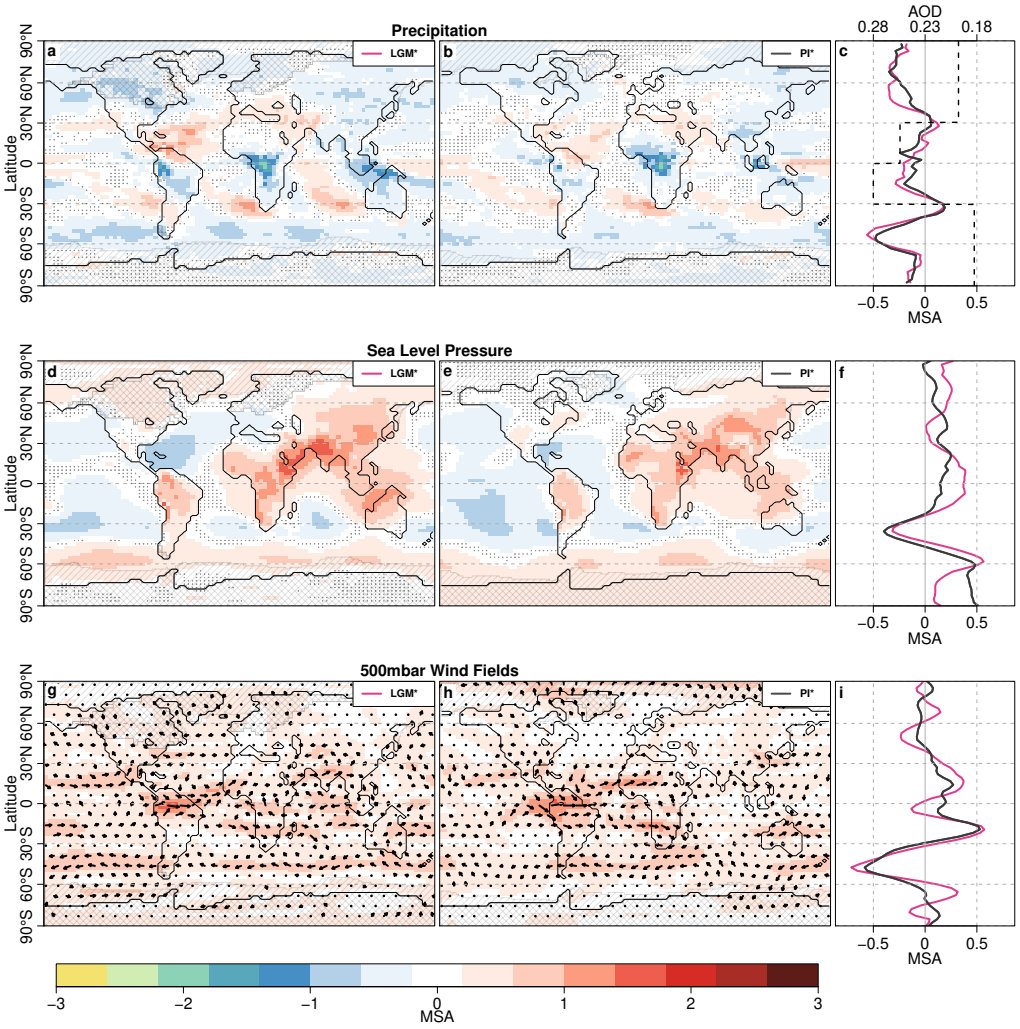


Figure S3. Mean standardized anomalies (MSA), as Figure 2 of the main manuscript, for precipitation rate (**a-c**), sea level pressure (**d-f**), and wind fields at 500mbar (**g-i**) from HadCM3. Dots indicate insignificant anomalies within the 99% quantile range of local variability. Grey shaded crosses and lines show land and sea ice, respectively. Mean anomalies over latitude intervals are displayed on the right-hand panels. The black dashed line shows the mean zonal Aerosol Optical Depth (AOD) imprint.

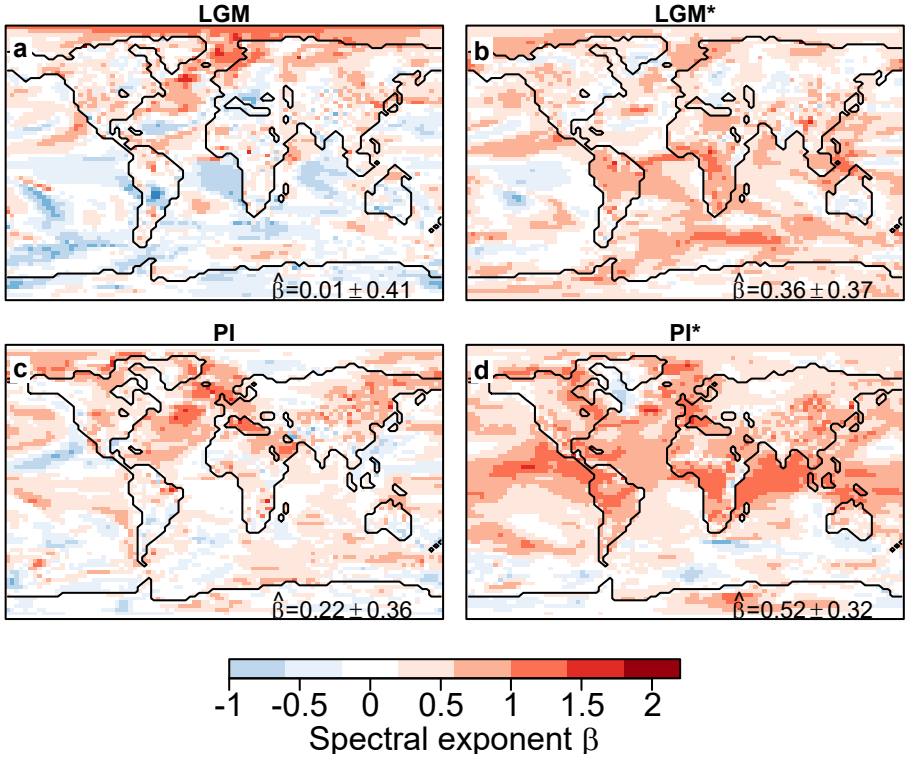


Figure S4. Scaling coefficient β of forced and unforced surface air temperature on the multidecadal-to-multicentennial scale (50-500yrs) as simulated by HadCM3 for the Last Glacial and Pre-Industrial. Surface air temperature variability was approximated by power-laws of the spectrum $S(\tau) \sim \tau^{\beta}$ with $50 \leq \tau \leq 500$ years and scaling coefficient β (Huybers & Curry, 2006; Fredriksen & Rypdal, 2017; Lovejoy & Varotsos, 2016). The area-weighted mean scaling coefficient is denoted by $\hat{\beta}$. Following Huybers and Curry (2006), we estimate β by linear regression after log-binning to prevent low-frequency biases.

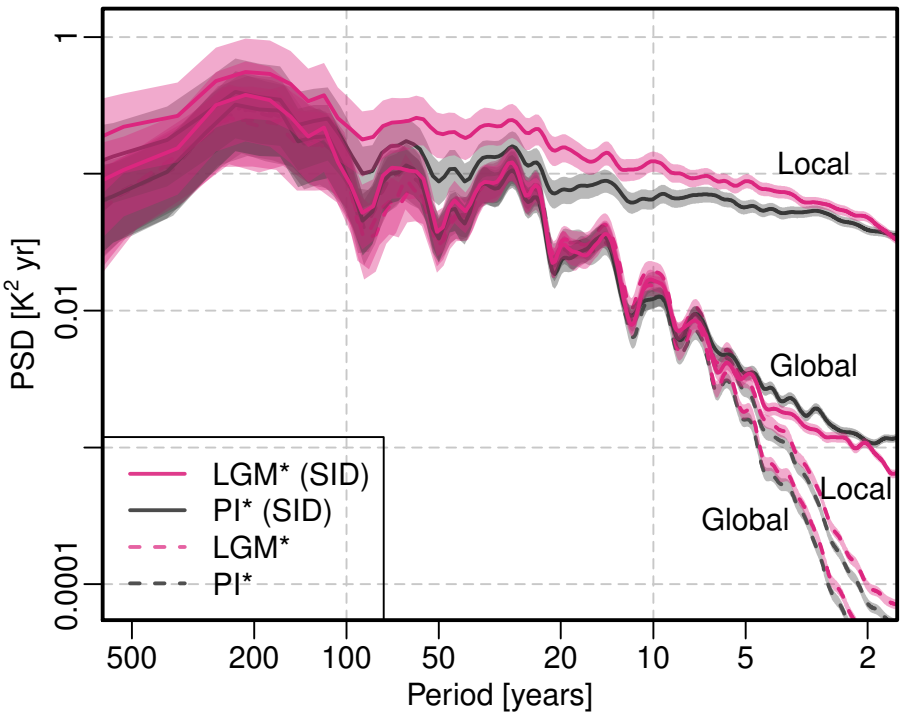


Figure S5. Global and local mean spectra of naturally forced surface air temperature as simulated by TransEBM with (solid lines) and without (dotted lines) time-varying sea ice dynamics (SID). TransEBM is a two-dimensional energy balance model with T42 resolution, as described by Ziegler and Rehfeld (2021) which draws on Zhuang et al. (2017). We run TransEBM with the same boundary conditions and time-varying forcing time series as the HadCM3 simulations, including the latitudinal-dependent volcanic forcing. Without loss of generality, we used the same constant CO2 values as in HadCM3 and neglected minor impacts from other greenhouse gases. The EBM is driven by yearly averaged solar and volcanic forcing. Dotted lines show the global and local mean spectra of the simulated temperature when sea ice extent is fixed. To mimic the sea ice dynamics in the two-dimensional model, we update the EBM's land-sea mask yearly based on the sea ice output from HadCM3 and repeat the simulations.

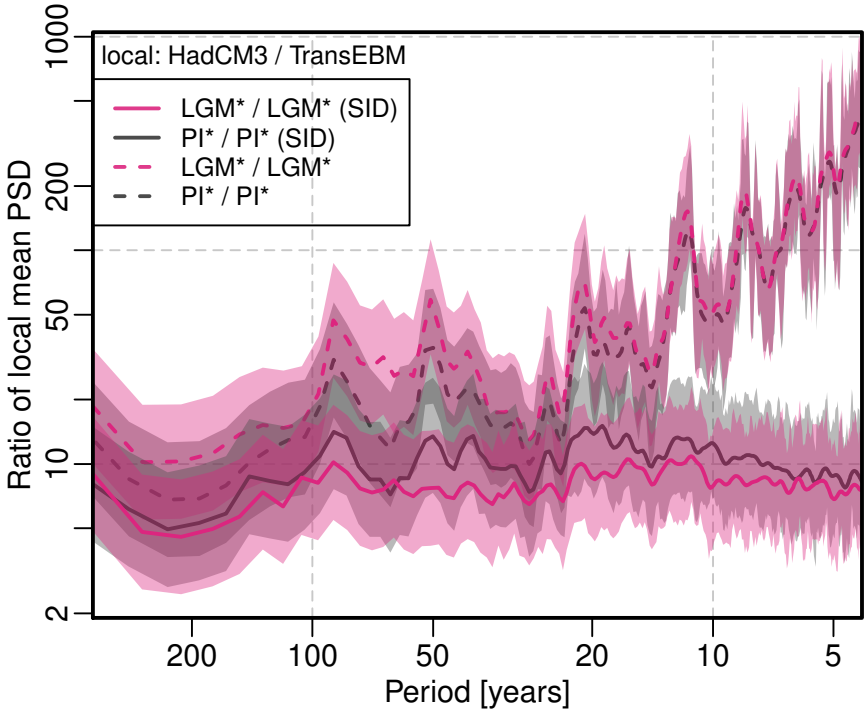


Figure S6. Ratio of HadCM3 (Figure 3) to TransEBM (Figure S5) local mean spectra (PSD) of naturally forced surface air temperature. LGM* / LGM* and PI* / PI* denote the ratios obtained from dividing the local mean spectrum of the naturally forced HadCM3 temperature by the one obtained from TransEBM with fixed sea ice. For the ratios of LGM* / LGM* (SID) and PI* / PI* (SID), time-varying sea ice dynamics in TransEBM was prescribed using the HadCM3 output. Hence, forming the ratio largely removes the linear response to naturally forcing and, for (SID), the contribution to variability from sea ice. The ratios therefore indicate the timescale-dependency of local variance simulated by HadCM3 that can be mainly attributed to internal dynamics excluding (solid lines) and including (dotted lines) sea ice dynamics. Shaded confidence intervals are computed from the F-distribution, based on the degrees of freedom of the spectral estimates.

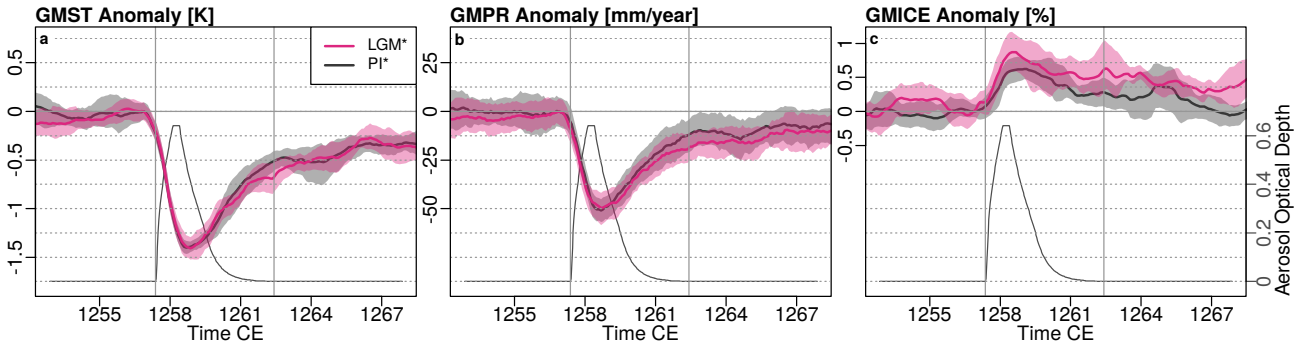


Figure S7. Superposed epoch analysis (see e.g., Sear et al. (1987); Brad Adams et al. (2003)) of globally averaged surface temperature (GMST, **a**), precipitation rate (GMPR, **b**), and sea ice concentration (GMICE, **c**) as simulated by HadCM3 using the reconstructed 1257 Samalas eruption. The lines represent the average value over all simulations in the LGM* and PI* state, and the shaded areas their respective ranges. The volcanic forcing from the Samalas eruption is shown in black. Anomalies are calculated against the three-year period before the eruption using the deseasonalized, detrended HadCM3 model output.

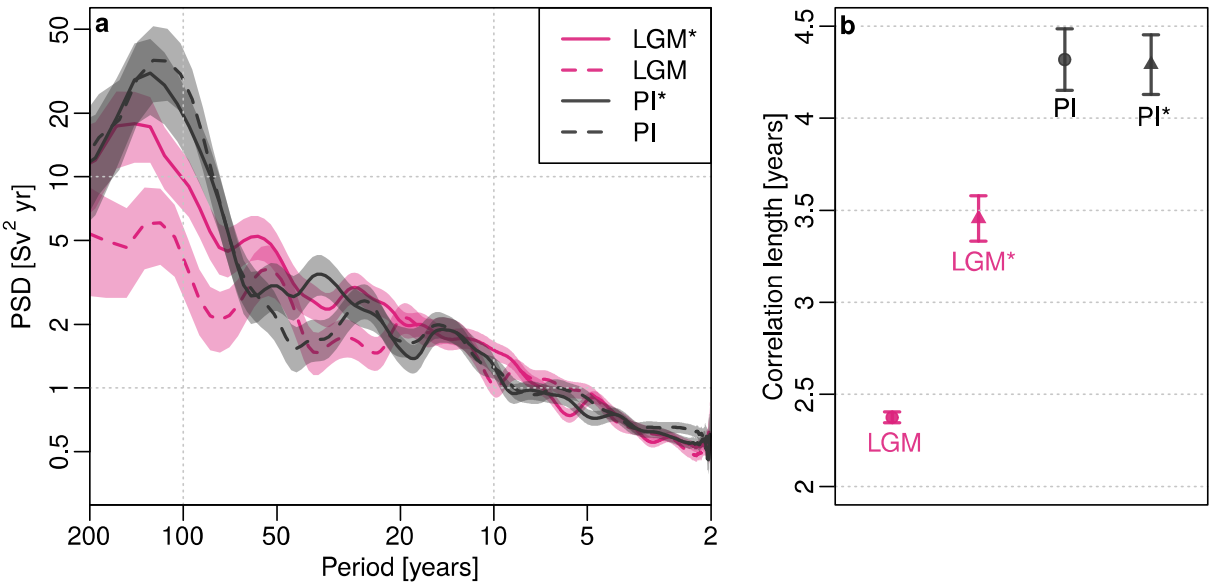


Figure S8. **a** Power spectral density (PSD) of the Atlantic Meridional Overturning Circulation (AMOC) strength from control and forced LGM and PI simulations using HadCM3. **b** Correlation length, defined as the lag at which the autocorrelation function first drops below 1/e and its standard error. Following Danabasoglu et al. (2012), we compute the AMOC strength as the maximum of the meridional ocean velocity field between 450 m and 2100 m depth and 20°N to 62.5°N at every timestep. Accordingly, the AMOC strength is given in Sv = 10^6 m^3 s^{-1}. The correlation length is an average of 3000 randomly sampled 100 year time slices of each state (1000 slices per run).

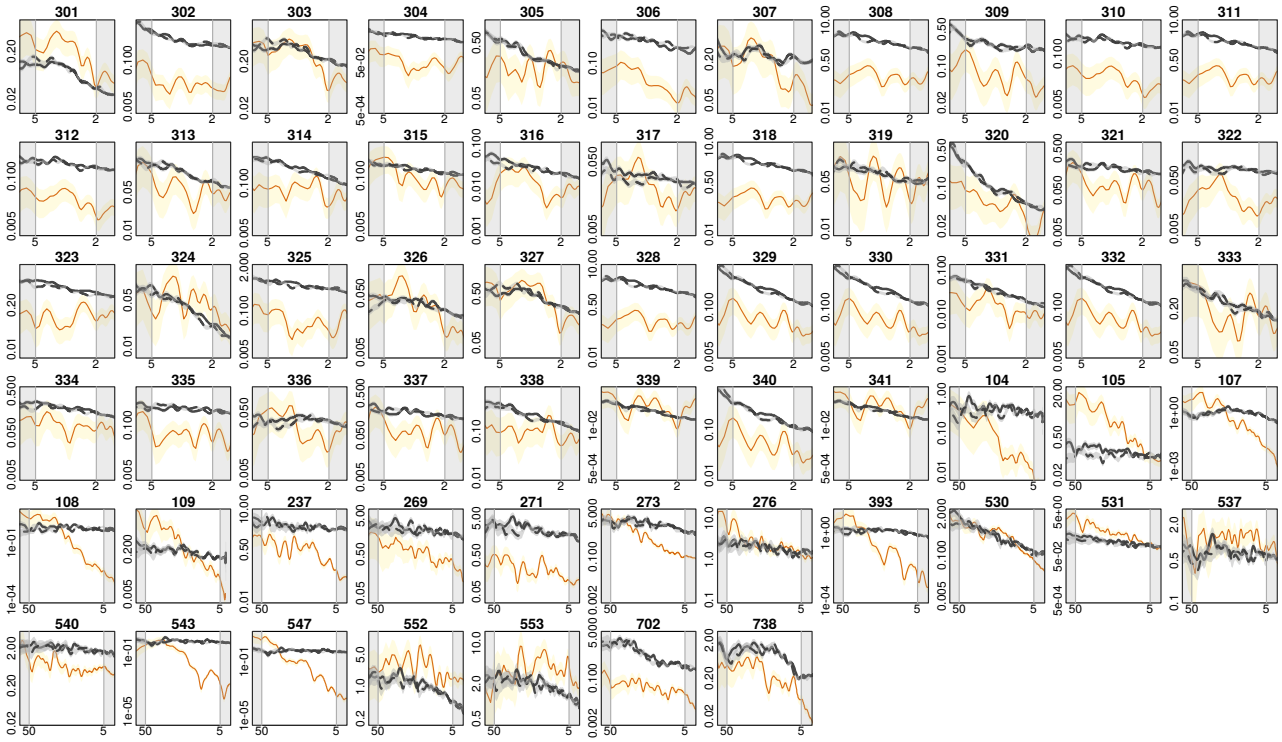


Figure S9. Temperature spectra from observations and proxy records (orange), and from HadCM3 simulations (grey) for the Pre-Industrial state, used for variance ratio estimates (Figure 4 of the main manuscript). The x-axis labels and background of each panel highlights the period (2-5 and 5-50 years) considered for timescale-dependent variance estimates. The y-axis denotes the power spectral density. Solid and dashed lines indicate forced and unforced runs. The title denotes the IDs from the separately uploaded dataset S1.

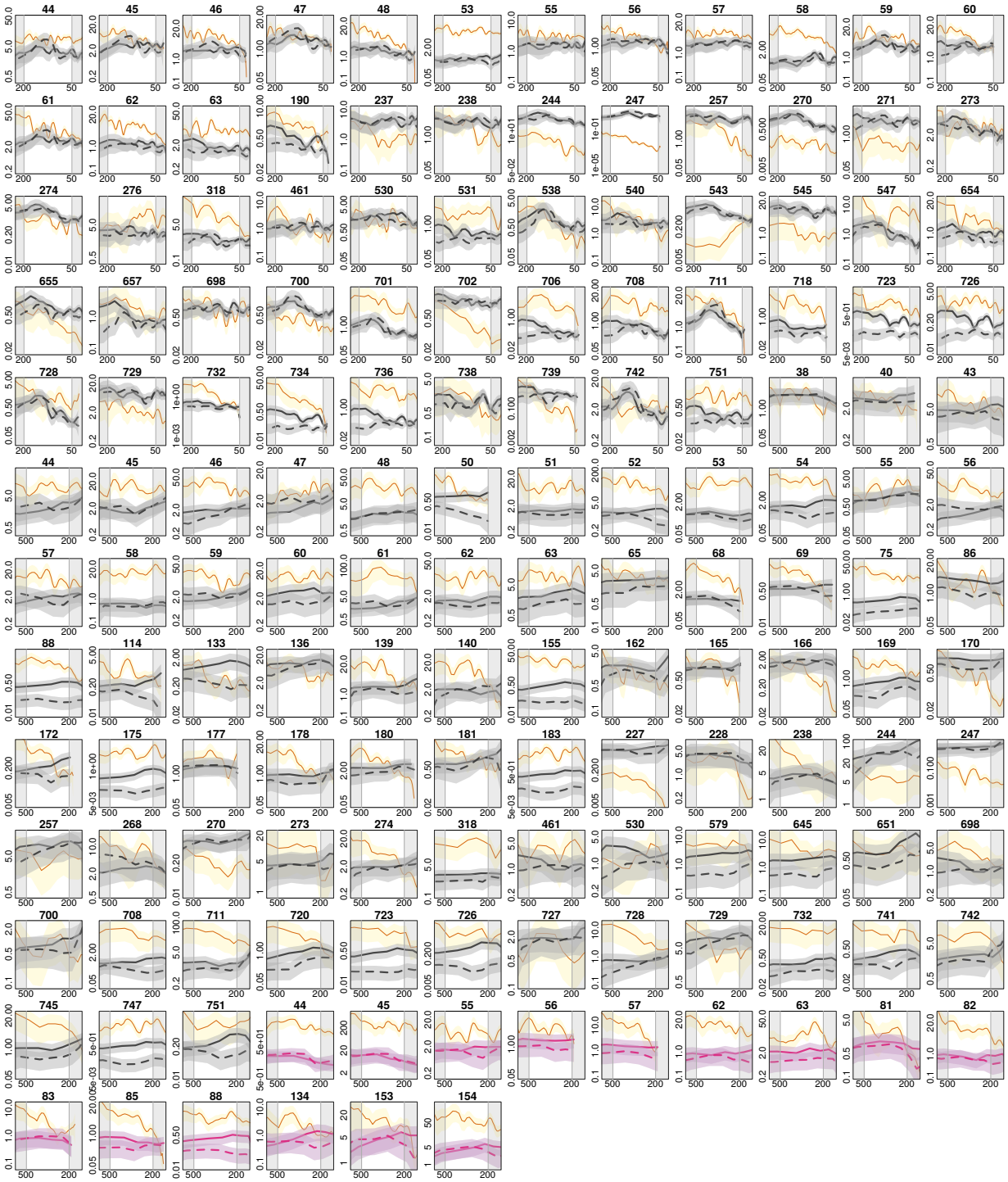


Figure S10. Same as Figure S9 but for multidecadal (50-200 years) and multicentennial (200-500 years) timescales. HadCM3 simulations under Last Glacial boundary conditions are shown in magenta.

Table S1. Boundary conditions of the HadCM3 simulation ensemble. Orbital parameters are internally calculated following (Berger, 1978) and Orography is taken from Singarayer and Valdes (2010) for 21 ka BP and 1850 CE. Greenhouse gas concentrations are taken from the protocols of the PMIP3 21ka and PI experiments (Schmidt et al., 2012). Vegetation is modeled with a 30-year timestep (Cox, 2001). Forced runs are driven by time-varying volcanic and solar (volc + sol) forcing as described in Table S2.

State	**Orography, Orb. Param.**	**CO_2, CH_4, N_2O**	**Forcing**	**#Runs**
Last Glacial Maximum (LGM)	21 ka BP	185ppm, 350ppb, 200ppb	–	3
Forced LGM (LGM*)	21 ka BP	185ppm, 350ppb, 200ppb	volc + sol	3
Pre-Industrial (PI)	1850 CE	280ppm, 650ppb, 270ppb	–	3
Forced PI (PI*)	1850 CE	280ppm, 650ppb, 270ppb	volc + sol	3

Table S2. Key specifications of the HadCM3 simulation ensemble. We provide the main references of the solar and volcanic forcing, whereby "-" indicates no forcing at all. The climate state defines the orography, orbital parameters, and greenhouse gas concentrations according to Table S1. The temporal resolution of each run is one month. The final column denotes the number of simulated years.

Name	Run ID	Climate state	Solar Forcing	Volcanic Forcing	Duration (yrs)
lgm1	xmzka	LGM	1365 W/m^2	–	1117
lgm2	xmzkd	LGM	1365 W/m^2	–	1066
lgm3	xmzki	LGM	1365 W/m^2	–	523
lgm4	xmzke	LGM*	Steinhilber et al. (2009); Wang, Lean, and Jr. (2005)	Crowley and Unterman (2013)	1064
lgm5	xmzkg	LGM*	Steinhilber et al. (2009); Wang et al. (2005)	Crowley and Unterman (2013)	1058
lgm6	xmzkh	LGM*	Steinhilber et al. (2009); Wang et al. (2005)	Crowley and Unterman (2013)	931
pi1	xnage	PI	1365 W/m^2	–	1055
pi2	xnagh	PI	1365 W/m^2	–	560
pi3	xnagb	PI	1365 W/m^2	–	1124
pi4	xnagd	PI*	Steinhilber et al. (2009); Wang et al. (2005)	Crowley and Unterman (2013)	1062
pi5	xnagf	PI*	Steinhilber et al. (2009); Wang et al. (2005)	Crowley and Unterman (2013)	1095
pi6	xnagg	PI*	Steinhilber et al. (2009); Wang et al. (2005)	Crowley and Unterman (2013)	1150

H Publication 3

M. Schillinger, B. Ellerhoff, R. Scheichl, and K. Rehfeld. "Separating internal and externally-forced contributions to global temperature variability using a Bayesian stochastic energy balance framework." *Chaos* 32, 113146. (2022). doi: 10.1063/5.0106123

Chaos ARTICLE scitation.org/journal/cha

Separating internal and externally forced contributions to global temperature variability using a Bayesian stochastic energy balance framework

Cite as: Chaos **32**, 113146 (2022); doi: 10.1063/5.0106123
Submitted: 27 June 2022 · Accepted: 1 November 2022 ·
Published Online: 29 November 2022

Maybritt Schillinger,[1] Beatrice Ellerhoff,[2,a)] Robert Scheichl,[3] and Kira Rehfeld[2]

AFFILIATIONS

[1]Seminar for Statistics, Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland
[2]Department of Physics and Department of Geosciences, Tübingen University, Schnarrenbergstr. 94-96, 72076 Tübingen, Germany
[3]Institute of Applied Mathematics and Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

Note: This article is part of the Focus Issue, Theory-informed and Data-driven Approaches to Advance Climate Sciences.
[a)]**Author to whom correspondence should be addressed:** beatrice-marie.ellerhoff@uni-tuebingen.de

ABSTRACT

Earth's temperature variability can be partitioned into internal and externally forced components. Yet, underlying mechanisms and their relative contributions remain insufficiently understood, especially on decadal to centennial timescales. Important reasons for this are difficulties in isolating internal and externally forced variability. Here, we provide a physically motivated emulation of global mean surface temperature (GMST) variability, which allows for the separation of internal and external variations. To this end, we introduce the "ClimBayes" software package, which infers climate parameters from a stochastic energy balance model (EBM) with a Bayesian approach. We apply our method to GMST data from temperature observations and 20 last millennium simulations from climate models of intermediate to high complexity. This yields the best estimates of the EBM's *forced* and *forced + internal* response, which we refer to as emulated variability. The timescale-dependent variance is obtained from spectral analysis. In particular, we contrast the emulated *forced* and *forced + internal* variance on interannual to centennial timescales with that of the GMST *target*. Our findings show that a stochastic EBM closely approximates the power spectrum and timescale-dependent variance of GMST as simulated by modern climate models. Small deviations at interannual timescales can be attributed to the simplified representation of internal variability and, in particular, the absence of (pseudo-)oscillatory modes in the stochastic EBM. Altogether, we demonstrate the potential of combining Bayesian inference with conceptual climate models to emulate statistics of climate variables across timescales.

Understanding the statistical properties and sources of the Earth's surface temperature variations is of great importance in climate science. To this end, we analyze the variability of global mean surface temperature (GMST) with a simple stochastic energy balance model (EBM). With Bayesian methods and spectral analysis, we separate internally generated and externally forced contributions to GMST variations on different timescales in state-of-the-art climate model simulations. Our results show that a stochastic EBM can emulate the variability of more complex climate models. The combined use of Bayesian inference and conceptual climate models, therefore, provides a versatile tool to advance the understanding of the internal and forced variability in the Earth's dynamical system.

I. INTRODUCTION

Climate variability describes the spatial and temporal variations in the mean and higher order statistics of climate parameters and

is of vital importance for living conditions on the Earth.[1] While many sources of natural variability exist, anthropogenic influences clearly dominate the recent trend in global mean surface temperature (GMST). To characterize variability, it is typically partitioned into internal and external components. Internal variability arises from intrinsic climate system processes such as oceanic and atmospheric circulation. External sources include changes in radiative forcing, for example, from solar irradiance, volcanic eruptions, and greenhouse gases. Despite a general agreement of the total simulated and observed GMST variability over the Common Era (0–2000 CE),[2,3] uncertainties remain about the mechanisms and magnitude of internal and external variations,[4–6] especially on decadal to centennial timescales.[3,7]

Simple mathematical models help understand climate variability[8–10] and can be used to emulate climate variables from more complex model simulations. Most general, the time evolution of a forced climate parameter $X(t)$ is described by $\dot{X}(t) = \mathscr{A}(t, X(t), F(t))$ for an arbitrary operator $\mathscr{A}$ and external driver $F(t)$. We consider X as the GMST, for which many studies have formulated physically motivated approximations of $\mathscr{A}$. One pivotal approach is centered around the idea of balancing incoming and outgoing radiations,[8,9] later extended to a stochastic energy balance model (EBM) by Hasselmann.[10] This approach assumes the climate system close to equilibrium, showing a linear and stationary response to perturbations. Then, $\mathscr{A}$ can be approximated by a linear stochastic operator[8,9]

$$C\frac{\mathrm{d}}{dt}T(t) = -\tilde{\lambda}T(t) + F(t) + \varepsilon(t). \tag{1}$$

Formula (1) describes the GMST anomaly $T(t)$ with respect to the equilibrium state, given the Earth's effective heat capacity C, a radiative forcing anomaly $F(t)$, and a term $\varepsilon(t)$, representing stochastic dynamics such as weather fluctuations. The response parameter $\lambda := \tilde{\lambda}/C$ is the reciprocal of the characteristic timescale $1/\lambda$. The response to radiative forcing $F(t)$ determines forced temperature variations. The response to the stochastic term $\varepsilon(t)$ approximates internal variability.

The stochastic EBM (1) is too simplistic to accurately represent long-term responses and, therefore, has been extended to the so-called multibox EBMs.[11–15] The latter are based on multiple ocean layers, referred to as boxes. The layers serve to approximate the vertical heat transfer and the integrated response to forcing over long periods. The EBM (1) laid the basis for attributing anthropogenic warming.[16,17] It was applied and modified to study climate sensitivity,[14,18–20] climate and ice cap stability,[21–24] regional temperatures,[25–27] glacial/interglacial cycles,[22,28] and future projections.[29,30] Key advantages of EBMs are their computational efficiency and comparatively easy interpretation.

To estimate uncertain parameters of conceptual climate models from data, Bayesian frameworks have become increasingly popular.[31–35] In comparison to other methods for inferring climate parameters, such as maximum likelihood estimation, Bayesian approaches have the advantage of providing full posterior distributions. The methods compute the posterior means and credible intervals (CIs) of uncertain parameters θ conditioned on target data y while including prior knowledge on θ. Central to this framework is applying the Bayes theorem

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{2}$$

with likelihood $p(y|\theta)$, prior $p(\theta)$, marginal $p(y)$, and posterior $p(\theta|y)$. We denote all probability densities by p and distinguish them by their arguments. Combining Bayesian inference with conceptual climate models typically also yields the posterior of the model's fit to the data.

With the ability to quantify fluctuations across timescales, power spectral analysis has improved the understanding of climate variability.[3,36–39] Simple climate models have been combined with spectral analysis to explain timescale-dependent variability. For example, Fredriksen and Rypdal[15] use a multibox EBM to study temporal scaling of temperature time series. Related works examine future projections[29] and climate sensitivity.[40] Soldatenko and Colman[41] study the sensitivity of the power spectrum on uncertainties in the parameters of a two-box EBM, considering stochastic noise but neglecting deterministic forcing. Yet, the potential of combining stochastic multibox EBMs, Bayesian inference and spectral analysis to study the magnitude of unforced and forced variability across timescales remains untapped.

Here, we examine and separate timescale-dependent internal and externally forced contributions to the GMST variations. In particular, we analyze GMST variability during the last millennium (850–1850 CE) as simulated by 20 climate models of intermediate to high complexity. To this end, we combine a stochastic two-box EBM (Sec. III A) with Bayesian inference (Sec. III B) and spectral analysis (Sec. III C). We present the "ClimBayes" software package[42] for Bayesian inference of climate parameters, which fits the stochastic EBM to GMST data. This results in the best estimate of the *forced* and samples of the *forced* + *internal* EBM's temperature response. First, we demonstrate our analysis on the example of historical observations (Sec. IV A) and then apply it to the considered set of last millennium simulations (Sec. IV B). Section IV C contrasts power spectra of the fitted EBM with and without internal variations. Comparing the internal and forced variance on interannual to centennial timescales (Sec. IV D), a stochastic two-box EBM captures most variations of more comprehensive model simulations. We summarize and discuss the potential for physics-informed emulation of GMST data and separation of variance contributions across timescales in Secs. V and VI.

II. DATA

Our study relies on annual GMST and corresponding radiative forcing time series. We use full-forced last millennium runs from climate models of varying complexity (Table I). We analyze 10 simulations with atmosphere-ocean general circulation models (AOGCMs), considered in the Coupled Model Intercomparison Project 5 (CMIP5).[43] Moreover, we use 10 simulations with Earth system models of intermediate complexity (EMICs) that are part of the IPCC's Fifth Assessment Report (AR5)[44] and described by Eby *et al.*[45] The AR5 EMICs represent single simulations, except for CLIMBER2 and LOVECLIM V.1.2, which are ensemble means and

TABLE I. Key forcing specifications and references of considered climate model simulations. The "Forcing" column gives the abbreviations from the PMIP3 protocol,[46] corresponding to the implemented land use, solar, and volcanic forcing reconstructions. The land use reconstruction PEA is taken from Pongratz *et al.*[49] Solar forcing reconstructions correspond to DB: Delaygue and Bard,[53] VSK: Krivova, Balmaceda, and Solanki,[54] Vieira and Solanki,[55] and SBF: Steinhilber, Beer, and Fröhlich.[51] They are calibrated to WLS modern values (1366.14 W/m^2) and continued by Wang, Lean, and Sheeley.[52] Volcanic forcing refers to CEA: Crowley *et al.*[50] and GRA: Gao, Robock, and Ammann.[56] Trace gases are prescribed in all simulations and follow the PMIP3 protocol.[46]

Climate model	Forcing	Reference
AR5 EMICs		
Bern 3D	PEA, DB, CEA	57
CLIMBER-3alpha	PEA, DB, CEA	58
CLIMBER2	PEA, DB, CEA	59
DCESS ESM v1	PEA, DB, CEA	60
IGSM 2.2	PEA, DB, CEA	61
LOVECLIM V.1.2	PEA, DB, CEA	62
MESMO 1.0	PEA, DB, CEA	63
MIROC3-lite	PEA, DB, CEA	47
UMD	PEA, DB, CEA	64
UVic v2.9	PEA, DB, CEA	65
CMIP5 models		
BBC-CSM1-1	–, VSK, GRA	66
CCSM4	PEA, VSK, GRA	67
CSIRO-Mk3L-1-2	–, SBF, CEA	68
FGOALS-s2	–, VSK, GRA	69
GISS-E2-R	PEA, SBF, CEA	70
HadCM3	PEA, SBF, CEA	71
HadGEM2-ES	PEA, SBF, CEA	72
IPSL-CM5A-LR	–, VSK, GRA	73 and 74
MIROC-ESM	–, DB, CEA	75
MPI-ESM-P	PEA, VSK, CEA	76 and 77

denoted by "(mean)" in the following. To compare variability in single ensemble members to that of the ensemble mean, we use the five available ensemble members LOVECLIM V.1.2 (E1–E5).

The transient radiative forcing applied to these simulations follows the Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) protocol.[46] For AR5 EMICs, we take the total estimated radiative forcing provided by Eby *et al.*[45] For CMIP5 simulations, we use the radiative forcing from reconstructions of well-mixed greenhouse gases (CO_2, CH_4, and N_2O), volcanic aerosols, total solar irradiance, and land use changes as provided by Schmidt *et al.*[46] We neglect orbital forcing, which is assumed to play a negligible role for GMST variability over the last millennium. To remove potential unforced drifts of the simulated background climate,[47] the simulated GMST is linearly detrended prior to analysis. For consistency, detrending is also applied to the corresponding forcing time series. This does not affect our results, as the simulations' forcing input exhibits no trend for the last millennium (850–1850 CE). Both temperature and forcing time series are considered as anomalies with respect to the starting year.

We use the GMST from HadCRUT5[48] observations (1850–2000 CE) to demonstrate the developed workflow of our Bayesian stochastic energy balance framework. As estimates for radiative forcing during the historical period, we consider the "PEA" land use (Pongratz *et al.*[49]), "CEA" volcanic forcing (Crowley *et al.*[50]), "SBF" solar irradiance reconstruction (Steinhilber *et al.*[51]) patched into Wang *et al.*,[52] and greenhouse gas concentrations from Schmidt *et al.*[46]

III. METHODS

Our analysis combines stochastic multibox EBMs, Bayesian inference, and spectral analysis. We introduce the approach implemented in "ClimBayes"[42] for the most generic case of a stochastic EBM with N boxes in Sec. III A. All results are obtained from the special case $N = 2$.

A. Stochastic two-box energy balance model (EBM)

The stochastic multibox EBM[13,15,29] extends the one-dimensional linear operator from Eq. (1) by multiple vertical layers, approximating the heat exchange between surface and deep ocean layers. In matrix notation, the model reads[15]

$$\boldsymbol{C}\frac{d\boldsymbol{T}}{dt}(t) = \boldsymbol{K}\boldsymbol{T}(t) + \boldsymbol{F}(t) + \boldsymbol{\varepsilon}(t). \tag{3}$$

For N boxes, $\boldsymbol{T}(t)$ is an N-dimensional vector, describing the temperature of each box. By convention, T_1 corresponds to the temperature of the uppermost and T_N to the temperature of the lowermost box. Accordingly, $\boldsymbol{C}$ is a diagonal matrix with the effective heat capacity C_{ii} of each layer ($i = 1, \ldots, N$). $\boldsymbol{K}$ is a N-dimensional tridiagonal matrix, parameterizing the surface temperature response and vertical heat transfer (Appendix A). The time-dependent radiative forcing $\boldsymbol{F}(t)$ is only applied to the uppermost box, such that $F_1 = F(t)$ and $F_k = 0$ for $k = 2, \ldots, N$. The stochastic forcing $\boldsymbol{\varepsilon}(t)$ is likewise implemented with non-zero entry $\varepsilon_1(t) = \sigma_W \xi(t)$. We motivate the white noise process $\xi(t)$ with standard deviation (SD) σ_W by the found impact of weather fluctuations.[10,14,78]

Integrating Eq. (3) yields the solution of the surface temperature $T_1(t)$, given by a sum of the forced response $T_{1,F}(t)$ and internal variations $T_{1,I}(t)$,

$$\begin{aligned} T_1(t) &= T_{1,F}(t) + T_{1,I}(t) \\ &= \int_{-\infty}^{t} R(t-s)\frac{1}{C_1}F(s)\,ds + \int_{-\infty}^{t} R(t-s)\frac{\sigma_W}{C_1}\,dW(s), \end{aligned} \tag{4}$$

where $C_1 = C_{11}$ is the heat capacity of the uppermost box. This assumes no interaction between forced and internal variability on a global scale. The response function

$$R(t) = \sum_{k=1}^{N} w_k\, e^{-\lambda_k t} \tag{5}$$

is uniquely defined[13,15] by the response parameters λ_k and weights w_k ($k = 1, \ldots, N$) with $\sum_{k=1}^{N} w_k = 1$, which depend on the entries of $\boldsymbol{C}$ and $\boldsymbol{K}$ (Appendix A). The internal fluctuations $T_{1,I}(t)$ in formula (4) represent an Itô-integral over the Wiener process $W(s)$. Therefore, $T_{1,I}(t)$ can be written as a weighted sum of Ornstein–Uhlenbeck

(OU) processes, where the kth OU process solves the stochastic differential equation $dU(t) = -\lambda_k U(t)\,dt + \frac{\sigma_W}{C_1} dW(t)$ and receives weight w_k. Accordingly, $T_{1,I}(t)$ is normally distributed with mean zero. Its covariance matrix is determined by σ_W, C_1, w_k, and λ_k (Appendix B).

B. Bayesian inference algorithm

1. Joint emulation of forced and internal variations

To separate the internal and forced contributions to the GMST variance, we introduce the Bayesian inference algorithm implemented in "ClimBayes." We fit the linear stochastic two-box EBM to GMST data (Table I), as illustrated in Fig. 1. The Bayesian inference algorithm relies on annually resolved temperature and forcing time series as input data. Moreover, it requires physics-informed prior information on $\theta = (\lambda_1, \lambda_2, w_1, C_1, T_0, F_0)$. The parameters λ_1, λ_2, and w_1 correspond to free parameters of the response function in Eq. (5). C_1 is the heat capacity of the upper ocean box. T_0 and F_0 are initial free parameters. T_0 allows for small deviations of the EBM solution to the input data in the starting year and is expected to be close to zero. Similarly, the additional parameter F_0 is needed to compensate for an initial forcing anomaly with respect to the equilibrium state.

We infer the posterior distributions of the uncertain parameters θ conditioned on target data y via Bayes theorem (2), using a Markov chain Monte Carlo (MCMC) algorithm. To this end, we assume that the target data can be described by a deterministic model Φ_F and stochastic measurement or intrinsic noise Z that is also allowed to depend on θ,

$$y = \Phi_F(\theta) + Z(\theta). \tag{6}$$

Formula (6) yields the likelihood $p(y|\theta)$. Combined with prior information $p(\theta)$, Bayes theorem (2) defines the posterior distribution $p(\theta|y)$. In our case, the deterministic model $\Phi_F(\theta)$ is given by a discretization of the temperature responses $T_{1,F}(t)$. The noise term $Z(\theta)$ corresponds to the internal fluctuations $T_{1,I}(t)$.

This approach provides a joint estimate of the internal and externally forced response, based on the same physics-informed response function and parameters. The best estimates of θ and the forced response $T_{1,F}(t)$ are defined as their posterior means $E[\theta|y]$ and $E[\Phi_F(\theta)|y]$. The SD σ_I of the internal variability $Z(\theta)$ determines σ_W (Appendix B) and is approximated by the residuals' SD, that is, the difference between the target data and the forced response. Samples of the internal variations $T_{1,I}(t)$ can be drawn from its covariance matrix, using the best estimates of θ (Appendix B). The *forced* + *internal* variations $T_{1,F}(t) + T_{1,I}(t)$ represent the full response of the stochastic two-box EBM and, thus, provide a model for the *target* variability. To streamline our discussion, we will refer to the modeled *forced* and *forced* + *internal* response as emulation.

2. Numerical implementation

The "ClimBayes" package provides the numerical implementation of our approach (Fig. 1) and allows for straightforward adjustments via a configuration file. Most importantly, this includes specifications of the number of boxes, prior distributions, MCMC

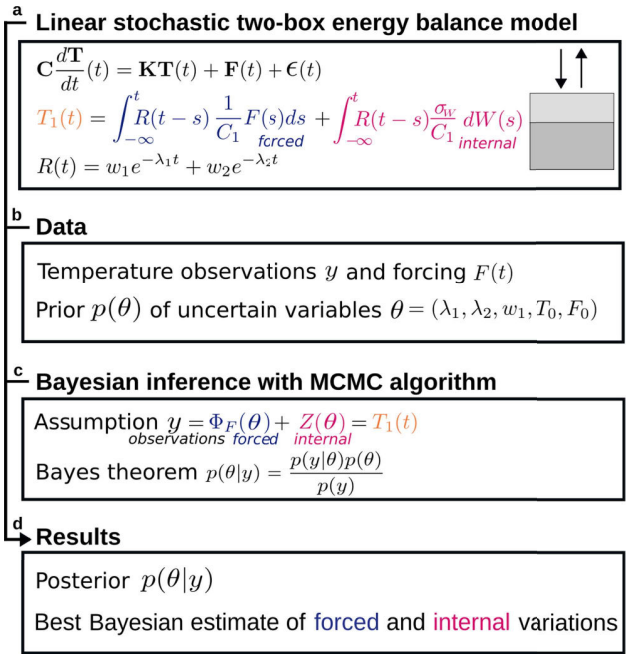


FIG. 1. Workflow of our Bayesian inference algorithm to emulate forced and internal GMST fluctuations. (a) The workflow builds on a linear stochastic two-box EBM, here in matrix notation. The solution for the surface temperature $T_1(t)$ is given by an integral with exponential response function $R(t)$. (b) Required input data include annually resolved GMST and time-dependent forcing, as well as physics-informed prior information $p(\theta)$ on uncertain parameters θ. (c) We infer the parameters of the two-box EBM using a Markov chain Monte Carlo (MCMC) algorithm and Bayes theorem, assuming that the target GMST can be described as a deterministic model $\Phi_F(\theta)$ and stochastic noise $Z(\theta)$. (d) The workflow yields posterior distributions $p(\theta|y)$ of uncertain parameters conditioned on temperature data y and a physically motivated emulation of the forced and internal variations.

sampling properties, and fixed parameters. We choose $N = 2$ boxes (Appendix C) in line with Held *et al.*[12] and Geoffroy *et al.*[13] Our experiments use independent prior distributions. We choose beta distributions with shape parameters $\alpha = \beta = 2$ for the marginal priors of λ_1, λ_2, as well as the initial parameters T_0 and F_0. This allows for a physics-informed, fixed parameter range, where the mean is preferred over boundary values. The algorithm's convergence is improved compared to uniform priors. The intervals for the response parameters $\lambda_1 : (0.2, 2)\,\mathrm{yr}^{-1}$ and $\lambda_2 : (0.005, 0.2)\,\mathrm{yr}^{-1}$ are similar to those from Fredriksen and Rypdal.[15] Tailored to our goal to emulate interannual to centennial variability from last millennium simulations, our choice of priors assumes characteristic response times $1/\lambda_k$ smaller than 200 years. These response times implicitly set a characteristic depth scale for the two ocean boxes of our stochastic EBM.

The prior of $C_1 : (4, 11)\ \mathrm{W\,yr\,m^{-2}\,K^{-1}}$ follows previous findings.[13,15,79] Initial values $T_0 : (-0.5, 0.5)\ \mathrm{K}$ and $F_0 : (-2, 2)\ \mathrm{W\,m^{-2}}$ are centered around zero. The weight $w_1 : (0, 1)$ has an uniform prior. We consider no measurement noise, which is inherently

fulfilled for simulated GMST. For observed GMST, we assume measurement errors to be small compared to internal fluctuations.[48] We verified that our findings are robust against reasonable variations of the MCMC and prior specifications.

To obtain the best parameter estimates, "ClimBayes" uses a Metropolis Hastings (MH) algorithm from the family of MCMC methods. To this end, we discretize the forward operator and compute annual temperature anomalies relative to the starting point $t = 0$ by a midpoint rule[29]

$$T_{1,F}(t_m) = T_0 + \frac{1}{C_1}\sum_{k=1}^{N} w_k \sum_{j=1}^{m} e^{-(m-j+1/2)\lambda_k}(F(t_j) + F_0). \quad (7)$$

Here, the time step t_j corresponds to the jth year, and $F(t_j)$ are forcing anomalies with respect to the starting point $t = 0$. F_0 represents an initial forcing anomaly. Discretizing $T_{1,I}$ leads to a normally distributed weighted sum of AR(1)-processes with covariance matrix entries depending on λ_k and w_k (Appendix B). The likelihood $p(y|\theta)$ is given by the normal distribution of $T_{1,F}(t) + T_{1,I}(t)$. This requires the calculation of the covariance matrix for each sample in the Markov chain. For numerical robustness and computational efficiency, however, we approximate the likelihood function using an iterative scheme (Appendix B).

The MCMC algorithm uses four chains with 25 000 samples each, from which the first 5000 are discarded as burn-in. The proposal distribution is initially set to a normal distribution with mean zero and variances $(0.2, 2, 1, 10, 1, 2) \times 10^{-5}$ for $\theta = (\lambda_1, \lambda_2, w_1, C_1, T_0, F_0)$. After 2500 samples, the proposals are distributed according to the weighted sum of the initial normal proposal distribution and the empirical covariance matrix of previous samples.

We check the convergence of the Markov chains following two performance measures: First, the Gelman–Rubin diagnostics[80,81] compares the inter-chain and between-chain variances. It is ≤ 1.1 for most models, complying to recommendations.[82,83] Second, we use the Monte Carlo standard error,[84] which constructs an asymptotic confidence interval for the posterior mean.[84,85] In our experiments, the half-width of this interval is smaller than 5% of the prior mean. We found that this criterion guarantees robustness of the parameter estimates when the same run is repeated multiple times or additional samples are added. DCESS ESM v1 is the only outlier, showing a tendency for bimodal posterior densities which lead to less stable estimates with wide CIs. However, we confirmed that convergence can be achieved by fixing the heat capacity to the estimated parameter.

C. Spectral analysis and variance ratios

Given a temperature time series $T(t)$, the power spectral density (PSD) at frequency f corresponds to the Fourier transform of the autocovariance

$$S(f) = \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi i f k} E\left[(T(t) - \mu)\left(T(t+k) - \mu\right)\right] dk, \quad (8)$$

with lag $k = t_2 - t_1$ and mean $\mu := E[T(t)]$. This assumes $X(t)$ to be weakly stationary, which is reasonably fulfilled after linear detrending the GMST data. Following Ellerhoff and Rehfeld,[3] we use the multitaper method with three windows to compute the PSD and χ^2-distributed uncertainties. Mean spectra are obtained after interpolation to the lowest resolution and binning into equally spaced log-frequency intervals.[86] The spectra are visualized over periods $\tau = 1/f$ and logarithmically smoothed using a Gaussian kernel of 0.04 dB.

We compare the PSD for three types of time series: (1) the *target* temperature data from historical observations or climate model simulations, AR5 EMIC and CMIP5 models, (2) the emulated *forced* variations $T_{1,F}(t)$, and (3) the emulated *forced* + *internal* temperature variations $T_{1,F}(t) + T_{1,I}(t)$ from the stochastic two-box EBM. To compute the *forced* + *internal* PSD, we first sample 1000 realizations of the internal response $T_{1,I}(t)$. We add these to $T_{1,F}(t)$ and compute the PSD for all samples. Mean spectra and 95% confidence bands are obtained from this ensemble.

Variance ratios are calculated by dividing the emulated by the *target* variance. Following Parseval's theorem, we determine the timescale-dependent variance by integrating the PSD over frequencies. We consider frequency bands corresponding to interannual (2–5 yr), decadal (5–20 yr), multidecadal (20–50 yr), and centennial (50–200 yr) scales.

IV. RESULTS

A. Example application to historical observations

We demonstrate the application of the Bayesian inference algorithm and our spectral analysis on the example of GMST observations from HadCRUT5.[48] Figure 2(a) shows the forcing and temperature time series together with the best estimate of the forced response $T_{1,F}(t)$. The forced response follows the global warming trend and shows cooling periods after volcanic eruptions. Credible intervals (CIs) capture the uncertainties of the forced response, but not those due to internal variability. As a result, observations partly lie outside CIs. Uncertainties of the two-box forced response are largest at the time series' start.

Figure 2(b) shows marginal prior and posterior distributions for the free parameters of the stochastic two-box EBM. The response parameters $\lambda_1 = 1.31\ (0.71, 1.86)\ \mathrm{yr}^{-1}$ and $\lambda_2 = 0.09\ (0.03, 0.16)\ \mathrm{yr}^{-1}$ (Table III) correspond to timescales of approximately 10 months and 10 years. The corresponding heat capacity C_1 of the upper ocean layer is 8.20 (5.68, 10.43) W yr m^{-2} K^{-1}. The initial values T_0 and F_0 are well constrained and close to zero. The weight $w_1 = 0.72\ (0.46, 0.90)$ tends to emphasize the fast response.

The power spectral density [Fig. 2(c)] of the *forced* response alone is smaller than that of the *target* temperature. Conversely, the magnitude of the emulated *forced* + *internal* PSD agrees with the *target* PSD within uncertainties, except for the interannual scale. While the emulated PSD constantly increases from interannual to multidecadal scales, HadCRUT5 shows a modulation with increased power on periods of two to ten years, which is not captured by the emulated response.

Figure 2(d) compares the variance on interannual to multidecadal timescales. The variance ratios are formed by dividing the emulated *forced* or *forced* + *internal* variance by the variance obtained from the HadCRUT5 *target*. The *forced* variance is smaller than the *target* variance on all timescales. Incorporating internal

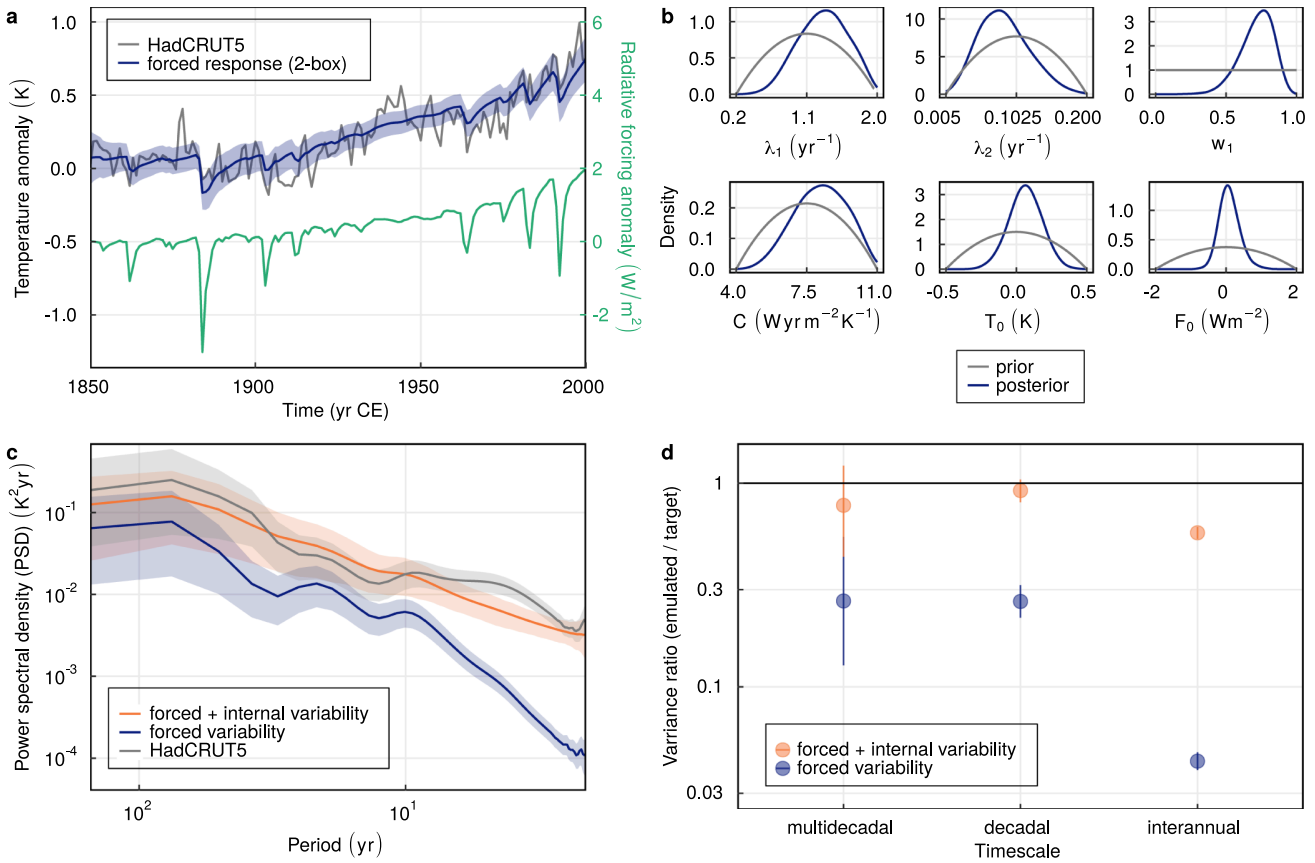


FIG. 2. Example application of our developed approach to historical GMST observations from HadCRUT5. (a) *Target* data y, forcing $F(t)$ (taken from Schmidt *et al.*,[46] Pongratz *et al.*,[49] Crowley *et al.*,[50] Steinhilber *et al.*,[51] Wang *et al.*[52]), and the best estimate of the *forced* response $T_{1,F}(t)$, that is the posterior mean $E[\Phi_F(\theta)|y]$. Shaded areas correspond to 95% CIs of $T_{1,F}(t)$. (b) Marginal prior and posterior densities for uncertain parameters θ. (c) PSD of the GMST observations, the *forced* response (both with χ^2-distributed confidence bands) and the sampled *forced + internal* variations. The sampled *forced + internal* PSD represents the mean and 95% confidence bands obtained from an ensemble of $T_{1,F}(t) + T_{1,I}(t)$ using 1000 realizations of $T_{1,I}(t)$. (d) Ratios of the emulated to observed GMST variance, computed by integration of the PSD on the multidecadal (20–100 yr), decadal (5–20 yr), and interannual (2–5 yr) scales. Uncertainties (95% CI) are calculated from a F-distribution based on the degrees of freedom of the variance estimate.

variability reduces this mismatch strongly, yet, is not enough to capture all fluctuations on interannual and multidecadal scales. On decadal scales, the emulated *forced + internal* variability agrees well with the observations.

B. Parameters estimated from last millennium simulations

We use Bayesian inference to fit the stochastic two-box EBM to GMST simulations from CMIP5 models and AR5 EMICs (Sec. II). Table II shows the best estimates that are the posterior means and 95% CIs of $\theta = (\lambda_1, \lambda_2, w_1, C_1, T_0, F_0)$ as well as the SD σ_I of the internal variability $T_{1,I}(t)$. Across all simulations, the short-term response parameter λ_1 varies between 1.91 and 0.31 yr^{-1}, spanning the full prior range between 6 months and 5 years. The long-term response λ_2 varies between 0.19 and 0.01 yr^{-1}, corresponding to characteristic timescales of approximately 5–100 years.

Differences between AR5 EMICs and CMIP5 models are most pronounced for λ_1. CMIP5 simulations exhibit larger CIs and inter-model differences, while AR5 EMICs show similar λ_1, except for CLIMBER 2 (mean) and LOVECLIM (E1 and mean). The weight $w_1 > 0.5$ is typically larger than $w_2 = 1 - w_1$, emphasizing the relative importance of the fast compared to the slow response. w_1 often is closer to unity for AR5 EMICs than for CMIP5 models. CLIMBER2 (mean) shows exceptionally large λ_1 and small w_1. The heat capacity C_1 varies from 4.7 to 10.96 $\mathrm{W\,yr\,m^{-1}\,K^{-1}}$. It spans the full prior range for both AR5 EMICs as well as CMIP5 models and shows no strong dependence on other parameters. The initial temperature T_0 is well constrained and close to zero. Inter-model differences are also large for F_0 and linked to varying temperature amplitudes at the beginning of the time series with respect to the mean (Fig. 7). The

TABLE II. Posterior means and 95% CIs of uncertain parameters θ for the stochastic two-box EBM fitted to GMST from climate model simulations. The SD σ_I of the internal variability equals the residual of the forced response (Appendix B) and is, therefore, given without CIs. Ensemble means are denoted by "(mean)." We exemplary show the LOVECLIM ensemble member "E1" as there are no major differences across the ensemble. Models with GRA[56] volcanic forcing are marked with an asterisk (*), all others use the CEA reconstruction[50] (Table I). The weight $w_2 = 1 - w_1$ (not shown) is uniquely defined by w_1.

Climate model	λ_1 (yr^{-1})	λ_2 (yr^{-1})	w_1 (unitless)	C_1 ($W\,yr\,m^{-2}\,K^{-1}$)	T_0 (K)	F_0 (Wm^{-2})	σ_I (K)
AR5 EMICs							
Bern 3D	0.74 (0.62,0.81)	0.08 (0.06,0.15)	0.91 (0.88,0.95)	4.76 (4.54, 5.43)	−0.09 (−0.50,0.04)	0.18 (−0.15, 1.48)	0.02
CLIMBER-3alpha	0.76 (0.72,0.80)	0.06 (0.05,0.07)	0.91 (0.90,0.93)	6.18 (5.99, 6.35)	−0.02 (−0.05,0.02)	−0.04 (−0.13, 0.04)	0.01
CLIMBER2 (mean)	1.43 (0.25,1.97)	0.19 (0.17,0.20)	0.05 (0.00,0.19)	10.96 (10.89,11.00)	0.01 (−0.11,0.13)	−0.19 (−0.45, 0.07)	0.05
DCESS ESM v1	0.31 (0.21,1.18)	0.08 (0.03,0.18)	0.64 (0.00,0.92)	10.25 (9.95,10.52)	0.02 (−0.05,0.08)	−0.18 (−0.27,−0.05)	0.03
IGSM 2.2	0.64 (0.57,0.73)	0.10 (0.07,0.15)	0.89 (0.82,0.93)	6.68 (6.38, 6.98)	0.01 (−0.08,0.11)	−0.10 (−0.36, 0.15)	0.03
LOVECLIM (E1)	1.32 (0.99,1.69)	0.13 (0.07,0.18)	0.90 (0.85,0.94)	6.03 (4.95,7.29)	−0.01 (−0.22,0.19)	0.28 (−0.58,1.21)	0.09
LOVECLIM (mean)	1.12 (0.93,1.34)	0.14 (0.09,0.19)	0.88 (0.84,0.92)	6.74 (6.06,7.42)	0.04 (−0.08,0.16)	−0.50 (−1.03,0.01)	0.04
MESMO 1.0	0.52 (0.49,0.56)	0.03 (0.02,0.04)	0.90 (0.88,0.92)	7.96 (7.72, 8.19)	0.02 (−0.03,0.06)	−0.32 (−0.39,−0.24)	0.02
MIROC-lite	0.44 (0.41,0.47)	0.02 (0.01,0.04)	0.96 (0.94,0.97)	6.75 (6.56, 6.96)	0.01 (−0.05,0.08)	−0.34 (−0.49,−0.25)	0.03
UMD	0.72 (0.68,0.77)	0.01 (0.01,0.01)	0.99 (0.99,0.99)	8.81 (8.46, 9.15)	0.00 (−0.04,0.02)	−0.09 (−0.21, 0.08)	0.02
UVic v2.9	0.45 (0.41,0.51)	0.05 (0.03,0.08)	0.88 (0.81,0.93)	10.88 (10.68,10.98)	−0.01 (−0.06,0.05)	−0.05 (−0.19, 0.09)	0.02
CMIP5 models							
BCC-CSM1-1*	1.91 (1.76,1.99)	0.06 (0.04,0.10)	0.94 (0.91,0.96)	10.49 (9.78,10.92)	−0.05 (−0.13,0.06)	1.31 (0.53, 1.87)	0.10
CCSM4*	1.62 (1.27,1.90)	0.12 (0.08,0.16)	0.88 (0.83,0.92)	4.70 (4.12, 5.60)	−0.06 (−0.34,0.18)	0.17 (−0.60, 1.05)	0.14
CSIRO-Mk2L-1-2	0.48 (0.29,1.12)	0.15 (0.07,0.19)	0.65 (0.31,0.96)	10.49 (9.78,10.92)	0.02 (−0.17,0.21)	−0.23 (−0.76, 0.30)	0.08
FGOALS-s2*	1.49 (1.13,1.88)	0.03 (0.01,0.10)	0.94 (0.89,0.97)	8.41 (6.71,10.38)	−0.05 (−0.20,0.11)	−0.58 (−1.42,−0.06)	0.14
GISS-E2-R	0.56 (0.50,0.65)	0.03 (0.01,0.15)	0.99 (0.96,1.00)	6.53 (5.94, 7.07)	0.12 (−0.43,0.32)	−1.19 (−1.77, 0.48)	0.09
HadCM3	0.49 (0.34,0.73)	0.13 (0.05,0.19)	0.76 (0.52,0.96)	7.99 (7.15, 8.86)	−0.06 (−0.31,0.20)	0.89 (0.27, 1.48)	0.12
HadGEM2-ES	0.64 (0.39,1.28)	0.14 (0.05,0.19)	0.74 (0.53,0.95)	7.79 (6.52, 8.81)	−0.06 (−0.30,0.19)	0.97 (0.36, 1.58)	0.12
IPSL-CM5A-LR*	1.69 (1.36,1.95)	0.09 (0.06,0.14)	0.85 (0.78,0.91)	9.49 (7.89,10.73)	−0.16 (−0.34,0.00)	−0.99 (−1.77,−0.14)	0.15
MIROC-ESM	1.49 (1.04,1.90)	0.14 (0.09,0.19)	0.71 (0.60,0.81)	9.62 (8.01,10.78)	−0.02 (−0.25,0.22)	0.10 (−0.83, 1.03)	0.11
MPI-ESM-P	0.66 (0.45,0.97)	0.14 (0.07,0.19)	0.77 (0.61,0.93)	6.11 (5.42, 6.80)	0.04 (−0.22,0.29)	−0.57 (−1.15,−0.01)	0.11

SD σ_I of the internal variability from CMIP5 models lies between 0.09 and 0.15 K and is larger than for AR5 EMICs (0.01–0.09 K).

C. Emulation of power spectral density

Figure 3 compares the *target* PSD to the emulated *forced* and *forced + internal* PSD. For most simulations, the emulated *forced + internal* PSD agrees with the *target* within uncertainties. AR5 EMICs show no major differences between *forced* and *forced + internal* PSD above decadal scales, except for the LOVECLIM ensemble members. Hence, the forced response is sufficient to emulate the long-term variability of most AR5 EMIC simulations. On interannual scales, considering internal variability in the emulation compensates for the mismatch between the *forced* and *target* PSD. Differences between emulated and *target* PSD are most pronounced for CLIMBER2 (mean), showing an overestimation on multidecadal to centennial scales by the emulation.

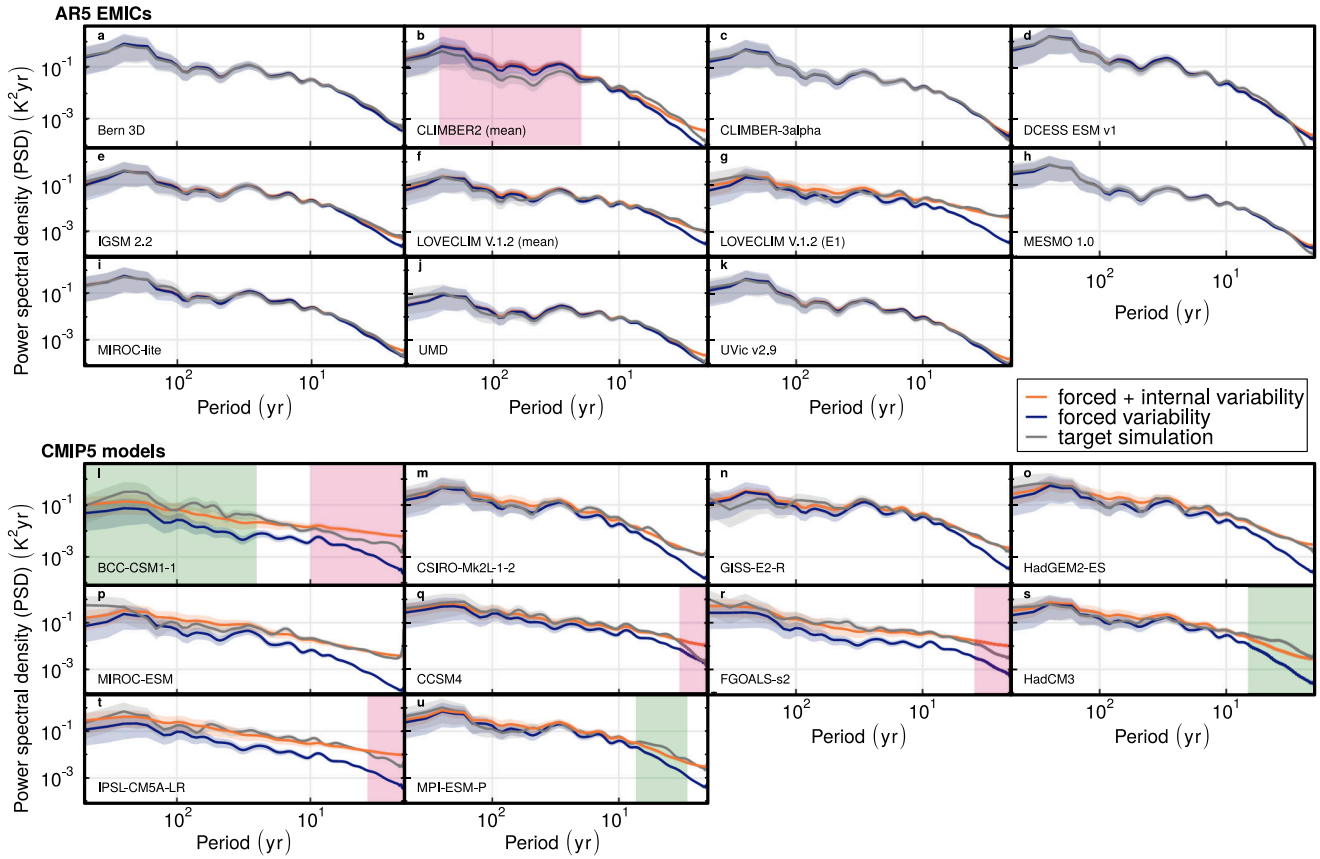


FIG. 3. *Target* and emulated *forced* and *forced + internal* power spectral density (PSD) and 95% confidence bands for AR5 EMICs [top: (a)–(k)] and CMIP5 models [bottom: (l)–(u)], as Fig. 1(c). Shaded intervals highlight examples of overestimation (pink) and underestimation (green) of the emulated PSD compared to the *target*, which are discussed in the main text. We show the PSD for LOVECLIM V.1.2 (mean) and the first ensemble member LOVECLIM V.1.2 (E1), as there are no major differences across ensemble members.

For CMIP5 simulations, the emulated *forced* PSD underestimates the *target* PSD on all timescales. Conversely, the *forced + internal* variability matches the *target* well for almost all models. Minor differences are found on interannual scales [Figs. 3(l) and 3(q)–3(u)]. Here, the *target* PSD of MPI-ESM-P and HadCM3 exhibit increased power on periods of 2–8 years. CCSM4, FGOALS-s2, and IPSL-CM5A-LR overestimate the PSD on the shortest timescales of approximately 2 years. The emulated *forced + internal* PSD for BCC-CSM1-1 deviates from that of the *target* by showing increased power on interannual and decreased power on multidecadal to centennial scales.

D. Separating internal and externally forced variance

Figure 4 shows the mean and spread of variance ratios on interannual to centennial scales for the considered model types. Variance ratios smaller than one indicate less emulated *forced* or *forced+internal* than *target* variance. We add a comparison of the variance from the five LOVECLIM V.1.2 ensemble members (E1–E5) [Fig. 4(b)] to that of the remaining AR5 EMICs [Fig. 4(a)] and CMIP5 models [Fig. 4(c)]. Here, AR5 EMICs explicitly include the LOVECLIM ensemble mean, but not its members. For all model types, the relative contribution of internal variability decreases with increasing timescale, as the ratios for *forced* and *forced + internal* variance become more similar. The contribution of internal variability is larger in LOVECLIM ensemble members and CMIP5 simulations compared to AR5 EMICs.

The emulated variance of AR5 EMICs [Fig. 4(a)] is dominated by *forced* variations and matches the *target* variance on interannual and decadal scales. On longer timescales, the emulated variance tends to overestimate the *target* variance. This is mostly due to outliers, which correspond to CLIMBER2 (mean), in line with the overestimated PSD in Fig. 3(b). Members of the LOVECLIM V.1.2. ensemble [Fig. 4(b)] exhibit more internal contributions to the variance on all timescales compared to AR5 EMICs. The ensemble's emulated *forced + internal* variance approximates the *target* well on

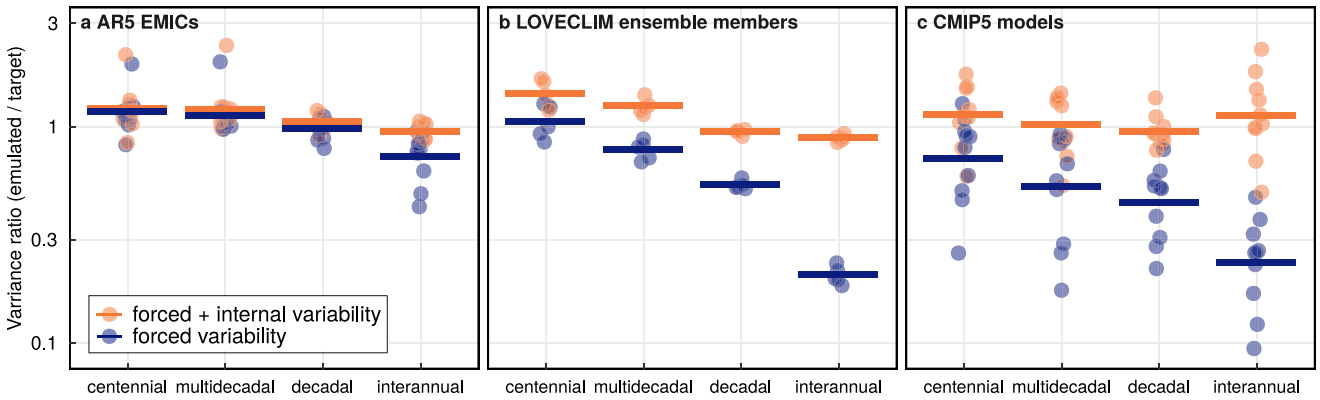


FIG. 4. Ratios of the emulated *forced* and *forced + internal* to the *target* variance for AR5 EMICs [panel (a)], LOVECLIM ensemble members [panel (b)], and CMIP5 simulations [panel (c)] on centennial (50–200 yr), multidecadal (20–50 yr), decadal (5–20 yr), and interannual (2–5 yr) scales [as Fig. 1(d)]. Bars indicate the mean variance ratio over the considered simulations for one model type. Circles correspond to individual simulations. Confidence bands for individual ratios (as in Fig. 2) are not shown for better visibility.

the interannual and decadal scale. On larger timescales, we find an overestimation of the *target* variance.

The mean emulated *forced + internal* variance of CMIP5 simulations [Fig. 4(c)] is close to one on all timescales. The relative contribution of internal compared to forced variations is similar to that of LOVECLIM ensemble members on interannual and decadal timescales. However, there is a larger spread of variance ratios. Moreover, we find a small tendency of the emulation to overestimate the interannual and centennial variance of CMIP5 models. BCC-CSM1-1 represents an outlier, with the uppermost variance ratio on the shorter and lowermost on the longer timescales, in line with the spectral analysis [Fig. 3(l)].

V. DISCUSSION

We demonstrate the emulation of GMST variability as simulated by state-of-the-art climate models using a linear stochastic two-box EBM and Bayesian inference. Our analysis builds on the same, physically motivated response function for internal and external processes and allows for consistent separation of internal and externally forced variability. Estimates of the timescale-dependent variance show that the relative contribution of internal variability increases with model complexity and decreases with timescale.

Building on previous studies,[29,31] the strength of our Bayesian framework is that it yields the posterior means and CIs for the uncertain parameters of the stochastic two-box EBM fitted to GMST simulations. Due to our choice of priors and the convergence of the fit, the estimated heat capacity C_1 agrees by construction with previous findings.[13,15,79] Our response parameters λ_1 and λ_2 are consistent with results from Fredriksen and Rypdal,[15] estimated from observational data. However, the estimates differ from those obtained in $4 \times CO_2$ experiments.[13,79] This is because our framework accounts for high-frequency pulses and, thus, estimates response parameters associated with faster dynamics. Furthermore, our findings reveal a dependence of the estimated response parameters on the imprint of intermittent volcanic eruptions on simulated temperatures. This is reflected in consistently high values for w_1, emphasizing the fast feedback. The inter-model spread of λ_1 in CMIP5 simulations suggests a link to the implemented volcanic forcing: CMIP5 simulations driven by a comparatively weak reconstruction ("GRA") tend to show higher values for λ_1 compared to those driven by "CEA" (Table I), which has greater forcing amplitudes. λ_1 and w_1 are particularly high for BCC-CSM1-1, indicating a fast and weak forced response of the fitted EBM (Fig. 7). This is consistent with a weak forced response in BCC-CSM1-1.[87] The parameter estimates for CLIMBER2 (mean) differ from the remaining AR5 EMICs. In particular, λ_1 is poorly constrained, as CIs span the full prior range. We find that the temperature response to volcanic eruptions in CLIMBER2 (mean) is delayed. The estimated λ_1 and w_1 can be reconciled with those of the other AR5 EMICs if the temperature data were shifted by 1 yr. Altogether, the sensitivity of the fit to intermittent volcanic forcing suggests that high-frequency forcing plays a crucial role for simulating temperature variability across scales correctly.

Different methods have been developed to isolate forced and internal variations based on detrending,[88] single-model ensembles,[5,89–91] and deterministic EBMs,[92,93] among others. The application of a Bayesian energy balance framework to the timescale-dependent quantification of forced and internal variance is novel. Our method provides a robust and joint separation of the variations at every step in time in a statistically sound way. Using data from the CESM Large Ensemble Community project,[94] we verify the robustness of our *forced* response (Fig. 8). The latter agrees well with the ensemble mean, which shows higher variability due to remaining internal variations that have not been averaged out over the 13 members. Hence, our method provides a robust tool to estimate the *forced* response when large ensembles are not available. Moreover, we find a wide agreement of the emulated *forced + internal* variance with that from CMIP5 simulations. The fact that the stochastic two-box EBM mimics the temperature variations well is in line with previous findings on a linear relation between external forcing and GMST.[3,13,15,95]

Small differences between the emulated *forced+internal* and *target* PSD for CMIP5 models on interannual timescales can be attributed to the simplified representation of internal variability as a weighted sum of AR(1) processes by the stochastic EBM. The latter cannot represent (pseudo-)oscillatory climate modes or modulations of internal variability by external forcing,[96,97] which could result in the observed underestimation of the PSD by the emulation on these timescales. For HadCM3 and MPI-ESM [Figs. 3(s) and 3(u)], deviations on the interannual scale are similar to those in HadCRUT5 and likely due to the spectral imprint of the El Niño–Southern Oscillation.[3,98] Moreover, our approach assumes that the covariance structure of the internal variations is determined by the estimated feedback parameters. In CMIP5 simulations with "GRA" volcanic forcing, particularly high values for the fast response parameter λ_1 and the weight w_1 lead to internal variations with autocorrelations on short timescales. This can cause an overestimation of the emulated PSD [Figs. 3(l), 3(q), 3(r), and 3(t)]. Similarly, the overestimation of the *target* PSD on longer timescales [Fig. 3(b)] in CLIMBER 2 can be explained by its estimated slow response. The latter is due to the biased delay between forcing and temperature time series (Table II). Additionally, mismatches on short timescales can propagate to longer timescales, as in the case of BCC-CSM1-1 [Fig. 3(l)]. Hence, explaining spectral properties of temperature time series not only requires consideration of stochastic noise,[41] but also precise knowledge of its correlation structure and the forced response. This highlights a need for simple, stochastic dynamical models[99] to simulate temperature fluctuations on long timescales.

The components of AR5 EMICs show a reduced number of scales compared to the AOGCMs, which simplifies the complexity of the processes contributing to variability. Therefore, the relatively strong forced variability in EMICs [Figs. 4(a) and 4(b)] is not a main deficiency but serves to explore the long-term coupling between different Earth system components in response to radiative forcing. Compared to other AR5 EMICs, LOVECLIM V.1.2 features a more complex, three-layered atmosphere, which likely explains increased internal variations in the ensemble members. However, this variability is predominately short-term correlated. Due to the EBM's covariance structure, this can lead to an overestimation of the emulated *forced+internal* variability [Fig. 4(b)] on longer timescales. On interannual and decadal scales, the variance ratios based on the emulated *forced* and *forced + internal* variability from the LOVECLIM ensemble members [Fig. 4(b)] are similar to CMIP5 simulations [Fig. 4(c)]. The similarity indicates that EMICs with a more realistic representation of atmospheric variability might better capture the relative contribution of forced and internal temperature variations on these timescales.

The contributions of internal variations on multidecadal scales and longer remain the largest in CMIP5 models, likely due to long-term variability mechanisms from the comprehensive ocean dynamics of AOGCMs. Compared to observations [Fig. 2(d)], however, internal variability on decadal and multidecadal scales is smaller in CMIP5 simulations [Fig. 4(c)]. This is particularly interesting given the agreement of observed and simulated total GMST variability on these scales.[2,3] Smaller low-frequency internal variability in climate models than in observations[100] could be offset by enhanced forced variability in response to volcanic eruptions,[101,102] such that the overall variance is largely conserved. We suspect that an incorrect ratio between internal and external variability could impact the long-term variability of simulated local temperatures. However, uncertainties in the interpretation of our HadCRUT5 findings arise due to the comparatively short time-span of the instrumental record and a possible change of the forced response under global warming.[103] Further investigating the spatial variability structure[104] and the link between local and global variability across climate states could help resolve mismatches between observed and simulated local variability on decadal and multidecadal scales.

One limitation of our study arises from the fact that the developed framework targets interannual to centennial timescales and is, therefore, designed for annually resolved GMST and forcing data. As a result, it cannot be readily applied to much shorter or longer timescales. Investigating the immediate effects of radiative forcing, for example, necessitates an extension to sub-annual resolution, and treatment of the seasonal cycle. An extension to coarser resolutions could be beneficial to study long-term changes such as millennial-scale variability. However, such applications require careful examination of the underlying assumptions of the current framework, and an extension and validation of the estimation algorithm. The forced response is likely sensitive to model-specific rapid adjustments[105] and uncertainties in the forcing. Applying the workflow to paleoclimate reconstructions could, therefore, be challenging. On the one hand, "ClimBayes" does not yet run at the best possible speed, as there are faster Bayesian algorithms.[29] On the other hand, "ClimBayes" represents an accessible, transparent, and well-documented numerical framework that can be easily adapted and extended, for example, by integrating "Rstan"[106] or multilevel delayed acceptance MCMC.[107] Similar to Fredriksen and Rypdal,[15] we use a response function of exponential form, solving the ordinary differential equation (1). Future research could investigate the potential of Bayesian methods to find response functions describing the effects of climate forcing on different observables.[108,109] Furthermore, future studies could test our findings with more advanced climate models including better representation of, for example, land surface processes, atmospheric dynamics and chemistry, and sea ice. The presented framework can be also applied to single forcing experiments[71] for quantifying the contribution of single forcings to the spectrum. This will help better understand the climate system's response and interplay of intrinsic and external components in driving climate variability.

VI. CONCLUSION

We presented a physically motivated emulation of GMST data using Bayesian inference and a stochastic energy balance model. Analyzing AR5 EMICs and CMIP5 simulations for the last millennium, we found that the power spectral density of the combined *forced + internal* response approximates the *target* spectrum well. We show that our emulation can be used to separate internal and forced contributions to GMST variability across timescales. The relative contribution of internal dynamics increases with model complexity and decreases with timescale. While AR5 EMICs predominately exhibit forced variations, simulations from CMIP5 models and the LOVECLIM ensemble members exhibit major contributions from the forced and internal response. This suggests that

EMICs with more realistic atmospheric variability can simulate statistical properties of interannual to decadal climate fluctuations more reliably. Our results show that precise knowledge of the forced response and correlation structure of internal variability is necessary to explain variability across scales, needed to assess future variability and potentially associated risks with long-term projections. Our developed framework is robust and readily available and can thus be widely applied to describe, emulate, and diagnose observed and simulated temperature variability.

ACKNOWLEDGMENTS

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, responsible for PMIP and CMIP. We thank the research groups listed in Table I and the Met Office for producing and making available their model output, measurements, and forcing reconstructions. We thank M. Eby for EMIC discussions, E. Myrvoll-Nilsen for discussion of Bayesian inference and the INLA package, and T. Gasenzer for discussion of conceptual climate models. We are grateful to N. Meinshausen, N. Weitzel, and E. Ziegler for helpful comments on the manuscript. We thank two anonymous referees for their constructive and valuable review. This study has been supported by funds of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 395588486, by the PalMod Project (Subproject No. 01LP1926C), the Heinrich-Böll-Stiftung (Heinrich Böll Foundation), and the Studienstiftung des deutschen Volkes (German Academic Scholarship Foundation). R. Scheichl is supported by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC No. 2181/1–390900948 (the Heidelberg STRUCTURES Excellence Cluster). The study benefited from discussions within the CVAS working group, a working group of the Past Global Changes (PAGES) project.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M.S. and B.E. contributed equally to this work. They carried out the analysis, created the figures and wrote the paper, supervised by K.R. and R.S. M.S. led the development of the ClimBayes package. B.E. led the variability analysis based on spectral methods. All authors designed the study, contributed to revisions, and approved the final version of the manuscript.

Maybritt Schillinger: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Beatrice Ellerhoff:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Robert Scheichl:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Kira Rehfeld:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The ClimBayes[42] software package in `R` is released at https://github.com/paleovar/ClimBayes. and published under the Zenodo identifier https://doi.org/10.5281/zenodo.7317984. Code and data to reproduce all figures is available at https://github.com/paleovar/EmulatingVariability. The data that support the findings of this study are openly available from the data holdings of the Climate Research Programme's Working Group on Coupled Modelling (e.g., https://esgf-node.llnl.gov/search/cmip5/), responsible for CMIP and PMIP, and from Schmidt *et al.*,[46] Eby *et al.*[45] (https://climate.uvic.ca/EMICAR5/participants.html), and Morice *et al.*[48]

APPENDIX A: SOLUTION TO THE MULTIBOX EBM

The tridiagonal matrix $\boldsymbol{K}$ of the multibox EBM in matrix notation (3) is given by

$$\boldsymbol{K} = \begin{pmatrix} -(\tilde{\lambda}+\kappa_2) & \kappa_2 & 0 & \cdots & 0 \\ \kappa_2 & -(\kappa_2+\kappa_3) & \kappa_3 & \ddots & \\ 0 & \kappa_3 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \\ & & & & \kappa_N \\ 0 & \cdots & & \kappa_N & -\kappa_N, \end{pmatrix}.$$

The parameter $\tilde{\lambda}$ controls the feedback of the surface layer. The coefficients $\kappa_2, \ldots, \kappa_N > 0$ describe the vertical heat transfer between ocean layers. The full solution to the multibox EBM[15] reads

$$\boldsymbol{T} = \int_{-\infty}^{t} e^{(t-s)\boldsymbol{C}^{-1}\boldsymbol{K}} \boldsymbol{C}^{-1} \boldsymbol{F}(s)\, ds.$$

$\boldsymbol{K}$ is symmetric and negative definite, and, thus, diagonalizable. Multiplication with the positive diagonal matrix $\boldsymbol{C}^{-1}$ does not change this property. Accordingly, the matrix exponential

$$e^{\boldsymbol{C}^{-1}\boldsymbol{K}} = \boldsymbol{V}^T \begin{pmatrix} e^{-\lambda_1} & 0 & 0 & \cdots & 0 \\ 0 & e^{-\lambda_2} & 0 & \cdots & 0 \\ & & & & \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & e^{-\lambda_N} \end{pmatrix} \boldsymbol{V}$$

exists for an orthonormal matrix $\boldsymbol{V}$ and eigenvalues $-\lambda_k$ of $\boldsymbol{C}^{-1}\boldsymbol{K}$. Since only the first component of the forcing vector $\boldsymbol{F}(t)$ is non-zero, the matrix entry $(e^{\boldsymbol{C}^{-1}\boldsymbol{K}})_{11}$ defines the surface temperature response. The response function reads

$$R(t) = (e^{t\boldsymbol{C}^{-1}\boldsymbol{K}})_{11} = \sum_{k=1}^{N} w_k\, e^{-\lambda_k t}.$$

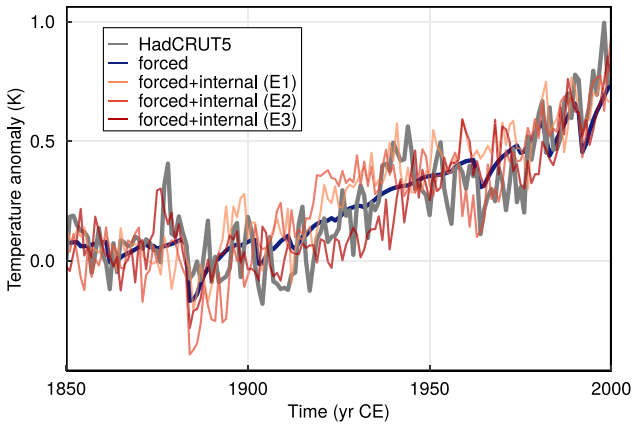


FIG. 5. HadCRUT5 *target* observation, the posterior mean of the *forced* response of the stochastic two-box EBM, and three example realizations of the emulated *forced + internal* variability.

The normalization of the weights results from $\sum_{k=1}^{N} w_k = \sum_{k=1}^{N} (V_{k1})^2 = 1$ for any orthonormal matrix $\boldsymbol{V}$.

APPENDIX B: COVARIANCE OF THE ORNSTEIN–UHLENBECK AND AR(1) PROCESS

1. Covariance matrix

The noise term in Eq. (4) is a weighted sum of Ornstein-Uhlenbeck (OU) processes. Consequently, its covariance structure results in

$$\mathrm{Cov}(T_{1,I}(t), T_{1,I}(t+s)) = \frac{\sigma_W^2}{C_1^2} \sum_{k=1}^{N} \sum_{l=1}^{N} w_k w_l \frac{e^{-\lambda_l |s|}}{\lambda_k + \lambda_l}. \quad \text{(B1)}$$

In the special case of $N = 1$, the covariance reduces to the covariance of a single OU process[40] $\mathrm{Cov}(T_{1,I}(t), T_{1,I}(t+s))_{N=1} = \frac{\sigma_W^2}{2C_1^2 \lambda_1} e^{-\lambda_1 |s|}$. Formula (B1) follows from a generalization of this special case to arbitrary N. Discretizing the noise term in Eq. (4) results in a weighted sum of AR(1) processes. This sum Z is normally distributed with mean zero and covariance matrix

$$\mathrm{Cov}(Z)_{ij} := \mathrm{Cov}(Z_i, Z_j) = \frac{\sigma_W^2}{C_1^2} \sum_{k=1}^{N} \sum_{l=1}^{N} w_k w_l \frac{e^{-\lambda_l |i-j|}}{\lambda_k + \lambda_l}. \quad \text{(B2)}$$

For given λ_k, w_k, and C_1, the SD of the stochastic forcing, σ_W, uniquely defines the SD of the internal variations $\sigma_I := \sqrt{\mathrm{Cov}(Z_i, Z_i)}$ and vice versa. It is possible to estimate σ_I within the Bayesian framework. However, additional uncertainties arise from the fact that the covariance matrix in Eq. (B2) is only an approximation to the true correlation structure of the residuals. As a result, the Bayesian estimation of σ_I might not preserve the total variance. Therefore, we determine σ_I from the residuals, that is, the data minus the estimated forced response.

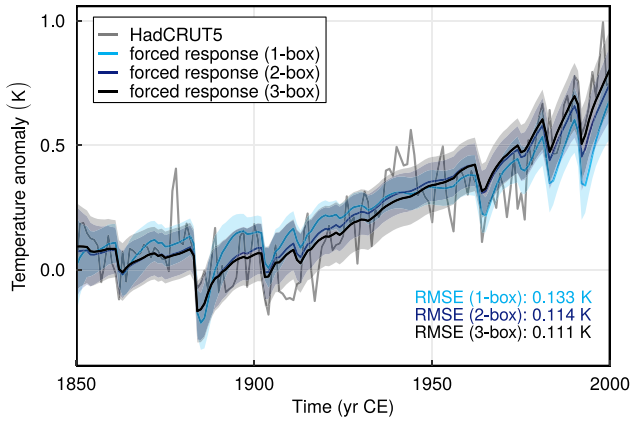


FIG. 6. HadCRUT5 *target* observation (grey), the *forced* response from the one-, two-, and three-box EBM fit to the data. We show their posterior means and the CIs (shaded) as well as their root mean square errors (RMSEs).

2. Iterative computation of the likelihood

Theoretically, the likelihood is given by a normal distribution with mean $T_{1,F}$ and covariance matrix $\mathrm{Cov}(Z)$, depending on θ. However, computing the covariance matrix dynamically for each sample in the Markov chain can lead to difficulties. In particular, $\mathrm{Cov}(Z)$ needs to be inverted for every sample, which is computationally expensive. Moreover, the determinant $\det(\mathrm{Cov}(Z))$ can be close to zero, which can make numerical calculations unstable. Potential biases include decreasing goodness of fit and accuracy of estimated posteriors.

To solve this problem, we propose an iterative approach. This keeps λ_k and w_k in the covariance matrix fixed for each iteration of the algorithm. The first iteration uses the prior means for λ_k and w_k as well as a starting value for the ratio σ_W/C_1. It is not necessary to consider σ_W and C_1 separately, since Eqs. (B1) and (B2) depend only on their ratio. This ratio is chosen such that $\sigma_I = 0.1$ K for CMIP5 simulations and LOVECLIM ensemble members, and $\sigma_I = 0.05$ K for AR5 EMICs. For the second iteration, the estimated posterior means of λ_k and w_k define the covariance matrix entries. Additionally, σ_I is set to the SD of the residuals, which defines σ_W/C_1. The results of this second iteration are the posterior distributions for $\lambda_1, \ldots, \lambda_N$, weights $w_2, \ldots, w_N$, heat capacity C_1, initial forcing F_0, and initial temperature T_0. These iterations can be repeated and adjusted with "ClimBayes." We find that two iterations are enough to fit the *forced + internal* response to the considered data well, and that further iterations do not improve the goodness of fit.

3. Sampling from internal variability

Sampling internal variations $T_{1,I}(t)$ requires the values of λ_k, w_k, and σ_W in $\mathrm{Cov}(Z)$ [Eq. (B1)]. λ_k and w_k are set to the posterior

TABLE III. Posterior means and 95% CI of estimated feedback parameters and weights for the 1-, 2-, and 3-box EBM fitted to the HadCRUT5 GMST.

Number of boxes	λ_1 (yr^{-1})	λ_2 (yr^{-1})	λ_3 (yr^{-1})	C_1 ($W\,yr\,m^{-2}\,K^{-1}$)	w_1 (unitless)	w_2 (unitless)
1-box	0.35 (0.21,0.63)	…	…	9.08 (6.83,10.68)	…	…
2-box	1.31 (0.71,1.86)	0.09 (0.03,0.16)	…	8.20 (5.68,10.43)	0.72 (0.46,0.9)	…
3-box	1.29 (0.70,1.86)	0.11 (0.04,0.18)	0.01 (0.01,0.02)	8.46 (5.86,10.57)	0.74 (0.46,0.90)	0.22 (0.02,0.53)

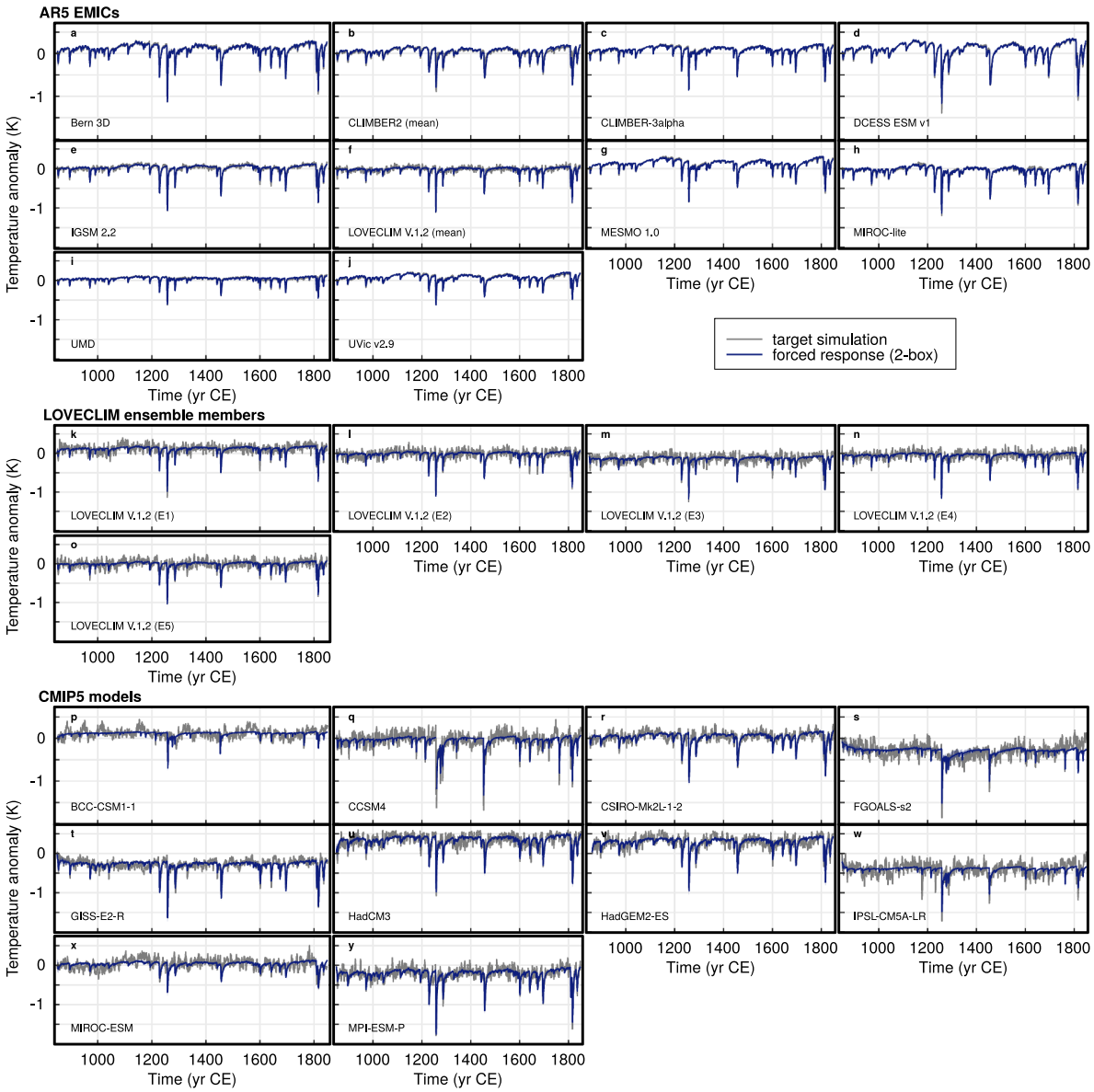


FIG. 7. Emulated *forced* response $T_{1,F}(t)$, that is the posterior mean, of the stochastic two-box EBM fitted to the GMST target data from AR5 EMICs [panels (a)–(j)], LOVECLIM V.1.2 ensemble members [panels (k)–(0)], and CMIP5 simulations [panels (p)–(y)].

means. We calculate σ_W from σ_I, which we assume to equal the SD of the residuals. As an example, Fig. 5 shows realizations of $T_{1,I}(t)$ drawn from the estimated covariance of HadCRUT5 observations.

APPENDIX C: COMPARISON OF ONE-, TWO- AND THREE-BOX EBM

We have verified that our results are robust against reasonable variations of the number of boxes. Here, we examine the difference between $N \in (1, 2, 3)$ boxes on the example of the HadCRUT5 GMST (Fig. 6). We choose the priors of the response parameters to cover the same overall range as for the two-box model [$N = 1 : \lambda_1 \in (1/200, 2)\,\mathrm{yr}^{-1}$ and $N = 3 : \lambda_1 \in (1/5, 2)\,\mathrm{yr}^{-1}, \lambda_2 \in (1/50, 1/5)\,\mathrm{yr}^{-1}, \lambda_3 \in (1/200, 1/50)\,\mathrm{yr}^{-1}$].

The stochastic two-box EBM fits the data more accurately (root mean square error: RMSE = 0.114 K) than the one-box EBM (RMSE = 0.133 K) (Fig. 6). The three-box EBM yields only minor improvements (RMSE = 0.0111 K). This pattern is consistent for AR5 EMICs and CMIP5 simulations and reflected in similar forced responses and power spectral densities for $N \in (1, 2, 3)$. Adding boxes, however, increases the risk of overfitting due to increasing degrees of freedom. This is reflected in increasing CIs for the forced response and parameters (Table III) with more boxes. That is why $N = 2$ represents the best compromise between goodness of fit, identifiability of parameters, and number of free parameters in our experiments.

APPENDIX D: EMULATED FORCED TEMPERATURE RESPONSE FOR CONSIDERED SIMULATIONS

Figure 7 shows the best estimate of the EBM's forced response, fitted to the target simulations from all considered models. CIs are much narrower and almost vanishing compared to HadCRUT [Fig. 2(a)]. This is due to the fact that with increasing length of the time series the posterior uncertainties of the parameters and the forced response decrease.

Figure 8 compares the emulated *forced* response against simulation data from the Last Millennium Ensemble of the Community Earth System Model (CESM).[110] Forming the ensemble mean over the available 13 members serves to average out uncorrelated internal variability. Our emulated *forced* response, fitted to the first ensemble member (E1), shows a large overlap with the ensemble mean despite remaining internal variability that has not been averaged out.

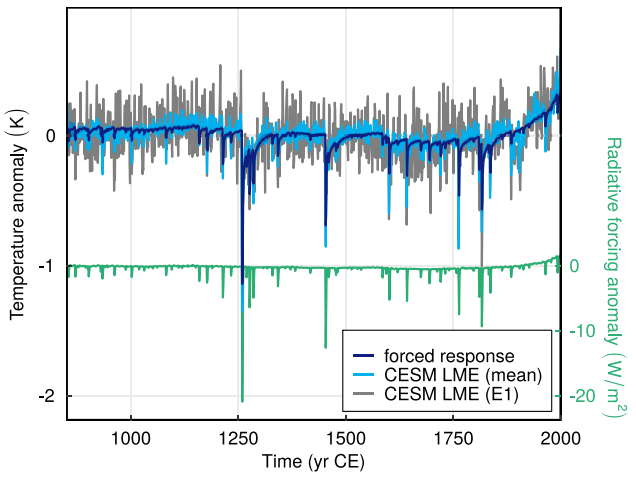


FIG. 8. Ensemble mean of the CESM Large Ensemble[110] (13 members) and the emulated *forced* response $T_{1,F}(t)$ of the stochastic two-box EBM fitted to GMST target data from one of the ensemble members (E1). The corresponding radiative forcing time series is shown in green.

REFERENCES

[1]R. W. Katz and B. G. Brown, "Extreme events in a changing climate: Variability is more important than averages," Clim. Change **21**, 289–302 (1992).
[2]PAGES 2k Consortium, "Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era," Nat. Geosci. **12**, 643–649 (2019).
[3]B. Ellerhoff and K. Rehfeld, "Probing the timescale dependency of local and global variations in surface air temperature from climate simulations and reconstructions of the last millennia," Phys. Rev. E **104**, 064136 (2021).
[4]E. Hawkins and R. Sutton, "The potential to narrow uncertainty in projections of regional precipitation change," Clim. Dyn. **37**, 407–418 (2011).
[5]L. M. Frankcombe, M. H. England, M. E. Mann, and B. A. Steinman, "Separating internal variability from the externally forced climate response," J. Clim. **28**, 8184–8202 (2015).
[6]R. Hébert and S. Lovejoy, "Regional climate sensitivity- and historical-based projections to 2100," Geophys. Res. Lett. **45**, 4248–4254, https://doi.org/10.1002/2017GL076649 (2018).
[7]T. Laepple and P. Huybers, "Global and regional variability in marine surface temperatures," Geophys. Res. Lett. **41**, 2528–2534, https://doi.org/10.1002/2014GL059345 (2014).
[8]M. I. Budyko, "The effect of solar radiation variations on the climate of the earth," Tellus **21**, 611–619 (1969).
[9]W. D. Sellers, "A global climatic model based on the energy balance of the earth-atmosphere system," J. Appl. Meteorol. **8**, 392–400 (1969).
[10]K. Hasselmann, "Stochastic climate models: Part I. Theory," Tellus **28**, 473–485 (1976).
[11]K. Fraedrich, U. Luksch, and R. Blender, "1/f model for long-time memory of the ocean surface temperature," Phys. Rev. E **70**, 037301 (2004).
[12]I. M. Held, M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, "Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing," J. Clim. **23**, 2418–2427 (2010).
[13]O. Geoffroy, D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, "Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments," J. Clim. **26**, 1841–1857 (2013).
[14]M. Rypdal and K. Rypdal, "Long-memory effects in linear response models of earth's temperature and implications for future global warming," J. Clim. **27**, 5240–5258 (2014).
[15]H.-B. Fredriksen and M. Rypdal, "Long-range persistence in global surface temperatures explained by linear multibox energy balance models," J. Clim. **30**, 7157–7168 (2017).
[16]K. Hasselmann, "Optimal fingerprints for the detection of time-dependent climate change," J. Clim. **6**, 1957–1971 (1993).
[17]G. C. Hegerl, H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, "Detecting greenhouse-gas-induced climate change with an optimal fingerprint method," J. Clim. **9**, 2281–2306 (1996).
[18]M. Ghil, "Climate sensitivity, energy balance models, and oscillatory climate models," J. Geophys. Res. **89**, 1280, https://doi.org/10.1029/JD089iD01p01280 (1984).

[19]P. M. Cox, C. Huntingford, and M. S. Williamson, "Emergent constraint on equilibrium climate sensitivity from global temperature variability," Nature **553**, 319–322 (2018).
[20]K. Rypdal, M. Rypdal, and H.-B. Fredriksen, "Spatiotemporal long-range persistence in earth's temperature field: Analysis of stochastic–diffusive energy balance models," J. Clim. **28**, 8379–8395 (2015).
[21]M. Ghil, "Climate stability for a sellers-type model," J. Atmos. Sci. **33**, 3–20 (1976).
[22]G. R. North, J. G. Mengel, and D. A. Short, "Simple energy balance model resolving the seasons and the continents: Application to the astronomical theory of the ice ages," J. Geophys. Res. **88**, 6576, https://doi.org/10.1029/JC088iC11p06576 (1983).
[23]T. Bódai, V. Lucarini, F. Lunkeit, and R. Boschi, "Global instability in the Ghil–Sellers model," Clim. Dyn. **44**, 3361–3381 (2015).
[24]L. Østvand, T. Nilsen, K. Rypdal, D. Divine, and M. Rypdal, "Long-range memory in internal and forced dynamics of millennium-long climate model simulations," Earth Syst. Dyn. **5**, 295–308 (2014).
[25]G. R. North, "Analytical solution to a simple climate model with diffusive heat transport," J. Atmos. Sci. **32**, 1301–1307 (1975).
[26]G. R. North and R. F. Cahalan, "Predictability in a solvable stochastic climate model," J. Atmos. Sci. **38**, 504–513 (1981).
[27]G. R. North, J. Wang, and M. G. Genton, "Correlation models for temperature fields," J. Clim. **24**, 5850–5862 (2011).
[28]B. Dortmans, W. F. Langford, and A. R. Willms, "An energy balance model for paleoclimate transitions," Clim. Past **15**, 493–520 (2019).
[29]E. Myrvoll-Nilsen, S. H. Sørbye, H.-B. Fredriksen, H. Rue, and M. Rypdal, "Statistical estimation of global surface temperature response to forcing under the assumption of temporal scaling," Earth Syst. Dyn. **11**, 329–345 (2020).
[30]R. Hébert, S. Lovejoy, and B. Tremblay, "An observation-based scaling model for climate sensitivity estimates and global projections to 2100," Clim. Dyn. **56**, 1105–1129 (2021).
[31]R. W. Bodman and R. N. Jones, "Bayesian estimation of climate sensitivity using observationally constrained simple climate models," WIREs Clim. Change **7**, 461–473 (2016).
[32]C. Proistosescu and P. J. Huybers, "Slow climate mode reconciles historical and model-based estimates of climate sensitivity," Sci. Adv. **3**, e1602821 (2017).
[33]A. Jonko, N. M. Urban, and B. Nadiga, "Towards Bayesian hierarchical inference of equilibrium climate sensitivity from a combination of CMIP5 climate models and observational data," Clim. Change **149**, 247–260 (2018).
[34]S. C. Sherwood, M. J. Webb, J. D. Annan, K. C. Armour, P. M. Forster, J. C. Hargreaves, G. Hegerl, S. A. Klein, K. D. Marvel, E. J. Rohling, M. Watanabe, T. Andrews, P. Braconnot, C. S. Bretherton, G. L. Foster, Z. Hausfather, A. S. Heydt, R. Knutti, T. Mauritsen, J. R. Norris, C. Proistosescu, M. Rugenstein, G. A. Schmidt, K. B. Tokarska, and M. D. Zelinka, "An assessment of earth's climate sensitivity using multiple lines of evidence," Rev. Geophys. **58**, e2019RG000678, https://doi.org/10.1029/2019RG000678 (2020).
[35]R. B. Skeie, T. Berntsen, M. Aldrin, M. Holden, and G. Myhre, "Climate sensitivity estimates—Sensitivity to radiative forcing time series and observational data," Earth Syst. Dyn. **9**, 879–894 (2018).
[36]J. M. Mitchell, "An overview of climatic variability and its causal mechanisms," Quat. Res. **6**, 481–493 (1976).
[37]P. Huybers and W. Curry, "Links between annual, Milankovitch and continuum temperature variability," Nature **441**, 329–332 (2006).
[38]A. S. von der Heydt, P. Ashwin, C. D. Camp, M. Crucifix, H. A. Dijkstra, P. Ditlevsen, and T. M. Lenton, "Quantification and interpretation of the climate variability record," Global Planet. Change **197**, 103399 (2021).
[39]C. L. E. Franzke, S. Barbosa, R. Blender, H.-B. Fredriksen, T. Laepple, F. Lambert, T. Nilsen, K. Rypdal, M. Rypdal, M. G. Scotto, S. Vannitsem, N. W. Watkins, L. Yang, and N. Yuan, "The structure of climate variability across scales," Rev. Geophys. **58**, e2019RG000657, https://doi.org/10.1029/2019RG000657 (2020).
[40]M. Rypdal, H.-B. Fredriksen, E. Myrvoll-Nilsen, K. Rypdal, and S. Sørbye, "Emergent scale invariance and climate sensitivity," Climate **6**, 93 (2018).
[41]S. A. Soldatenko and R. A. Colman, "Power spectrum sensitivity analysis of the global mean surface temperature fluctuations simulated in a two-box stochastic energy balance model," Tellus A: Dyn. Meteorol. Oceanogr. **74**, 68 (2022).
[42]M. Schillinger, B. Ellerhoff, R. Scheichl, and K. Rehfeld, "The ClimBayes package in R," Zenodo, V. 0.1.1, https://doi.org/10.5281/zenodo.7317984 (2022).
[43]J. H. Jungclaus, E. Bard, M. Baroni, P. Braconnot, J. Cao, L. P. Chini, T. Egorova, M. Evans, J. F. González-Rouco, H. Goosse, G. C. Hurtt, F. Joos, J. O. Kaplan, M. Khodri, K. Klein Goldewijk, N. Krivova, A. N. LeGrande, S. J. Lorenz, J. Luterbacher, W. Man, A. C. Maycock, M. Meinshausen, A. Moberg, R. Muscheler, C. Nehrbass-Ahles, B. I. Otto-Bliesner, S. J. Phipps, J. Pongratz, E. Rozanov, G. A. Schmidt, H. Schmidt, W. Schmutz, A. Schurer, A. I. Shapiro, M. Sigl, J. E. Smerdon, S. K. Solanki, C. Timmreck, M. Toohey, I. G. Usoskin, S. Wagner, C.-J. Wu, K. L. Yeo, D. Zanchettin, Q. Zhang, and E. Zorita, "The PMIP4 contribution to CMIP6—Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 *past1000* simulations," Geosci. Model Dev. **10**, 4005–4033 (2017).
[44]G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S. C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, and C. Forest, "Evaluation of climate models," in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2014), pp. 741–866.
[45]M. Eby, A. J. Weaver, K. Alexander, K. Zickfeld, A. Abe-Ouchi, A. A. Cimatoribus, E. Crespin, S. S. Drijfhout, N. R. Edwards, A. V. Eliseev, G. Feulner, T. Fichefet, C. E. Forest, H. Goosse, P. B. Holden, F. Joos, M. Kawamiya, D. Kicklighter, H. Kienert, K. Matsumoto, I. I. Mokhov, E. Monier, S. M. Olsen, J. O. P. Pedersen, M. Perrette, G. Philippon-Berthier, A. Ridgwell, A. Schlosser, T. Schneider von Deimling, G. Shaffer, R. S. Smith, R. Spahni, A. P. Sokolov, M. Steinacher, K. Tachiiri, K. Tokos, M. Yoshimori, N. Zeng, and F. Zhao, "Historical and idealized climate model experiments: An intercomparison of earth system models of intermediate complexity," Clim. Past **9**, 1111–1140 (2013).
[46]G. A. Schmidt, J. H. Jungclaus, C. M. Ammann, E. Bard, P. Braconnot, T. J. Crowley, G. Delaygue, F. Joos, N. A. Krivova, R. Muscheler, B. L. Otto-Bliesner, J. Pongratz, D. T. Shindell, S. K. Solanki, F. Steinhilber, and L. E. A. Vieira, "Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1)," Geosci. Model Dev. **5**, 185–191 (2012).
[47]K. Tachiiri, J. C. Hargreaves, J. D. Annan, A. Oka, A. Abe-Ouchi, and M. Kawamiya, "Development of a system emulating the global carbon cycle in earth system models," Geosci. Model Dev. **3**, 365–376 (2010).
[48]C. P. Morice, J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. H. Dunn, T. J. Osborn, P. D. Jones, and I. R. Simpson, "An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set," J. Geophys. Res.: Atmos. **126**, e2019JD032361, https://doi.org/10.1029/2019JD032361 (2021).
[49]J. Pongratz, C. Reick, T. Raddatz, and M. Claussen, "A reconstruction of global agricultural areas and land cover for the last millennium," Global Biogeochem. Cycles **22**, GB3018, https://doi.org/10.1029/2007GB003153 (2008).
[50]T. J. Crowley, G. Zielinski, B. Vinther, R. Udisti, K. Kreutz, J. Cole-Dai, and E. Castellano, "Volcanism and the little ice age," PAGES News **16**, 22–23 (2008).
[51]F. Steinhilber, J. Beer, and C. Fröhlich, "Total solar irradiance during the holocene," Geophys. Res. Lett. **36**, L19704, https://doi.org/10.1029/2009GL040142 (2009).
[52]Y.-M. Wang, J. L. Lean, and J. N. R. Sheeley, "Modeling the sun's magnetic field and irradiance since 1713," Astrophys. J. **625**, 522–538 (2005).
[53]G. Delaygue and E. Bard, "An Antarctic view of beryllium-10 and solar activity for the past millennium," Climate Dyn. **36**(11–12), 2201–2218 (2011).
[54]N. A. Krivova, L. Balmaceda, and S. K. Solanki, "Reconstruction of solar total irradiance since 1700 from the surface magnetic flux," Astron. Astrophys. **467**, 335–346 (2007).
[55]L. E. A. Vieira and S. K. Solanki, "Evolution of the solar magnetic flux on time scales of years to millenia," Astron. Astrophys. **509**, A100 (2010).
[56]C. Gao, A. Robock, and C. Ammann, "Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models," J. Geophys. Res. **113**, D23111, https://doi.org/10.1029/2008JD010239 (2008).
[57]S. P. Ritz, T. F. Stocker, and F. Joos, "A coupled dynamical ocean–energy balance atmosphere model for paleoclimate studies," J. Clim. **24**, 349–375 (2011).
[58]M. Montoya, A. Griesel, A. Levermann, J. Mignot, M. Hofmann, A. Ganopolski, and S. Rahmstorf, "The earth system model of intermediate complexity CLIMBER-3α. Part I: Description and performance for present-day conditions," Clim. Dyn. **26**, 327–328 (2006).

[59]V. Petoukhov, M. Claussen, A. Berger, M. Crucifix, M. Eby, A. V. Eliseev, T. Fichefet, A. Ganopolski, H. Goosse, I. Kamenkovich, I. I. Mokhov, M. Montoya, L. A. Mysak, A. Sokolov, P. Stone, Z. Wang, and A. J. Weaver, "EMIC intercomparison project (EMIP–CO_2): Comparative analysis of EMIC simulations of climate, and of equilibrium and transient responses to atmospheric CO_2 doubling," Clim. Dyn. **25**, 363–385 (2005).
[60]G. Shaffer, S. Malskær Olsen, and J. O. Pepke Pedersen, "Presentation, calibration and validation of the low-order, DCESS earth system model (Version 1)," Geosci. Model Dev. **1**, 17–51 (2008).
[61]A. P. Sokolov, C. A. Schlosser, S. Dutkiewicz, S. Paltsev, D. W. Kicklighter, H. D. Jacoby, R. G. Prinn, C. E. Forest, J. M. Reilly, C. Wang, and B. S. Felzer, "MIT integrated global system model (IGSM) version 2: Model description and baseline evaluation," Tech. Rep. (MIT Joint Program on the Science and Policy of Global Change, 2005).
[62]H. Goosse, V. Brovkin, T. Fichefet, R. Haarsma, P. Huybrechts, J. Jongma, A. Mouchet, F. Selten, P.-Y. Barriat, J.-M. Campin, E. Deleersnijder, E. Driesschaert, H. Goelzer, I. Janssens, M.-F. Loutre, M. A. Morales Maqueda, T. Opsteegh, P.-P. Mathieu, G. Munhoven, E. J. Pettersson, H. Renssen, D. M. Roche, M. Schaeffer, B. Tartinville, A. Timmermann, and S. L. Weber, "Description of the earth system model of intermediate complexity LOVECLIM version 1.2," Geosci. Model Dev. **3**, 603–633 (2010).
[63]K. Matsumoto, K. S. Tokos, A. R. Price, and S. J. Cox, "First description of the minnesota earth system model for ocean biogeochemistry (MESMO 1.0)," Geosci. Model Dev. **1**, 1–15 (2008).
[64]N. Zeng, "How strong is carbon cycle-climate feedback under global warming?," Geophys. Res. Lett. **31**, L20203, https://doi.org/10.1029/2004GL020904 (2004).
[65]A. J. Weaver, M. Eby, E. C. Wiebe, C. M. Bitz, P. B. Duffy, T. L. Ewen, A. F. Fanning, M. M. Holland, A. MacFadyen, H. D. Matthews, K. J. Meissner, O. Saenko, A. Schmittner, H. Wang, and M. Yoshimori, "The UVic earth system climate model: Model description, climatology, and applications to past, present and future climates," Atmos. Ocean **39**, 361–428 (2001).
[66]X. Xiao-Ge, W. Tong-Wen, and Z. Jie, "Introduction of CMIP5 experiments carried out with the climate system models of Beijing climate center," Adv. Clim. Change Res. **4**, 41–49 (2013).
[67]L. Landrum, B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng, "Last millennium climate and its variability in CCSM4," J. Clim. **26**, 1085–1111 (2013).
[68]S. J. Phipps, L. D. Rotstayn, H. B. Gordon, J. L. Roberts, A. C. Hirst, and W. F. Budd, "The CSIRO Mk3L climate system model version 1.0—Part 2: Response to external forcings," Geosci. Model Dev. **5**, 649–682 (2012).
[69]Q. Bao, P. Lin, T. Zhou, Y. Liu, Y. Yu, G. Wu, B. He, J. He, L. Li, J. Li, Y. Li, H. Liu, F. Qiao, Z. Song, B. Wang, J. Wang, P. Wang, X. Wang, Z. Wang, B. Wu, T. Wu, Y. Xu, H. Yu, W. Zhao, W. Zheng, and L. Zhou, "The flexible global ocean-atmosphere-land system model, spectral version 2: FGOALS-s2," Adv. Atmos. Sci. **30**, 561–576 (2013).
[70]G. A. Schmidt, R. Ruedy, J. E. Hansen, I. Aleinov, N. Bell, M. Bauer, S. Bauer, B. Cairns, V. Canuto, Y. Cheng, A. Del Genio, G. Faluvegi, A. D. Friend, T. M. Hall, Y. Hu, M. Kelley, N. Y. Kiang, D. Koch, A. A. Lacis, J. Lerner, K. K. Lo, R. L. Miller, L. Nazarenko, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, G. L. Russell, M. Sato, D. T. Shindell, P. H. Stone, S. Sun, N. Tausnev, D. Thresher, and M.-S. Yao, "Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data," J. Clim. **19**, 153–192 (2006).
[71]A. P. Schurer, S. F. B. Tett, and G. C. Hegerl, "Small influence of solar variability on climate over the past millennium," Nat. Geosci. **7**, 104–108 (2014).
[72]C. D. Jones, J. K. Hughes, N. Bellouin, S. C. Hardiman, G. S. Jones, J. Knight, S. Liddicoat, F. M. O'Connor, R. J. Andres, C. Bell, K.-O. Boo, A. Bozzo, N. Butchart, P. Cadule, K. D. Corbin, M. Doutriaux-Boucher, P. Friedlingstein, J. Gornall, L. Gray, P. R. Halloran, G. Hurtt, W. J. Ingram, J.-F. Lamarque, R. M. Law, M. Meinshausen, S. Osprey, E. J. Palin, L. Parsons Chini, T. Raddatz, M. G. Sanderson, A. A. Sellar, A. Schurer, P. Valdes, N. Wood, S. Woodward, M. Yoshioka, and M. Zerroukat, "The HadGEM2-ES implementation of CMIP5 centennial simulations," Geosci. Model Dev. **4**, 543–570 (2011).
[73]J.-L. Dufresne, M.-A. Foujols, S. Denvil, A. Caubel, O. Marti, O. Aumont, Y. Balkanski, S. Bekki, H. Bellenger, R. Benshila, S. Bony, L. Bopp, P. Braconnot, P. Brockmann, P. Cadule, F. Cheruy, F. Codron, A. Cozic, D. Cugnet, N. de Noblet, J.-P. Duvel, C. Ethé, L. Fairhead, T. Fichefet, S. Flavoni, P. Friedlingstein, J.-Y. Grandpeix, L. Guez, E. Guilyardi, D. Hauglustaine, F. Hourdin, A. Idelkadi, J. Ghattas, S. Joussaume, M. Kageyama, G. Krinner, S. Labetoulle, A. Lahellec, M.-P. Lefebvre, F. Lefevre, C. Levy, Z. X. Li, J. Lloyd, F. Lott, G. Madec, M. Mancip, M. Marchand, S. Masson, Y. Meurdesoif, J. Mignot, I. Musat, S. Parouty, J. Polcher, C. Rio, M. Schulz, D. Swingedouw, S. Szopa, C. Talandier, P. Terray, N. Viovy, and N. Vuichard, "Climate change projections using the IPSL-CM5 earth system model: From CMIP3 to CMIP5," Clim. Dyn. **40**, 2123–2165 (2013).
[74]F. Hourdin, M.-A. Foujols, F. Codron, V. Guemas, J.-L. Dufresne, S. Bony, S. Denvil, L. Guez, F. Lott, J. Ghattas, P. Braconnot, O. Marti, Y. Meurdesoif, and L. Bopp, "Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model," Clim. Dyn. **40**, 2167–2192 (2013).
[75]T. Sueyoshi, R. Ohgaito, A. Yamamoto, M. O. Chikamoto, T. Hajima, H. Okajima, M. Yoshimori, M. Abe, R. O'ishi, F. Saito, S. Watanabe, M. Kawamiya, and A. Abe-Ouchi, "Set-up of the PMIP3 paleoclimate experiments conducted using an earth system model, MIROC-ESM," Geosci. Model Dev. **6**, 819–836 (2013).
[76]M. A. Giorgetta, J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornblueh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K. D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, and B. Stevens, "Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5: Climate changes in MPI-ESM," J. Adv. Model. Earth Syst. **5**, 572–597 (2013).
[77]J. H. Jungclaus, N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. Storch, "Characteristics of the ocean simulations in the max planck institute ocean model (MPIOM) the ocean component of the MPI-earth system model," J. Adv. Model. Earth Syst. **5**, 422–446 (2013).
[78]D. P. Cummins, D. B. Stephenson, and P. A. Stott, "Optimal estimation of stochastic energy balance model parameters," J. Clim. **33**, 7909–7926 (2020).
[79]O. Geoffroy, D. Saint-Martin, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéca, "Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs," J. Clim. **26**, 1859–1876 (2013).
[80]A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," Stat. Sci. **7**, 457–472 (1992).
[81]S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," J. Comput. Graph. Stat. **7**, 434–455 (1998).
[82]A. Gelman, "Bayesian data analysis," in *Chapman & Hall/CRC Texts in Statistical Science*, 3rd ed. (CRC Press, Boca Raton, FL, 2014).
[83]M. Plummer, N. Best, K. Cowles, K. Vines, and M. M. Plummer, "R package 'CODA'" https://cran.r-project.org/package=coda (2020).
[84]J. M. Flegal, M. Haran, and G. L. Jones, "Markov chain Monte Carlo: Can we trust the third significant figure?," Stat. Sci. **23**, 250–260 (2008).
[85]M. Haran and J. Hughes, "R package 'batchmeans: Consistent batch means estimation of Monte Carlo,'" https://cran.r-project.org/package=batchmeans (2020).
[86]T. Laepple and P. Huybers, "Ocean surface temperature variability: Large model data differences at decadal and longer periods," Proc. Natl. Acad. Sci. U.S.A. **111**, 16682–16687 (2014).
[87]O. Bothe, J. H. Jungclaus, and D. Zanchettin, "Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble," Clim. Past **9**, 2471–2487 (2013).
[88]R. Zhang, T. L. Delworth, R. Sutton, D. L. R. Hodson, K. W. Dixon, I. M. Held, Y. Kushnir, J. Marshall, Y. Ming, R. Msadek, J. Robson, A. J. Rosati, M. Ting, and G. A. Vecchi, "Have aerosols caused the observed atlantic multidecadal variability?," J. Atmos. Sci. **70**, 1135–1144 (2013).
[89]D. Olonscheck and D. Notz, "Consistently estimating internal climate variability from climate model simulations," J. Clim. **30**, 9555–9573 (2017).
[90]F. Lehner, C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, "Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6," Earth Syst. Dyn. **11**, 491–508 (2020).
[91]M. E. Mann, B. A. Steinman, D. J. Brouillette, A. Fernandez, and S. K. Miller, "On the estimation of internal climate variability during the preindustrial past millennium," Geophys. Res. Lett. **49**, e2021GL096596, https://doi.org/10.1029/2021GL096596 (2022).

[92]M. E. Mann, B. A. Steinman, and S. K. Miller, "On forced temperature changes, internal variability, and the AMO," Geophys. Res. Lett. **41**, 3211–3219, https://doi.org/10.1029/JD089iD01p01280 (2014).
[93]T. J. Crowley, "Causes of climate change over the past 1000 years," Science **289**, 270–277 (2000).
[94]J. E. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. M. Arblaster, S. C. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J.-F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein, "The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability," Bull. Am. Meteorol. Soc. **96**, 1333–1349 (2015).
[95]D. G. MacMynowski, H.-J. Shin, and K. Caldeira, "The frequency response of temperature and precipitation in a climate model," Geophys. Res. Lett. **38**, L16711, https://doi.org/10.1029/2011GL048623 (2011).
[96]N. Maher, D. Matei, S. Milinski, and J. Marotzke, "ENSO change in climate projections: Forced response or internal variability?," Geophys. Res. Lett. **45**, 11390–11398, https://doi.org/10.1029/2018GL079764 (2018).
[97]R. Bonnet, D. Swingedouw, G. Gastineau, O. Boucher, J. Deshayes, F. Hourdin, J. Mignot, J. Servonnat, and A. Sima, "Increased risk of near term global warming due to a recent AMOC weakening," Nat. Commun. **12**, 6108 (2021).
[98]B. Ellerhoff, M. J. Kirschner, E. Ziegler, M. D. Holloway, L. Sime, and K. Rehfeld, "Contrasting state-dependent effects of natural forcing on global and local climate variability," Geophys. Res. Lett. **49**, e2022GL098335, https://doi.org/10.1029/2022GL098335 (2022).
[99]S. Lovejoy, "The half-order energy balance equation—Part 1: The homogeneous HEBE and long memories," Earth Syst. Dyn. **12**, 469–487 (2021).
[100]X. Yan, R. Zhang, and T. R. Knutson, "Underestimated AMOC variability and implications for AMV and predictability in CMIP models," Geophys. Res. Lett. **45**, 4319–4328, https://doi.org/10.1029/2018GL077378 (2018).
[101]A. P. Schurer, G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps, "Separating forced from chaotic climate variability over the past millennium," J. Clim. **26**, 6954–6973 (2013).
[102]P. Chylek, C. Folland, J. D. Klett, and M. K. Dubey, "CMIP5 climate models overestimate cooling by volcanic aerosols," Geophys. Res. Lett. **47**, e2020GL087047, https://doi.org/10.1029/2020GL087047 (2020).
[103]P. O. Hopcroft, J. Kandlbauer, P. J. Valdes, and R. S. J. Sparks, "Reduced cooling following future volcanic eruptions," Clim. Dyn. **51**, 1449–1463 (2018).
[104]T. Kunz and T. Laepple, "Frequency-dependent estimation of effective spatial degrees of freedom," J. Clim. **34**, 7373–7388 (2021).
[105]R. S. Smith, J. M. Gregory, and A. Osprey, "A description of the FAMOUS (version XDBUA) climate model and control run," Geosci. Model Dev. **1**, 53–68 (2008).
[106]J. Guo, D. Lee, K. Sakrejda, J. Gabry, B. Goodrich, J. De Guzman, E. Niebler, T. Heller, and J. Fletcher, "Rstan: R interface to stan," https://cran.r-project.org/package=rstan (2022).
[107]M. B. Lykkegaard, T. J. Dodwell, C. Fox, G. Mingas, and R. Scheichl, "Multilevel delayed acceptance MCMC," arXiv:2202.03876 (2022).
[108]V. Lucarini, "Revising and extending the linear response theory for statistical mechanical systems: Evaluating observables as predictors and predictands," J. Stat. Phys. **173**, 1698–1721 (2018).
[109]G. L. Torres Mendonça, J. Pongratz, and C. H. Reick, "Identification of linear response functions from arbitrary perturbation experiments in the presence of noise—Part 1: Method development and toy model demonstration," Nonlinear Processes Geophys. **28**, 501–532 (2021).
[110]B. L. Otto-Bliesner, E. C. Brady, J. Fasullo, A. Jahn, L. Landrum, S. Stevenson, N. Rosenbloom, A. Mai, and G. Strand, "Climate variability and change since 850 CE: An ensemble approach with the community earth system model," Bull. Am. Meteorol. Soc. **97**, 735–754 (2016).